Article

Virtual Urban Field Studies: Evaluating Urban Interaction Design Using Context-Based Interface Prototypes

Robert Dongas [1,*] **, Kazjon Grace** [1] **, Samuel Gillespie** [1] **, Marius Hoggenmueller** [1] **and Martin Tomitsch** [2] **and Stewart Worrall** [3]

1 Sydney School of Architecture, Design and Planning, The University of Sydney, Sydney, NSW 2006, Australia
2 TD School, University of Technology Sydney, Sydney, NSW 2007, Australia
3 Australian Centre for Field Robotics, The University of Sydney, Sydney, NSW 2006, Australia; stewart.worrall@sydney.edu.au
* Correspondence: robert.dongas@sydney.edu.au

Abstract: In this study, we propose the use of virtual urban field studies (VUFS) through context-based interface prototypes for evaluating the interaction design of auditory interfaces. Virtual field tests use mixed-reality technologies to combine the fidelity of real-world testing with the affordability and speed of testing in the lab. In this paper, we apply this concept to rapidly test sound designs for autonomous vehicle (AV)–pedestrian interaction with a high degree of realism and fidelity. We also propose the use of psychometrically validated measures of presence in validating the verisimilitude of VUFS. Using mixed qualitative and quantitative methods, we analysed users' perceptions of presence in our VUFS prototype and the relationship to our prototype's effectiveness. We also examined the use of higher-order ambisonic spatialised audio and its impact on presence. Our results provide insights into how VUFS can be designed to facilitate presence as well as design guidelines for how this can be leveraged.

Keywords: user studies; virtual reality; prototyping; autonomous vehicles; external human–machine interfaces; presence; spatial audio

1. Introduction

Ever since computing began to emigrate from the desktop environment, testing in real-world contexts (i.e., "field" testing) has been argued as critical [1]. The more complex and dynamic the environment in which a system will be deployed, the more important it is that testing be conducted in the physical, spatial, temporal, sensory, and social surroundings of that context. This is particularly relevant in urban interactive systems: there are few environments more complex and dynamic than cities.

The challenge for any ubiquitous computing research project has always been that field testing is typically unavailable for a wide variety of economic, temporal, technological, and safety reasons. From limitations with testing medical devices with real patients, to limitations with testing AVs on real roads, to the simple limitation of taking a prototype out of the lab before it is mature enough...the reasons to avoid field testing may seem overwhelming, regardless of how critical it is known to be. In some industries, particularly those where safety regulations are paramount (such as vehicle interfaces), there has been continued development of high-fidelity physical interfaces—sit-in simulators—as part of the prototyping process [2]. However, for those products and services without the budget of a car or aeroplane manufacturer, real field testing has always been limited in both scope and frequency.

Virtual Field Studies (VFS) is an emerging area of HCI work that present a technology-mediated solution to the high cost of real field trials [3]. VFS, also known as "context-based interface prototypes" [4]) leverage mixed-reality technologies (i.e., AR, VR, etc.) to place users in an immersive, interactive virtual environment which mimics both the deployment

context and interaction model of the product or service being tested. We see this as particularly appropriate for exploring emerging urban technologies, as the complex urban environment will strongly mediate user interaction in ways that cannot be studied in a usability lab. At the same time, the urban environment also places acute technical, financial, and safety constraints on real field trials.

VFS for pedestrian–AV interaction [5] in urban environments are becoming more commonplace, but to date there have been very few published papers specifically focused on audio. In this article, we present a virtual field study in the domain of sound design for pedestrian–AV interaction. Given the novelty of the method, we then develop a set of considerations for conducting such studies in urban environments, which we refer to as Virtual Urban Field Studies (VUFS). Our design considerations include a new approach to validating the capacity of VUFS to stand in for the real environment it is intended to reproduce. Specifically, our study has the following research objectives:

RO$_1$: *To explore the efficacy, affordances and characteristics of VUFS with a sound design and audio focus.*

RO$_2$: *To investigate the impact on user experience of overlaying virtual sounds over recorded imagery in VUFS.*

We also present a new approach to validating the efficacy of VUFS in its capacity to mirror the real-world context it is intended to represent—its verisimilitude—based on psychological studies of the notion of "presence". This validation is intended to sit alongside the actual measures of prototype efficacy (at whatever its intended purpose is—in our case communicating vehicle intent sonically) to ensure that the virtual setup is externally valid.

2. Related Work

Our study draws on and contributes to human–computer interaction (HCI) research investigating the use of virtual reality (VR) for prototyping and evaluating complex interactive urban systems. We examine and expand upon claims that virtual environments offer a high degree of realism, fidelity, and sense of presence. Our study also contributes to research in interfaces between autonomous vehicles (AVs) and pedestrians. Accordingly, we review below the literature on (1) virtual environments, (2) presence, (3) prototyping and evaluation methods in HCI, and (4) AV–pedestrian interfaces.

2.1. Virtual Environments for Developing VUFS

Virtual environments (VEs) strive to depict digital representations of spaces and places, allowing users to engage with and become absorbed in their surroundings [6]. Although these environments cannot fully replicate real-life experiences, they serve as effective platforms for simulating comparable situations by fostering a sense of immersion. In this research, we adopt the perspective of McMahan [7], as well as Nilsson et al. [8], suggesting immersion can be both perceptual (i.e., based on objective capabilities of the devices used to interact with the VE) and psychological (i.e., based on subjective experiences of the user while interacting with the VE). We acknowledge the difference between the broader notion of immersion and the more distinct concept of presence, which we explore in detail in Section 2.2.

VR technology offers users the opportunity to interact with VEs from a first-person perspective, completely enveloping them with the sensory stimuli of the digital experience. This enhances the perceptual immersion for the user, more closely resembling actual audiovisual input from the real world by correlating with their head movements.

The development of high-quality, computer-generated VEs, especially for use in VR, is not without its challenges. The creation of such environments often demands considerable investment in terms of time, financial resources, and expertise [9]. Researchers

and developers must consider the potential benefits and drawbacks of employing VEs in their specific context, ensuring that they balance the associated costs and complexities. By carefully selecting the sources of visual and audio stimuli for virtual environments, some of the investment burden can be reduced, allowing for more efficient content creation without compromising the overall quality and user experience. We discuss several specific VE technologies relevant to our study below: 360-degree video and spatial audio.

2.1.1. 360-Degree Video

One method of creating VEs for VR is through the use of 360-degree video. This approach involves recording video content and spherically projecting it in the virtual space. Users are able to observe the entire environment through three degrees of freedom, determined by their head rotation. Prior research has highlighted the immersive qualities of 360-degree video in areas such as storytelling [10], education [11], and health/wellbeing [12]. These studies point to the inherent realism of the medium, which can be attributed to its ability to capture real-world locations and situations through video recordings. A major advantage of 360-degree video is that there is no lengthy asset design, development and integration process, as there is with more-traditional 3D modelled VEs [9].

A notable limitation of 360-degree video is the restricted perspective, as users are confined to the vantage point of the camera and cannot actively navigate or move around the VE. To prevent motion-induced sickness, it is generally advised not to alter the camera position during recording [13]. Despite this constraint, 360-degree video remains a valuable medium, particularly in scenarios where the user is stationary and the primary focus is more on observation rather than direct interaction. Our use of this approach balances the immersive qualities of VR and the practical considerations of realistic content generation, enabling faster and more cost-effective production.

2.1.2. Spatial Audio and Soundscapes

The sense of hearing plays a crucial role in human perception, providing context and situational awareness beyond the scope of our vision. Spatial audio, characterised by its ability to convey the direction and distance of sound sources around a listener, has been utilised in various mediums such as film, television, video games, and exhibitions. In VR specifically, it can elevate users' sense of perceptual and psychological immersion within a virtual environment, especially when compared to non-spatial audio [14,15].

Spatial audio can be leveraged in VR using a variety of methods. One such method is ambisonics, a technique for capturing and reproducing sound fields in three dimensions [16]. This allows physical acoustic properties of the sound field (e.g., directionality and pressure) to be encoded to varying orders of spatial resolution. While first-order (i.e., lower-quality approximate) ambisonics systems have seen renewed interest in recent years, Zotter [17] argues that the advantages in interactivity offered by higher-order ambisonics as well as the limitations of first-order recording techniques are both compelling reasons to increase the directional resolution of recordings. These higher-order systems are now possible due to improvements in microphone technologies and signal processing methods [18], although this comes with added cost and complexity. Ambisonics offer a flexible platform for designing and producing immersive auditory soundscapes for use in VEs. Hong et al. [19] highlights the potential of spatial audio for creating immersive and ecologically valid soundscapes in VR, which we intend to explore further in this research. One of the objectives of this study is determining what, if any, advantage in the subjective user experience is conveyed by the incorporation of ambisonic spatial audio technologies in VR experiences.

2.2. Presence

At its most simple, any prototype designed for a virtual field study should be judged by its ability to make participants feel as if they really are in the "field" in question. Throughout the literature, researchers have presented multiple explanations of this notion of "being there" in order to conceptualise the complex psychological construct usually referred to as "presence" [20–25]. Of particular relevance to this study, Slater [26] defines the terms place illusion ("the strong illusion of being in a place in spite of the sure knowledge that you are not there") and plausibility illusion ("the illusion that what is apparently happening is really happening") as key components of presence.

Witmer and Singer [27] posit that presence occurs when users do not notice the artificiality of objects in a simulated environment. Expanding upon this is Lee's theory of presence, which defines it as "a psychological state in which virtual objects are experienced as actual objects in either sensory or nonsensory ways" [28]. Lee promotes a way of categorising the virtual experience by making the distinction between artificiality and para-authenticity. Objects with no real-world counterparts represent artificiality, while para-authenticity describes the extension of actual objects in a virtual environment. We agree with Gilbert's [29] notion that "authenticity is the human-based factor that influences presence, as measured by whether it aligns with the expectations of users". Lee's model also breaks presence down into three distinct subdomains: physical (presence arising from the environment), social (presence arising from others), and self (presence arising from the body) [28]. We find this to be a suitable contextual framing, particularly within the domain of interaction design in urban settings, where there is interplay between all three subdomains.

Measuring Presence

Although the immersiveness of VEs has been widely promulgated, translating these qualities into quantifiable and operational measures can be difficult. Evaluating the effectiveness of these environments presents a unique challenge, especially when utilising the technological medium of VR. We propose elevating the psychological measure of presence as a key quality in designing and developing VUFS (as we argue in this paper), as it has been well-operationalised in that literature and is a strong driver of immersion. These measures of presence can be applied through a variety of psychometrically validated instruments, which we apply to VE users as a way of measuring the efficacy of those VEs at recreating the desired environment.

We adopt the definition of presence as "the extent to which something (environment, person, object, or any other stimulus) appears to exist in the same physical world as the observer" [30]. Understanding how to effectively measure presence according to this definition is crucial in assessing the success and validity of VEs. Presence as a subjective psychological construct requires the user to self-report their experience [31]. These data are commonly collected through post-experience questionnaires, which are low-cost and easy to administer. They also "embody self-reports and therefore gather the participants' subjective experiences" [32]. When utilising this method, the reliability and validity of the questionnaires need to be carefully designed for consistent research to avoid potential biases such as the "design of the question, questionnaire design, and administration" [33].

The Multimodal Presence Scale (MPS) encompasses each of Lee's presence constructs: physical, social, and self [34]. It has been validated in the context of VR learning simulations through confirmatory factor analysis and item response theory. The ability of the MPS to effectively measure attributes of presence such as realism, attention and mediation strengthens its suitability for our purposes. These attributes are good indicators of involvement in the virtual experience, but also help to establish the ecological validity of the environments for VUFS. For these reasons, we adopt the MPS for use in our study.

Souza et al. [35] provides a review of the various measurements of presence, categorising them into objective and subjective methods—the latter being mostly the questionnaire-based methods discussed above. Objective methods are "captured through data collection

or observation, and can be divided into physiological and behavioural measures" [35]. Examples include measuring externally observed responses, heart rate, eye movements and skin temperature [6]. While these objective measures can be said to provide data that are more reliable than self-reporting from a participant bias standpoint, the literature is not clear on whether they are more accurate as standalone measures of the sense of presence [36]. While these physiological tools are an area worth watching for VUFS evaluation, we administer the more well-established survey instruments in this work.

2.3. Prototyping and Evaluation Methods in Human Computer Interaction

The creation of prototypes is an important activity in any interaction design process [37], as it allows designers to explore and refine ideas before the envisioned product or service is developed in its final manifestation. Prototypes can fulfil different purposes, such as enabling designers to traverse a design space in a more generative manner, and evaluating manifestations of design ideas with stakeholders [38]. For screen-based interfaces, at an early stage of the design process, designers typically start with low-fidelity prototypes (e.g., sketches, paper prototype) to validate concepts, before moving on to more high-fidelity prototypes (e.g., realistic mockups, interactive click-throughs).

Over the past two decades, the use of interactive products and services has become increasingly interwoven in people's everyday lives. As a consequence, designers have had to adapt how they prototype in order to consider interactions in context, in addition to traditional task-based measures of usability. Buchenau and Suri [39] coined the term "experience prototype", which they define as "any kind of representation, in any medium, that is designed to understand, explore or communicate what it might be like to engage with the product, space or system we are designing". In a similar vein, related terms such as "context-based interface prototyping" have been used to describe approaches of prototyping interfaces and interactions in the context that they are expected to be used [4,40]. This incorporates a plethora of prototyping and design techniques, including bodystorming, enactments, physical prototyping, and more.

With this turn towards the contextual in HCI, researchers have stressed the importance of evaluating prototypes "in the wild" rather than relying entirely on tests in laboratory environments [41]. This has been identified as particularly relevant in the design of urban technologies [42,43]. However, conducting field studies is expensive and time-consuming, since prototypes have to be safe, stable, movable, and robust. In the context of prototyping and evaluating interfaces for autonomous vehicles, conducting user studies in the real-world could even cause harm to participants [44].

To reduce costs and risks, HCI researchers increasingly turned to video- and simulation-based approaches for evaluating prototypes through contextualised setups, also referred to as virtual field studies [3]. This includes video recordings which are displayed on a conventional screen [4], as well as computer-generated simulations presented to participants through projection-based VR environments (also known as CAVE) [45,46] or VR headsets [47]. Several studies across different contexts (e.g., public displays [3], wearable technology [48] and smart home devices [49]) have demonstrated that the evaluation of interfaces in immersive VR environments holds comparable usability results and user behaviour to that of real-world settings. Researchers have further found that the use of 360-degree video recordings of real-world prototypes can further improve perceived representational fidelity and sense of presence [50]. In the context of AV–pedestrian interfaces, Hoggenmueller et al. [4] have shown that the use of real-world representations in VR is highly effective for the evaluation of contextual aspects and to assess overall trust. We propose that these advantages of higher visual fidelity means of prototyping suggest that a similar effect will be observable for technologies offering higher aural fidelity (i.e., ambisonic recordings). However, to the best of our knowledge, previous research has not established the efficacy of ambisonics in the context of VR prototypes.

2.4. AV–Pedestrian Interfaces

With the uptake of autonomous vehicles expected to continue, research has increasingly turned to addressing human factor challenges in order to increase public acceptance and trust towards AV technologies [51]. One such challenge is the lack of interpersonal communication cues (e.g., hand gestures, facial expressions, and eye contact) between AVs and other road users. Researchers have therefore proposed the use of external human–machine interfaces (eHMIs) to augment automated driving [52]. For instance, eHMIs can be used to communicate the AV's intent to yield when negotiating right-of-way with pedestrians in street crossing scenarios [53], or to signal the AV's intent and awareness in shared spaces where they operate in close proximity to other road users [5].

There now exist a wide range of eHMI design concepts, a majority of which use a visual communication modality [52]. These range from projections on the road [44] to displays attached to the vehicle itself, including light band bars [5,54] as well as screens capable of displaying text or symbols [53]. VR simulation studies have shown that eHMIs can significantly reduce the risk of collisions [55,56] and increase pedestrians' subjective feeling of safety [57]. However, there is also increasing evidence that pedestrians mainly rely on implicit cues [58–60], such as a vehicle's movement, when deciding whether to cross in front of it. Consequently, sound has been suggested as an implicit communication modality, with the latest regulation frameworks requiring engine-like noise emissions for electric and hybrid cars to protect vulnerable road users [61]. For example, Moore et al. [62] found that adding synthetic engine sound to a hybrid autonomous vehicle led to increased interaction quality and clarity around the vehicle's intent to yield. Pelikan et al. [61] adopted a research-through-design approach to explore sound designs for autonomous shuttle buses in close-proximity interactions with other road users. Using voice-overs to video recordings of the buses, followed by Wizard of Oz testing in the real world, they discovered that prolonged jingles drew attention to the bus and encouraged interaction, whereas repeated short beeps and bell noises could be used to direct people away from the bus. At the same time, they emphasised the impact of situational and sequential contexts (e.g., the movements of the bus and other people) in determining the interaction that a specific sound may accomplish.

3. Methods

We combined research-through-design, an approach which establishes design knowledge through the generation and evaluation of artifacts [63], with a summative mixed-methods user study of the efficacy of our VUFS. To address RO1, we developed a context-based interface prototype, iteratively evaluating and improving it using a real-world use-case. Our final user study provided results and analysis of how effectively our virtual sounds preserved participants sense of presence in our VUFS (addressing RO2). We integrate this analysis with the insights from our design process [64] to develop a set of considerations for future VUFS.

In this section we describe both the design of our prototype VUFS (see Section 3.1) and the methodology of our summative user study (see Section 3.2). In the former, we detail technical specifications, design decisions, and where and what we depicted in our virtual environment. In the latter section we describe the participant demographics, study procedures and data analysis methods.

3.1. Designing Our AV–Pedestrian Context-Based Interface Prototype

We designed and developed a prototype to evaluate auditory cues for AV–pedestrian interaction. The prototype utilised recorded 360-degree video and ambisonic audio to create a VE to simulate urban shared spaces. We chose to focus on an area with minimal segregation between transport modes, i.e., concurrent walking, cycling and driving, a known challenge for AVs [65]. Users engaged with this VE through a VR headset, adopting the perspective of the pedestrian.

3.1.1. Virtual AV–Pedestrian Scenarios

Our prototype presented a variety of scenarios in which an AV approached a stationary pedestrian, depicting potential encounters in a shared space. In each scenario, the vehicle performed one of three possible behaviours: avoiding the pedestrian by changing path; slowing down to a gradual stop in front of the pedestrian; or continuing at the same speed and theoretically colliding with the pedestrian. The scenarios were recorded with a real vehicle, autonomously performing the avoid and stop behaviours listed above. For the collision behaviour, the vehicle was operated manually, maintaining a constant speed until emergency braking as close as possible to the camera. The selected scenarios and behaviours in the study were designed to represent realistic potential AV–pedestrian interactions, particularly in a shared walking/driving/cycling multi-use space.

To record the scenarios for our prototype, we used a fully functional research AV, developed by the Intelligent Transport Systems Group from the Australian Centre for Field Robotics [66]. The small vehicle (see Figure 1) is specifically engineered to operate in urban environments at low speeds, aiming to manoeuvre safely through pedestrians using advanced sensors and control systems.

Figure 1. The autonomous vehicle during technical testing, having navigated through a shared space and come to a stop. The scenario is being recorded by a 360-degree video camera and spatial microphone to create an immersive virtual environment for a Virtual Urban Field Study.

The avoid and stop behaviours implemented in the prototype represent the default programming of the AV depending on its mode of operation. These behaviours are designed to prioritise pedestrian safety and comfort, placing the responsibility of movement and adaptability on the vehicle. The collision behaviour was created for the purpose of researching the impact of sound design in such a scenario. Although it is (obviously) not a default behaviour of the vehicle, its inclusion in the study acknowledges the possibility that an AV may become aware of a potential collision with a pedestrian, but be unable to react in time to avoid it. Our choice of scenarios and behaviours is informed by our design

to align the study with real-world urban situations, including both normal operation of an AV as well as potential exceptional cases.

3.1.2. Virtual AV Sound Design

The soundscape design for our prototype followed a hybrid approach (see [67]), integrating recorded ambient environmental sounds with designed and produced vehicle sound effects. This combines the conceptualisations of soundscapes as "perceptual constructs that emerge when a listener hears sounds from a physical environment" and as "a collection of sounds that are organised in an intentional way and played to listeners" [67]. In other words, sounds are concurrently both sensory and socio-cultural experiences. A hybrid approach such as this is extremely relevant for conducting VUFS, particularly for auditory interfaces, given the need for sound composition as an intervention within the rich acoustic context of an urban environment. Our ambient recordings captured the live environmental audio for each video of each scenario that made up the virtual experience. This included many different sound sources: wind, bird song, people walking and talking, aeroplane noises, distant traffic sounds and consequential sounds of the AV (e.g., tyres on the ground). We employed fourth-order ambisonic recording techniques to capture a high-resolution spatial representation of the sound field (combining 32 discrete directional recording channels). This was mapped into a headphone mix for the VR experience, which (when combined with head tracking) enables a highly immersive audio field.

For each of the three possible AV behaviours (avoiding, stopping and colliding), three distinct sound variations were designed (S1, S2, and S3). These sound designs differed in characteristics, ranging from simpler to more complex. The stop sound aimed to convey the vehicle's intent to decelerate to a halt without requiring any pedestrian response, while the avoid sound was designed to indicate the AV's intent to manoeuvre around the stationary pedestrian. The collide sound, on the other hand, was intended to warn pedestrians of an impending collision, requiring them to act. Various sound characteristics (including timbre, pitch, and tempo) were manipulated to design the desired effects for each behaviour (see Table 1). For instance, the stop sound exhibited decreasing intensity, pitch, and tempo, whereas the collide sound portrayed increasing intensity, pitch, and tempo to elicit a sense of urgency in pedestrians.

Table 1. Overview of sound variations (S1, S2 and S3) designed for each of the three possible autonomous vehicle behaviours: avoiding, stopping and colliding. The sound designs exhibit distinct characteristics, with various attributes such as timbre, pitch and tempo manipulated to communicate each specific behaviour.

Behaviour	Characteristic	S1	S2	S3
Avoid	Notes in pattern	One	Two	Three
Stop	Falling pitch	No	Yes	Yes
	Diminishing tempo	Yes	No	Yes
	Reducing timbrel intensity	Yes	Yes	Yes
Collide	Rise in pitch	No	Yes	Yes
	Increasing tempo	Yes	No	Yes
	Intensifying Timbre	Yes	Yes	Yes

Our sound production approach focused on applying just enough processing to situate the sound effectively in the VE. We employed some reverb to introduce approximations of diffused reflections based on suggestions by [68], which decreased as the autonomous vehicle approached the user. More accurate simulation of the acoustic reverberation conditions would be possible through measuring multiple spatial impulse responses in the testing environment [69]. However, given the additional complexity this would add to the audio recording, processing and implementation, we opted against this higher-fidelity processing. We also implemented an equalisation (EQ) shift and used compression to minimise the

variation in dynamic range, ensuring a consistent audio experience throughout the virtual environment. Our goal here was to make the (introduced) vehicle sound effects parallel the sonic environment of the ambisonic recording as much as feasibly possible.

3.1.3. Virtual Environment Implementation

360-degree videos were recorded using an Insta360 Pro 2 camera (Maufacturer: Insta360, Shenzhen, China) set to a high frame rate (120 fps) in stereoscopic mode for capturing high fidelity images (3840 × 3840 per eye). For ambient audio recording, an Eigenmike EM-32 (32-channel spherical microphone array – Manufacturer: mh acoustics, New Jersey, United States) was utilised to capture and encode 4th-order ambisonic output. Figure 2 shows the configuration of the recording devices in the real-world location. The filming took place on a well-lit sunny day, which provided ample natural light to enhance clarity and details. The recording equipment was strategically placed within a shadowed area so its own shadow was not visible and could not distract users or detract from immersion. To mitigate microphone noise, we also recorded during periods of minimal wind.

Figure 2. The camera and microphone setup used for capturing 360-degree video and spatial audio, illustrating the equipment configuration and arrangement for creating immersive virtual environments used in Virtual Urban Field Studies.

Unreal Engine, a versatile game development platform, was used to develop the virtual environment for interaction in VR by combining the 360-degree video recordings, ambient audio recordings, and overlaid vehicle sound effects.

3.2. User Study Design

3.2.1. Participant Demographics

We conducted a between-subjects user study of 48 participants, with 12 users in a control group and 36 users in the experimental group. The control group experienced the virtual scenarios with no overlaid vehicle sound effects, divided into two conditions of spatial ($n = 6$) and non-spatial ($n = 6$) ambient audio recordings. The control group was used to assess if the vehicle behaviours were predictable from the visuals alone, which would have confounded the evaluation of the sound's ability to communicate vehicle

intent. The experimental group completed the experience with overlaid vehicle sounds and were similarly divided into two conditions for spatial (*n* = 18) and non-spatial (*n* = 18) audio recordings.

We recruited participants by distributing flyers and advertisements around a university campus, targeting a broad spectrum of potential participants. The demographics of the participants covered a wide age range, from 19 to 53 (M = 26.8, SD = 7.5), thus providing varied perspectives in evaluating the AV–pedestrian interactions. The participants also represented diverse cultural and ethnic backgrounds, although we did not collect specific demographics in this respect. Due to recruitment limitations, our age ranges were skewed towards younger adult participants (<=25: 54%; 26–40: 38%; >=41: 8%), which limits the generalisability of this study to older adults. The majority of participants (over 90%) were familiar with the environment presented in the virtual experience, as they were staff or students at the university.

3.2.2. Study Procedures

Prior to each session, we informed participants of the study aims relating to the sound design of AV–pedestrian interactions, and secured their consent to participate in the study. The experience was deployed on a Vive Pro 2 headset (Manufacturer: HTC, New Taipei, Taiwan), which was adjusted and calibrated for each participant (including volume levels in the integrated headphones to account for individual hearing ability). Participants were given a short introduction on how to interact with the system using the controllers, with specific instructions to stop and remove the headset if they felt any sickness or discomfort during the study. All participants completed the full VR experience (which totalled about 15 min) without reporting any sickness or discomfort.

The experience consisted of nine total scenarios, with three unique variations of each of the three possible AV behaviours. Given that there were nine scenarios, 162 participants would be required for a balanced Latin square design, which was a prohibitively high number. We instead paired spatial- and non-spatial-condition participants and assigned each pair a random order, counterbalancing between conditions but not fully controlling for all ordering effects. While we only sampled a subset of the orderings, the control condition results show that it was not possible to predict vehicle intent from just the video recording, which means that any learning effects could only have come from the sounds.

The participants watched each 360-degree video play for approximately 20–30 s, to understand the context of the scenario as the vehicle drove towards them. The videos were paused before the behaviours of the vehicle became obvious (either through changes in speed or direction). In the experimental condition, the designed AV sounds were played at the appropriate time, synchronised to end just as the video paused. At this point an interface appeared asking the participant what they expected the vehicle behaviour to be: "Change it's path and avoid me"; "Slow down to a stop"; or "Continue and collide with me". The interface then displayed the question "Why did you think the vehicle was going to [chosen behaviour]?", prompting participants to answer aloud. This flow was repeated for each of the nine scenarios. Figure 3 shows an example of what the user would see in the virtual experience.

Both quantitative and qualitative measures were employed to assess user responses and evaluate the study outcomes. Following the scenarios, participants were asked to complete a short questionnaire in VR to evaluate their sense of presence throughout the experience. The Multimodal Presence Scale (MPS) [34] was used based on its validated capacity to measure users' subjective feelings of presence within virtual environments. Evaluating presence is important to validate that our prototype was effectively leveraging the immersive qualities of the virtual environment, eliciting realistic responses from users when interacting with the AV–pedestrian scenarios. Two additional audio-specific items were added, adapted from Witmer and Singer's [27] Presence Questionnaire (PQ), as the MPS lacks any audio-specific questions. Table 2 shows the full list of questions. These

questions aim to specifically investigate the influence of audio on user experience and overall immersion within the virtual environment.
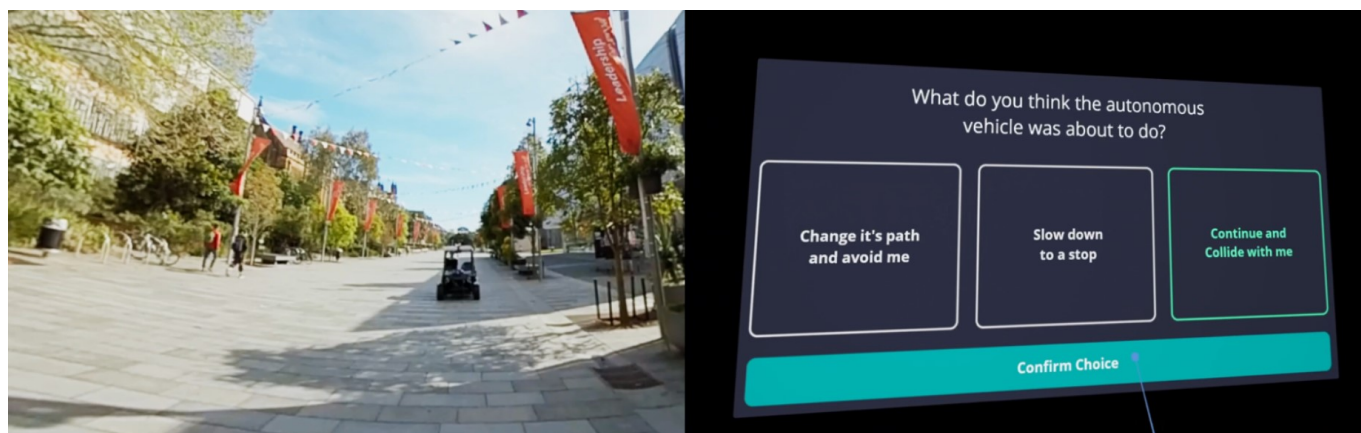
Figure 3. Screenshot of the 360-degree video recording from a user's perspective in virtual reality (**left**). The panel interface enabling a user to select their prediction for the behaviour of the autonomous vehicle (**right**).

Table 2. Statements from the Multimodal Presence Scale [34], organised by subscales for physical, social and self presence. Additional audio-specific items adapted from Witmer and Singer's [27] Presence Questionnaire. These statements were used in a questionnaire to evaluate users' sense of presence throughout the Virtual Urban Field Study.

Item	Source (Subscale)	Statement
1	MPS (Physical)	The virtual environment seemed real to me.
2	MPS (Physical)	I had a sense of acting in the virtual environment, rather than operating something from outside.
3	MPS (Physical)	My experience in the virtual environment seemed consistent with my experiences in the real world.
4	MPS (Physical)	While I was in the virtual environment, I had a sense of being there.
5	MPS (Physical)	I was completely captivated by the virtual world.
6	MPS (Social)	I felt like I was in the presence of another person in the virtual environment.
7	MPS (Social)	I felt that the people in the virtual environment were aware of my presence.
8	MPS (Social)	The people in the virtual environment appeared to be sentient (conscious and alive) to me.
9	MPS (Social)	During the simulation there were times where the computer interface seemed to disappear, and I felt like I was working directly with another person.
10	MPS (Social)	I had a sense that I was interacting with other people in the virtual environment, rather than a computer simulation.
11	MPS (Self)	I felt like my virtual embodiment was an extension of my real body within the virtual environment.
12	MPS (Self)	When something happened to my virtual embodiment, it felt like it was happening to my real body.
13	MPS (Self)	I felt like my real arm was projected into the virtual environment through my virtual embodiment.
14	MPS (Self)	I felt like my real hand was inside of the virtual environment.
15	MPS (Self)	During the simulation, I felt like my virtual embodiment and my real body became one and the same.
16	PQ (Audio [1])	I was able to identify different sounds and distinguish them from one another.
17	PQ (Audio [1])	I was able to localise and determine the position of where sounds were coming from.

[1] These items were in the audio subscale in the original Presence Questionnaire; however this subscale was not utilised in our study.

Following the virtual experience we conducted an open-ended, semi-structured interview with the participants (outside of VR). This began with several questions to gather demographic data, before discussing the responses to the prototype, probing for further information as needed. A reflective approach was taken, showing the participants a screen recording of their experiences and how well they performed. Appendix A includes a list of indicative questions asked during the interview.

3.2.3. Data Analysis Methods

By triangulating both qualitative analysis of the interviews and quantitative analyses of the survey responses we aimed to gain a comprehensive understanding of user experiences within the virtual environments.

The MPS comprises three distinct subscales, each focusing on a specific aspect of presence: physical, social and self-presence. These subscales were used as the basis for our quantitative analysis, comparing differences between the experimental and control sound groups as well as between the spatial and non-spatial conditions. The two audio-specific questions were treated independently without being grouped into any subscale.

Given that survey responses are inherently subjective and may not follow a normal distribution, it was necessary to apply nonparametric statistical tests for evaluating differences between samples. By employing the Mann–Whitney U test, we could assess potential differences in user experiences of presence and audio while accounting for variability in individual perceptions.

The qualitative data from the in-experience questions and the post-experience interview were transcribed and evaluated using thematic analysis [70]. Using the 'storybook' inductive approach allowed for a more flexible and constructionist examination of the data, which aligns well with design contexts where subjective interpretations are crucial to understanding user experiences. Through this approach, we identified patterns and themes emerging from participants' responses, seeking to reveal common trends, preferences, and concerns among users that would not be apparent through quantitative methods alone.

4. Findings

In this section, we present our quantitative results from the in-experience presence questionnaire, followed by our qualitative results from our thematic analysis of open-ended user responses.

As expected, users were not able to differentiate the intended behaviours of the AV in the control (no sound) group. This validated our experimental setup in which participants were asked to predict vehicle behaviour before seeing any indicative movement cues. While the efficacy of the specific vehicle sounds in helping users understand the intended vehicle behaviours is not the focus of this article, the lack of predictive power from the videos alone shows that the audio evaluation component of our methodology was (for lack of a better word) sound.

4.1. Quantitative Results

We conducted an assessment of normality in the quantitative data using the Shapiro–Wilk test due to the small sample sizes in our groups (control = 12; experiment = 36). This revealed that the data were not normally distributed, leading us to use the lower-powered but nonparametric Mann–Whitney U test to identify any differences between the samples.
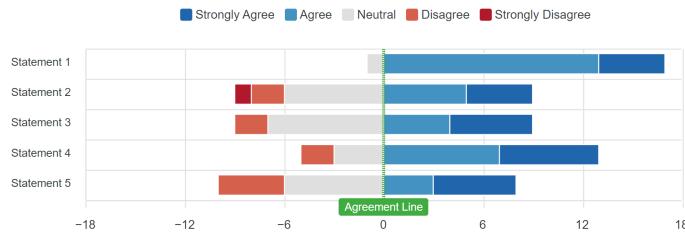
Within the experimental group, presence levels were found to be relatively high, particularly for the physical subscale (see Figure 4). Specifically, for items 1 and 4 of the MPS, the total mean was greater than or equal to 4—indicating a high level of presence. This accords with the literature on 360-degree videos, which suggests that they can induce feelings of presence in VEs. No significant difference was observed between the spatial and non-spatial conditions for any of the MPS items or the additional audio questions (see Table 3 and Figure 5). This indicates that the ambisonic spatialisation of ambient audio and vehicle sounds did not detectably influence presence levels. We were not able
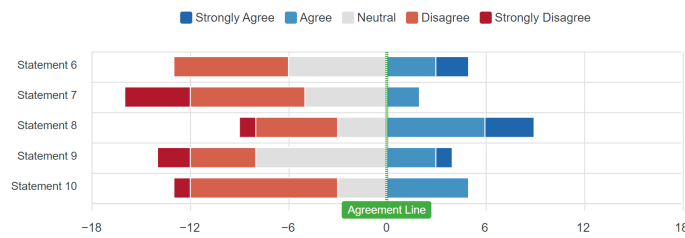
to detect an effect of age on these results (Mann–Whitney U test results, <= 25 vs. >26), with no significant difference between age groups in presence ratings (including audio-specific questions). However, due to the low median age, this test had low statistical power, making it infeasible to test the older group (>40) due to the small number of participants in this range.
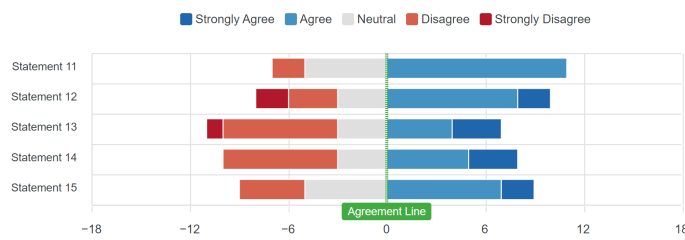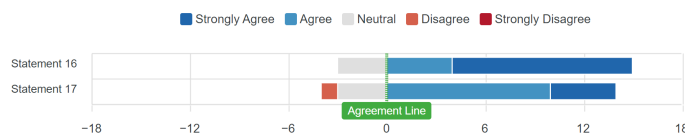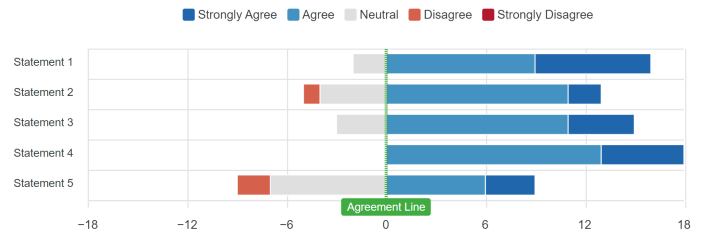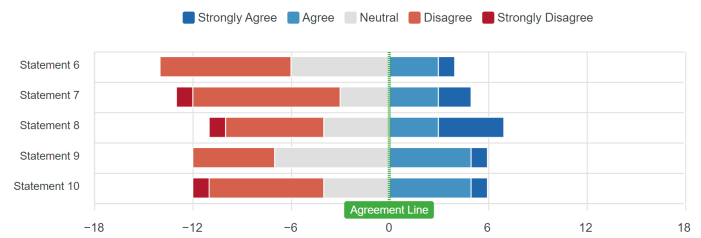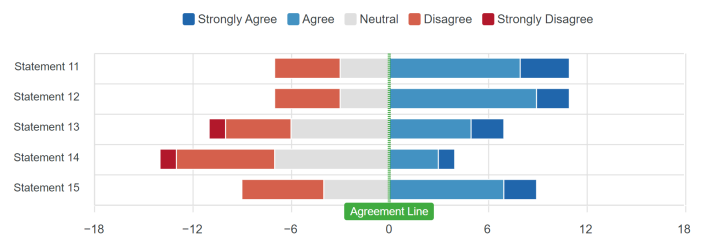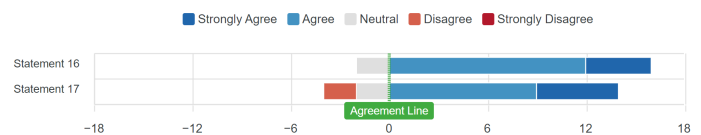

Figure 4. Visualisation of the responses for the Multimodal Presence Scale and additional audio questions, illustrating the levels of agreement among participants. These charts provide insights into the various dimensions of presence experienced by users during the study, highlighting particularly high physical presence.

Table 3. Differences between mean ratings from the Multimodal Presence Scale (MPS) subscales or the additional audio questions. No significant difference was observed between the non-spatial and spatial conditions in the experimental group with overlaid vehicle sounds, suggesting that the spatialisation of ambient audio and vehicle sounds did not impact presence.

Subscale/Item	Non-Spatial	Spatial	p Value	Z Value
Physical Presence	3.76	3.99	0.3232	−0.9878
Social Presence	2.81	2.96	0.6794	−0.4132
Self Presence	3.29	3.28	0.9873	0.0159
Audio Differentiation	4.44	4.11	0.0937	1.6764
Audio Localisation	3.94	3.94	0.8762	−0.1558
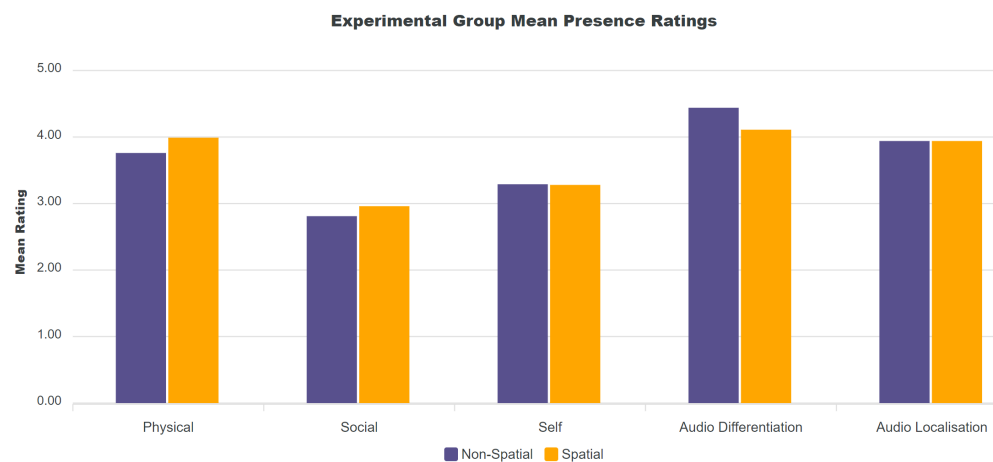
Figure 5. Comparison of mean presence ratings from the Multimodal Presence Scale (MPS) subscales and additional audio questions for the experimental group. This shows no significant difference between the conditions, which indicates that the para-authentic stimuli in the virtual environment did not degrade users' feelings of presence, demonstrating their potential as ecologically valid platforms for Virtual Urban Field Studies.

Comparing the control and the experimental groups, there was no significant difference between the presence ratings from the MPS and additional audio questions (see Table 4). This suggests that the augmentation of the virtual environment with para-authentic virtual vehicle sounds did not detract from the users' feelings of presence, validating our sound design and production choices in integrating them. Subtle increases in scores for certain items may hint at the potential enhancement of presence due to the inclusion of vehicle sounds, although these observations were not statistically significant.

Table 4. Comparison of control and experimental groups, showing no significant difference in mean presence ratings from the Multimodal Presence Scale (MPS) subscales and additional audio questions. This indicates that the para-authentic stimuli in the virtual environment did not degrade users' feelings of presence, demonstrating their potential as ecologically valid platforms for Virtual Urban Field Studies.

Survey Subscale/Item	Control Group			Experimental Group		
	Non-Spatial	Spatial	Total	Non-Spatial	Spatial	Total
Physical Presence	3.43	3.73	3.58	3.76	3.99	3.86
Social Presence	2.60	2.87	2.73	2.81	2.96	2.90
Self Presence	3.03	2.70	2.87	3.29	3.28	3.28
Audio Differentiation	4.00	3.50	3.75	4.44	4.11	4.29
Audio Localisation	3.50	3.67	3.58	3.94	3.94	3.94

4.2. Qualitative Results

The qualitative results from the thematic analysis are presented in Tables 5 and 6, outlining the organising themes, corresponding themes, and codes derived from the participants' responses during the interview.

Table 5. Organising themes and corresponding themes relating to virtual environments and feelings of embodiment in a Virtual Urban Field Study. Illustrative codes and occurrence counts, generated through thematic analysis.

Organising Theme	Sub Theme	Illustrative Codes	Occurrences
Virtual Environments	Realism	Felt like the real world	27
		People were real and alive	16
		People didn't feel real within the environment	15
		Real world elements strengthened sense of presence	13
		People walking around enhanced plausibility	7
		Feels like a video is being played	6
		Environment feels real whilst also feeling simulated	3
	Familiarity	Familiarity with the location	36
		Real life experiences reinforced sense of being there	5
		People enhanced the familliar environment	2
	Interactivity	Lack of social interaction with other people	15
Embodiment	Motor Activation	Lack of interactivity impacted a sense of presence	13
		Not aware you could look around in the space	4
		Lack of kinaesthetic feedback	3
		Unaware if they should use arms or look around	3
		Lack of locomotion	2
	Body Representation	Not seeing hands causing lower sense of embodiment	12
		Arm wasn't projected into the scenarios	5
		Lack of virtual representation of body	4
	Hybrid Awareness	Virtual embodiment was an extension of their body	14
		Conscious of both internal and external worlds	8
		External environment was still perceivable	7
		Feels like they're operating something from the outside	5
		Felt that their real body was in the scene	3
		Accuracy of height increased sense of embodiment	2
		Felt inorganically placed within the scene	2
	Reactions	Felt as if the AV was going to impact with them	18
		Feelings of self-preservation and concern of being hit	4
		Feeling real emotions made experience more real	2

Table 6. Organising themes and corresponding themes relating to perception and response to audio in a Virtual Urban Field Study. Illustrative codes and occurrence counts, generated through thematic analysis.

Organising Theme	Sub Theme	Illustrative Codes	Occurrences
Soundscapes	Ambience	Generally could hear ambient sounds	24
		Background sounds of people were audible	6
		Background sounds made experience immersive	5
		Background sounds felt natural	3
		Background sounds improved awareness of the environment	2
	Spatialisation	Ability to localise sounds improved immersion	9
		Correctly recognised spatialising of audio	6
		Falsely perceived recognition of directional audio	4
		Hard to recognise where sound were coming from	3
		Spatial audio made distinguishing sound sources easier	2
		Non-spatialised sound detracted from immersion	2
Complexity	Focus	Lack of focus on background sounds	22
		Not hearing the consequential car sounds	7
		Background sound was a distraction to task	3
	Confusion	Vehicles with distinct sounds would be easily recognisable	15
		Confusion of AV sound with plane	6
		Sounds from multiple AVs can confuse pedestrians	2
		Sound was too ambiguous	3
		Context influences the interpretation of vehicle sounds	2

Table 6. *Cont.*

Organising Theme	Sub Theme	Illustrative Codes	Occurrences
Responses	Emotion	Rising sound escalated tension/intensity	7
		Loud/sudden sounds that are close can cause fear	5
		Softer sounds are more calming and comfortable	4
	Association	Existing sounds in vehicles	33
		Sounds very reminiscent of video game sounds	5
		Sound associated/reminiscent with particular environment	3
		Sound was reminiscent of horror movies	3
		Reminiscent of emergency vehicle sounds	3

4.2.1. 360-Degree Video VR Environments

The majority of participants across all age-ranges found the 360-degree videos to be highly realistic and experienced them as (unsurprisingly) captures of real-life places and events. This sense of realism was further accentuated by the physical and material elements in the environment, solidifying a sense of scale through the presence of trees and architectural features. The visual phenomena, such as subtle movements in the trees and sunlight reflecting off glass buildings, also contributed to a sense of physical presence for some participants. The qualitative data suggest that this visual realism may explain the high agreement with statement 1 of the MPS across all groups and conditions ("The virtual environment seemed real to me").

A significant number of users expressed their familiarity with the location (which was on our campus), an expected outcome given the sampling of participants from our University. This para-authenticity with a real-world environment seemed to facilitate the feeling of 'being there', in conjunction with the realism from the 360-degree videos. Even the few participants who had not ever been on the university campus described how the environment was typical of an urban shared-use space, which provided them with a sense of familiarity. These feelings may potentially explain the relatively positive ratings for statement 4 of the MPS ("My experience in the virtual environment seemed consistent with my experiences in the real world").

Several users noted the lack of physical and social interaction within the environment, leading to a diminished sense of presence as the experience "felt like it was a video". While acknowledging the environment's realism, these users perceived the experience as a pre-recorded simulation (which—other than the vehicle sounds—it was). This made them feel like they were perceiving the situation as an outsider without any agency to control events. It is likely this partially contributed to the slightly lower ratings for statement 2 of the MPS ("I had a sense of acting in the virtual environment, rather than operating something from outside"). The ratings for statement 5 ("I was completely captivated by the virtual world"), might have also been lowered by the relatively short and repetitive nature of the task scenarios.

As anticipated (since there was no interaction with other humans), the social presence ratings were generally low, which was reinforced through the qualitative data. Participants had mixed opinions about the realism of other pedestrians in the scene, resulting in differing levels of agreement to statement 8 in the MPS ("The people in the virtual environment appeared to be sentient (conscious and alive) to me"). Participants who agreed with that statement reasoned that "having different people in different outfits just going about their day" felt real and that "they were minding their own business and everything feels natural". Those who disagreed likened the other pedestrians to "non-player characters" or "bots", who were not real because they had no awareness of the user and could not respond to their actions.

4.2.2. Embodiment: Perception, Awareness and Representation

Several themes emerged concerning embodiment, which provided insights into users' personal experiences in the virtual environment that remained consistent irrespective of participants' age. One of the most prevalent themes was the lack of motor activation, referring to the limited amount of direct action with the body. This lack of motor activation likely contributed to the low self-presence ratings in the MPS. Factors such as the lack of locomotion and kinaesthetic feedback, as well as the limited need to look around (at least, once the car was spotted) may have contributed to this.

Another prevalent theme was the absence of a visual representation of the body or any of its parts. This likely impacted responses to statements 13 and 14 of the MPS, as users did not see their arms or hands during the virtual scenarios. Additionally, the limited body representation and motor activation may have influenced statement 2, resulting in users perceiving themselves as static observers rather than agents "acting in the virtual environment".

Despite the limitations above, the qualitative data revealed some positive indications of embodiment and self-presence. A theme emerged regarding the hybrid awareness experienced by a few participants, reflecting a dual perception of the virtual and real worlds simultaneously. This sensation was expressed by participants as being "somewhere in the middle" or having "one foot in reality and one foot in virtual reality". The relatively high agreement with statement 11 of the MPS ("I felt like my virtual embodiment was an extension of my real body within the virtual environment") could be attributed to users acknowledging the involvement of their visual sensory apparatus in the virtual world despite the lack of bodily representation. This was summarised by one user who felt that "my body was here [in the real world] but I think my head and my eyes were more in the [virtual] environment".

For some users, this hybrid awareness was primarily characterised by a continued perception of the real world where they were "aware of their outside surroundings" while still engaging immersively with the virtual scenarios. However, for some this hybridity was not as positive for immersion, due to an ongoing awareness they were using VR and operating something from the outside. These kinds of users suggested "it still just did feel like I was doing stuff for the computer" and "I'm in the VR, this is not real, this is a virtual world". It is possible this also contributed to the relatively low scores for statement 2 of the MPS ("I had a sense of acting in the virtual environment, rather than operating something from the outside").

One of the most significant indications of embodiment was the participants' mental and, in some reported cases, physiological reaction to the apparent threat of the approaching vehicle. Even in the control group without vehicle sounds, the visual perception of the car driving directly towards the user elicited feelings of fear and nervousness. While the intensity of these sensations was probably less than what would be experienced in real life, and most participants did not "flinch" or "tense up", most were able to "relate to actually being there" in that moment. This increased their sense of having a body in the virtual space rather than just watching a recording, further enhancing their feelings of presence.

4.2.3. Ambient Soundscapes

Approximately half of the participants mentioned some aspect of the ambient audio in the soundscapes, including bird song, wind, pedestrians talking, planes and distant traffic sounds. Users felt that these sounds were "natural" and "accurate", contributing to their feeling of being "immersed in the actual environment". One user also claimed that this gave them a better sense of "spatial awareness". In particular, the talking and plane sounds appeared to reinforce the familiarity with the environment (as discussed in Section 4.2.1), likely strengthening the feelings of realism and physical presence.

Participants' recollections of the ambient audio were often vague or uncertain, with many referring only to "background noise" or identifying one or two distinct sounds within the overall soundscape. This was true for both the experimental conditions, suggesting the

audio spatialisation did not make the ambient audio any more noticeable. When prompted, a few users did not perceive or remember any background audio at all (see Section 4.2.5).

The qualitative data revealed mixed opinions regarding the spatialisation of the background audio. Some users in the spatial condition accurately perceived the sound as spatialised, specifying that "it all sounded like it was coming from an actual source that I could identify" (e.g., people talking to the left, or a plane flying overhead). In contrast, some of the users in the non-spatial condition falsely perceived localisation of the sounds, despite the audio being flattened to a single channel with no directionality. This occurred with the people talking ("I do remember a comment coming towards more of the right, I believe") and the plane sound ("It sounded like it was coming from above me I guess"). These participants were generally less confident in the way they described perception than those in the spatial condition, even though no difference was visible in the quantitative data. This does not clarify why the non-spatial users perceived the directionality of the sounds, given the similarity in results for statement 16 of the presence questionnaire (see Section 4.1).

Some minor effects of the background audio were mentioned by individual participants. One person mentioned that the background audio distracted them from the task at hand. Another participant felt similarly, suggesting that too noisy an environment would make it harder to determine the vehicle position. No significant relationship was found between participants' age and the responses relating to the ambient soundscapes.

4.2.4. User Responses to Para-Authentic Vehicle Sound Effects

Several themes from our analysis characterised the participants' responses to the designed sound effects for the AV. This is a positive indication of the efficacy of para-authentic virtual sounds embedded within 360° video as an approach, although analysing the comparative efficacy of specific sound designs is out of scope for this article.

One thing that emerged from our analysis was the highly varied emotional response to sounds, reinforcing the subjective nature of individual reactions to auditory stimuli. Negative emotions were associated with certain sounds, such as the fear caused by sudden sounds or the "escalation of tension" resulting from rising pitch. An increase in volume and deeper bass were also linked to the feeling that "something bad was going to happen". Conversely, some participants experienced positive emotional responses to specific sounds. Softer sounds were perceived as more "calm" and "comfortable", while more melodic sounds were described as "passive and non-confrontational". Additionally, one user referred to the high speed of some sounds as "exciting". All of these responses suggest that sounds can prompt different "emotional responses based on [their] intensity".

Participants also drew associations with sounds they had encountered in the past. The most common comparisons were with existing vehicle sounds, such as engine noises, horns and sirens in emergency vehicles, as well as warning beeps for systems such as reversing sensors or pedestrian detection. These familiar sounds seemed to provide appropriate mental models for participants, increasing familiarity and potentially supporting presence. Older participants (>26) exclusively referenced existing vehicle sounds, while younger participants (<26) also made reference to media, such as sounds from films and video games. Some of these participants suggested that the higher pitch sounds reminded them of the "stabbing"-style soundtracks used in classic horror movies, contributing to the fear response mentioned previously. Lastly, a few participants associated the sounds with non-vehicular warning systems, such as fire alarms.

4.2.5. Sound Complexity and Ambiguity

Another organising theme highlighted the intricate nature of sounds and the challenges that a dynamic real-world setting can pose to communicating clear meaning. Some participants experienced confusion regarding the source and characteristics of certain sounds. For example, some people were unable to distinguish between the para-authentic virtual vehicle sounds and the noise of a plane passing overhead. In a more traditional HCI

user study context, it might be prudent to remove incidental interruptive noises such as aeroplanes from experiments; the challenge of VUFS prototyping is that such infrequent-but-loud noises are part of the urban soundscape. While the plane and the car sounds did conflict, the possibility of them doing so both shows the realism and fidelity of our overlaid sounds and suggests a challenge for AV sound designers to address. Despite potential differences arising from age (e.g., hearing ability), responses about confusion resulting from ambiguity did not appear to be influenced by these variations.

5. Discussion

Here we discuss the two key contributions of this study: the use of presence while developing VUFS, and the use of VUFS more broadly as a way to evaluate urban interaction designs.

5.1. Presence as a Proxy for Verisimilitude in VUFS

The first objective of this research was to explore the efficacy and characteristics of VUFS (particularly with a sound design focus) as a way to approximate field studies. We found that the notion of presence serves as a valuable indicator of verisimilitude—by which we mean the capacity of the virtual environment to stand in for a real field study—for VUFS. When conducting virtual field studies, it is essential not only to evaluate the effectiveness of a prototype in achieving its intended goals (design efficacy), but also to measure how well the prototype actually reflects the intended environment (VUFS efficacy/verisimilitude). In some HCI studies that utilise virtual environments, presence is clearly desirable but neither measured nor explicitly designed for. By making presence an explicit goal of our iterative design process, we aimed to maximise the presence dimensions that were relevant for our specific context-based interface prototype.

The significance of maximising the various dimensions of presence depends on the specific application and context in which it is employed. For instance, architectural contexts might necessitate a high level of physical presence, ensuring that users have an authentic perception of factors such as space and scale. By contrast, social robotics applications would likely benefit from enhanced social presence, fostering a more realistic relationship between the user and the robot. Wearable technology might rely on high self-presence, where the user experiences a strong sense of embodiment and connection with the device. In urban settings, particularly when designing technologies for automated cities, there may be good reasons why all three dimensions of presence become relevant.

The sense of physical presence in our study was primarily achieved through realistic recordings and familiarity with the environment contributing to both a "place illusion" and "plausibility illusion" [26]. The quantitative results indicate that the fidelity of the virtual environment was sufficient to create a sense of physical presence for many users. Although some users mentioned the blurriness of the video due to resolution limitations, it did not seem to strongly impact the realism or sense of 'being there'. Participants not only felt they were in a real place, but also believed that the scenarios could genuinely have occurred. This plausibility was predominantly evidenced in the qualitative data, where users perceived realism through audiovisual recording fidelity and comparisons to their past experiences.

Participants who had frequently visited the campus exhibited a specific familiarity with the space used in recordings, while the few that had never visited still reported a general familiarity with similar spaces (i.e., shared pedestrian/vehicle urban areas). Due to the limitations of our recruitment methods, most of our participants were well-acquainted with this environment. Despite this high level of familiarity, our users still found the experience realistic and plausible, which is a good indication of the validity of the method (as this group would be more likely to find potentially presence-breaking inconsistencies in the virtual experience). Given this limitation, our findings are most applicable to the context of on-campus AVs (such as for mobility), rather than more general AV use-cases. Designers of future VUFS should consider the levels of both specific and general familiarity which are

appropriate for their use cases. For example, in a study on augmented-reality-facilitated navigation in urban environments, general (but not specific) familiarity may be desirable, to prevent effects on task efficacy.

The lack of interaction with the environment may have slightly detracted from physical presence. However, this did not appear to negatively impact the study, as users were still primarily focused on the vehicle and were able to engage with the scenarios presented. Design contexts which involve more movement or physical action must carefully consider these interactions to successfully leverage context-based interface prototypes. In these cases, it may be better to leverage alternative approaches to creating VEs (such as 3D computer generated worlds), for more complex interactivity.

Despite the lack of social interactions, other pedestrians in the scenarios contributed to the plausibility, likely somewhat enhancing social presence. Limited embodiment resulting from minimal physical engagement and the absence of a visual representation of the body may have impacted self-presence in relation to motor function. Nevertheless, there was still some level of self-presence at a sensory and cognitive level, further validating the use of this prototyping method for conducting studies of user perceptions.

5.2. Virtual Urban Field Studies for Evaluating Interaction Design

Our study supports the notion that VR is a highly suitable medium for simulating field studies testing interaction design, supporting what has been suggested before in similar studies [3]. This is particularly relevant for situations that could be considered dangerous, such as our case of an AV driving towards a pedestrian with the possibility of collision.

The quantitative data showed there was no significant deterioration in feelings of presence when para-authentic virtual sounds were used to augment the real ambient soundscape recordings. There was actually a small (although not significant) increase in all subscale ratings as well as most individual items of the MPS in the experimental group when compared to the control, potentially indicating increased presence due to the layered sounds. This could partly be due to emotional responses to the sound (discussed below and further in Section 5.3) making people feel more present. Perhaps users suspended their disbelief upon hearing sounds from the car, further immersing themselves in the scenarios without critically examining the unreal elements. We believe this is an area that warrants further investigation in context-based interface prototypes, including other sensory contexts beyond sound—what is the impact on presence of elements in the environment seeming to "respond to" or "address" the participant?

When the para-authentic virtual vehicle sounds were overlaid on the recorded ambient soundscape there was no significant difference in presence between the spatial and non-spatial conditions. This is actually a positive, as it suggests that for these kinds of experiments, a simple level of spatialisation is enough to generate presence and appropriate task outcomes. Ambisonic recording hardware and software is expensive, and requires specialised skillsets to operate. For VUFS developers not to have to engage with it is a benefit, leaving them free to focus on the design of the sounds themselves, rather than on sound production.

As suggested by Jain et al. [68], "designers make creative decisions for sounds in VR to be more pleasant and manageable". In our design process, we iterated the vehicle sounds multiple times to test different characteristics and how they interacted with the virtual environment. This resulted in tweaking qualities such as pitch, tempo, timbre and melody. We tuned features such as positionality and reverb for simplicity not realism (mostly to avoid having to take detailed acoustic scans of the environment). For example, the attenuation of sound in the space was not physically realistic, but was apparently good enough to situate the sound well enough that users perceived no difference in their sense of "being there", even in the audio-specific questions.

Background ambient soundscapes played a crucial role in the overall cohesiveness of the virtual environment. As one participant stated, "it's easier to report about an experience that feels off than about an experience that is natural", suggesting they would have noticed if the sounds were not appropriate. This aligns with prior research [71], demonstrating that including a plausible soundscape has a high impact on presence and perceptions of realism. For the ambient audio, the level of spatialisation appeared not to significantly impact feelings of presence. In fact, the layering of para-authentic virtual sounds caused users to focus less on or even completely ignore the background audio. It is possible a lower-order (thus cheaper and more convenient to use) microphone could have been utilised, without sacrificing the presence-inducing effects of having an appropriate soundscape.

5.3. Generating Authentic User Responses with Para-Authentic Stimuli

The second objective of this research was to investigate the impact of overlaying virtual sounds on recorded imagery on user experience in VUFS. Our qualitative results suggest that the characteristics of the para-authentic vehicle sounds did influence users' responses to the virtual stimuli. For example, some participants claimed to feel fear and nervousness in response to the intense crescendos with rising pitch and oscillation. This anticipatory fear reaction is consistent with the literature, which states that "individuals make behavioural, cognitive, and physiological adjustments that facilitate the process of coping with the upcoming stressor" [72]. The users did not get to see the impact of the vehicle before making their decisions as the video was paused before the outcome was shown, yet these sensations still arose. That VR could generate these anticipatory stress responses, even in the absence of any real danger, supports the efficacy of VUFS for these kinds of experiments. Other users reported feeling a sense of calmness from the more melodic audio patterns with steady rhythms. Both of these responses indicate the ecological validity of our VEs, where the users are affectively engaging with the sensory stimuli as if they were real. Our results reinforce that sounds can influence cognitive processing and potential emotional responses by leveraging associations. These associations can be both domain-specific (e.g., similar sounds in vehicles) or draw on outside references (e.g., game sound effects). Our qualitative data reinforce that our users interpreted our sound designs to impart meaning and (potentially) emotional responses, often drawing on past experiences (e.g., horn-like sounds captured attention and then induced caution/alertness) to do so.

When designing a VUFS prototype to maximise physical presence, sounds intended to generate a response should adhere to certain guidelines regarding realism. They should at least be somewhat spatialised to situate them within the environment, with some directionality to enhance plausibility. In some cases, higher-order ambisonics may offer additional advantages that offset the increased effort, specialised equipment, and expert knowledge required. The sounds should support existing mental models, as overly fantastical and inappropriate sounds may detract from the plausibility. Our study focused on notification-style communication in small electric AVs, as opposed to simulated motor sounds (see [62]). Both kinds of sounds require further research to explore potential sound signatures for different vehicles (considering factors such as vehicle size, velocity and pedestrian proximity).

Ensuring a plausible and appropriate multisensory environmental context is crucial for inducing authentic user behaviours within virtual settings. For instance, participants may have perceived our car sounds very differently if they did not originate from an approaching vehicle, and instead were presented without video or background audio. Incorporating realistic auditory stimuli within virtual environments effectively elicits accurate user responses and enhances the overall immersive experience.

5.4. Considerations for Designing Immersion-Supporting Sounds

While ambient soundscapes are often designed to be unnoticeable (or at least unobtrusive), there are instances where these sounds may attract the user's attention and further ground their experience. While the background sounds in our study were recorded, it could benefit some VUFS to utilise composed soundscapes. As we discovered in our study, care should be taken to not distract the user from the task at hand, perhaps by synchronising the immersion-supporting background sounds with specific lulls in the scenarios. This could help to sustain immersion in both the physical and fictional elements of a virtual environment, potentially preventing users from becoming distracted when not actively engaged in a task. These sounds should align with the visuals and be appropriately calibrated for the scene (i.e., if a vehicle noise is heard, the vehicle should be visible somewhere).

When designing interface sounds, designers should also consider possible confusion of sound sources and characteristics. In our study, one of the scenarios contained the audio of a plane flying overhead, which coincided with the layered para-authentic vehicle sound. Although we considered the potential of overlapping sound frequencies in our design process, we did not account for the general similarity of our sound with other environmental sounds in the environment (such as a gradual decrease in pitch indicating the AV slowing down). This led to an interesting situation in which participants confused and sometimes conflated the vehicle sound with the plane sound. This reinforces the notion of utilising sounds with prior associations to leverage users' potential existing mental models (see Section 4.2.4). However, it is essential to design these sounds in a way that minimises any negative interaction between audio interfaces and background sounds. Striking a balance between realism and experiment design is essential in designing effective auditory stimuli for VUFS in particular. Retaining realistic elements, such as the sound of a plane, contributes to the immersive experience, but may also increase the scope of design issues that might be identifiable. For example, if designing sounds for vision-impaired users at bus stops, it would be valuable to know if those sounds might be confused with a bus engine. The same would be true if LED lights on a vehicle interface were confused with reflections from other light sources. Ultimately, our opinion is that VUFS verisimilitude should win out, as the rigour of controlled experiments conducted as part of the evaluation of a design prototype should be secondary to the design quality of that prototype. That said, one should be just as careful making claims about a VUFS experiment that lacked certain environmental controls as a real-world experiment with the same issue.

6. Conclusions

In this study, we explored the potential of virtual reality for testing interactive systems, specifically focusing on the context of AV–pedestrian interaction in shared spaces. We proposed Virtual Urban Field Studies (VUFS) as the label for studies using context-based interface prototypes to approximate the context of urban interaction. Engaging in research-through-design, we developed a prototype for AV auditory interfaces, which we evaluated with a user study. Our prototype employed 360-degree video and spatial audio recordings to create a virtual environment (VE), then augmented it by overlaying para-authentic vehicle interface sounds.

We also examined the role of presence in determining the verisimilitude and real-world applicability of the simulated urban environments. By investigating how the dimensions of presence shaped user experiences in VUFS, we identified ways that the immersive affordances (both perceptual and psychological) of VR can be leveraged to provide an ecologically valid testing environment. This was quantitatively validated by our data showing that the addition of para-authentic vehicle sounds led to no deterioration in presence.

The results of our study indicated a high level of physical presence in our virtual experience, partly due to the realism of recorded 360-degree video and spatial audio soundscapes. This was represented in the quantitative data (see Section 4.1) as well as the qualitative data (see Section 4.2.1). This was true whether the participant experienced the "spatial" audio or the simple stereo version, suggesting that expensive and specialised

ambisonic technology is not necessary for VUFS development, at least not when sound localisation is not the focus. We did not find any effects of age on the efficacy of VUFS, but we acknowledge the limited power of our study due to a lack of older users (only 8% were over 40). Additionally, our recruitment from a university setting might have introduced a bias towards higher levels of education and socioeconomic status, thus more exposure to new technologies such as VR. Further studies would be required to assess the viability and effectiveness of the VUFS approach across a more representative sample of urban inhabitants.

Physical presence was also bolstered by the experience of a "place illusion" (resulting from familiar and authentic locations) and "plausibility illusion" (resulting from scenarios that could genuinely occur). Participants reported feelings of social and self-presence, just not as high as physical presence. While there was no direct interaction with other social actors in the experience, having other pedestrians in the recordings contributed to the plausibility of the scenarios. The moderate ratings for self-presence suggest that minimal embodiment can be experienced from just experiencing 360-degree sensory input, even without bodily representation in the VE.

Future research should explore the use of VUFS to explore different applications for urban technology. Use-cases such as pedestrian interaction with AVs are particularly well suited, given the risk and costs associated with the vehicles. However, we believe this method can be more generally applied for the design and evaluation of many other urban interactions and interfaces.

We have discussed the advantages and potential disadvantages of employing VEs, using approaches such as 360-degree video and spatial audio. Based on our findings, we have offered design considerations for operationalising presence for the purpose of VUFS. We hope this method can reduce the time, cost and development effort associated with traditional field studies. By harnessing the power of VEs and measuring presence as a proxy for verisimilitude, designers and researchers can create more engaging and authentic experiences, advance urban HCI, and develop innovative solutions for real-world urban challenges.

Abbreviations

The following abbreviations are used in this manuscript:

AV	Autonomous Vehicles
eHMIs	External Human–Machine Interfaces
HCI	Human–Computer Interaction
VE	Virtual Environment
VR	Virtual Reality
VFS	Virtual Field Studies
VUFS	Virtual Urban Field Studies

Appendix A

- What do you remember focusing on during this part of the experience?
- Comparing that scenario to the last scenario, how do you think they were similar or different?
- Why did you agree/disagree with that statement from the questionnaire?
- What were you hearing during the experience?
- What kind of sounds were you able to distinguish?
- Were there any sounds that stand out in your mind? Either on the positive or the negative side?
- Was there anything about the sound that either contributed to immersion or your relationship with the vehicle?
- What kind of cues were you getting that helped you localise the sound?
- Was there anything that felt consistent/inconsistent with your real-world past experiences?
- Can we dive further into what about this experience made it feel real to you?
- What were you thinking about the other people around in the environment?
- How do you think that would compare if it was a more stylised rendered 3D representation with virtual characters?
- How did you feel about just generally physically being inside that experience?
- Which parts of your body felt involved inside that experience?
- What kind of things were you thinking were happening to your body?

References

1. Duh, H.B.L.; Tan, G.C.B.; Chen, V.H.H. Usability evaluation for mobile device: A comparison of laboratory and field tests. In Proceedings of the 8th Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '06), Helsinki, Finland, 12–15 September 2006; Association for Computing Machinery: New York, NY, USA, 2006; pp. 181–186.
2. Weir, D.H. Application of a driving simulator to the development of in-vehicle human–machine-interfaces. *IATSS Res.* **2010**, *34*, 16–21. [CrossRef]
3. Mäkelä, V.; Radiah, R.; Alsherif, S.; Khamis, M.; Xiao, C.; Borchert, L.; Schmidt, A.; Alt, F. Virtual Field Studies: Conducting Studies on Public Displays in Virtual Reality. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20), Honolulu, HI, USA, 25–30 April 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 1–15.
4. Hoggenmüller, M.; Tomitsch, M.; Hespanhol, L.; Tran, T.T.M.; Worrall, S.; Nebot, E. Context-Based Interface Prototyping: Understanding the Effect of Prototype Representation on User Feedback. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan, 8–13 May 2021; Association for Computing Machinery: New York, NY, USA, 2021; Number Article 370 in CHI '21, pp. 1–14.
5. Hoggenmueller, M.; Tomitsch, M.; Worrall, S. Designing Interactions with Shared AVs in Complex Urban Mobility Scenarios. *Front. Comput. Sci.* **2022**, *4*, 866258. [CrossRef]
6. Hale, K.S.; Stanney, K.M. (Eds.) *Handbook of Virtual Environments: Design, Implementation, and Applications*, 2nd ed.; CRC Press: Boca Raton, FL, USA, 2014.
7. McMahan, A. Immersion, engagement, and presence: A method for analyzing 3-D video games. In *The Video Game Theory Reader*; Routledge: Abingdon, UK, 2013; pp. 67–86.
8. Nilsson, N.C.; Nordahl, R.; Serafin, S. Immersion Revisited: A review of existing definitions of immersion and their relation to different theories of presence. *Hum. Technol.* **2016**, *12*, 108–134. [CrossRef]
9. Jones, P.; Osborne, T.; Sullivan-Drage, C.; Keen, N.; Gadsby, E. Creating 360° imagery. In *Virtual Reality Methods*, 1st ed.; A Guide for Researchers in the Social Sciences and Humanities; Bristol University Press: Bristol, UK, 2022; pp. 98–116.

10. Škola, F.; Rizvić, S.; Cozza, M.; Barbieri, L.; Bruno, F.; Skarlatos, D.; Liarokapis, F. Virtual Reality with 360-Video Storytelling in Cultural Heritage: Study of Presence, Engagement, and Immersion. *Sensors* **2020**, *20*, 5851. [CrossRef] [PubMed]

11. Blair, C.; Walsh, C.; Best, P. Immersive 360° videos in health and social care education: A scoping review. *BMC Med. Educ.* **2021**, *21*, 590. [CrossRef]

12. Waller, M.; Mistry, D.; Jetly, R.; Frewen, P. Meditating in Virtual Reality 3: 360° Video of Perceptual Presence of Instructor. *Mindfulness* **2021**, *12*, 1424–1437. [CrossRef]

13. Patrão, B.; Pedro, S.L.; Menezes, P. How to Deal with Motion Sickness in Virtual Reality. In Proceedings of the Encontro Português de Computação Gráfica e Interação (EPCGI2015), Coimbra, Portugal, 12–13 November 2015.

14. Poeschl-Guenther, S.; Wall, K.; Döring, N. Integration of spatial sound in immersive virtual environments an experimental study on effects of spatial. In Proceedings of the Virtual Reality (VR), Lake Buena Vista, FL, USA, 18–20 March 2013; pp. 129–130.

15. Potter, T.; Cvetković, Z.; De Sena, E. On the Relative Importance of Visual and Spatial Audio Rendering on VR Immersion. *Front. Signal Process.* **2022**, *2*, 904866. [CrossRef]

16. Fellgett, P.B. Ambisonic reproduction of directionality in surround-sound systems. *Nature* **1974**, *252*, 534–538. [CrossRef]

17. Zotter, F.; Frank, M. *Ambisonics*; Springer International Publishing: Cham, Switzerland, 2019.

18. Bertet, S.; Daniel, J.; Moreau, S. 3D sound field recording with higher order ambisonics—Objective measurements and validation of a 4th order spherical microphone. In Proceedings of the 120th AES Convention, Paris, France, 20–23 May 2006.

19. Hong, J.Y.; He, J.; Lam, B.; Gupta, R.; Gan, W.S. Spatial Audio for Soundscape Design: Recording and Reproduction. *Appl. Sci.* **2017**, *7*, 627. [CrossRef]

20. Steuer, J. Defining Virtual Reality: Dimensions Determining Telepresence. *J. Commun.* **1992**, *42*, 73–93. [CrossRef]

21. Slater, M.; Usoh, M. Representations systems, perceptual position, and presence in immersive virtual environments. *Presence Teleoperators Virtual Environ.* **1993**, *2*, 221–233. [CrossRef]

22. Biocca, F. The Cyborg's Dilemma: Progressive Embodiment in Virtual Environments. *J. Comput. Mediat. Commun.* **1997**, *3*, JCMC324. [CrossRef]

23. Lombard, M.; Ditton, T. At the Heart of It All: The Concept of Presence. *J. Comput. Mediat. Commun.* **1997**, *3*, JCMC321. [CrossRef]

24. Ijsselsteijn, W.A.; de Ridder, H.; Freeman, J.; Avons, S.E. Presence: Concept, determinants and measurement. *Proc. SPIE Int. Soc. Opt. Eng.* **2000**, *3959*, 520–529.

25. Riva, G.; Waterworth, J.A.; Waterworth, E.L. The layers of presence: A bio-cultural approach to understanding presence in natural and mediated environments. *Cyberpsychol. Behav.* **2004**, *7*, 402–416. [CrossRef]

26. Slater, M. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2009**, *364*, 3549–3557. [CrossRef]

27. Witmer, B.G.; Singer, M.J. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence Teleoperators Virtual Environ.* **1998**, *7*, 225–240. [CrossRef]

28. Lee, K.M. Presence, Explicated. *Commun. Theory* **2004**, *14*, 27–50. [CrossRef]

29. Gilbert, S.B. Perceived Realism of Virtual Environments Depends on Authenticity. *Presence Teleoperators Virtual Environ.* **2016**, *25*, 322–324. [CrossRef]

30. Felton, W.M.; Jackson, R.E. Presence: A Review. *Int. J. Hum.-Comput. Interact.* **2022**, *38*, 1–18. [CrossRef]

31. Insko, B.E. Measuring presence: Subjective, behavioral and physiological methods. In *Being There: Concepts, Effects and Measurements of User Presence in Synthetic Environments*; American Psychological Association: Washington, DC, USA, 2003; Volume 321, pp. 109–119.

32. Putze, S.; Alexandrovsky, D.; Putze, F.; Höffner, S.; Smeddinck, J.D.; Malaka, R. Breaking The Experience: Effects of Questionnaires in VR User Studies. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 1–15.

33. Choi, B.C.K.; Pak, A.W.P. A catalog of biases in questionnaires. *Prev. Chronic Dis.* **2005**, *2*, A13.

34. Makransky, G.; Lilleholt, L.; Aaby, A. Development and Validation of the Multimodal Presence Scale for Virtual Reality Environments: A Confirmatory Factor Analysis and Item Response Theory Approach. *Comput. Human Behav.* **2017**, *72*, 276–285. [CrossRef]

35. Souza, V.; Maciel, A.; Nedel, L.; Kopper, R. Measuring Presence in Virtual Environments: A Survey. *ACM Comput. Surv.* **2021**, *54*, 1–37. [CrossRef]

36. Grassini, S.; Laumann, K. Questionnaire Measures and Physiological Correlates of Presence: A Systematic Review. *Front. Psychol.* **2020**, *11*, 349. [CrossRef] [PubMed]

37. Buxton, B. *Sketching User Experience: Getting the Design Right and the Right Design*; Morgan Kaufmann: San Francisco, CA, USA, 2007.

38. Lim, Y.K.; Stolterman, E.; Tenenberg, J. The anatomy of prototypes. *ACM Trans. Comput. Hum. Interact.* **2008**, *15*, 1–27. [CrossRef]

39. Buchenau, M.; Suri, J.F. Experience Prototyping. In Proceedings of the Conference on Designing Interactive Systems, New York, NY, USA, 17–19 August 2000; pp. 424–433.

40. Flohr, L.A.; Wallach, D.P. The Value of Context-Based Interface Prototyping for the Autonomous Vehicle Domain: A Method Overview. *Multimodal Technol. Interact.* **2022**, *7*, 4. [CrossRef]

41. Chamberlain, A.; Crabtree, A.; Rodden, T.; Rogers, Y. Research in the wild: Understanding 'in the wild' approaches to design and development. In Proceedings of the Designing Interactive Systems Conference, Newcastle upon Tyne, UK, 11–15 June 2012; pp. 795–796.

42. Paulos, E.; Jenkins, T. Urban probes: Encountering our emerging urban atmospheres. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Portland, OR, USA, 2–7 April 2005; ACM: New York, NY, USA, 2005.

43. Hoggenmüller, M.; Wiethoff, A. LightSet: Enabling urban prototyping of interactive media façades. In Proceedings of the 2014 Conference on Designing Interactive Systems (DIS '14), Vancouver, BC, Canada, 21–25 June 2014; Association for Computing Machinery: New York, NY, USA, 2014; pp. 925–934.

44. Nguyen, T.T.; Holländer, K.; Hoggenmueller, M.; Parker, C.; Tomitsch, M. Designing for projection-based communication between autonomous vehicles and pedestrians. In Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Utrecht, The Netherlands, 21–25 September 2019; ACM: New York, NY, USA, 2019.

45. Sutcliffe, A.; de Bruijn, O.; Gault, B.; Fernando, T.; Tan, K. Comparing Interaction in the Real World and CAVE Virtual Environments. In *Proceedings of the People and Computers XVIII—Design for Life*; Springer: London, UK, 2005; pp. 347–361.

46. Sutcliffe, A.G.; Poullis, C.; Gregoriades, A.; Katsouri, I.; Herakleous, K. Reflecting on the Design Process for Virtual Reality Applications. *Int. J. Hum.-Comput. Interact.* **2018**, *35*, 168–179. [CrossRef]

47. Flohr, L.A.; Janetzko, D.; Wallach, D.P.; Scholz, S.C.; Krüger, A. Context-Based Interface Prototyping and Evaluation for (Shared) Autonomous Vehicles Using a Lightweight Immersive Video-Based Simulator. In Proceedings of the 2020 ACM Designing Interactive Systems Conference (DIS '20), Eindhoven, The Netherlands, 6–10 July 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 1379–1390. [CrossRef]

48. Tran, T.T.M.; Parker, C.; Hoggenmüller, M.; Hespanhol, L.; Tomitsch, M. Simulating Wearable Urban Augmented Reality Experiences in VR: Lessons Learnt from Designing Two Future Urban Interfaces. *Multimodal Technol. Interact.* **2023**, *7*, 21. [CrossRef]

49. Voit, A.; Mayer, S.; Schwind, V.; Henze, N. Online, VR, AR, Lab, and In-Situ: Comparison of Research Methods to Evaluate Smart Artifacts. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19), Glasgow, UK, 4–9 May 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 1–12. [CrossRef]

50. Yeo, D.; Kim, G.; Kim, S. Toward Immersive Self-Driving Simulations: Reports from a User Study across Six Platforms. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20), Honolulu, HI, USA, 25–30 April 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 1–12.

51. Kaur, K.; Rampersad, G. Trust in driverless cars: Investigating key factors influencing the adoption of driverless cars. *J. Eng. Technol. Manag.* **2018**, *48*, 87–96. [CrossRef]

52. Dey, D.; Habibovic, A.; Löcken, A.; Wintersberger, P.; Pfleging, B.; Riener, A.; Martens, M.; Terken, J. Taming the eHMI jungle: A classification taxonomy to guide, compare, and assess the design principles of automated vehicles' external human-machine interfaces. *Transp. Res. Interdiscip. Perspect.* **2020**, *7*, 100174. [CrossRef]

53. Holländer, K.; Colley, A.; Mai, C.; Häkkilä, J.; Alt, F.; Pfleging, B. 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 (MobileHCI '19), Taipei, Taiwan, 1–4 October 2019; Association for Computing Machinery: New York, NY, USA, 2019. [CrossRef]

54. Dey, D.; Habibovic, A.; Pfleging, B.; Martens, M.; Terken, J. Color and Animation Preferences for a Light Band EHMI in Interactions Between Automated Vehicles and Pedestrians. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI'20), Honolulu, HI, USA, 25–30 April 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 1–13. [CrossRef]

55. Faas, S.M.; Kao, A.C.; Baumann, M. A Longitudinal Video Study on Communicating Status and Intent for Self-Driving Vehicle – Pedestrian Interaction. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI'20), Honolulu, HI, USA, 25–30 April 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 1–14.

56. Faas, S.M.; Kraus, J.; Schoenhals, A.; Baumann, M. Calibrating Pedestrians' Trust in Automated Vehicles: Does an Intent Display in an External HMI Support Trust Calibration and Safe Crossing Behavior? In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan, 8–13 May 2021; Association for Computing Machinery: New York, NY, USA, 2021; Number Article 157 in CHI '21, pp. 1–17.

57. Holländer, K. A pedestrian perspective on autonomous vehicles. In Proceedings of the 24th International Conference on Intelligent User Interfaces: Companion (IUI '19), Marina del Ray, CA, USA, 16–20 March 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 149–150.

58. Risto, M.; Emmenegger, C.; Vinkhuyzen, E.; Hollan, J. Human-Vehicle Interfaces: The Power of Vehicle Movement Gestures in Human Road User Coordination. In Proceedings of the Driving Assessment Conference, Manchester Village, VT, USA, 26–29 June 2017; pp. 186–192.

59. Dey, D.; Terken, J. Pedestrian Interaction with Vehicles: Roles of Explicit and Implicit Communication. In Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '17), Oldenburg, Germany, 24–27 September 2017; Association for Computing Machinery: New York, NY, USA, 2017; pp. 109–113.

60. Moore, D.; Currano, R.; Strack, G.E.; Sirkin, D. The Case for Implicit External Human-Machine Interfaces for Autonomous Vehicles. In Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '19), Utrecht, The Netherlands, 21–25 September 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 295–307.

61. Pelikan, H.R.M.; Jung, M.F. 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), Stockholm, Sweden, 13–16 March 2023; Association for Computing Machinery: New York, NY, USA, 2023; pp. 172–182.

62. Moore, D.; Currano, R.; Sirkin, D. Sound Decisions: How Synthetic Motor Sounds Improve Autonomous Vehicle-Pedestrian Interactions. In Proceedings of the 12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '20), Virtual Event, 21–22 September 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 94–103.

63. Zimmerman, J.; Forlizzi, J.; Evenson, S. Research through Design As a Method for Interaction Design Research in HCI. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07), San Jose, CA, USA, 28 April–3 May 2007; ACM: New York, NY, USA, 2007; pp. 493–502.

64. Keyson, D.V.; Alonso, M.B. Empirical Research through Design. In Proceedings of the 3rd IASDR Conference on Design Research, Seoul, Republic of Korea, 18–22 October 2009; pp. 4548–4557.

65. Wang, Y.; Hespanhol, L.; Worrall, S.; Tomitsch, M. Pedestrian-Vehicle Interaction in Shared Space: Insights for Autonomous Vehicles. In Proceedings of the AutomotiveUI '22: 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Seoul, Republic of Korea, 17–20 September 2022; pp. 330–339.

66. Olaverri, C. The ACFR Centre: ITS Group [ITS Research Lab]. *IEEE Intell. Transp. Syst. Mag.* **2019**, *11*, 235–240. [CrossRef]

67. Johansen, S.S.; van Berkel, N.; Fritsch, J. Characterising Soundscape Research in Human-Computer Interaction. In Proceedings of the Designing Interactive Systems Conference (DIS '22), Virtual Event, 13–17 June 2022; Association for Computing Machinery: New York, NY, USA, 2022; pp. 1394–1417.

68. Jain, D.; Junuzovic, S.; Ofek, E.; Sinclair, M.; Porter, J.; Yoon, C.; Machanavajhala, S.; Ringel Morris, M. A Taxonomy of Sounds in Virtual Reality. In Proceedings of the Designing Interactive Systems Conference 2021 (DIS '21), Virtual Event, 28 June–2 July 2021; Association for Computing Machinery: New York, NY, USA, 2021; pp. 160–170.

69. Murphy, D.T.; Stevens, F.K.M. Spatial impulse response measurement in an urban environment. In Proceedings of the Spatial Audio, AES 55th International Conference, Helsinki, Finland, 27–29 August 2014; p. 8.

70. Braun, V.; Clarke, V. Using thematic analysis in psychology. *Qual. Res. Psychol.* **2006**, *3*, 77–101. [CrossRef]

71. Kern, A.C.; Ellermeier, W. Audio in VR: Effects of a Soundscape and Movement-Triggered Step Sounds on Presence. *Front Robot AI* **2020**, *7*, 20. [CrossRef] [PubMed]

72. Pulopulos, M.M.; Baeken, C.; De Raedt, R. Cortisol response to stress: The role of expectancy and anticipatory stress regulation. *Horm. Behav.* **2020**, *117*, 104587. [CrossRef] [PubMed]