PEDESTRIAN INTERACTIONS WITH LEVEL 4 AUTOMATED DRIVING SYSTEMS: EVALUATION OF A NOVEL AUDITORY EXTERNAL HUMAN MACHINE INTERFACE

Gabrial T. Anderson[1], Srinath Sibi[2], Sadie Cooke[1], Jeremy Yang[2], Lowell Pickett[2], Ellen Xing[2], Kyle Evans[2] , Jasmine Kim[2], Huadong Wang[2] , Rossitza Kotelova[2], Rafael Patrick[3], Tanner Upthegrove[3], Stephen A. Ridella[2], Luke Neurauter[1], & Sheila G. Klauer[1]

[1]Virginia Tech Transportation Institute, United States of America
[2]Zoox, United States of America
[3]Virginia Tech, United States of America

Abstract Number **ESV26-295**

ABSTRACT

With the advent of SAE Level 4 Automated Driving Systems (ADS) there is no longer a human driver to provide supplemental cues to pedestrians. To replicate this driver-to-pedestrian communication, ADSs may need additional external methods of communication to aid pedestrians and other road users with understanding and anticipating vehicle actions. The current study examined the effect of novel sounds on pedestrian behavior. Forty participants were balanced across gender and vision status (i.e., 20 low-vision/visually impaired and 20 normal sighted participants). Participants experienced two scenarios: (1) a jaywalking scenario and (2) a pedestrian crosswalk scenario. In each scenario, an external speaker on the automated vehicle played non-speech-based, auditory sounds to communicate with participants. Participants provided objective crossing decisions and subjective ratings about the sounds they experienced. For the jaywalking scenario, time to not cross (TTNC; i.e., time to decide "I will not cross the road"; faster is better) showed that the effect of sound was significant as compared to the baseline vehicle sound. Further, all test sounds resulted in similar TTNC metrics as the horn sound. This suggests that the test sounds were as effective as the horn sound in influencing TTNC for jaywalking scenarios. For the pedestrian crosswalk, results suggested that sound did not influence crossing decisions. This suggests that for pedestrian-ADS interactions at crosswalks, repeated sound exposure may be needed for learning what a sound is communicating or that additional external communication may not be as necessary above that of the overall motion of the vehicle.

Keywords: Pedestrian, SAE Level 4 , Automated Driving Systems (ADS), road crossing decisions

INTRODUCTION

Pedestrians are among the most vulnerable road users. According to the Insurance Institute for Highway Safety (2025), 7,314 pedestrians died in motor-vehicle crashes in 2023, representing 18% of all crash fatalities. Of these fatalities, 23% occurred at intersections and 77% occurred at non-intersections (Insurance Institute for Highway Safety, 2025). Intersections and crosswalks without traffic signals—or with ambiguous right-of-way—can produce serious conflicts when pedestrians and drivers misinterpret one another's intent. In conventional, human-driven vehicles, drivers often rely on eye contact, facial expressions, and hand gestures to communicate awareness and yielding behavior (Dam et al., 2023). These nonverbal cues support negotiation of traffic priority, signaling of intent, resolution of ambiguity, acknowledgement of other road users, responses to violations, and management of hazardous situations (Rouchitsas & Alm, 2019).

Automated Driving Systems (ADSs) introduce interaction dynamics that may differ substantially from those of traditional vehicles (Larsson et al., 2019). ADSs leverage sensors, computation, and control to partially or fully replace the human driver's role in navigation and traffic response. By reducing or eliminating human error, ADSs have the potential to lower traffic-related injuries and fatalities (Center for Sustainable Systems, University of Michigan, 2024). However, ADS-equipped vehicles do not provide a visible human driver with whom pedestrians can establish eye contact or interpret facial expressions and gestures. Consequently, pedestrians may have greater difficulty inferring ADS intent than the intent of a human driven vehicle. Communicating vehicle intent to other road users is therefore both critical and challenging for safe pedestrian–ADS interactions (Lee et al., 2019; Tran et al., 2021). Although some companies are conducting ADS trials in multiple cities worldwide, these deployments have primarily emphasized the onboard user experience rather than communication with other road users (Lee et al., 2021).

For human driven vehicles, intent is communicated through both explicit cues (e.g., turn signals, brake lights) and implicit cues (e.g., vehicle motion, eye contact, gestures). Selecting appropriate mechanisms to replace or augment these implicit cues in ADSs is central to external human-machine interface (eHMI) design (Roundtree et al., 2020). Effective eHMI design must also account for road user demographics (e.g., age, gender, and visual, auditory, and/or cognitive impairments) and environmental conditions, which directly influence what information an ADS should convey and how that information is perceived.

Modalities of eHMI

eHMIs have been developed to bridge communication gaps between pedestrians and ADS-equipped vehicles using a range of modalities, including LED light bands, digital displays, auditory signals, and haptic cues using wearable devices (Lee et al., 2019; Verma et al., 2019; Schmitt et al., 2022; Tran et al., 2021). These interfaces are commonly categorized as visual, auditory, or vehicle body language (Verma et al., 2019; Colley et al., 2020; Dey, 2020). To date, the research emphasis has been disproportionately placed on visual eHMI, including digital depictions (e.g., smiles, waving gestures, pedestrian-style signals) and LED-based color/motion patterns intended to convey yielding or intent (Schmitt et al., 2022). Researchers have proposed visual modalities to support interactions between pedestrians and ADSs (Roundtree et al., 2020). However, visual cues and vehicle body language may unintentionally produce ambiguous or misleading signals (Larsson et al., 2019). Despite these concerns, comparatively less effort has been directed toward developing non-visual eHMI that are inclusive of pedestrians with visual impairments (Colley et al., 2020; Roundtree et al., 2020; Asha et al., 2022).

Auditory eHMIs remain less studied than visual eHMI despite their relevance for accessibility and attention capture (Colley et al., 2020; Roundtree et al., 2020; Asha et al., 2022). Moreover, existing evaluations of auditory eHMI have been conducted predominantly in controlled or semi-controlled contexts; outdoor, naturalistic evaluations—where open-air acoustics and ambient masking noise materially affect detectability and interpretation—remain rare. This limitation is consequential because real-world performance of auditory cues is inherently context dependent.

Sound, in general, can provide continuous, low-effort information, capture attention, and influence emotional and physiological states—properties that align with alerts, warnings, and alarms, which are among the most common sonic interactions in vehicles. Auditory signals are consistently identified as particularly beneficial for pedestrians who are blind or have low vision (Asha et al., 2022; Larsson et al., 2019). Accordingly, auditory signals are often proposed to communicate ADS presence and intent, especially to support pedestrians with visual impairments (Larsson et al., 2019; Dey, 2020; Roundtree et al., 2020; Tran et al., 2021). Mahadevan et al. (2018) further reported that participants preferred auditory signals as a means for the vehicle to communicate awareness.

Auditory signals may also be advantageous for capturing the attention of distracted pedestrians. In dense urban environments, competing sounds and visual stimuli, other road users, distractions, and general inattention may degrade pedestrians' ability to detect and interpret vehicle intent cues. Although distraction countermeasures have been explored, direct comparisons across mitigation strategies (e.g., technology-based versus infrastructure-based) remain underdeveloped (Arafat et al., 2023). Within infrastructure-based approaches, pedestrians with typical vision and hearing have been shown to prefer auditory intervention (Larue & Watling, 2021). Additional work is needed to clarify the relative benefits of vehicle-integrated technology-based interventions and to establish performance across contexts.

Considerations for Auditory eHMI Design

Auditory eHMI design requires careful consideration before broader adoption (Larsson et al., 2019; Dey, 2020). First, attention-capturing sounds tied to discrete events requiring immediate responses can be intrusive and irritating. Second, ambient noise can mask signals, reducing detectability and increasing the likelihood of misses or misinterpretation. In high-noise contexts, mixed-frequency auditory content such as speech may be particularly vulnerable to masking. In addition, speech-based warnings introduce a need for language proficiency. As of 2018, over 8% of people living in the United States reported limited confidence in their English-speaking ability (U.S. Census Bureau, 2021). This may leave nearly 30 million individuals with reduced ability to interpret complex speech-based warnings if messages are delivered only in spoken or written English.

Although abstract tones (e.g., beeps, trills, tonal motifs) can improve detectability in some environments, they may be less intuitive due to limited inherent semantic meaning. Overall, auditory eHMI could improve inclusivity for pedestrians with visual impairments (Larsson et al., 2019; Dey, 2020; Roundtree et al., 2020; Tran et al., 2021), and limited evidence suggests pedestrians and cyclists may prefer simple non-speech sounds over spoken content or textbased messaging (Lee et al., 2019). To be effective, nonverbal auditory eHMI should (1) capture attention without undue intrusiveness, (2) remain perceivable under masking noise, and (3) communicate meaning in an intuitive and learnable manner (Dey, 2020; Larsson et al., 2019; Mahadevan et al., 2018).

Characteristics of Auditory Warnings

Auditory warning design has been examined across domains to characterize how acoustic parameters influence intended and perceived urgency (Edworthy et al., 1991). Perceived urgency is strongly affected by temporal characteristics such as speed, acceleration/deceleration, and repetition; however, repetition can also increase annoyance. Melodic parameters also influence urgency, typically to a lesser extent. For example, very wide or very narrow pitch ranges (relative to moderate ranges), as well as atonal or unpredictable patterns, can increase perceived urgency. Systematic manipulation of temporal and melodic parameters is therefore foundational for scalable warning systems that can adapt to situational demands (Edworthy et al., 1991).

Perceived situational urgency varies across individuals and depends on perceived safety and personal risk. Accordingly, warnings may require scalable urgency levels. Werneke et al. (2014) found that across driving contexts, drivers preferred informative messages approximately 20–30 seconds prior to a potentially critical event, whereas in-vehicle warnings were acceptable about 8–14 seconds in advance. This discrepancy suggests that hypothetical urgency judgments may differ from in situ urgency. Matching warning characteristics to assessed urgency may improve the effectiveness of communication and support safer interaction with the environment.

Annoyance is also a critical outcome of warning design. A substantial number of studies have examined in-vehicle warnings and their relationship to driver annoyance (e.g., Marshall et al., 2007; Zaki et al., 2021). In many contexts, intended users are also the owners/operators of the alarm, and false alarm rates are a central predictor of annoyance over time (Block et al., 1999). For pedestrians, exposure is typically brief (e.g., during a single crossing), so repeated false alarms may be less relevant. A more pertinent long-exposure stakeholder may be the vehicle occupant, who could experience repeated auditory cues over extended trips. Therefore, auditory warnings should be matched to the urgency of the environmental context but also designed to limit false alarms.

Pedestrian Behavior

Pedestrian interaction with human drivers commonly relies on vehicle movement and implicit visual signaling, including gestures and gaze-based cues (Dam et al., 2023). This reliance is reflected in pedestrian gaze behavior: as vehicles approach, pedestrians often shift their gaze from the bumper toward the windshield, consistent with an expectation that the driver is a key communication channel (Dey, 2019). Such implicit communication patterns have influenced eHMI designs for pedestrian-ADS interactions (Lownes et al., 2022). However, without standardization or prior exposure, pedestrians may struggle to interpret unfamiliar eHMI signals. Understanding pedestrian behavior and perceptions across scenarios is therefore essential for characterizing interaction dynamics (Lownes et al., 2022). Miscommunication can lead to unsafe crossing decisions and unpredictable behavior around ADS-equipped vehicles.

Blind and Low-vision Pedestrians

Blind and low-vision pedestrians constitute a population for whom ADS communication design has heightened safety relevance, as street-crossing risk is amplified by reduced access to visual information in a motion-based environment. Highly detectable, localizable, and intuitive audio may also benefit sighted pedestrians when attention is divided or confirmation is needed. Although standardization efforts have recommended visual cues for vehicle-to-pedestrian communication (International Organization for Standardization, 2018), visual-only approaches do not support pedestrians with low or no vision. Best practices for inclusive communication for low-vision road users, therefore, recommend auditory cues as explicit indicators of vehicle intent, with sufficient information content and sufficient time for detection and comprehension (Colley et al., 2020; Lanzer & Baumann, 2023). For these reasons, auditory cues are necessary for inclusive vehicle-to-pedestrian (V2P) communication.

Study Purpose and Research Questions

This paper investigates the effectiveness of auditory eHMIs for pedestrian interaction with a Level 4 ADS-equipped vehicle. Because auditory eHMIs are less studied than visual eHMIs and have rarely been evaluated in outdoor, naturalistic contexts, this study evaluates an auditory eHMI in a closed-course, test-track roadway environment. Participants with normal and low vision made objective crossing decisions and provided subjective ratings of the experienced sounds. The research questions were:

1. How did the different sound conditions affect their decision to cross?

2. How are different sound conditions rated subjectively?

METHODS

Participants

Forty individuals from the New River Valley region of Virginia participated in this study. The sample included 19 females and 21 males, with an average age of 43.1 years (SD = 19.1; range = 18–72 years). Participation was strictly voluntary, and all participants received a compensation award of $250 in the form of a pre-paid Mastercard upon completion of the 2.5-hr study session. Twenty participants reported normal or corrected-normal vision, 20 participants self-reported non-correctable vision or an acuity of 20/200 or worse. Hearing acuity was assessed using a clinical, calibrated MAICO MA 28 audiometer. Pure-tone air conduction thresholds were measured at octave frequencies between 250 Hz and 8,000 Hz. Testing was performed in a quiet environment with minimal distractions. The presentation order began with higher frequencies (1,000 Hz to 8,000 Hz) to mitigate auditory fatigue, followed by a 1,000-Hz spot check to assess response reliability and potential bias, and concluded with lower frequencies (500 Hz to 250 Hz). The overall average hearing level across all participants was 24.3 dB HL, indicating a mild threshold elevation that slightly exceeds the upper limit of adult normal hearing.

Materials/Equipment

<u>Vehicle</u>

The study used a retrofitted Toyota Highlander hybrid that was provided by Zoox (see Figure 1).

Figure 1. Test vehicle.

This test vehicle was equipped with a proprietary external speaker system along the roof of the vehicle. This external speaker system was used to present the various test sounds used in this experiment. Although this vehicle was equipped with sensors (e.g., cameras and radar) that support automated driving features, the vehicle was driven manually during trials. Participants were told the vehicle was equipped with automated features, and to give the appearance of operating with those features engaged, a seat suit was worn to conceal the driver’s presence in the vehicle (see Figure 2; the window was rolled up during trials).

Figure 2. Driver wearing seat suit viewed from the driver side window.

<u>Test and masking sounds</u>

Seven sound conditions, previously curated for effectiveness, were tested for two roadway scenarios; however, no individual sound was evaluated for both scenarios. It is important to note the test vehicle is a hybrid vehicle with an internal combustion engine (ICE) and electric powertrain. At low speeds, the vehicle emits a FMVSS 141 compliant sound. Engine noise or FMVSS 141 sound or both were present during all scenarios in addition to any sounds emitted from the external speaker system. This was controlled for by having a trial in each scenario where the vehicle presented without any external speaker sound as a baseline. The FMVSS 141 sound and all test sounds were developed by Zoox. Jaywalking (JW)-1, 2, and 3 were designed to convey a sense of urgency, with the Horn Sound representing a conventional horn sound. The stock horn found on the test vehicle was not used because that would have varied the method of sound presentation compared to the other test sounds. For this reason, the Zoox-designed horn sound was presented through the external speaker system. Pedestrian Crossing (PC)-Sound 1 and 2 were designed to communicate that the vehicle was waiting for the pedestrian, with the Friendly Horn representing a gentle “double tap” of the horn. The baseline sound was included in both the jaywalking and pedestrian crossing scenarios (scenarios described in the procedure section). See Table 1 for a summary of all sounds and scenarios.

Table 1.

Presented sounds across scenarios

Sounds	Jaywalking	Pedestrian Crossing
JW-Sound 1	X	
JW-Sound 2	X	
JW-Sound 3	X	
Horn Sound	X	
PC-Sound 1		X
PC-Sound 2		X
Friendly Horn		X
Baseline	X	X

For jaywalking, a designed masking noise was used during trials to control the environmental noise floor. The masking noise was created by Zoox engineers and could be characterized as pink noise with boosts in the frequency spectrum below 1,000 Hz. The intent of this noise was to represent sound levels typical of an urban environment. The spectral content of the masking noise met or exceeded the National Highway Traffic Safety Administration (NHTSA) profile for background noise specified in the Notice of Proposed Rule Making (NPRM) for Federal Motor Vehicle Safety Standard (FMVSS) No. 141 (FMVSS Minimum Sound Requirements for Hybrid and Electric

Vehicles, 2013). The masking noise signal was presented at 60 dB(A), “which pedestrians would no longer be able to reasonably rely on hearing to detect approaching vehicles” . This masking noise was not used for the pedestrian crossing scenario because the participant was actively engaged in the crossing decision (i.e. not also distracted with task) thus masking noise was not critical.

Trial Equipment

Two data acquisition systems (DASs) were used to collect data for all trials. These DASs collected GPS data, video data, microphone data, and button press data. GPS data were collected continuously through trials in UNIX time. This afforded the ability to sync all data collected with the DAS to the data collected through the test vehicle. Video data were collected at various angles to record participant behavior during trials. The DAS microphone was used to collect scratch audio that gave context to the video data (cameras did not record audio). The DAS microphone was a miniDSP UMIK-2 Class 1 microphone with its signal delivered over a class-compliant Universal Serial Bus. The button press data was used to collect participant crossing behavior for various trials.

A high-end microphone was used to collect spectral and sound pressure auditory data of the presented conventional, experimental, and masking sounds during each trial. This high-end microphone was an Earthworks M23 IEC 60942:2003 Class 1 omnidirectional measurement microphone. A NoiseMeter IEC 60942:2003 Class 1 calibrator was used to calibrate microphones before each jaywalking scenario with a 94-dB, 1-kHz sine tone. During trials for the jaywalking scenarios, participants were required to be distracted by a task to mimic a pedestrian who is not fully attentive to the roadway environment. With half of the sample involving low-vision participants, a distraction task was designed to be vision agnostic. A tactile search task was employed that involved participants wearing a shoulder bag that contained a number of shapes with varying edges. The target in this bag was a single triangle that was presented to each participant before the start of trials.

Finally, at the end of each trial, participants responded to a questionnaire to assess participants’ ratings of the characteristics of the sounds presented. For simplicity, this questionnaire was administered verbally, with responses being collected primarily by an audio recorder pinned to a participant’s shirt, with redundant recordings found on VTTI-managed cell phones using a recorder app, the DAS microphone, and the high-end microphone.

Procedure

Setup

At the beginning and end of each test day, one of the DAS microphones was placed approximately 1 m away from the external speaker system on the test vehicle. A sample of experimental sounds was played into that microphone to confirm they were being presented at the specified sound pressure level. The microphone used a calibration file, and sound pressure level data was measured and logged in Room EQ Wizard.

Before each jaywalking scenario, the high-end microphone would be calibrated using the NoiseMeter calibrator. This tool was placed on the high-end microphone with a fitted adapter and played a constant tone at 94 db SPL. This tone was played until the microphone displayed a sound pressure level of 93-95 dbA by automating the preamplifier’s gain control. After this calibration, the high-end microphone would be placed in the middle of the external speaker system (see Figure 3). The masking noise would then be played out of the speakers located at each participant station.

Participant procedures

Procedures for this study were approved by the Virginia Tech Institutional Review Board (IRB# 25-169). Two participants were recruited for each approximately 2.5-hr session. Session participants were intentionally matched on vision status (i.e., both were sighted or low vision) with an effort to match across age and gender. Upon arrival, participants were given an informed consent form, a W-9 form, pre-session surveys and a hearing test. Participants signed the consent form and W-9 prior to completing the survey or hearing test. The hearing test was a pure tone audiometry that followed the procedures outlined in Zhao & Mayr (2021), which are based on recommendations of the British Society of Audiology (BSA; 2018). This test was performed in an individual room with ambient noise levels typical for an office setting (approximately 30–35 dBA). Following the intake procedures, participants were transported to the Virginia Smart Roads – Rural Section. Two scenarios were tested per participant session: jaywalking and pedestrian crosswalk.

Jaywalking scenario

In the design of the study, the interaction between jaywalking pedestrians and ADS-equipped vehicles was chosen for investigation. Prior evidence (IIHS, 2025) indicates that non-intersection encounters account for 77% of total pedestrian fatalities. Although ADS-equipped vehicles possess the capability for continuous environmental sensing and can execute emergency braking to avoid collisions, such maneuvers significantly elevate the risk of rear-end impacts from trailing vehicles. Furthermore, the integration of auditory signaling may serve as a critical mitigation strategy to alert inattentive pedestrians who might otherwise enter the roadway without watching for proximal traffic. Hence, a jaywalking interaction scenario was chosen to study the effect of sounds on participants' decision to cross the street. In this scenario, participants were instructed to imagine themselves as a distracted pedestrian crossing the street. Note that experimental procedures were developed to ensure that participants did not actually cross the street in front of a moving vehicle; rather, they held a button connected to the DAS and were instructed to release the button whenever they felt it was no longer safe for them to cross the street. While holding the button, participants were instructed to search through the distraction bag for a triangle. At the end of each trial, participants would respond verbally to a subjective questionnaire assessing their impression of the presented sound. See Figure 3 for an overview of the equipment used in this scenario.

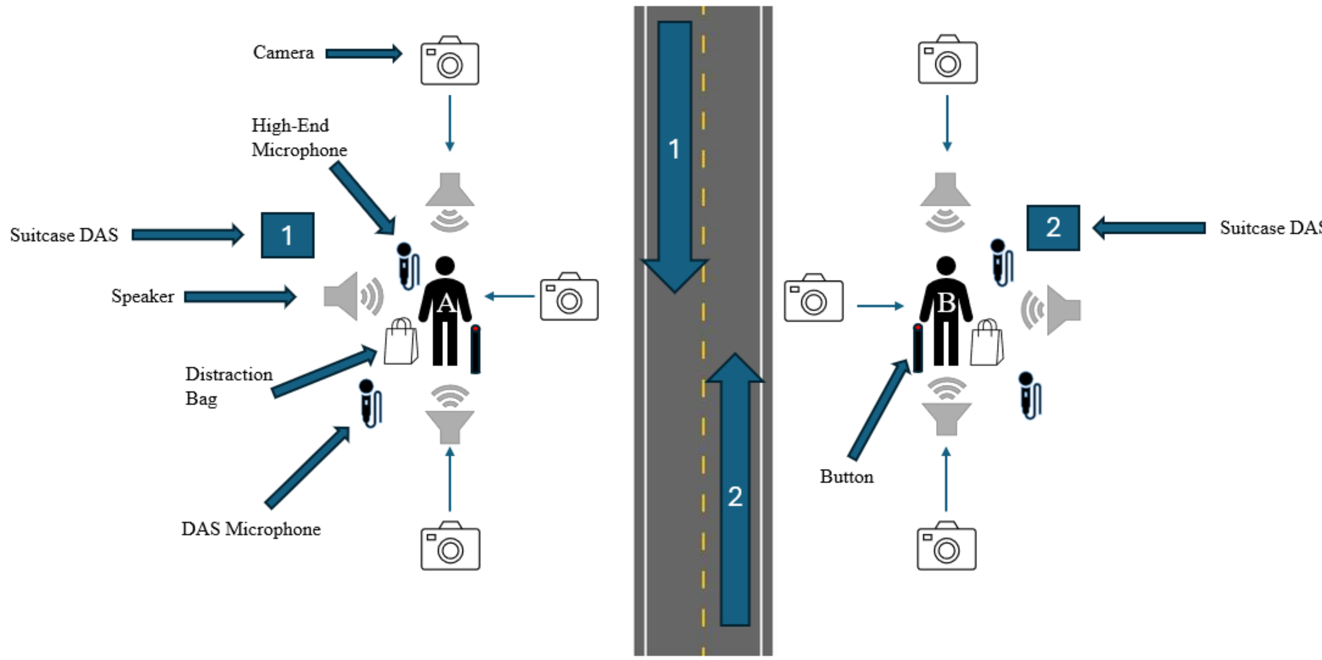

Figure 3. Overview of equipment setup for jaywalking trials.

During jaywalking trials, the test vehicle would come from either Direction 1 or Direction 2 (going west to east "1" or east to west "2" in Figure 3) at 25 mph. The test vehicle was hidden from view in the first few seconds of its approach from either direction. The test sounds (i.e., sounds emitted from the external speaker system) were triggered using a tablet inside the test vehicle cabin (the baseline condition involved the test vehicle driving by the participant with no additive sound). Test sounds were manually selected prior to each trial and automatically triggered at 60 m from the participant locations; however, due to a slight delay from trigger initiation to sound presentation, it would actually present at 55 m distance. This distance was selected because it represented a 5-s time-to-collision (TTC), which is the 50^{th} percentile of acceptable TTC for a pedestrian to still cross with approaching traffic (Wang et al., 2010).

Participants stayed in their location for the entirety of this scenario (Figure 4).

Figure 4. Jaywalking Direction 2, east to west, with Participant A on the right side of the roadway and Participant B on the left side of the roadway (researchers depicted).

<u>*Pedestrian Crosswalk scenario*</u>
Human drivers employ explicit communication cues to establish right-of-way for pedestrians intending to enter a crosswalk, particularly during ambiguous traffic scenarios. Conversely, pedestrian interactions involving autonomous vehicles that lack communicative interfaces, such as actions taken by a human driver, may lead to protracted and inefficient exchanges. Such navigational ambiguities significantly degrade the operational efficiency of a larger autonomous fleet. To this end, a crosswalk interaction scenario was chosen to understand the effect of the test sounds on a participant's crossing behavior in a crosswalk when the right-of-way is ambiguous and can cause a navigational ambiguity. Only one participant was tested per trial in this scenario, and they were restricted from actually crossing the street for safety. During trials, participants stood at the edge of the road with the intent to cross the street as the test vehicle approached. To indicate they would cross, they were asked to march in place and stop marching when they would not cross. See Figure 5 for the test track setup; the participant is facing south with the test vehicle approaching from east to west as depicted in the crosswalk location marked in Figure 5.

Figure 5. Pedestrian crosswalk scenario with test vehicle and participant visible (researcher depicted).

A key factor for this scenario was creating a confusion state between the test vehicle and participant as to who had the right-of-way to proceed. This was achieved by having the test vehicle come to a stop and then move forward slightly (i.e., nudge) to create confusion on the part of the participant as to whether the vehicle was going to continue through the intersection or not. Once this confusion state was created, the sound for that trial was presented. Once presented, the driver would wait for approximately 4 s to allow the participant to respond, then continue driving through the intersection. The subjective questionnaire was verbally administered after each trial.

Experimental Design

All test sounds that were presented (with the exception of the engine and tire noise in the baseline condition) during the two scenarios were developed and designed by Zoox to test their effectiveness on pedestrian behavior. All sound condition orders were counterbalanced within scenarios. The scenario order was evenly split between participants experiencing the jaywalking scenario first or the pedestrian crosswalk scenario first. The direction of approach for jaywalking was also counterbalanced. The independent and dependent variables varied by scenario. See Table 2 for a list of variables shared across scenarios and Table 3 for variables unique to specific scenarios.

Table 2.
Variables shared across scenarios

Shared Across Scenarios			
	Description	**Levels**	**Designation**
Age	Self-reported age of the participant	Young/Old (split at median age 48.5)	Between-Subjects
Gender	Self-reported gender of the participant	Male/Female	Between-Subjects
Vision Status	Vision status of the participant	Sighted/Low Vision	Between-Subjects
Participant ID	ID assigned to participants	Various ID numbers	Within-Subjects
Subjective Ratings	Subjective rating	Urgency, Assertiveness, Friendliness, Aggression, Clarity of Meaning, Loudness, Appropriateness	Within-Subjects; each level rated 1–5, 5 being the highest

Table 3.
Variables unique to each scenario

Jaywalking Scenario			
	Description	**Sounds**	**Designation**
Sound	The test sound presented during a given trial	JW-Sound 1 JW-Sound 2 JW-Sound3 Horn Sound Baseline	Within-Subjects Independent Variable
Time to Not Cross (TTNC)	The time difference between a sound presentation and button release	Seconds	Dependent Variable

Pedestrian Crosswalk Scenario			
	Description	**Levels**	**Designation**
Sound	The test sound presented during a given trial	PC-Sound 1 PC-Sound 2 Friendly Horn Baseline	Within-Subjects Independent Variable
Time to Cross	The time difference between a sound presentation and crossing decision	Seconds	Dependent Variable

Age, gender, and vision status were collected during the recruitment process, and participant ID was assigned at the beginning of a given test session. The subjective ratings were collected at the end of each trial using one of the recorded audio sources

TTNC (jaywalking) was derived from DAS button data. As mentioned above, all DAS data were collected continuously and in parallel with GPS timestamps (GPS_{TS}). To calculate those button data metrics, the GPS_{TS} for each behavior (i.e., the first GPS_{TS} with a button press or release) was extracted from each relevant trial. The sound trigger times were collected using hardware in the test vehicle. Those data points were used in the formulas below:

$$(1)\ Time\ to\ Not\ Cross = Button\ Release\ GPS_{TS} - Sound\ Trigger\ GPS_{TS}$$
$$(2)\ Response\ Time = Button\ Press\ GPS_{TS} - Sound\ Trigger\ GPS_{TS}$$

The pedestrian crosswalk outcome and Time to Cross required researchers to review video and code participant crossing behavior. This process focused on participant behavior after the vehicle nudge, as this marked the beginning of the "crossing confusion state" that was of interest. Researchers would code their crossing status (participant marching or standing still) before and after the sound presentation for a given trial. Active marching indicated they would cross, and lack of marching indicated they would not cross. The GPS_{TS} of the first movement in response to the sound presentation (e.g., lifting foot) was coded as the crossing decision time. Two researchers independently coded each pedestrian crosswalk trial with an agreement rate of 88%. Disagreements were discussed and resolved. The Time to Cross variable was calculated using the first movement GPS_{TS} and the sound trigger times using the following formula:

$$(3)\ Time\ to\ Cross = First\ Movement\ GPS_{TS} - Sound\ Trigger\ GPS_{TS}$$

Across all scenarios, the baseline condition was unique as it did not have a "trigger" for that sound to present as the test vehicle and its ambient noise was always "presenting." However, a GPS_{TS} was needed to calculate the various behavior metrics for each scenario's baseline condition. Therefore, the in-vehicle operators marked when the test vehicle passed the trigger point for that given scenario to indicate when the vehicle sound was "triggered," recording a GPS_{TS} for that trial. For trials that did not have that GPS_{TS} marked (e.g., missed by mistake), the video from the test vehicle was reviewed for when the front bumper of the vehicle passed the trigger point for that scenario. For jaywalking, this was indicated by cones on the shoulder of the road that marked the trigger points. For pedestrian crosswalk, this moment was marked as after the vehicle nudge ended. Once coded, they were used in Formulas 1–3.

RESULTS

Research questions were related to the sounds and their effect on the variables outlined in Table 2 and Table 3. Each scenario had specific research questions of interest that are detailed below.

Jaywalking

1. How did the sound conditions compare on the decision to cross?

Of the 200 trials performed for the jaywalking scenario, participants described hearing the target sound in 184 trials (92%), did not describe the target sound in five (3%) of trials, and had missing data in 11 (5%) trials. Of the 11 trials where a sound was not heard, Participant 13 accounted for five of those trials; all other participants who did not hear the sound had only one missed trial each. This suggests that not hearing a sound was overrepresented in this scenario by Participant 13.

Analyses for TTNC focused on the 184 trials where the participant described hearing the target sound, as it was assumed that hearing the target sound influenced their crossing decision in some way. Of those 184 trials, 16 trials involved participants deciding not to cross before a sound was presented. For this reason, it was assumed the sound did not influence their decision to cross and were excluded from analyses (168 final trial count). The decision to cross was analyzed by investigating variance in TTNC across the presented sound conditions. A mixed-effect analysis of variance included age, gender, sound, and vision status as independent variables and participant ID as a random effect. Results showed that the main effects of Sound, Vision Status, and Age were significant (see Table 4).

Table 4.
Mixed-effect model output for decision to cross

Variable	Metric	Effect	*F*(df 1, df 2)	*P*	Level	*M*	*SE*
Decision to Cross	Seconds	**Sound (A)**	**17.87(4, 126.6)**	**<0.001**	Baseline[a]	3.68	0.26
					Horn Sound[b]	2.02	0.25
					JW-Sound 1[b]	1.98	0.25
					JW-Sound 2[b]	2.14	0.25
					JW-Sound 3[b]	1.91	0.25
		Gender (B)	0.001(1, 29.0)	0.85	Male	2.34	0.29
					Female	2.35	0.29
		Age (C)	**4.47(1, 28.9)**	**<0.05**	Young[a]	2.79	0.29
					Old[b]	1.90	0.29
		Vision (D)	**19.42(1, 28.7)**	**<0.01**	Sighted[a]	1.46	0.29
					Low Vision[b]	3.23	0.29
		A * B	0.87(4, 128.2)	0.48	-	-	-
		A * C	1.08(4, 128.1)	0.37	-	-	-
		A * D	2.05(4, 127.7)	0.09	-	-	-
		B * C	2.33(1, 29.1)	0.14	-	-	-
		B * D	0.93(1, 29.1)	0.34	-	-	-
		C * D	0.03(1, 29.1)	0.86	-	-	-

Note. For significant factors, levels that do not share the same letter are significantly different.

Tukey's honest significant difference (HSD) post hoc test was used to compare significant differences due to the large number of comparisons being made. Tukey's HSD adjusts p-values to account for family-wise comparisons and lowers the likelihood of Type I error. For the main effect of sound, the baseline condition had significantly longer TTNC than all other sound conditions (mean differences ranged from 1.54 to 1.77 s). TTNC for all other sound conditions (excluding the baseline) was *not* significantly different from the horn sound. In other words, all sound conditions had a similar effect on TTNC as the horn sound. Younger participants had longer TTNC than older participants, and low-vision participants had longer TTNC than sighted participants.

2. How did the sound conditions compare on participants' subjective ratings?

This analysis focused on the 184 participants who heard the target sound. Due to the large number of sound conditions and subjective rating scales, the model was restricted to only understanding differences in ratings across sound conditions. The main effects of age, gender, and vision or their interactions with the sounds were not of interest for this analysis. A mixed-effect analysis of variance was used to investigate subjective ratings across sound conditions (ratings were on a scale of 1 -5, 5 being the highest). The model included sound as the independent variable and participant ID as a random effect. Tukey's HSD post hoc test was used to compare differences across sounds. Results suggest JW-Sound 3 was rated lower on urgency, assertiveness, and clarity of meaning compared to the horn condition. Compared to the baseline condition, sounds differed on urgency, assertiveness, friendliness, and loudness ratings. See Table 5 for all comparisons.

Table 5.

Subjective ratings for the jaywalking sound conditions

Sound	Rating Scale						
	Urgency M(SE)	Assertiveness M(SE)	Friendliness M(SE)	Aggression M(SE)	Clarity M(SE)	Loudness M(SE)	Appropriateness M(SE)
Baseline	3.09(0.21)[b]	2.67(0.21)[b]	2.00(0.20)[b]	2.68(0.21)	3.69(0.23)	2.49(0.19)[b]	3.19(0.23)[b]
Horn	4.08(0.19)[ac]	3.98(0.20)[ac]	2.70(0.19)[ac]	3.18(0.20)[a]	4.11(0.21)[a]	3.80(0.18)[ac]	3.94(0.21)[a]
JW-Sound 1	4.18(0.19)[c]	4.19(0.20)[c]	2.96(0.20)[c]	3.34(0.20)	4.06(0.21)	3.73(0.18)[c]	3.83(0.22)
JW-Sound 2	3.85(0.19)[c]	3.45(0.20)[c]	3.05(0.19)[c]	2.73(0.20)	3.47(0.21)	3.56(0.18)[c]	3.75(0.21)
JW-Sound 3	3.45(0.20)[b]	3.22(0.20)[b]	3.18(0.20)[c]	2.50(0.20)[b]	3.29(0.22)[b]	3.43(0.18)[c]	3.35(0.22)

Note. Comparisons to the horn sound are shown. Sound conditions with a "b" are significantly different from the horn sound; sound conditions with a "c" are significantly different from the baseline.

Pedestrian Crosswalk Scenario

1. How did the presented sounds compare on the decision to cross?

One hundred sixty trials were performed for the pedestrian crosswalk scenario. Of the 160 trials, participants described hearing the target sound in 143 trials (89%), did not describe the target sound in 10 trials (6%), and had missing data in 7 trials (4%). The 17 trials with missing descriptions or a description of something other than the target sound were excluded from analyses as it is assumed the target sound for a given trial did not influence their behavior. The following results focused on the remaining 143 trials.

After the test vehicle nudge, participants did not change their crossing decision in 76 trials (53%), went from no to yes in 67 trials (47%), and from yes to no in one trial (<1%). The Yes to No trial was excluded from analyses as a count of one is not enough to perform inferential tests on. The distribution of crossing decisions by sound condition is shown in Figure 6.

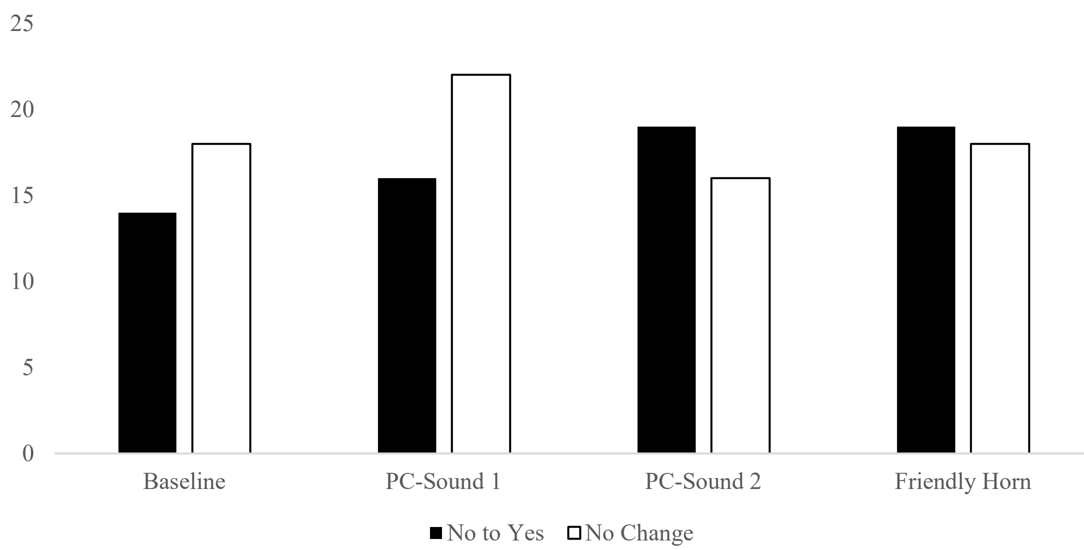

Figure 6. ***Distribution of pedestrian crossing decisions.***

A binomial generalized linear mixed model was used to understand the odds of a participant transitioning from no to yes. Gender, age, sound condition, and vision status were included as fixed effects, participant ID as a random effect, and "crossing decision outcome" as the dependent variable. Interactions were excluded from the model due to low cell counts in those stratified variable combinations. Results for the model are shown in Table 6. The estimate column shows the odds of transitioning from no to yes in a given condition specified in the term row compared to the condition specified in the reference column.

Table 6.
Binomial GLMM results

Target	Reference	Term	Estimate*	95% CI Lower	95% CI Upper	P
Pedestrian Crosswalk Outcome "No to Yes"	-	Intercept	0.06	0.003	1.22	0.07
	Baseline Sound	Friendly Horn	1.30	0.28	5.96	0.74
		PC- Sound 1	0.88	0.19	4.04	0.87
		PC-Sound 2	1.78	0.38	8.46	0.46
	Female	Gender	0.71	0.06	8.63	0.79
	Old	Age	0.93	0.08	11.38	0.96
	Sighted	**Vision**	**89.43**	**2.68**	**2982.18**	**0.01**

*Note. *Log-odds ratios and confidence intervals have been exponentiated to convert to odds ratios.*

For example, the odds of going from no to yes in the Friendly Horn condition compared to the Baseline condition were 1.30 but were not significant. The only significant effect was that sighted participants had 89.43 times higher odds of going from no to yes than low-vision participants. However, this result should be interpreted with caution and considered non-significant due to the extreme width of the confidence interval. There was no difference in sound, gender, or age on likelihood to go from no to yes.

Of those 67 trials where the crossing decision outcome was no to yes, six participants started marching before the sound was presented. Therefore, the presented sounds were assumed to not have influenced their crossing decision and were excluded from the time to cross analysis. For the remaining 61 trials, a mixed-effect model was built using gender, age, sound, and vision status as fixed effects, participant ID included as a random effect, and time to cross as the dependent variable. Due to the low trial count, higher order interactions (e.g., three-way and above) were excluded from the model. Results suggest none of the main effects or interactions on time to cross were significant (see Table 7).

Table 7.
Mixed model results for time to cross

Variable	Metric	Effect	*F*(df 1, df 2)	*P*	Level	*M*	*SE*
Time to Cross	Seconds	Sound (A)	0.86(3, 29.5)	0.48	Baseline	1.55	0.54
					Friendly Horn	2.41	0.56
					PC_Sound 1	2.34	0.65
					PC_Sound 2	2.21	0.48
		Gender (B)	0.01(1, 19.0)	0.91	Male	2.02	0.48
					Female	2.24	0.53
		Age (C)	2.90(1, 18.7)	0.11	Young	1.43	0.54
					Old	2.83	0.49
		Vision (D)	0.75(1, 17.3)	0.40	Sighted	2.17	0.37
					Low Vision	2.09	0.60
		A * B	0.56(3, 30.2)	0.64	-	-	-
		A * C	0.34(3, 30.0)	0.80	-	-	-
		A * D	1.76(3, 30.8)	0.18	-	-	-
		B * C	0.02(1, 17.5)	0.89	-	-	-
		B * D	0.50(1, 17.3)	0.49	-	-	-
		C * D	2.28(1, 18.6)	0.15	-	-	-

2. How did the presented sounds compare on subjective ratings?

Participant mean subjective ratings of each sound condition are shown in (ratings were on a scale of 1–5, 5 being the highest). A mixed-effect analysis of variance was used to investigate each rating scale across sound conditions. Sound condition was included as the independent variable and participant ID as a random effect. Tukey's HSD post hoc test was used to compare differences across sounds. Results suggest the Friendly Horn was rated as more urgent, assertive, and aggressive than all other sound conditions. PC-Sound 1 and PC-Sound 2 were rated as more friendly than the Friendly Horn and baseline condition. PC_Sounds 1 and 2 were rated as having less clarity of meaning than the Friendly Horn and baseline conditions. Friendly Horn was rated as louder than PC_Sound 1 and the baseline condition. PC_Sound 1 was rated as less appropriate for the traffic situation than the baseline condition.

Table 8.
Overview of subjective ratings for each sound condition in the pedestrian crosswalk scenario

	Rating Scale						
	Urgency	Assertiveness	Friendliness	Aggression	Clarity	Loudness	Appropriateness
Sound	M(SE)	M(SE)	M(SE)	M(SE)	M(SE)	M(SE)	M(SE)
Baseline	2.89(0.21)^{b}	2.69(0.20)^{b}	2.67(0.20)^{d}	2.17(0.19)^{b}	3.65(0.24)^{d}	2.92(0.20)^{b}	3.69(0.25)^{b}
Friendly Horn	3.69(0.20)^{a}	3.90(0.19)^{a}	2.50(0.19)^{a}	3.29(0.18)^{a}	3.71(0.23)^{a}	3.70(0.19)^{a}	3.50(0.24)
PC_Sound 1	2.53(0.20)^{b}	2.68(0.19)^{b}	3.26(0.19)^{b,c}	1.97(0.18)^{b}	2.52(0.23)^{b,c}	3.20(0.19)^{b}	2.85(0.24)^{a}
PC_Sound 2	2.66(0.20)^{b}	2.57(0.19)^{b}	3.78(0.19)^{b,c}	1.87(0.19)^{b}	2.67(0.24)^{b,c}	3.27(0.20)	2.95(0.24)

Note. “a” is significantly different from “b”; “c” is significantly different from “d.”

DISCUSSION

The current study sought to understand the effect of non-speech-based sounds presented through an auditory eHMI on naïve pedestrian behavior. These sounds were novel and have not been used in a roadway environment previously. Participants were not trained nor were they told how to interpret the sounds. Thus, the results of this study should be considered a “baseline” understanding of how naïve pedestrians would respond to an auditory eHMI in these types of scenarios. With further exposure and/or training, pedestrian responses may improve. The sounds used in this study were presented to inform and communicate the intent of the L4 ADS-equipped vehicle to these naïve pedestrians. The sounds were tested across two unique scenarios: jaywalking and pedestrian crossing. Each scenario had unique sounds that were tested. Participant behavioral response and subjective ratings of those sounds were recorded after each trial. The following section will summarize the findings from each scenario.

Test Sounds and Pedestrian Behavior

Jaywalking

The jaywalking scenario sought to test the impact of sounds emitted from the external speaker system on pedestrian jaywalking behavior within a 10-s window. Results suggest TTNC was faster for all sound conditions compared to the baseline condition. Further, participants’ TTNC was similar in response to the horn sound compared to the other test sound conditions. This suggests that the test sounds were as effective as the horn sound in influencing TTNC.

Beyond effectiveness on TTNC, participants’ subjective rating in their perception of each sound was investigated. Conventional horn use can provoke aggressive responses (McGarva & Steiner, 2000), presumably because it is associated with aggressive drivers (Hennessy & Wiesenthal, 2001; Turner, Layton, & Simons, 1975). Therefore, sounds that were as effective as a horn sound on influencing TTNC but were perceived in a more favorable way are viable candidates for further development. In the current study, most test sounds were rated similarly on urgency and aggression compared to the horn sound; however, JW-Sound 3 was the only test sound rated as less assertive. Although rated as slightly less urgent than the horn sound, this suggests that participants perceived JW-Sound 3 as being less aggressive and less assertive than the horn sound in its communication. Urgency and aggression may be essential to communication given the safety-critical context of this scenario; however, a high ranking of assertiveness might not be necessary to effectively communicate. This rating, in combination with the TTNC results, suggests JW-Sound 3 is as effective as a horn sound on pedestrian behavior but is perceived in a more favorable manner.

Results suggested that older participants had faster TTNC than younger participants. In other words, older participants were faster to indicate they would not cross than younger participants. This behavior is supported in previous research that suggests older pedestrians prefer a larger gap between themselves and an approaching vehicle when deciding to cross the street compared to younger pedestrians (e.g., Lobjois & Cavallo, 2009). Further, low vision participants had longer TTNC than sighted participants. This means that low vision participants waited longer to make a crossing decision than sighted participants. Variation in crossing behavior by sight-status is supported in previous research (Ashmead et al., 2005; Hassan et al., 2012) and supports the need for vision-agnostic eHMI.

Overall, participants had no physical reaction (e.g., startle) to the sounds. This suggests there was no triggering of hypersensitivities when presenting these novel sounds in a safety-critical context. Previous research suggests that ambient vehicle noise can raise stress levels (Sygna et al., 2014; LaJeunesse et al., 2021) and that horn use can startle and irritate some pedestrians (Takada et al., 2017).

Pedestrian Crosswalk

The pedestrian crosswalk scenario sought to test the effectiveness of sounds in resolving a right-of-way confusion state between a pedestrian and a yielding L4 ADS-equipped vehicle. Pedestrians tend to focus their gaze on the driver of an approaching vehicle (Dey et al., 2019) because nonverbal communication (e.g., hand signals) is common in vehicle-to-pedestrian interactions (Dam et al., 2023). The sounds in this scenario attempted to replace those nonverbal methods of communication due to the lack of a driver in an L4 ADS-equipped vehicle.

Results suggest that the sounds did not influence pedestrian crossing decisions. The likelihood of a participant crossing did not improve with the addition of sounds compared to only the baseline. For those that did decide to cross, their time to cross did not change with the presentation of sounds compared to the baseline condition. This trend is represented in previous research with a visual eHMI not affecting crossing decision times compared to a control condition (Clamann et al., 2017). Previous research suggests novel eHMI communication requires more time for pedestrians to process than conventional methods (Goonetilleke, Shih, & Fritsch, 2001; Huang & Bias, 2012). Table 7 shows a numeric trend, though not statistically significant, for this, with each sound's time to cross being ~1 s longer than the baseline condition.

Previous research suggests that pedestrian interactions with a novel eHMI improve over time as learning takes effect (de Clercq et al., 2019). Vehicle motion is often cited as the primary source of information used in pedestrian crossing decisions (Soares et al., 2021; Liu & Tung, 2014; Mahadevan et al., 2018). Repeated exposure to the sounds in combination with the L4-ADS equipped vehicle performing a yielding action may improve participant understanding of the sound, its intent, and the time required to resolve the confusion state. It is important to note the current study investigated a confusion state where the test vehicle was not in motion at the time of sound presentation. However, the precursor to this confusion state involved the test vehicle decreasing speed as it approached the crosswalk.

Participants' subjective ratings of each sound condition were also investigated. As with the jaywalking scenario, it was important to understand participant perception of the sounds. Friendly Horn was rated as more urgent, assertive, and aggressive than PC-Sounds 1 and 2, and the baseline condition. PC-Sounds 1 and 2 were rated as more friendly but less clear in meaning than Friendly Horn and baseline sound. This suggests that PC-Sounds 1 & 2 were perceived more favorably to Friendly Horn, but their meaning was not as clear. Future research is needed to see if repeated exposure to PC-Sounds 1 and 2 improves clarity of meaning as well as crossing decision and time to cross.

LIMITATIONS

The current study was not without limitations. The test sounds in this study are intended for use in an urban environment. Efforts were made to replicate this environment by using masking sounds to raise the environmental noise floor in participant locations; however, the testing environment did not replicate ambient traffic or other pedestrians. In an urban environment, there is a greater density of vehicles and pedestrians (Carlson et al., 2018; National Center for Statistics and Analysis, 2023) that could influence vehicle-to-pedestrian interactions (Obeid et al., 2017). Further research is needed to investigate the potential impact of ambient noise on vehicle-to-pedestrian interactions.

Similarly to environmental noise, the baseline condition for this test was a combination of FMVSS 141 sound and/or the engine sound , depending on the speed of the vehicle. The test vehicle used an ICE/electric powertrain; thus, at lower speeds, participants could hear the engine and/or the FMVSS 141 sound. While these sounds are a limitation of this study, it was less so for the jaywalking condition where speakers produced a masking sound and the vehicle was traveling at a consistent 25 mph. For the pedestrian crosswalk condition, we believe these sounds had less of an impact on decision to cross compared to vehicle motion.

The effect of jaywalking sounds on pedestrian TTNC was significant. However, those sounds were triggered at a TTC of 5s. It is important to note that their effectiveness at any lower TTC should not be extrapolated from these results, and more testing is needed. Further, the horn sound used in this scenario was played out of the external speaker system and not the stock OEM horn found on the test vehicle. This difference in presentation could have impacted results. Although a horn sound was used, participant response to a traditional horn presentation compared to the test sounds is unknown and would require further testing.

Participants were local to southwest Virginia and may not be representative of pedestrians that live in metropolitan areas (e.g., Northern Virginia). Those that live in metropolitan areas may have more experience with the vehicle-to-pedestrian interactions tested in this study. Future research is needed to understand differences in locality on pedestrian-to-sound interactions.

CONCLUSIONS

An ADS equipped vehicle may benefit from supplemental external communication to communicate vehicle intent to pedestrians as compared to a vehicle with a human driver. A human driver can use methods of communication (e.g., hand gestures) with pedestrians that do not rely on external modalities (Rasouli et al., 2017; Dam et al., 2023). Such driver-based communication is not currently implemented/available in ADSs (Rasouli & Tsotsos, 2020). The current study tested the effectiveness of auditory sounds in replacing driver-based communication. It is important to note that this study can be listed as one of the first to test the effectiveness of an auditory eHMI as a replacement of driver-based communication in ADS-to-pedestrian interactions on a test track (i.e., an actual roadway environment). Additionally, given this unique experimental design, all participants were naïve, having never heard any of the sounds prior, nor were they trained on how to interpret the sounds. Thus, this experiment and its results represent "baseline" pedestrian behavior and/or opinions on their ratings of these sounds or their understanding of what the sounds are communicating to them.

Jaywalking pedestrians decided not to cross the street in a similar time, whether responding to a horn sound or the test sounds. In the pedestrian crosswalk right-of-way scenario, the test sounds did not influence crossing behavior, but this could improve with learning through repeated exposure.

Overall, the jaywalking sounds promoted favorable participant responses. Repeated exposure may improve the understanding of right-of-way sounds and the implications for improved safety for pedestrian interactions with ADSs.

DISCLAIMER

Include any necessary disclaimers or declare conflict(s) of interest.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the larger team (aside from the authors of this paper) from both Zoox and VTTI that made this study possible, including the engineers and test drivers that assisted with the data collection and the research associates who assisted with all aspects of data collection, including recruitment, experimental set up, data collection, and equipment maintenance.

REFERENCES

Arafat, M. E., Larue, G. S., & Dehkordi, S. G. (2023). Effectiveness of interventions for mobile phone distracted pedestrians: A systematic review. *Journal of Safety Research, 84*, 330-346.

Asha, A. Z., Brierley, O., Somanath, S., Finn, P., & Sharlin, E. (2022). *Towards designing audio interactions with autonomous vehicles: A hearing-enhanced pedestrian story*. ACM Conference on Human Factors in Computing Systems (CHI 2022), New Orleans, LA, USA, April 30, 2022.

Ashmead, D. H., Guth, D., Wall, R. S., Long, R. G., & Ponchillia, P. E. (2005). Street crossing by sighted and blind pedestrians at a modern roundabout. *Journal of Transportation Engineering*, *131*(11), 812-821

Block, F. E., Nuutinen, L., & Ballast, B. (1999). Optimization of alarms: a study on alarm limits, alarm sounds, and false alarms, intended to reduce annoyance. *Journal of Clinical Monitoring and Computing, 15*, 75-83.

British Society of Audiology (BSA). (2018). Recommended procedure: Pure-tone air-conduction and bone-conduction threshold audiometry with and without masking. https://www.thebsa.org.uk/ resources/pure-tone-air-bone-conduction-threshold-audiometry-without-masking/

Carlson, S. A., Whitfield, G. P., Peterson, E. L., Ussery, E. N., Watson, K. B., Berrigan, D., & Fulton, J. E. (2018). Geographic and urban–rural differences in walking for leisure and transportation. *American Journal of*

Preventive Medicine, *55*(6), 887-895.

Center for Sustainable Systems, University of Michigan. (2024). *Autonomous vehicles factsheet* (No. CSS16-18). https://css.umich.edu/publications/factsheets/mobility/autonomous-vehicles-factsheet

Clamann, M., Aubert, M., & Cummings, M. L. (2017). *Evaluation of vehicle-to-pedestrian communication displays for autonomous vehicles* (No. 17-02119).

Colley, M., Walch, M., Gugenheimer, J., Askari, A., & Rukzio, E. (2020). Towards inclusive external communication of autonomous vehicles for pedestrians with vision impairments. *CHI '20: Proceedings of the 2020 Conference on Human Factors in Computing Systems*, April 25–30, 2020, Honolulu, HI, USA.

de Clercq, K., Dietrich, A., Núñez Velasco, J. P., de Winter, J., & Happee, R. (2019). External human-machine interfaces on automated vehicles: Effects on pedestrian crossing decisions. *Human Factors*, *61*(8), 1353-1370. https://doi.org/10.1177/0018720819836343.

Dey, D. (2020). *External communication for self-driving cars: Designing for encounters between automated vehicles and pedestrians* [Doctoral Thesis 1 (Research TU/e / Graduation TU/e), Industrial Design. Technische, Universiteit Eindhoven].

Dey, D., Walker, F., Martens, M., & Terken, J. (2019, September). Gaze patterns in pedestrian interaction with vehicles: Towards effective design of external human-machine interfaces for automated vehicles. In *Proceedings of the 11th international conference on automotive user interfaces and interactive vehicular applications* (pp. 369-378).

Edworthy, J., Loxley, S., & Dennis, I. (1991). Improving auditory warning design: Relationship between warning sound parameters and perceived urgency. *Human Factors, 33*(2), 205-231.

Endsley, M. R. (1995). A taxonomy of situation awareness errors. *Human factors in aviation operations*, *3*(2), 287-292.

Federal Motor Vehicle Safety Standards: Minimum Sound for Hybrid and Electric Vehicles, 78 FR 2798 (proposed January 14, 2013). https://www.govinfo.gov/content/pkg/FR-2013-01-14/pdf/2013-00359.pdf

Goonetilleke, R. S., Shih, H. M., & Fritsch, J. (2001). Effects of training and representational characteristics in icon design. *International Journal of Human-Computer Studies*, *55*(5), 741-760.

Hassan, S. E. (2012). Are normally sighted, visually impaired, and blind pedestrians accurate and reliable at making street crossing decisions?. *Investigative ophthalmology & visual science*, *53*(6), 2593-2600.

Hennessy, D. A., & Wiesenthal, D. L. (2001). Gender, driver aggression, and driver violence: An applied evaluation. *Sex Roles*, *44*(11), 661-676.

Huang, S. C., & Bias, R. G. (2012). Without context, icons are significantly worse than texts to convey meanings in terms of accuracy and efficiency. In *Proceedings of the ASIS&T 75th Annual Meeting*.

Insurance Institute for Highway Safety. (2025). *Fatality Facts 2023*. https://www.iihs.org/research-areas/fatality-statistics/detail/pedestrians#yearly-snapshot

International Organization for Standardization. (2018). *Road Vehicles — Ergonomic aspects of external visual communication from automated vehicles to other road users* (ISO/TR 23049:2018). ISO OBP. https://www.iso.org/

LaJeunesse, S., Ryus, P., Kumfer, W., Kothuri, S., & Nordback, K. (2021). Measuring pedestrian level of stress in urban environments: Naturalistic walking pilot study. *Transportation Research Record*, *2675*(10), 109-119.

Lanzer, M., & Baumann, M. (2023). It matters who is crossing: Users' trust when the automated vehicle stops for children and blind pedestrians. *Transportation Research Part F: Psychology and Behaviour*, 96, 222–232.

Larsson, P., Maculewicz, J., Fagerlönn, J., & Lachmann, M. (2019) Auditory displays for automated driving: Challenges and opportunities. *The 25 the International Conference on Auditory Display (ICAD 2019)*, June 23–27, Northumbria University. https://doi.org/10.21785/icad2019.038

Larue, G. S., & Watling, C. N. (2021). Acceptance of visual and audio interventions for distracted pedestrians. *Transportation Research Part F: Traffic Psychology and Behaviour, 76*, 369-383.

Lee, Y. M., Madigan, R., Garcia, J., Tomlinson, A., Solernou, A., Romano, R., Markkula, G., Merat, N., & Uttley, J. (2019). Understanding the messages conveyed by automated vehicles. *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular*, *19*, September 21-25. Utrecht, Netherlands. https://doi.org/10.1145/3342197.3344546

Lee, M. Y., Madigan, R., Giles, O., Garach-Morcillo, L., Markkula, G., Fox, C., Camara, F., Rothmueller, M., Vendelbo-Larsen, S., A., Rasmussen, P. H., Dietrich, A., Nathanael, D., Portouli, V., Schieben, A., & Merat, N. (2021). Road users rarely use explicit communication when interacting in today's traffic: Implications for automated vehicles. *Cognition, Technology & Work*, *23*, 367–380. https://doi.org/10.1007/s10111-020-00635-y.

Liu, Y. C., & Tung, Y. C. (2014). Risk analysis of pedestrians' road-crossing decisions: Effects of age, time gap, time of day, and vehicle speed. *Safety Science*, *63*, 77-82.

Lobjois, R., & Cavallo, V. (2009). The effects of aging on street-crossing behavior: from estimation to actual crossing. *Accident Analysis & Prevention*, *41*(2), 259-267.

Lownes, N., Rezwana, S., Shaon, M. R. R., & Jackson, E. (2022). *Pedestrian behavior and interaction with autonomous vehicles* (CAMMSE-UNCC-2021-UTC-Project-06). Center for Advanced Multimodal Mobility Solutions and Education. University of North Carolina at Charlotte. https://rosap.ntl.bts.gov/view/dot/64494/dot_64494_DS1.pdf

Marshall, D. C., Lee, J. D., & Austria, P. A. (2007). Alerts for in-vehicle information systems: Annoyance, urgency, and appropriateness. *Human factors*, *49*(1), 145-157.

McGarva, A. R., & Steiner, M. (2000). Provoked driver aggression and status: A field study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *3*(3), 167-179.

Mahadevan, K., Somanath, S., & Sharlin, E. (2018). Communicating awareness and intent in autonomous vehicle-pedestrian interaction. *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, *429*, 1–12. https://doi.org/10.1145/3173574.3174003

Obeid, H., Abkarian, H., Abou-Zeid, M., & Kaysi, I. (2017). Analyzing driver-pedestrian interaction in a mixed-street environment using a driving simulator. *Accident Analysis & Prevention*, *108*, 56-65.

Rasouli, A., Kotseruba, I., & Tsotsos, J. K. (2017, June). Agreeing to cross: How drivers and pedestrians communicate. In *2017 IEEE Intelligent Vehicles Symposium (IV)* (pp. 264-269). IEEE.

Rasouli, A., & Tsotsos, J. K., (2020). Autonomous vehicles that interact with pedestrians: A survey of theory and practice. *IEEE Transactions on Intelligent Transportation Systems*, *21*, 3.

Rouchitsas, A., & Alm, H. (2019). External human–machine interfaces for autonomous vehicle-to-pedestrian communication: A review of empirical work. *Frontiers in Psychology, 10*, 2757. https://doi.org/10.3389/fpsyg.2019.02757

Roundtree, K. A., Hattrup, S., Swaminathan, J., Zerbel, N., Klow, J., Shitole, V., Fallatah, A., Khanna, R., & Adams, J. A. (2020). Inclusive design guidance: External autonomous vehicle interfaces. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *64*(1), 1054-1058. https://doi.org/10.1177/1071181320641253

Schmitt, P., Britten, N., Jeong, J., Coffey, A., Clark, K., Kothawade, S. S., Grigore, E. C., Khaw, A., Konopka, C., Pham, L., Ryan, K., Schmitt, C., & Frazzoli, E. (2022). Can cars gesture? A case for expressive behavior within autonomous vehicle and pedestrian interactions. *IEEE Robotics and Automation Letters*, *7*(2). https://doi.org/10.3929/ethz-b-000530000

Soares, F., Silva, E., Pereira, F., Silva, C., Sousa, E., & Freitas, E. (2021). To cross or not to cross: Impact of visual and auditory cues on pedestrians' crossing decision-making. *Transportation Research Part F: Traffic Psychology and Behaviour*, *82*, 202-220.

Sygna, K., Aasvang, G. M., Aamodt, G., Oftedal, B., & Krog, N. H. (2014). Road traffic noise, sleep and mental health. *Environmental Research*, *131*, 17-24.

Takada, M., Suzuki, S., Kim, K. H., Shin, J. H., & Iwamiya, S. I. (2017). Driver's vehicle horn use and its effects on other drivers and pedestrians: A case study in South Korea. *Mechanical Engineering Journal*, *4*(2), 16-00433.

Tran, T. T. M., Parker, C., & Tomitsch, M. (2021). A review of virtual reality studies on autonomous vehicle-pedestrian interaction. *IEEE Transactions on Human-Machine Systems*, *51*(6), 641-652. https://doi.org/0.1109/THMS.2021.3107517

Turner, C. W., Layton, J. F., & Simons, L. S. (1975). Naturalistic studies of aggressive behavior: Aggressive stimuli, victim visibility, and horn honking. *Journal of Personality and Social Psychology*, *31*(6), 1098.

U.S. Census Bureau. (2021, October 8). *People that speak English less than "very well" in the United States*. https://www.census.gov/library/visualizations/interactive/people-that-speak-english-less-than-very-well.html

Verma, H., Pythoud, G., Eden, G., Lalanne, D., & Evequoz, F. (2019). Pedestrians and visual signs of intent: Towards expressive autonomous passenger shuttles. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 3, 3*, Article 107, 31. https://doi.org/10.1145/3351265

Wang, T., Wu, J., Zheng, P., & McDonald, M. (2010, September). Study of pedestrians' gap acceptance behavior when they jaywalk outside crossing facilities. In *13th International IEEE Conference on Intelligent Transportation Systems* (pp. 1295-1300). IEEE.

Werneke, J., Kleen, A., & Vollrath, M. (2014). Perfect timing: Urgency, not driving situations, influence the best timing to activate warnings. *Human Factors, 56*(2), 249-259.

Zaki, N. M., Husin, S. C., Husain, M. A., Husain, N. A., Ma'aram, A., Marmin, S. A., ... & Kassim, K. A. (2021). Auditory alert for in-vehicle safety technologies: a review. *Journal of the Society of Automotive Engineers Malaysia*, *5*(1), 88-102

Zhao, F., & Mayr, R. (2021). Pure tone audiometry and speech audiometry. In *Manual of Clinical Phonetics* (pp. 444-460). Routledge.