X-Road: Virtual Reality Glasses for Orientation and Mobility Training of People with Visual Impairments

LAUREN THEVIN, Inria Bordeaux & LMU Munich
CARINE BRIANT, IRSA
ANKE M. BROCK, ENAC, Université Toulouse

Orientation and Mobility (O&M) classes teach people with visual impairments how to navigate the world; for instance, how to cross a road. Yet, this training can be difficult and dangerous due to conditions such as traffic and weather. Virtual Reality (VR) can overcome these challenges by providing interactive controlled environments. However, most existing VR tools rely on visual feedback, which limits their use with students with visual impairment. In a collaborative design approach with O&M instructors, we designed an affordable and accessible VR system for O&M classes, called X-Road. Using a smartphone and a Bespoke headmount, X-Road provides both visual and audio feedback and allows users to move in space as in the real world. In a study with 13 students with visual impairments, X-Road proved to be an effective alternative to teaching and learning classical O&M tasks, and both students and instructors were enthusiastic about this technology. We conclude with design recommendations for inclusive VR systems.

7

CCS Concepts: • **Human-centered computing** → **Virtual reality**; **Accessibility systems and tools**; *Participatory design*;

Additional Key Words and Phrases: Accessibility, visual impairment, virtual reality, augmented reality, mobility training

ACM Reference format:
Lauren Thevin, Carine Briant, and Anke M. Brock. 2020. X-Road: Virtual Reality Glasses for Orientation and Mobility Training of People with Visual Impairments. *ACM Trans. Access. Comput.* 13, 2, Article 7 (April 2020), 47 pages.
https://doi.org/10.1145/3377879

1 INTRODUCTION

Developing spatial skills is a well-known challenge for people with visual impairments (PVI) [17, 29]. In Orientation and Mobility (O&M) classes, instructors teach. students with visual impairments how to analyze an environment using audio-tactile cues, how to prepare a journey, and how to cross a street in the absence of vision, for instance. As this training is often done in real outdoor environments, the O&M instructors have to supervise students to avoid accidents (e.g.,

This project was conducted with the support of the Erasmus+ Program of the European Union Pr. no 2016-1-EL01-KA201-023731.
Authors' addresses: L. Thevin, Inria Bordeaux & LMU Munich, Frauenlobstr. 7a, Muenchen, Germany, 80337; email: lauren.thevin@ifi.lmu.de; C. Briant, IRSA, 12, rue Alfred Musset Ambarès et Lagrave, Carbon Blanc cedex, France, 33565; email: c.briant@irsa.fr; A. M. Brock, ENAC, Université Toulouse, 7 Avenue Edouard Belin, Toulouse, France, 31055; email: anke.brock@enac.fr.

1936-7228/2020/04-ART7 $15.00
https://doi.org/10.1145/3377879

getting hit by cars) and deal with unpredictable conditions (e.g., traffic, weather, other pedestrians' behavior). Virtual Reality (VR) systems may help to overcome these issues, as VR environments allow users to experience spaces in an interactive manner without exposure to safety risks. Indeed, VR systems are environments created by a computer that the user is totally immersed in and able to interact with [51]. Therefore, VR environments are controllable (both for static elements and dynamic interactive elements) and can provide a safe and comfortable alternative to training in real traffic situations. Such an environment can provide students with the opportunity for trial and error (e.g., getting hit by a car in the simulation when crossing at the wrong time) to better understand the impact of their actions without getting hurt. VR has already been successfully used for teaching and training, e.g., for surgery and physical and spatial skills [16, 65]. These Virtual Environments (VE) can have different form factors and sizes. Most current devices rely on visual feedback. Multimodality is used for audio, haptics, and movement combined with vision [52]. Only recently has research emerged about accessible VR for people with visual impairments, using either audio [41] or haptics [77] or both [38]. Yet, these solutions are then not necessarily inclusive for sighted people, as they do not provide any visual feedback at all. Using VR with and without vision may not be equivalent, as visual information is widely used in VR [52], leading to unknown differences between people with low vision (who might use visual cues) and blind people (who cannot access visual information). Therefore, it is not trivial to propose a VR system that is inclusive for low-vision and blind people.

For VR simulations to be natural and immersive, it is necessary for the user to move in the physical space. A lack of (natural) movement can be a problem for training or education [24], even though it has the advantage of avoiding fatigue [52]. Kreimeier and Götzelmann [38] highlight the need for natural movement in terms of both scale and movement mechanisms. However, in existing VE for PVI, the physical space for movement is limited (for instance, 5m*10m in Reference [41] and 22m^2 in Reference [77]), and it is difficult to represent virtual outdoor scenarios. We argue that movement in VE at a scale of 1:1 compared to reality is important. First, it allows the user to naturally understand the spatial organization of the virtual world by moving in the real world as the world is directly mapped and manipulated [24]. Second, it avoids the classical VR locomotion technique of "Point & Teleport" (i.e., pointing at a distant virtual place with a virtual laser to "appear" at this position [13]), which is difficult for people with visual impairments to understand and use.

Current accessible VR systems require a large amount of specialized material that can be costly and hinder adoption in an educative context [16, 65]. An important goal for our project is therefore to provide a mainstream or low-cost solution. Based on these requirements, we designed X-Road, a visual and audio street simulator application for smartphones. Users move in the virtual world with the same scale as in the real world without any additional devices. Our technical solution uses the Augmented Reality (AR) framework ARCore. ARCore usually allows users to move around a virtual object positioned in the real world. In our solution this virtual object measures several hundred meters and is the entire virtual world in which the user moves (a virtual street or crossroad in scale 1:1). In a study with 13 students with visual impairments and 3 O&M instructors, X-Road proved to be an effective alternative to teaching and learning classical O&M tasks, e.g., crossing or walking parallel to a street. Both students and instructors were enthusiastic about this technology. To summarize, the main contributions of this article are:

(1) The design and development of X-Road, an inclusive VR system using a mainstream smartphone with audio and visual feedback, as well as the possibility to move in space at the real scale by walking. The proposed system is portable, self-sufficient, and requires no specific calibration.

(2) Its evaluation in the classroom with 13 students with visual impairments and 3 O&M instructors.
(3) The presentation of design recommendations for inclusive VR systems on the basis of this experience.

2 RELATED WORK

Our work is multidisciplinary, touching on work in both computer science and cognitive sciences. In this section, we will discuss the development of spatial skills of PVI. We then discuss the technical aspects of VR devices, feedback, movement, and usability issues for PVI. Finally, we present related work on accessible systems for spatial skills for PVI, including VR.

2.1 Visual Impairments and Spatial Cognition

2.1.1 Spatial Abilities as a Challenge for People with Visual Impairments. Spatial skills are important for many activities, such as mathematics, geography, and mobility and orientation. The latter in particular strongly impacts social life, autonomy, and inclusion in society. Vision is an important vector of information in the context of mobility and orientation, and therefore PVI face challenges related to navigating the world [17]. For people with low visual acuity, it is indeed more difficult than for sighted people to perceive objects and avoid them, to recognize points of interest and to update their spatial situation (e.g., to understand where they are and to correct their path). It is also more difficult to prepare trips, as most geographic maps are visual. However, it is not impossible to navigate in the absence of vision [17, 29]. Objects can be perceived through different modalities at once (for instance, a car is visible and emits sound) and potentially through deduction (the sidewalk is parallel to the road) [29]. Several studies argue that a spatial representation is ultimately a unique encoding, not dependent on whether a person has full vision, residual vision, or no vision at all [29]. Spatial representations can be created through multiple modalities (vision, audio, tactile, language). They are amodal, in particular for spatial images in working memory [43]. These representations exist in the form of cognitive maps, defined by Golledge [30] as "internal representation of perceived environmental features or objects and the spatial relations among them." PVI develop cognitive maps for spaces [67] based on landmark, route-based, and survey-based knowledge [64].

2.1.2 Tools for Supporting Orientation and Mobility among People with Visual Impairments. O&M instructors rely on tactile tools to represent spatial environments, such as tactile maps, small-scale models, or magnet boards [2, 21]. Tactile maps and diagrams are widely used to represent graphical information for PVI, in particular for teaching O&M and mathematics in school. They are made using various methods (e.g., laser cutting, swell-paper [68], or 3D printing [28, 31]) or are composed of objects with different textures [70]. These maps represent environments at a small scale (size of a paper map or model) and allow the effective acquisition of spatial knowledge in a safe environment [72]. Interactive audio-tactile maps have been developed in recent years to augment tactile tools with interactivity [21]. These devices propose novel features such as audio feedback instead of Braille annotations [15], the possibility to dynamically update content, quiz modes with the possibility to provide audio-based feedback [70], or map construction by PVI (to go beyond simple map exploration) [2, 22]. While maps are helpful tools for acquiring knowledge about unknown environments, they also have limitations. For instance, they represent spatial information from a survey representation, while many people with visual impairments tend to prefer representations from a route-based egocentric perspective [17]. Moreover, while maps preserve relations between landmarks in space, they provide this information in a smaller scale that can be explored manually but not by moving physically [72]. Consequently, the user then has to make a cognitive effort to adapt the acquired information to a bigger scale.

Another approach that has been investigated in previous research is to support PVI through in situ Electronic Travel Aids (ETA). On example is the NAVIG project, in which an ETA provided the user not only with guidance instructions but also additional cues about the environment [37]. In the NavCog project, a smartphone-based system has been developed that provides turn-by-turn navigation assistance based on accurate real-time localization over large spaces [1].

2.1.3 Teaching Orientation & Mobility to People with Visual Impairments. Besides providing PVI with tools for orientation and mobility support, it is important to teach them techniques to obtain information about the current environment and to orient themselves in the absence of vision. First, PVI need to be able to navigate safely and avoid obstacles, even in an unknown place (Mobility). This relates to alignment (e.g., being perpendicular to a street before crossing, following a curve or a line [29]) or configuration recognition (e.g., intersection type, relative angles between streets, traffic light statuses, flow analysis [29]), i.e., egocentric information for immediate travel. Second, they need to be able to recognize places and points of interest, and their position relative to this environment, and to develop relevant strategies to reach a goal in this environment (Orientation). This relates to integrating a goal, adopting strategies to reach this goal (e.g., correction strategies, shortcuts, and re-orienting strategies) and updating a representation of the environment (e.g., cognitive maps with one's current position) [29]. To acquire these skills and learn how to navigate safely and autonomously, PVI attend O&M classes.

During these classes, maps [49] or applications [59] can be used to gain knowledge of unknown environments, as described in the previous section. Alternatively, students learn how to analyze (urban) environments in situ and how to navigate within them [29]. For instance, instructors teach the students how to move in urban environments using a white cane or how to identify the characteristics of a crossing. For this part of the training, the teacher accompanies the student traveling in the real environment. To experience and learn from various situations, the instructors and the students need to move to different locations. However, real environments are exposed to many uncontrollable conditions, such as weather, traffic, obstacles, or other pedestrians, and can create anxiety and fear [72]. Moreover, in urban environments, and in particular when crossing streets, safety is a key problem when the student is learning the appropriate strategies within a trial-and-error process.

Interactive technologies can help teach students with visual impairments to use all available information in an appropriate and testable way [29]. VR can be a solution to address this problem. Indeed, it has been shown that PVI can learn when to cross a street based on O&M training in VR [12]. As stated by Friedman [25], one of the reasons for using VR is to explore and practice tasks in a dynamic and realistic yet controlled and safe environment.

2.2 Virtual Reality

Arnaldi et al. defined VR as the capacity given to users to carry out real tasks in a VE, this simulation being based on the immersion in this environment through the use of interactive feedback from and interaction with the system [4]. "Immersion" refers to the objective level of sensory fidelity of how the system provides virtual signals to the user's senses [11]. "Presence" refers to the user's subjective experience of "being there" in the virtual world [11, 63, 66] and forgetting the technology [40]. Most people imagine VR systems as visual systems. But the V of VR means the feedback is virtual, not visual. Indeed, VR and more widely Mixed Reality do not necessarily apply to the visual sense only, but potentially to all senses (vision, audition, touch, and even smell and taste) [5]. People's representation of VR systems being visual is grounded in the fact that VR applications today rely mostly on visual feedback. However, multimodality is included in definitions of Virtual Reality. For instance, Gigante defines VR as "the illusion of participation in a synthetic environment rather than external observation of such an environment." [27]. VR relies on three-dimensional (3D), stereoscopic, head-tracked displays, hand/body tracking and binaural sound. "VR is an immersive, multi sensory experience" [27]. While vision is often the main support to represent 3D spaces in the literature, in most cases vision is not the only modality. Feedback is often at least audio-visual and concerns interaction outputs (such as display and sound) and inputs (such as movement and controls).

2.2.1 Visual Feedback. Visual feedback in VR is provided through displays [51]. These displays can have diverse forms, including screens offering a "window on the world" (WoW), i.e., a small screen displaying the image of the virtual world in the real world [23]. In other words, the screen represents a window revealing the virtual world, depending on the position and orientation of the user. Other display-based VR devices exist, e.g., stereoscopic visual feedback with VR goggles, such as Head-Mounted Displays (HMDs). Some HMDs, such as HTC Vive[1] or Oculus Rift,[2] provide the ability to interact using hand controllers similar to the controllers or gamepads of a gaming console. HMDs typically display two slightly different images to the right and the left eye to recreate depth perception of the image [53]. Alternative techniques consist in larger, collaborative environments called CAVE with projections on the walls and potentially on the ceiling and on the floor [52]. The effects of using these visual displays with low-vision people have not been extensively studied. For instance, the usability of such VE might be lower for people with myopia (nearsightedness), light sensitivity, or diplopia (double vision).

2.2.2 Audio Feedback. In VR, audio feedback is often complementary to vision. For example, CAVE stands for "audio-visual experience automatic virtual environment" [19]. In comparison with vision, audio feedback has the advantage of being omnidirectional [17]. Like vision, within a certain limit it can be parallel, i.e., several audio sources can be perceived at the same time [33].

3D sound rendering (also known as virtual acoustics or spatialized sound) makes (pseudo) source-located audio feedback possible. An observer can perceive 3D sound to the left or right (azimuth), up or down (elevation), closer or further (distance), and from a smaller or bigger source (audio-image size) [6]. 3D sound potentially positions a sound at any place in a virtual space. Three solutions exist for creating 3D sound. First, 3D sound can be created from a mono audio source; for instance, by reproducing a delay and difference in magnitude in the signals between the two ears. Second, binaural sound is created by recording sound as it would be perceived by the eardrums; for instance, by placing two microphones on an observer's head[3] or an artificial plastic head.[4] This creates sound with realistic delays and distance variations as well as skeleton deformation effects, but which is valid for only one position of the listener's head (i.e., the position of the record). Third, ambisonic recorded sound[5] is used to render the perceived binaural sound consistent with the orientation of the head of the listener in terms of azimuth and elevation. If the listener changes her/his location (in the virtual world), then the sounds will follow her/him. Classical binaural and ambisonic records are thus not suitable for VR, as typically six degrees of freedom are required in VR: the listener's head moves in 3D (height, width, and depth) and can look in different directions around three axes (raw, pitch, and roll) over time. Some techniques to compute realistic spatialized sound with six degrees of freedom exist, but they require significant computing effort and access to the pure mono sound source [6]. It is possible to create 6D spatialized sounds in VR (e.g., Unity plugin Oculus Spatializer by Facebook/Oculus, Resonance Audio by Google, and Steam Audio by

[1] https://www.vive.com/fr/.
[2] https://www.oculus.com/rift/.
[3] e.g., https://hookeaudio.com/.
[4] e.g., https://3diosound.com/ and the "dummy head" of Neumann. https://en-de.neumann.com/ku-100.
[5] For instance, with this microphone: https://en-us.sennheiser.com/microphone-3d-audio-ambeo-vr-mic.

Valve), but the calculation of the approximation of audio effects in complex world modeling (reverberation, occlusion, etc.) is costly. Partially due to the relatively low speed of sound, audio effects do not instantly impact large spaces.

Imprecise 3D audio rendering can be corrected with vision, e.g., the effect of visual capture serves to map sound sources to visual elements (such as the moving lips of an actor) [6]. As VR often combines audio and vision, we have only little knowledge about the perception and usability of 3D real-time sound rendering by PVI (i.e., low-vision and blind people). This is true in particular for audio sources that are not reachable without walking [20]. In the case of inclusive VR for PVI, there is currently no knowledge on the minimum audio quality that needs to be reached for spatialized audio simulation. However, it is known that stereo feedback in VE is promising [62]. In visual environments, the "illusion of VR" works for a long time with unrealistic, low-definition and downgraded visual effects (for instance, for shadows and occlusion). We hypothesize that there might be a similar tolerance for low-quality audio feedback. More studies are needed to investigate the necessary realism of audio feedback in VR for PVI. This question is discussed more broadly in the "Question of the Realism" section.

2.2.3 Haptic and Proprioceptive Feedback. Besides auditory and visual clues, haptic, tactile, and proprioceptive feedback can be provided in VR. Haptic feedback can be perceived using the mechanico-receptors of the skin and body. A variety of feedback types exist and can be defined from types of stimulation (e.g., vibration, pressure, shape) and the simulated part of the body (e.g., hand, fingertip, forearm) [54]. Vibrotactile feedback is easy to produce and is already used in VR. Vibrations on an observer's seat can improve the perception of self motion [57] in VE. Gloves or wearable haptic systems for the fingertips and the hand can provide information about touching or grasping an object in VR, with vibrations either on the fingertips [54] or on the palm of the hand [46]. An alternative to vibrations is force-feedback. It is possible to provide force-feedback on the fingertips ("pinching" the finger) [61] or to use a hand exoskeleton on the inner side [10] or on the outer side of the palms [26]. Moreover, dedicated force-feedback devices exist, such as joysticks, pen-based devices [36], and the Geomagic Touch X (formerly the Phantom) [47]. Some devices such as the haptic revolver [73] combine textured tactile feedback with small displacements of the textures under the index. It is also possible to use an inflatable element on the palm to obtain dynamic feedback when holding an object [69]. Another approach is to create the sensation of proprioception through stimulation of the skin receptors and muscles. With electricity, a weight sensation can be created by stimulating the muscles of a user carrying a virtual object [44] as well as a perception of movement and body acceleration through galvanic vestibular stimulation in the inner ear [3]. Finally, passive haptics aim to provide haptic feedback in the virtual world, which is consistent with a real object that is not interactive.

Visuo-tactile multimodality is used extensively in VR. Most of these approaches use haptics to confirm and intensify visual information and may not work without vision, because the feedback would not be understandable (for instance vibrating gloves) or because the surface to explore without vision would be much bigger than that provided by the haptic system (for instance, in the case of the haptic revolver or Phantom) [35, 71, 75].

We conclude that more research is needed to understand the effect of haptics in VR for PVI.

2.2.4 Movement in the Virtual World. One advantage of VR over the traditional screen, mouse, and keyboard interfaces is that the user can move in the physical space to explore the virtual world. VR goggles track the head's orientation (raw, pitch, roll) and adapt the visual and auditory feedback.

Head orientation can be tracked with gyroscopic measures, e.g., using Google Cardboard. Some devices, such as HTC Vive and Occulus Rift, track the movement of the HMD and of the hand controllers in a limited space (for instance, 3m*4m). Besides head orientation, the head (or HMD) position is tracked with external sensors positioned on the corners of the virtual space, e.g., infrared sensors on the wall for the HTC Vive HMD.

The "Point & Teleport" mechanism [13] is widely used to move to remote locations in the virtual world without moving in the real one by pointing with a virtual laser to the next virtual place to move to. The user then virtually jumps to this new place, i.e., is "teleported." This allows users to cover larger distances in the virtual world than in the physical one. However, it relies entirely on visual information for selecting distant places and updating the new spatial position.

For large interactive spaces (16 feet*16 feet, i.e., approximately 6m*6m) [14] and large virtual worlds [8], techniques exist to redirect the user to remain inside the physical space, such as the walking mechanism, walk-in-place, stepping machine, and flying. Redirection using the walking mechanism consists in changing the perceived orientation in the virtual space so the user changes his/her trajectory in the physical world and remains inside the area. Walk-in-place (the user mimics walking by moving his/her tracked feet or his/her arms left and right), a stepping machine (a specific platform, e.g., a stepper or a bowl counting the user's steps), and flying (animation in the virtual world launched by a movement or a controller button) require visual feedback for updating the new virtual spatial position. Another possibility is to ask the user at the edge of the zone to turn 180 degrees; however, this will interrupt the interaction [74]. Most of these mechanisms require visual feedback and may be inaccessible to PVI.

2.3 Accessible Virtual Reality for People with Visual Impairments

While we have shown that many questions regarding VR use by PVI are still unanswered, a few projects have specifically investigated VR for PVI. Lahav and Mioduser showed that a multisensory VE can be used by PVI to acquire spatial skills [42]. In their study, users navigate a VE with a force-feedback joystick and perceive environmental cues through audio-tactile feedback. Connors et al. [18] conducted a study where blind participants explored an audio-based VE using the keyboard. Participants explored this VE following game instructions ("get the jewels") or spatial instructions ("learn the building's spatial organization"). Both types of instructions allowed the acquisition of spatial knowledge in a VE. Moreover, the video-game-based learning allowed a more flexible spatial representation to be encoded. Guerreiro et al. developed two virtual smartphone applications that allow PVI to build mental maps prior to travelling in the real world [32]. These applications provide a description of a sequence of points of interest on a route. Users can modify their virtual walking speed by tilting the smartphone, while feedback is provided in the form of step sounds. Guerron et al. [34] proposed smartphone-based exploration of virtual environments using multisensory feedback (speech, beeps, and gestures). Users do not walk in the physical space, but this application only requires a smartphone and headphones.

Other systems allow the user to walk in the physical and virtual space. Zhao et al. developed a virtual cane simulator using HTC Vive goggles and a controller with a brake to provide haptic feedback while manipulating a white cane [77]. Kunz et al. [41] designed and tested a system that helps PVI to navigate in a virtual space with audio feedback working like the beep of the radar-based parking aids in cars. The user wears a camera, and the user's movement in physical space is computed through detection of QR markers positioned on the walls and the ceiling of the room. This movement is then transposed in the virtual world. Kreimeier and Götzelmann [39] propose a system using walk-in-place with a Cyberith Virtualizer to navigate in an indoor environment. The system enables bipedalism and a virtual white cane through vibration on a hand controller while navigating in large spaces. Regal et al. [56] created location-based indoor games to make O&M classes more entertaining to increase motivation and learning. Students could either navigate by physically moving in the real world, which was equipped with NFC tags to trigger text-to-speech information, or by using a keyboard in a virtual world to trigger the tags with the "space" key. The study shows that games with real movement in the real world were preferred to the use of the keyboard. The space tracked in the VE of these systems is limited, and they are mostly used to represent indoor environments. They rely on specific devices, which means that users have to buy dedicated hardware and to have specific knowledge to use them.

Surprisingly, none of the presented applications provides visual feedback at all, although this would be interesting for low-vision people, as even low-vision people seem to have a preference for using visual information in navigation tasks [78]. Visual feedback might also be interesting for potential collaborative use with sighted people; for instance, when visually impaired students work with sighted instructors. We suggest that an inclusive system should be usable both with access to vision and without. We identified two use cases for using accessible VR with PVI. First, they may enable users to acquire spatial representations of particular places. Second, they can be used to simulate and test new navigation systems, as in Reference [41]. Bowman and Liu [12] used VR for O&M classes and demonstrated skill transfer to a real situation in the specific content of analyzing street environments. To the best of our knowledge, none of the existing projects has addressed the trial-and-error practices of O&M classes, i.e., analyzing environments, crossing the street, and observing if the timing is right or if a (virtual) collision occurs. Moreover, while Bowman and Liu [12] used a projection-based large-scale VE, we propose a portable system.

3 DESIGN PROCESS: CHALLENGES, SOLUTIONS, AND MATERIAL

In the previous sections, we have underlined the importance of spatial abilities for PVI that O&M instructors aim to develop. We believe that the potential of VR for O&M training for PVI has not yet been fully investigated and address this gap in the current article. In this section, we present the origin of the project and the design choices with an overview of the final hardware and software and the results of the participative process.

3.1 Illustrative Scenario

At the origin of this project was a request from an O&M instructor who approached us about the possibility of designing an accessible crossing simulator for O&M classes. This request specifically addressed the Mobility side of O&M classes in which students learn how to analyze and navigate safely and consistently in any spatial environment (even unknown places). This includes analyzing streets and crossings and then navigating in these environments. Students travel with instructors to real places to learn how to navigate and cross.

As an illustrative scenario, let us imagine Alex, a student with visual impairments, having an O&M class with an O&M instructor called Max. Today, Max wants to teach Alex how to cross a street. Max and Alex take the bus to go to the city center and reach a first Y-shaped crossing (three branches) with stop signs. Max asks Alex how many branches he can identify. Max helps Alex to understand the angles between the three branches using the direction of engine sounds. Indeed, at a Y-crossing, the branches are not perpendicular, as is the case for a T-crossing. For this purpose, Max asks Alex to move to be parallel to the branches. Alex uses the engine noise to get a distant representation of the branches' locations. Then, Max asks Alex about the priority rules of the crossing. Alex learns to identify a stop branch by sound: All cars stop before continuing. On the instructor's instructions, Alex finds the start of the pedestrian crossing with a white cane, thanks to the tactile paving in front of it, and chooses when to initiate the crossing. Alex hears a car stopping and decides to cross. Max stops Alex: The car starts again, because it was stopping for the traffic priority rules and not because of pedestrian priority. Alex waits for a moment without traffic and crosses when the audible cars are far away. As there are a lot of cars, Alex waits for several minutes. After this first part, Max and Alex walk 15 mins to reach a new crossing. This is the first time Alex learns how to cross a traffic light crossing. There are not many cars, so Alex needs time before recognizing the configuration. Max helps with indications such as "How would you describe the starting moment?" (massive engine sounds are emitted when a light turns green) and "On which branches are cars driving?." This has to be repeated, because in traffic-light configuration, driving lanes alternate. Alex finally identifies the crossing. Max checks if Alex knows the crossing rules in this situation: First count the time to find out the duration of each traffic light, then cross parallel to the cars emitting sound. Alex is positioned in front of the pedestrian crossing, which is easy to locate. When the car in front of him starts, Alex initiates the crossing. Max stops Alex immediately to avoid an accident. Max and Alex then go to another crossing but it starts raining. Alex and Max now need to take the bus to go back to school.

3.2 Current Limitations and Challenges

This scenario illustrates the limitations of current O&M training: (1) It is necessary to move from one point to another to study different configurations, which takes time. (2) The number of cars driving is not controllable and may not be adapted to a novice or to understanding and crossing new environments. (3) Students can only analyze the environment from the sidewalk without experiencing other perspectives, e.g., the middle of the street. (4) Instructors have to stop students before they commit an error to avoid accidents, so students cannot learn from their mistakes. (5) Teachers can only bring students to places that are easily reachable. (6) Students and teachers are exposed to the conditions of the environment, e.g., rain.

From the initial request of the O&M instructor, we identified several design challenges: First, our application is intended to be used by students in school. Second, the learning is related to standard situations (e.g., traffic-light crossings in general) and not a particular place (e.g., the traffic-light crossing at a specific intersection). A VE for this use case should therefore not represent a specific "existing" space, but propose a generic space with the relevant representation of spatial elements. Third, our system is intended to represent outdoor environments rather than indoor environments.

Based on these challenges for O&M classes, we designed and implemented "X-Road," a VR street simulator for O&M training in a scale 1:1 (i.e., comparable to the scale of a real street), which can be used to control the number of vehicles and the weather and to change the environment without changing the physical location. Moreover, it is possible to let the student try the simulator, even if this results in the student "getting hit" by a virtual car. We suggest that this might make the consequences of certain actions clearer to the students.

3.3 Users

As illustrated in the scenario with Alex and Max, there are two types of users of the system: O&M instructors and students. This impacted the design process, with iterations with the two types of users. It also impacted the design choices, since teachers and students should be able to use the system even though they have different levels of vision. The system should be usable with and without vision, for manipulation and for feedback. The materials should be compatible with a school infrastructure, from the device to the Internet network (no Wi-Fi in the school and limited connection in the countryside). The design should bear in mind the fact that the instructors may not have technical skills: settings, calibration, and bug resolution should require minimal and acceptable actions. The system should support O&M tasks, as well as interactions between instructors and students in a class.

3.4 Design Choices

To design an accessible VR system for O&M training of students with visual impairments, we needed to address different design choices, which we will explain below.

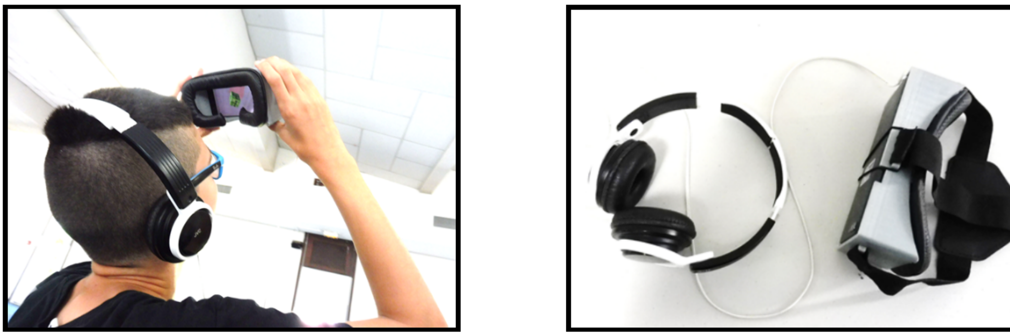


Fig. 1. Left: a virtual cube emitting a sonar sound can be localized using audio (and optionally visual) cues. The same approach is used to represent vehicles in virtual streets. Right: the X-Road prototype consists of a headset, smartphone, and 3D-printed holder. The prototype can be held in the hand or worn on the head.

Fig. 2. Overview of the configuration of the virtual worlds. Right: top view of VR#1 representing a street with a pedestrian crossing. Left: top view of VR#2 representing a traffic light crossing.

3.4.1 Overview of the Hardware. X-Road consists of VR goggles (a smartphone and a 3D-printed DIY support) and headphones through which it provides audio feedback (Figure 1). The smartphone presents the user with a WoW (Window on the World) [51]. We have chosen this solution over HMD goggles because the latter are heavy and can cause physical discomfort or specific discomfort related to the visual impairments (glare, external residual vision on the distorted image by the lenses, diverging eyes). Moreover, this choice can be justified pedagogically: VR goggles completely cut off the student from the teacher in the real world; while with a WoW solution, the teacher remains in the student's perceived environment. Finally, for security reasons, it is important to keep awareness of the physical surroundings.

3.4.2 Overview of the Software. We used the ARCore[6] framework, an AR toolkit for Android. ARCore uses virtual "anchors" that instantiate virtual objects in the world and compute the movement of the user in the world relative to the virtual object. This offers real-time feedback by generating virtual items respecting perspective and position in the world without requiring any supplementary material. We used the AR framework as a technical solution to instantiate the virtual world as one single object measuring several hundred meters (see Figure 2) at a specific location in the real world to generate the VE.

[6] https://developers.google.com/ar/discover/.

3.4.3 Movement in VR. As discussed in the Related Work section, several mechanisms can be used to represent movement in VR. We decided that the movement in the virtual world should be induced by the movement of the user in the physical world at a 1:1 scale. We suggest that techniques such as teleportation, flying, or passive haptics, which rely heavily on visual information, are challenging for PVI to use. Techniques that include redirection of trajectories also make use of visual information and, besides, might not lead to a correct spatial representation of a journey. However, movement in 1:1 scale requires access to a large space that allows the user to walk 50 to 100 m in every direction. Many of the special education schools we work with have access to large rooms; for instance, plenary work rooms, hallways, or outdoor spaces.

Computing the movement of the user in the real world to provide feedback about virtual objects is already used in AR. AR technologies such as Vuforia[7] compute a movement relative to markers or images that are tracked. The relative orientation, relative position, and perspective are computed with image processing. Virtual elements can be associated with the tracked physical elements. ARCore computes the position in the environment relative to the starting point. It then builds a representation of the environment by tracking movement through camera data and inertial data. ARCore creates virtual 3D anchors in the environment to position virtual objects relatively to the user. We use the same technologies in our final X-Road prototype, although the virtual object is not a small AR object but an entire VR world. Instead of walking around the object as in AR, the user walks inside the object: a virtual crossing. An alternative to using ARCore would be the use of fiducial markers. For such a solution to work, at least one fiducial marker should be visible at all times to the camera. As a consequence, the number of markers needs to be quite high for an environment as big as a street, and markers need to be placed on the ceiling, walls, and floor. This solution is not appropriate for outdoor use and for limited time in class due to its cumbersome installation. Using ARCore requires no markers and no installation at all, and is therefore low-cost and easy to install.

3.4.4 Multimodal Feedback in VR. As presented in the Related Work section, multimodal VR systems exist. Multimodality increases the complexity of the system, as there are multiple interactions between the modalities. Hence, we decided to focus on two virtual modalities (audition and vision) and to mostly omit haptic feedback. We decided to maintain visual feedback to make the system inclusive for sighted people (e.g., O&M trainers) and usable by low-vision people, similar to Reference [2].

Virtual haptic feedback (e.g., in the form of vibrating gloves) did not appear necessary for exploring a virtual crossing, since it does not correspond to the way a crossing is explored in reality. Moreover, the position of the intersection can be deduced from the audio feedback from each branch. Yet, some elements of the world cannot be easily deduced from audio feedback, such as cars that are not moving or a sidewalk located parallel to the cars. Such elements are then not directly accessible to blind people. Mimicking a real situation, our system can be coupled with passive haptic feedback in the form of real tactile paving fixed on the floor at the position of the pedestrian crossing on a virtual street (see Figure 3). In the real world, such tactile pavings can be found in public buildings (such as train stations) to indicate paths, pedestrian crossings, train platforms, and stairs. In our project, the tactile paving we used belongs to the school and is used in O&M classes to train students to detect tactile markers on the ground.

3.4.5 Audio Feedback in VR. In X-Road, audio feedback is associated with moving objects (vehicles) in the same way as in the real world. As presented above, stereo feedback has already been used and has proved to be promising in VE for PVI [62]. However, little is known about

[7] https://www.vuforia.com/.

Fig. 3. Physical tactile paving aligned with the virtual pedestrian crossing. In the virtual world, the student is in front of the virtual pedestrian crossing, and at 50 centimeters of the road, parallel to the road on the sidewalk.

the minimum level of realism of audio feedback required to enable spatial representations in VR for PVI. In our prototype, we used the native 3D audio feedback proposed in Unity 3D,[8] one of the most widely used engines for VR and Mixed Reality development.[9] Our iterative design process and pretests have shown that the sound of the cars alone may provide sufficient spatial feedback. We also experimented with the Resonance Audio plugin for Unity. However, for long-distance sound from cars (see participative design iteration), the audio rendering was worse with Resonance Audio than with the native audio management in Unity according to feedback from O&M Instructors and a person with visual impairments.

3.4.6 Sharing the Point of View. The street and crossing are situated in a known position in the physical world, so it is not necessary to share the view with the instructor to know where the crossing has to take place. It is, however, helpful for checking the right time to cross. We proposed two solutions for screen sharing so the teacher can see what the student sees. The first one was to set up with a WLAN network between smartphones (one used by the student as VR device, the other one by the teacher to follow). The second solution was to use an existing internet network, using ARCloud or, in a similar manner, implementing a network game. The first solution was not validated by the teachers, as it was too technical to set up by themselves. The second was not possible due to the poor internet connection in the school. A pre-test with a blind student, who was starting O&M classes, confirmed that it was possible to run an O&M class by looking at the student's screen instead of a screen on a separate device either during or after the movement (Figure 7, bottom). Finally, in practice, the students realize incidents themselves, which makes it possible to conduct the O&M classes even if the device is not visible to the teacher, e.g., when worn as a headset (Figure 3, right).

3.4.7 Continuum of AR and VR. Milgram introduced the concept of the "reality virtuality continuum" [51]. Different degrees of reality and virtuality distinguish Real Environment, Augmented Reality, Augmented Virtuality, and Virtual Environment [50]. Our work is situated on the virtuality side, as the "principal environment" is the virtual street or crossing. In our final prototype, a blind person hears the virtual street and is fully immersed in a Virtual Environment. For low-vision

[8] https://unity3d.com/unity.

[9] According to https://uploadvr.com/vr-developers-unity-unreal/, "59% of VR developers use Unity, but devs make more money with Unrealand"; and https://unity.com/solutions/xr, "Learn why over 60% of all XR content is created with Unity.".

and sighted people, the real world is visible through the virtual screen. Thus, they perceive this application as Augmented Virtuality. However, in our application, the Real Environment has no meaning in the Virtual Environment and is not integrated into it. Considering that it is a residual perception and not a view-port on the real world, it is equivalent to hearing the Real World audio with headphones or touching a Real World floor while in the Virtual Environment. By adding tactile paving (Figure 3), we were able to connect the physical and the virtual world as Augmented Virtuality even for blind people. This definition also makes it possible to specify the role of the size of the virtual object in AR or VR. For VR, the virtual object needs to be big enough for the user to be immersed into it. We also suggest that congruent controls (egocentric displays and ego-referenced controls) [51] are required to define a Virtual Environment as a VR. These concepts have been studied in applications with zooming features, e.g., References [7, 58], switching between AR and VR within the same virtual object.

3.5 Participatory Design Iterations

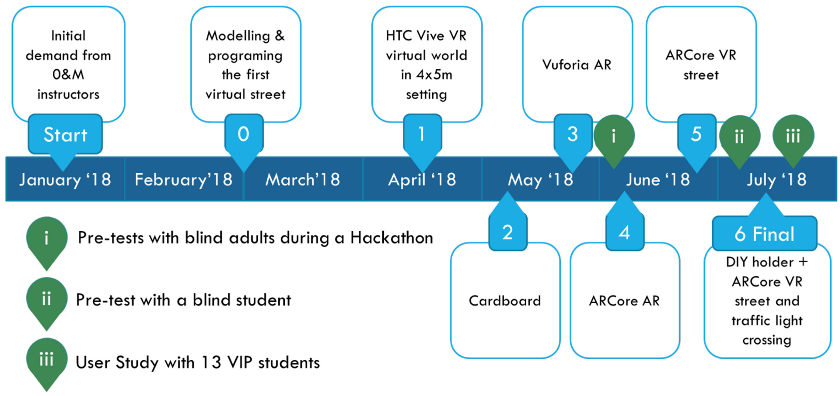
Participatory design approaches serve to consider target users by involving them actively in an iterative approach [60]. This approach generally distinguishes several phases: observing use, generating ideas, prototyping design, and evaluating the system [45]. As noted by Giudice [29], human- and user-centered approaches are needed to develop technologies for PVI "as sighted designers frequently have many fears and misconceptions about blindness that may have little to do with the actual experience." Including people with special needs during the design process may increase adoption of Assistive Technology [55]. Participatory design methods have been adopted in the design with and for people with visual impairments in previous projects (e.g., References [2, 48]).

In our case, the starting point of the project was a request from O&M instructors. In the "observing use" phase, we followed O&M classes to understand current practices and user needs. Part of this analysis is reported in Reference [2], and the illustrative scenario presented earlier is extracted from this knowledge. In the "generating ideas" phase, we talked with O&M teachers about technologies and acceptable and usable feedback that we verified during pretests. We also defined elements that should be present in the virtual world. In the prototyping phase, we conducted several iterations to reach the acceptable level of design so our tool could be used in O&M classes. These iterations are summarized in Figure 4, and the detailed description can be found in the Appendix. The iterations were possible because the implementation of X-Road in Unity was versatile and could be adapted to different hardware settings and VR/AR technologies (iterations 0 to 4) as well as features in the software (iterations 4 to final): computer video-game version (iteration 0), HTC Vive (iteration 1), smartphone with Cardboard (iteration 2), with Vuforia instantiating scene elements with picture markers (iteration 3), with ARCore instantiating virtual objects such as cars in AR (iteration 4), then with ARCore instantiating the virtual world (iterations 5 and final). Finally, the evaluation was conducted with O&M instructors and students with visual impairments.

3.6 Lessons Learned

As shown in Figure 4, during our design process, we used several devices and technologies (HTC Vive, Google Cardboard, Vuforia, ARCore) and implemented various uses of it (instantiation of virtual objects by targeting images or by anchoring objects by modeling the VR world or by designing a VR world as 400×200m AR objects). This process allowed us to gain insights into developping accessible VR and AR systems.

3.6.1 Requirements. Regular encounters with target users (O&M instructors and PVI) allowed us to identify user needs. To sum up, we identified the following requirements:



Itera-tion	Prototype	Main idea	Outputs for iterative improvement (sources in brackets)
1	HTC Vive VR 4mx5m	Move in scale 1:1 in the real and virtual world	Unusable device in school: too expensive and specific (O&M instructors) Space for movement too small (O&M instructors)
2	Card-board	Have a portable device	Usable device in school: smartphone (O&M instructors) No usable technology without vision : no movement in real time, delayed or false detection of steps with pedometer (O&M instructors). The sounds of the cars should not start at the same time when the simulation is launched in order to avoid engine sounds being synchronized (Scientific team)
3	Vuforia AR	Use the computation of movement around object in AR. Dynamic instantiation of objects with A3 printed paper (for instance lane and cars generator)	Instantiated objects not stable enough in space (O&M instructors) Basic features of sound can be enough (O&M instructors, Hackathon sighted and VIP users) Specific behavior when a collision with car occurs can be stressful (O&M instructors, Hackathon sighted and VIP users)
4	ARCore AR	Find a robust solution dealing with visual and audio virtual feedback (object still audible when not in the screen)	Usable and robust virtual object feedback and position in space (Scientific team) Usable for moving object with a defined trajectory, e.g. single car (Scientific team)
5	ARCore VR street	Test if the robustness of the AR (with objects) is transposable with an big object representing the virtual world	Usable concept for VR (O&M instructors, pre-test blind student) Engine sound to be improved (O&M instructors, pre-test blind student) Need for a holder around the smartphone to handle it without touching the touchscreen and to avoid covering the camera (pre-test blind student)
Final	DIY holder + ARCore VR#1 and VR#2	Created a more complex virtual world with 2 street crossings and various car behaviors (driving, stop etc.). Create a 3D printed holder	Usable concept for VR (O&M instructors, pre-test blindfold student)

Fig. 4. Timeline of our design process over five iterations. The timeline is presented at the top of the figure, while the bottom table describes the technologies used at each step, the reason for the technological choice (*Main idea* column), and the feedback about the proposed system (last column). AR means Augmented Reality and VR means Virtual Reality.

- Technology working on smartphones is preferable to technology working on VR devices. First, smartphones are more affordable. Second, those mainstream devices are perceived as easier to use by O&M instructors than dedicated VR devices.
- Movement in physical space should be transcribed with a scale 1:1 in the virtual world.
- It is necessary to have stable virtual objects (visual and auditory) or a stable virtual world at a stationary position in the real world.
- Having a holder around the smartphone improves usability for PVI who cannot see whether their fingers are blocking the smartphone camera.

3.6.2 The Question of Realism. The realism of the virtual world is a key question in our design. Two sub-questions exist: (1) Is realism necessary for VR? and (2) Is X-Road realistic enough? In the context of spatial learning, it is important for spatial information to be precise. However, the required realism of audio and visual feedback is less clear. Regarding the first question, we first analyzed the list of the most downloaded VR games on Steam in 2018[10] and the 12 games bringing in the most money ("Platinum" category). In this list unrealistic, low-definition and downgraded visual effects (for instance, regarding shadows and occlusion) are observable: four games show unrealistic visual environments (Beat Saber, SUPERHOT VR, Job Simulator, OrbusVR[11]), four games show low-definition visual effects (Budget Cuts, Hot Dogs Horseshoes & Hand Grenades, GORN, Fallout 4 VR[12]), and four other games, despite having more realistic visual effects, show downgraded visual effects, in particular for floor textures and trees (The Elder Scrolls V: Skyrim VR, Onward, Pavlov VR, Arizona Sunshine[13]).

For VR games, at least, realism is not a requirement. But for a street simulator, this is less obvious, since it relates to real-life situations. We ensured that the realism of our simulation was sufficient by acquiring early feedback about our system regarding the sounds used (versions 2 and 3) that was confirmed for the use in an O&M context in the pretest. At the early stage, two adults with visual impairments tried the system during a hackathon. They found the sound usable and were able to analyze the virtual world, the car positions, movements, and collisions. They explicitly encouraged us to continue with this solution. With these tests, we hypothesize that there might be a similar tolerance to low-fidelity audio feedback compared to visual feedback. One specificity of our system is the audio feedback associated with a spatialized world. In particular, the virtual streets are at a specific and fixed position in the virtual world. The audio feedback is spatialized (e.g., sound from cars), and it is possible to identify the spatial layout (i.e., street and sidewalk location) of the virtual world through it.

Moreover, we controlled for the realism through an evaluation with standardized questionnaires (see Sections 4.3.2, 5.3) about O&M tasks to verify that the representation was appropriate (see Sections 4.3.1, 5.1), and that the students and instructors could use it for O&M classes (see Sections 4.3.3, 4.3.4, 5.4, and 5.5). Since we co-designed the system with O&M instructors and PVI users, we believe that the achieved level of realism is sufficient for a system that is usable in the classroom in the context of O&M training. We also observed that realism might be a disadvantage. For instance, after testing with blind adults, we decided to remove the braking behavior of the cars, as it was

[10]On 14/02/2019 https://store.steampowered.com/sale/2018_so_far_top_vr_titles/.
[11]https://store.steampowered.com/app/620980/Beat_Saber/, https://store.steampowered.com/app/617830/SUPERHOT_VR/, https://store.steampowered.com/app/448280/Job_Simulator/, https://store.steampowered.com/app/746930/OrbusVR/.
[12]https://store.steampowered.com/app/400940/Budget_Cuts/, https://store.steampowered.com/app/450540/Hot_Dogs_Horseshoes__Hand_Grenades/, https://store.steampowered.com/app/578620/GORN/, https://store.steampowered.com/app/611660/Fallout_4_VR/.
[13]https://store.steampowered.com/app/611670/The_Elder_Scrolls_V_Skyrim_VR/, https://store.steampowered.com/app/496240/Onward/, https://store.steampowered.com/app/611670/Pavlov_VR/, https://store.steampowered.com/app/342180/Arizona_Sunshine/.

too realistic and therefore scary. VR also allows unrealistic behavior. Some students used VR to reach the middle of the crossing and observe the cars' movements from there. In real situations, the crossing is described from this point but never observed for security reasons.

Our work with the O&M instructors and PVI suggests that the problem to solve is "how to make the virtual world spatialized?" and not "how to make the virtual world more realistic?" This hypothesis is consistent with the theory of amodal but spatialized mental models of the world [43]. Moreover, if realism was a more crucial problem than spatialization, small scale models, 3D printed models, and raised line graphics would not be relevant. The question of making the VE more realistic is, however, an interesting question to pursue in the future (perhaps by researchers in other domains such as computer graphics and acoustics).

3.6.3 The Role of O&M Instructors. The proposed system was designed in collaboration with the O&M instructors. During the experiments, the instructors were conducting a typical O&M class with one student: asking questions, helping with the analysis if needed, and giving feedback on the student's actions. The purpose of our system is to be used during such classes. Information thus needs to be accessible for the O&M instructors. In the real world, the instructor sees what the student sees. In our final prototype, the instructors do not have their own independent access to the virtual world, as discussed in Section 3.4.6. The instructors preferred to have a system that is easy to use rather than a more complex system that allows them to share the screen. The instructors used the following techniques to overcome this limitation: The world is positioned at one specific place, in a specific orientation, and in the launch configuration. Consequently, the instructors know the state of the world on the smartphone (Figure 7, top). They know the answers to most of the analysis questions: "Is it a street or a crossroad?," "Are the cars circulating one way or both ways?," "Where are the cars relative to you?," "Can you walk parallel to the street?," "Are there traffic lights?." When a student initiates a crossing, the instructors inquire about the circulation of cars. In the virtual world, the student can cross whether the timing is correct or incorrect without any possible harm. The instructors look at the screen when the student is crossing (Figure 7, middle and bottom) or take the smartphone after the crossing to verify the cars' positions and provide feedback to the student.

3.7 Material: The X-Road Final Prototype

3.7.1 Hardware. In its final version, X-Road (see Figure 1) consists of a smartphone OnePlus 5 A5000 128 GB (8 GB RAM), an audio-headset JVC HA-S200 (12 Hz–22 KHz) for the sound rendering, and a 3D printed holder (Figure 6). As described above, the holder allows users to hold the X-Road smartphone without touching the touchscreen or hiding the camera (Figure 5, top). The holder can also be worn on the head with a rubber band (Figure 5, bottom). While in this latter setting the screen is too close or not visible, the audio feedback is directly related to the head movement of the user for a more precise spatial sound, and the user can hold a white cane at the same time as X-Road (Figure 3).

3.7.2 Software. The application runs on Android 7.1.1. and was developed with Unity 2017.3.1f1 in C#. We designed each 3D world with virtual objects having behaviors (traffic light states through C# scripts) as single parent objects. More specifically, we defined each 3D world (VR#1 and VR#2) as a prefab in Unity. In Unity, the system of prefabs allows users to store any virtual object as a template, so modifying the prefab modifies all other instances of this object. The ARCore plugin (arcore-unity-sdk-v1.2.1) was used to instantiate, in the environment and at precise locations, virtual elements with visual properties (mesh and texture) and audio properties (3D audio sources).

Fig. 5. Ways of wearing the VR system: (top left) hand-held without seeing the screen, (top right) hand-held seeing the screen, (bottom left) head-worn in front of eyes, and (bottom right) on forehead.

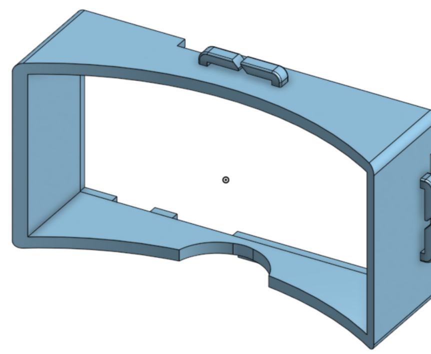


Fig. 6. 3D sketch of the holder. Once 3D-printed, it allows the users to handle the smartphone without touching the screen and to plug in the headset. It can be used head-worn or hand-held.

3.7.3 Familiarization Application: AR Cube Hunter. We designed an AR familiarization application so the students could learn how to manipulate X-Road and tested the virtual feedback (audio and visual). The goal for users is to find the location of a virtual cube in the room using either the screen (green cube visible as in Figures 1 and 8) or using 3D audio sound (direction, and intensity for the distance). The cube can be positioned anywhere in the room by the instructors (Figure 8, top). O&M instructors found it helpful to place a real object in the position of the virtual object to increase understanding; for example, a chair (Figure 9).

3.7.4 X-Road VR Applications. The goal of X-Road is for users to analyze the scene to understand the type of situation (street, crossroad, traffic lights, stops, number of lanes, and direction of driving on these lanes) and to cross when appropriate. For the final X-Road application, we designed two worlds: VR#1, a street with a pedestrian crossroad (Figure 2, left), and VR#2, a traffic light crossing (Figure 2, right). These two worlds represent typical traffic situations in cities in France.

The circulation of the cars in VR#1 was achieved by creating a new car for each lane at the two extremities of the street. The goal of the students was to find a time where no car was coming along either lane, so they would be able to cross the street safely. To avoid predictability, the cars were created at random times. Our algorithm first decides randomly on a break without cars (25% probability) or to generate a new car (75% probability). Then, our program picks a time for the break (random between 25 and 50 seconds) or a time before re-generating a new car (random between 3 and 6 seconds after the previous car is 10 meters past). Once this time is over, the program again decides randomly between a break or generating a car.

Fig. 7. Top left: The instructor asks the student where the street is. Top right: the student answers. Middle: the instructor asks the student to find the pedestrian crossing, (left) the student stands in front of it, and (right) the instructor can verify on the screen. Bottom, left to right: a student stands in front a the pedestrian crossing. From the instructor's point of view, on the screen, we can see the pedestrian crossing, then the front and the rear wheel of the car.

The circulation in VR#2 was achieved by simulating the traffic light status (red, green, and orange). The traffic lights have a collider on the relevant lane, so a car will brake to stop if it enters the collider and the status of the light is orange or red. Once the light turns green, the cars inside the collider of the traffic light start again. To simulate the wave of braking and starting when a traffic light turns red or green (not all cars stop or start at the same time), the cars avoid collision by braking if they are too close to the previous one. Once the vehicle in front of the car is far enough ahead, the car starts up again and accelerates until the maximum speed is reached. If there is a traffic jam around the crossing, then the cars wait until the car in front of them has left before restarting and continuing.

The VR application is used by instantiating a virtual road-world in the room in which the experiment takes place, as in the case of the AR Cube Hunter (see Figure 8, top). The students can

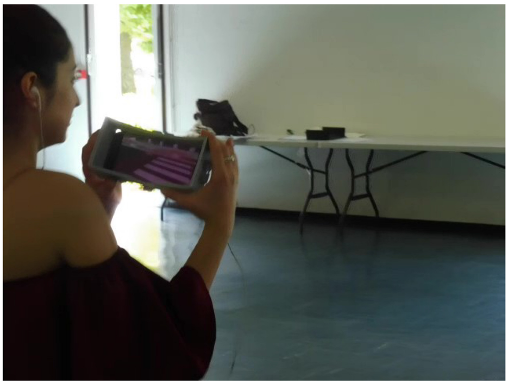
Fig. 8. Top left: The instructor touches the screen to position the virtual cube anywhere in the room. Top right: When touching the screen, a green cube is instantiated in the environment. Middle left: A student follows spatial audio feedback depending on his/her position relative to the cube. Middle right: The student identifies the cube's position. Bottom left: The student shows the teacher the screen for verification. Bottom right: The cube is visible in the screen, i.e., the position is correct.

Fig. 9. Example of embodiment of the cube with a chair. The cube is instantiated at the same location as a chair. A student moves around the chair (left, face and back to the chair) to understand the sound variations depending on his/her position relative to the object.

analyze the virtual world using the information about cars both on the screen and using spatial sound. The other parts of the world are only visually represented (houses, streets).

3.7.5 Use of the Software. Upon the first use, the applications have to be installed on a smartphone that is compatible with ARCore.[14] The smartphone can be positioned in the 3D-printed holder with a rubber band. As the support is designed to provide access to all buttons and plugs of the smartphone, it is easy to connect the audio-headset, and it can be plugged into an external battery while in use. AR Cube Hunter, VR#1 (street crossing) or VR#2 (traffic lights crossroad) is launched on the smartphone. The camera detects plane surfaces where the object (virtual cube, VR#1, or VR#2) can be instantiated by touching the screen (Figure 8, top). The object is located in the world at the location the user has designated on the screen, with the same orientation as the smartphone. The user moves in the virtual world or around the cube. The movement of the user in the real world is the same as that calculated in the virtual world. If the user walks one meter in the real world in a particular direction, then he or she walks one meter in the same direction in the virtual street (Figure 10). The students can access, without vision, the direction of streets, the traffic, the spatial configuration, branches, and sidewalks. The only spatial element that is inaccessible without vision is the position of the pedestrian crossing. Thus, as in a real street, tactile

[14]List available at: https//developers.google.com/ar/discover/supported-devices.



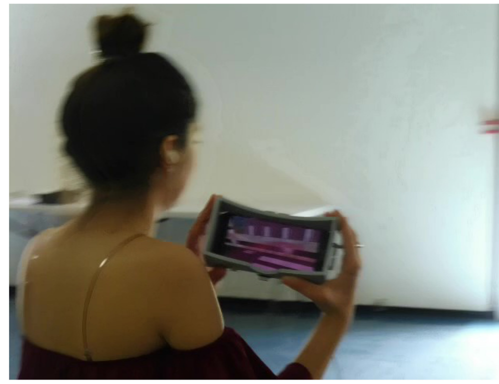


Fig. 10. Left: in the VR#1 representing a street, the student is at the beginning of the pedestrian crossing. Right: the student walks almost four meters in the real world as well as in VR#1. On the screen, we can see she is now in the middle of the pedestrian crossing.

paving can be used to indicate the position of the pedestrian crossing and is perceivable through the feet and with the white cane (Figure 3).

4 METHOD AND PARTICIPANTS

4.1 Participants

Our panel of participants was composed of both professionals who work with PVI and students with visual impairments. When not specified below, "participants" refers to "students with visual impairments," while we refer to the professionals from the specialized school mainly as "instructors/teachers." The member of the research team conducting the experiment is referred to as the "experimenter."

4.1.1 Instructors and Professionals. For all participatory design stages (observing use, generating ideas, prototyping, and evaluation), we worked with three O&M instructors from IRSA, a specialized school for children and young adults with visual impairments. We were in contact with one orthoptist and three psychologists from the same specialized school. This helped us acquire feedback about potential problems with using screens or stereoscopic vision with PVI, as well as gave us insights into the use of VR for students with visual impairments.

The evaluation (i.e., the user study) was conducted as an O&M class. One of the O&M instructors actively conducted the O&M class with the student participating in the experiment. The questions in the O&M class were designed jointly between the experimenter and the instructors to be sure they corresponded to the teaching.

4.1.2 Visually Impaired Students. 13 students (5 F, 8 M) participated in the experiment. They were aged between 12 and 20 years ($M = 15$, SD = 2.4); 7 students were low vision and 6 were blind (2 with light perception, and 4 without light perception). Ten were right-handed, 2 left-handed, and 1 ambidextrous. The gender, age, description of level of vision, and order of presentation of the virtual worlds are presented in Table 1. The details are presented in Figure 11.

4.2 Protocol

The experimental protocol included seven steps:

Table 1. Table of Participants (Students with Visual Impairments): Characteristics of Gender, Age, Description of Level of Vision and Order of Presentation for the Virtual Worlds VR#1 (Street) and VR#2 (Traffic Light Crossing)

Id	Gender	Age	Description of the level of vision	Order of presentation
P1	F	13	Low vision	V#1 - V#2
P2	M	13	Low vision, tunnel vision with low acuity	V#1 - V#2
P3	F	12	Legally blind from birth, perceives some shapes	V#1 - V#2
P4	M	15	Blind from 3 years old	V#1 - V#2
P5	M	12	Low vision from birth	V#2 - V#1
P6	M	17	Blind from 5 years old	V#2 - V#1
P7	F	19	Low vision, 1.5/10 with correction from 13 years old	V#2 - V#1
P8	M	14	Early-blind	V#1 (no VR#2)
P9	M	14	Early-blind	V#1 - V#2
P10	M	20	Low vision, only near vision	V#2 - V#1
P11	F	12	Low vision, 1/20 on the best eye from 6 years old	V#2 - V#1
P12	M	19	Early-blind with light perception	V#2 - V#1
P13	F	16	Low vision from birth, glare	V#1 - V#2

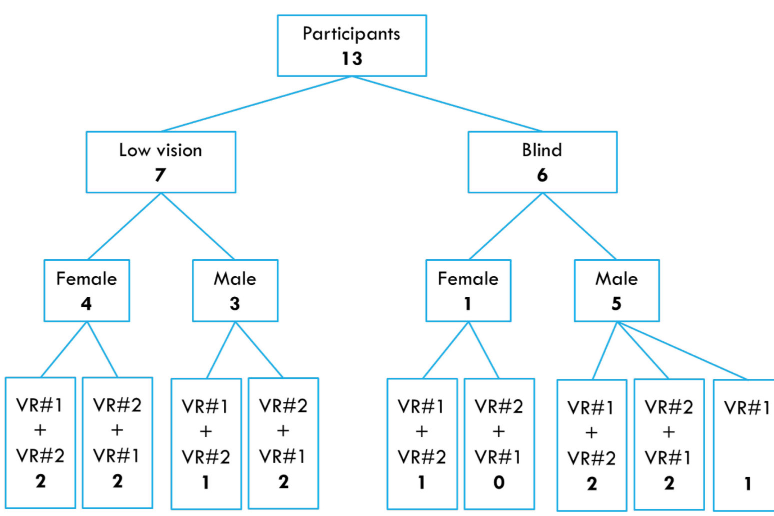


Fig. 11. Distribution of the participants over visual conditions and order of presentation.

(1) **Introduction:** verification and collection of the consent and image forms sent in advance to the families for information and signature, presentation of the experiment and questionnaire about apprehension, and comfort when crossing streets in real life.
(2) **Familiarization:** familiarization with the manipulation of X-Road with the AR Cube Hunter application.
(3) **Phase 1 -** ***world 1*****:** first virtual world presented (counterbalanced between participants, VR#1 street or VR#2 traffic light road-crossing).
(4) **Phase 1 -** ***world 2*****:** second virtual world presented (counterbalanced between participants, VR#2 traffic light road-crossing or VR#1 street).

(5) **Phase 2 - questionnaires about VR:** the questions are described in the following section.
(6) **Phase 2 - personal questionnaire:** personal questions about age, school level, or profession, vision level, audio analysis skills, passive echo-location skills, body positioning (straightened spine or coiled spine), experience with video games, experience with new technologies.
(7) **Phase 3 - VR goggles case study:** (optional, depending on the time and the agreement of participants) presentation of the same virtual worlds with the cardboard goggles (stereoscopic vision, but no possible movement in space), and same evaluation questionnaires.

The questionnaires and tasks are described in the next section. In practice, the introduction and familiarization lasted 20 minutes in total, Phase 1 lasted around 30 minutes (in general, 15 minutes for each virtual world), and Phase 2 lasted around 20 minutes. This adds up to a duration of 1 hour and 10 minutes, approximately. The optional Phase 3 lasted, in general, 20 additional minutes, as the schedule of the experiment was calculated to last 1.5 hours per participant.

4.3 Observed Variables and Statistics

Our principal scientific objective is to evaluate a new VR system showing a virtual crossroad that is used by students with visual impairments in O&M classes. The associated research question is "Is it possible to propose an inclusive VR system for training PVI, i.e., a system that is usable by blind, low-vision, and sighted people?." We asked the students and teachers to conduct their regular O&M classes using X-Road. We present in detail the measures in the following sections.

4.3.1 Effectiveness: Task-specific Assessment Using Realistic O&M Tasks.

Principle: To assess the VR simulator, we asked the students to perform the same tasks with our simulator as the ones they usually perform in O&M classes outdoors. We evaluated the effectiveness of the system by asking the students to analyze the environment for the two virtual worlds.

Design: We wanted to answer the question "Can an audio-visual VR crossing simulator provide elements for analyzing the environment as in a classical O&M class?." The questions for analyzing the virtual worlds (described in the next paragraph) were designed in collaboration with the O&M instructors to correspond to classical O&M classes. They were the same for worlds VR#1 and VR#2 to avoid bias, and some questions depended on the student's previous answers.

Activity: The instructors conducted an O&M class with the students with VR#1 and VR#2. The two worlds were presented in counterbalanced order. During the class, the O&M instructors asked the students questions. Some questions required an analysis of the environment, while others were evaluated by performing activities. The students were free to move in the virtual world by moving in the physical world to reply to the questions. The list of questions was as follows:

- **OM1: Do you hear the cars?** This question allows us to verify that the virtual world is launched, the audio settings are working, and the student recognizes the virtual cars.
- **OM2: In your opinion, is it a street or a crossroad?** Analysis question. It requires participants to obtain a spatial representation of the cars driving and to perceive if there are one or two axes.
- **OM3: If you think this is a crossroad, how many streets are crossing, in which configuration are they crossing, and are there traffic lights?** Analysis questions. They require a precise spatial perception and understanding of the cars driving and perception of the number of axes, the angles of the axes, and the traffic regulation mechanism. This question was asked only if the student replied "crossroad" to the previous question.

- **OM4: Are the vehicles driving one way or both ways?** Analysis question. It requires perception of the direction of most co-located audio sources moving along the same axis.
- **OM5: Where are the cars located relative to your position?** Analysis question. It requires situating the audio-source relative to oneself.
- **OM6: Can you walk parallel to the traffic lane?** Action. It requires walking parallel to the virtual street. To achieve this, the student needs to situate the audio source movement relative to himself/herself (to feel the angle with the vehicle movement) and to update his/her perception while walking to remain parallel.
- **OM7: Can you initiate crossing the street?** Action. It requires crossing the virtual street without hitting a car. To achieve this, the student needs to perceive whether there is enough distance between cars when he/she wants to cross in a two-way traffic situation. For VR#1, this is done by identifying no close car coming from the two sides, possibly by crossing just after the last car in a lane when there are no cars for long enough in the second lane. For VR#2, the student needs to count the time for each traffic light status and to start to walk when the cars in the parallel lane start moving. The student needs to have a good representation of the virtual space: identify where to cross, identify the orientation for crossing (cross perpendicularly to the road to make it shorter), and identify when the street is actually crossed and when he/she is on the opposite sidewalk (the student can use their remaining visual capacities and their position relative to the cars in one or two directions, as all the cars from the same road are on the same side once they have crossed). The instructors verify that there are no virtual cars and that the student has actually completely crossed the street.

Assessment: The O&M instructors evaluated the students during the activity by assigning a score to each question using the following rating system: For each question, we considered three cases for evaluation: (1) the student was not able to answer the question correctly (success rate of 0%), (2) the student needed some time to answer the question correctly or temporarily gave a wrong answer before correcting it (success rate of 50%), (3) the student immediately answered the question correctly (success rate of 100%). The score was then between 0% and 100% for each of the seven questions.

4.3.2 Presence and Immersion: Assessing Immersion and the Realism of the VR System.

Principle: We wanted to assess how participants perceived the proposed VR. Two key elements of the presence experience in VR identified are immersion and realism of the virtual environment.

Design: We used a 7-point Likert scale questionnaire adapted to PVI, as presented by Zhao et al. [77]. These authors based this questionnaire on an existing questionnaire [63] and adapted it to assess the perceived immersion and realism of VR systems among people with visual impairments. We translated the questionnaire into French.

Activity: In Phase 2, we asked the student to answer this questionnaire [77]. The questions are identified by the dimension they refer to, with G for "general sense of being there," SP for spatial presence, INV for involvement, REAL for experienced realism:

- G: In the VR world I had a sense of "being there" (1 "I fully disagree," 7 "I fully agree").
- SP1: Somehow I felt that the virtual world surrounded me (1 "I fully disagree," 7 "I fully agree").
- SP2: I felt present in the virtual space (1 "I fully disagree," 7 "I fully agree").
- INV: I was completely captivated by the virtual world (1 "I fully disagree," 7 "I fully agree").
- REAL1: How real did the virtual world seem to you? (1 "not real at all," 7 "completely real")
- REAL2: How much did your experience in the virtual environment seem consistent with your real-world experience? (1 "not consistent at all," 7 "very consistent")

- REAL3: How real did the virtual world seem to you? (1 "about as real as the imagined world," 7 "indistinguishable from the real world")
- REAL4: The virtual world seemed more realistic than the real world (1 "I fully disagree," 7 "I totally agree").

The questions were asked by the same experimenter for all participants. The O&M instructors were the same as those present during the O&M activities of Phase 1.

Assessment: As proposed by Zhao et al. [77], we evaluated immersion and presence in VR by assigning a quantitative score to a subjective evaluation with a 7-point Likert scale.

4.3.3 Satisfaction: Assessing the Usability of the System (Interest, Comfort, Re-use).

Principle: We wanted to assess the use of the system and the way participants interacted with the hardware, the application, and the virtual environment. More specifically, we were interested in its usability in a classroom setting. Moreover, we wanted to evaluate satisfaction with regard to the VR environment (independently of the hardware).

Design: We identified the following points for the acceptability of the system and the VR environment: interest (in the sense of "appeal"), comfort, possibility of re-use. We used a subjective measurement through questionnaires, as this concerns the users' perception of the acceptability of the system. We wanted to assess whether the system appeared interesting to the students. In particular, the presented VR environment might be rated as "not good enough" for now in terms of realism, immersion, or usability by the students, but the idea of using it could remain interesting to them. Conversely, the VR environment could be rated highly in terms of realism, immersion, or usability by the students, although they show little interest in using it. We assessed the comfort of the hardware when used during an entire O&M class. We also wanted to assess whether users were ready to re-use this system, as the classroom setting involves a repetition of activities.

Activity: We evaluated acceptability for the students by asking questions after the activity:

- on interest
  - Do you see the value of using this system in O&M classes?
  - Do you see the value of using this system in other contexts?
  - Do you see the value of using this system in video games?
- on comfort
  - Have you found this device comfortable (including physical fatigue, e.g., holding it)?
  - Have you found this system easy to use?
  - Have you found this system tiring to use (including workload and fatigue of eyes)?
- on possible re-use of the system
  - Would you like inclusive VR and AR to become more widespread?
  - Would you be interested in re-using similar systems?

The questions were asked by the same experimenter for all participants. The same O&M instructors as in Phases 1 and 2 were present.

Assessment: We evaluated the answers to the questions by assigning a score on a 7-point Likert scale. For all questions, the participants were asked to give a rating from 1 "I totally disagree" to 7 "I totally agree."

4.3.4 Satisfaction: Assessing Whether the System Is Compatible with O&M Classes from the Instructors' Point of View.

Principle: Besides acceptance by the students, we wanted to assess the acceptance of the system by the instructors. Indeed, students and instructors may have different opinions about the usability of a system and it is important to satisfy both. For instance, a student might enjoy the entertainment value of X-Road, while an instructor might have a preference for more serious training.

Design: We assessed three different elements. First, the instructors' appreciation of the system. Second, their comparison of this system with a classical O&M class in the real world and the advantages or disadvantages of our system. Third, feedback on each student's appreciation of the system, as the instructors know to what extent the students like O&M tasks in classical settings.

We identified a possible bias of "fantasy about the technology" that may lead the instructors to rate the perspectives of the system or an imaginary representation of VR rather than the current system. In particular, the instructors might rate an imaginary future or improved version of X-Road after new apps and features are offered. This could also lead to a strong confirmation bias during the rating of the system by the instructors, who might highlight or minimize the students' success with the system. We asked the instructors to answer the question for each student individually after his or her session.

Activity: Immediately after each participant's session, the instructors answered a questionnaire and were asked to answer only for the case "of the current participant" and "after this session specifically," to limit evaluation bias on an imaginary use of VR as well as confirmation bias. The questions were as follows:

- possible use in O&M classes
  - Do you think this system is usable in O&M classes—for instance, considering time constraints and integration with classical practices?
  - Do you think you can use this system alone and autonomously in O&M classes?
  - Do you think the student has enjoyed using this system in an O&M class?
- added value in O&M classes
  - Do you think this system is complementary with classical O&M classes?
  - Do you think this system presents advantages compared to traditional resources used in O&M classes?
  - Do you think the student can benefit from the new functionalities made possible by Virtual Reality?
- downsides of the system
  - Do you think this system is not complementary with classical O&M classes and does not add anything new?
  - Do you think this system presents downsides compared to traditional resources used in O&M classes?
  - Do you think the student may have difficulties using Virtual Reality?
- finally, the instructors were asked to provide any other remarks.

The questions were printed on paper and given to the instructors once the student had finished the experiment. They were (yes/no) questions with free space to provide more details. This method was chosen to prevent the instructors from being influenced in their answers. It also enabled clear answers (yes or no) but with qualitative explanations.

Assessment: We evaluated the answers to the questions by assigning a score to them: yes (success rate of 1) or no (success rate of 0).

4.3.5 Case-study: Assessing Possible Effects of VR Goggles on PVI.

Principle: As the use of VR by PVI is a relatively new field, there is not much knowledge about the usability of VR goggles and stereoscopic vision. Therefore, we wanted to obtain insights into the possible effects of VR goggles (in particular, with stereoscopic devices close to the eyes).

Design: For this part of the case-study, we first evaluated the effects on vision. We compared with the problems using screens mentioned by the O&M instructors and orthoptists: diplopia (i.e., double vision) inducing fatigue of the eyes, headaches, and blurred vision; and glare inducing discomfort, pain, or impossibility of using the visual information on the screen. We also wanted to verify the effect of wearing VR goggles on the body positioning, e.g., the spine.

For that purpose, we decided to compare these effects when the participant used the two devices (X-Road with a smartphone and the VR goggles) when "visiting" the same virtual world. This is possible, as Unity can be used to compile VR#1 and VR#2 as a Google Cardboard application on the same smartphone as that used in the X-Road prototype. Thus, conditions are comparable, as the screen, audio-feedback, computational power, and weight of the smartphone are identical. The main difference concerns movement in space. Using the VR goggles application with the cardboard system, it is not possible to move in the virtual world as in the real world. The classical mechanism to move in the cardboard application is teleportation and, as discussed earlier, this is not applicable to PVI. There is also a skybox in these apps, which in Unity is a cube wrapping around the virtual scene, typically a sky, representing the horizon in all directions. The six faces of the skybox cube have textures (for instance, pictures of the sky with clouds) that are visible on the screen when there is no object hiding them. In X-Road, one can see the physical environment at the same time as the visual, as there is no "virtual sky." Finally, the VR goggles were always presented at the end of the experiment, so the presentation was never counterbalanced.

Activity: Students were optionally invited to visit the same virtual worlds VR#1 and VR#2 with the Google Cardboard system. Some students had time limitations due to transport that had been organized, and therefore some of them were not able to test the prototype or to answer all the questionnaires as described below.

We asked the students to recognize the virtual space they were visiting, i.e., VR#1 or VR#2, and asked them questions about effectiveness. As movement in space was not possible, we did not invite them to cross the street but to position themselves with the right orientation for crossing.

When possible, we asked the previous questions about satisfaction (interest, comfort, and re-use). Then, we asked the participants which system they preferred and why.

We finished with the questionnaire on presence in VR. We did this questionnaire at the end, because we wanted to know more about interest, comfort, re-use, and preferences than about presence in the same virtual world, in case there was not enough time to answer both.

Assessment: We evaluated the answers to the questions by assigning a 7-point Likert scale to them. For all questions, the participants were asked to answer from 1 "very bad" to 7 "very good."

5 RESULTS

5.1 Effectiveness: Task-specific Assessment Using Realistic O&M Tasks

5.1.1 Overview of Performances. As described above, participants were asked to perform O&M tasks by replying to a series of questions (OM1 to OM7). All participants were able to identify and hear the cars in the two virtual worlds (OM1). No student failed entirely. Out of the 13 participants, some needed more time or gave a wrong answer before eventually changing to a correct answer: 0 participants (questions OM1 and OM6 in VR#1 and VR#2), 1 participant (OM4 in VR#1 and VR#2, OM5 and OM7 in VR#1, and OM3 "Question: which configuration?" in VR#2), 2 participants (OM2 in VR#1 and VR#2, and OM5 and OM3 "Question: how many streets?" in VR#2) or 3 participants (OM7 and OM3 "Question: traffic lights?" in VR#2). The average success in the tasks, as shown in Figure 12, ranges from 82% to 100%.

5.1.2 Performances in VR#1. One student became bored of the O&M tasks after testing only VR#1 and asked to stop the experiment. We decided to keep the results for VR#1 even though this

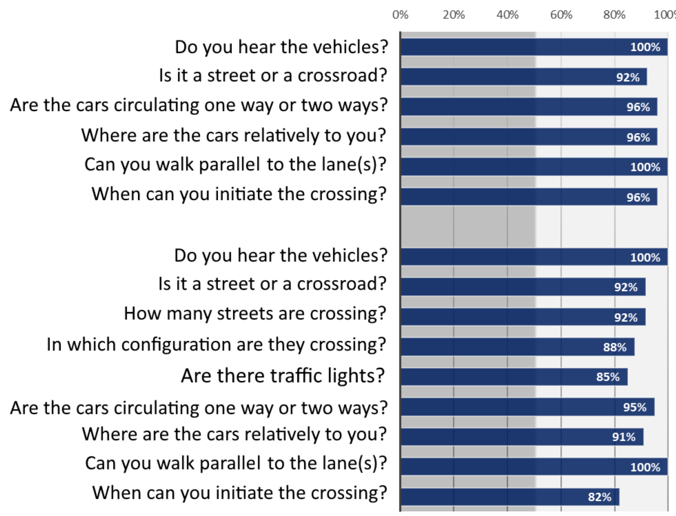


Fig. 12. Effectiveness of performing O&M tasks: success rate for the questions and tasks on average across the students (n = 13) for each question. Grey part (left half of the graphic): values from 0% to 50% on average, i.e., failure or delay in response. White part (right half of the graphic): values from 50% to 100% on average.

student was not included in the VR#2 result. The reason for not excluding this participant from the analysis was to avoid keeping only the participants that enjoy O&M classes.

In the virtual street world VR#1, all participants were able to walk parallel to the street (OM6). This is surprising, given that not all were able to respond immediately to the question "Where are the cars relative to you?" (OM5). We identified two reasons for this. The first one is that the question of positioning the cars was asked and answered before walking parallel to the street. Second, the discrimination of front and back using spatial sound is not easy, as the perceived sound is the same in terms of intensity, latency, and left/right balance [76]. However, to walk parallel, the students kept the sound on their left or right ear, which avoids such audio artifacts.

For the question "Is it a street or a crossroad?" (OM2), some students first said a crossroad and corrected it to street after being asked "How many streets are crossing?" (OM3), as they could only identify one street. A participant also confused a mower in the real world with a car in the VE.

For the question "Are the cars circulating one way or both ways?" (OM4), students temporarily made errors when there currently were no cars circulating in one of the lanes, i.e., when the sound was identical to a one-way road. As explained above, VR#1 uses a random generation of time without cars for each lane to make the right moment for crossing unpredictable.

All students crossed the street at the right time at least once. However, some students crossed but "hit" a virtual car at the end of the crossing. They identified the time without cars but underestimated the crossing time, so the time was not long enough even in a break without cars. Some students said that in real life, the driver of the car would have stopped, since he/she would have seen somebody crossing.

5.1.3 Performances in VR#2. For VR#2, we had 12 participants, since one student dropped out, as explained above. VR#2 is a traffic light crossing with two streets crossing perpendicularly. Due to the traffic light, there is only one street with cars moving at once. For this reason, some students confused the setting with a street initially, but changed when they heard the change of sound as the cars on the second street started moving. For the question "How many streets are crossing?" and "In which configuration?" (OM3), the correct answers were four branches in a cross configuration.

In a real-world setting, cars would also turn left or right at the crossing. However, in our VR world, the cars came from only two branches out of four at the same time (for instance, from south and west) and only went in one direction (south to north, west to east). No car came from east to north or south, no car from south to east or west, and so on. The students needed some time to identify precisely how the cars from the two "source" branches continued along two different branches, in particular when they were not driving at the same time.

For the question "Are there traffic lights?" (OM3), there were two strategies depending on vision level. For the blind students, the first O&M strategy was to count seconds to verify whether the changes in lane driving were regular or not. Some students needed some time to figure this out. One other O&M strategy was to use an audio model of the car, as learned in O&M classes: massive engine sounds are emitted when a light becomes green. As in other studies [78], low-vision students tended to use their remaining vision. They first put forward a hypothesis about the traffic light when they saw the cars alternating, then they looked for elements to confirm it—in particular, the traffic light color. Without this confirmation, they concluded that there was no traffic light.

This strategy by low-vision people was also observable with the question "Are the cars circulating one way or both ways?" (OM4). They saw two lanes but only cars circulating in one direction. They observed that there was no "prohibited direction" sign in the virtual world. Thus, the low-vision students first concluded that this was a two-way street. When they observed for longer, they noticed that no car ever came from the other direction. For all students (low vision and blind), the rare circulation configuration of traffic lights with only two lanes instead of three or four led to longer search times. The same audio-related factors as in VR#1 led to temporary errors in answer to the question "Where are the cars relatively to you?" (OM5). Moreover, as in #VR2, there were two perpendicular streets with cars to position, so this effect may always have concerned one of the streets (as one street may always be facing the student's front or back).

5.1.4 Instructors' Observations. The instructors confirmed that the behavior using X-Road was similar to that in real situations. In particular, the errors reflected the errors made in real O&M classes. For example, the instructors explained that there is a common error when PVI cross a traffic light crossing that also occured in #VR2. Indeed, without vision it is not natural to cross a street when the cars start in the parallel street (which means that this is the right moment to cross).

Unlike in real O&M classes, VR allowed the instructors to let the students finish their actions and experience the consequence of them. For instance, in understanding traffic light road crossing mechanisms, VR allowed the students to cross and check if there had been a collision with a car. This helped participants to understand why the crossing has to be initiated when the parallel cars start to drive (i.e., when the traffic light just switched). Some students used VR to observe the crossing from the middle of the street, which is not possible in real situations. The instructors argued that this opportunity for trial and error enhances the spatial representations of the students.

5.2 Strategies to Improve Spatial Understanding

We observed that students employed different strategies to improve their understanding of space. Some students applied a perception extension using the smartphone. Instead of making the same movement of the head and the smartphone to get consistent feedback, some students moved the smartphone only and were then able to position the virtual feedback in the world depending on the hand movement.

One student had trouble locating himself in the crossroad. He used a magnet board (Figure 13) to locate the cars and lanes, as in classical O&M classes. The idea is to position magnets to represent the branches of the streets and thus construct a map. The student can then point on the map where he thinks he is located and the teacher corrects him if needed. Using this, he was able to analyze and understand the crossing.

Fig. 13. A blind student uses a magnet board as a complement to improve his understanding of the VR#2 world. Left: The student constructs the representation of the crossing with magnets on the metal surface. Right: The student explores his own tactile representation to understand where he is on the crossing.

5.3 Assessing Immersion and the Realism of the VR System

5.3.1 Results. The students rated X-Road as immersive and comfortable, and neither realistic or unrealistic. The realism average values are comparable to Reference [77]. The results are presented in detail in Figure 15. Regarding the global picture, the students were in general captivated by the virtual world (Mean = 6, SD = 1.42 on a 7-point Likert scale), found the system easy to use (Mean = 6.15 SD = 0.78), and would be interested in re-using a similar system (Mean = 6.31, SD = 0.96). The virtual world is not perceived as more real than the physical world (Mean = 1.23, SD = 0.56). The mean score for distinguishing the real world from the virtual world is slightly negative (Mean = 3.62 and SD = 1.92), and the consistency experienced in the virtual world compared to the real world is slightly positive (Mean = 4.54, SD = 1.76). Students felt present in the virtual space (Mean = 5, SD = 1.92) and felt the virtual environment surrounding them (Mean = 5, SD = 1.82). However, the students found the standardized questionnaire strange and redundant and the vocabulary unclear.

5.3.2 Instructors' Observations. The instructors noticed that the students were forgetting the real world due to a lack of attention or perception of the real world. Indeed, students had to be stopped before running into the walls or obstacles, as they were focusing too much on the virtual world or not perceiving the real world due to the headset. The instructors also noticed some events that were positively experienced by the students. For instance, collisions with cars were not traumatic experiences. The students also tolerated bugs such as flying cars when the camera lost the scene, as the system is not too realistic and not as immersive as VR goggles. In particular, a car drove through a second car as if there was no object at all at this position. A low-vision student saw this bug in X-Road and pointed it out to the instructors as a funny experience.

5.4 Satisfaction: Assessing the Usability of the System (Interest, Comfort, Re-use)

5.4.1 Results. As can be seen in Figure 14, X-Road was rated as comfortable (Mean = 5.08 out of 7). However, participants also mentioned a few shortcomings. Three participants said that the synthetic engine sound was not pleasant, and hearing the same sound during the whole simulation was not very comfortable. Two participants reported fatigue in the arms while holding the device in their hands (top-left configuration in Figure 5). One student stated that he or she sometimes accidentally covered the camera with his/her finger, so the position of the virtual world in the environment was lost. One participant estimated he/she would not use X-Road for more than 1.5 hours (he/she used it for one hour during the experiment), but did not specify the reason. Three students would prefer earphones instead of the headset, including one participant who really

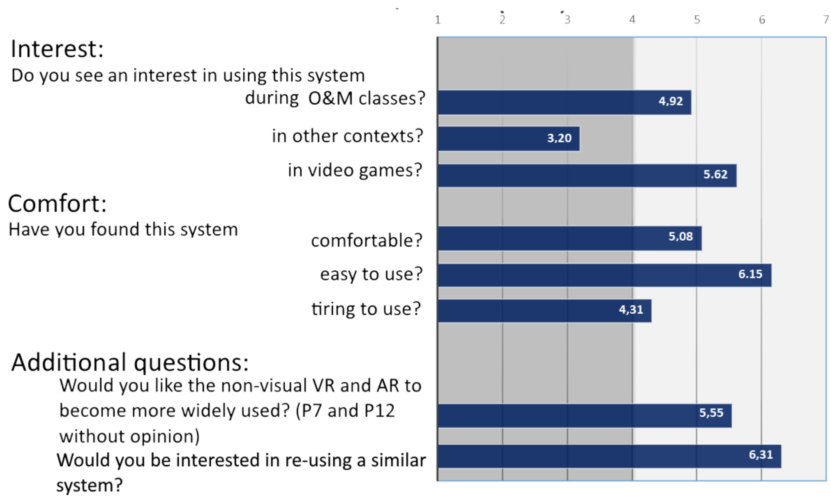


Fig. 14. Results of the questionnaire about X-Road in terms of interest, comfort, and additional questions regarding VR. Average rating over all students (n = 13) on a 7-point Likert scale. Responses above 4 (white area) are positive except for the question "tiring to use."

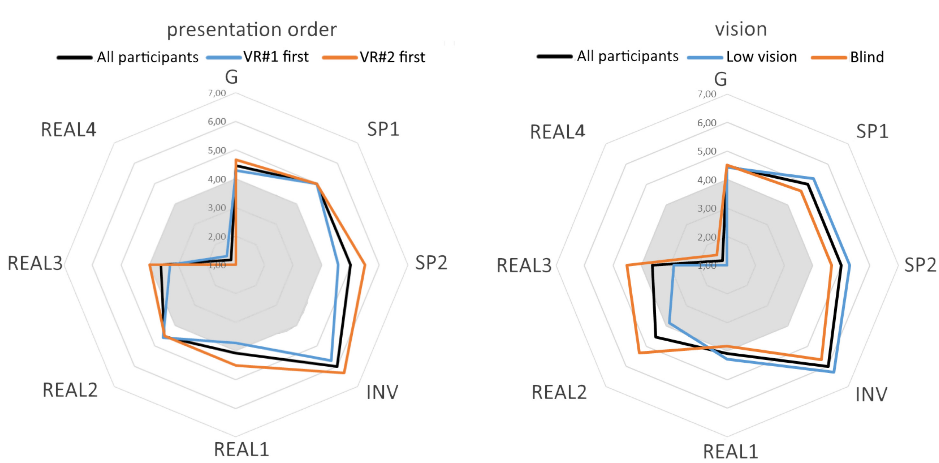


Fig. 15. Results of the VR Immersion and Presence Questionnaire, depending on presentation order (left) and vision (right).

disapproved of the headset, as he felt cut off: "It obscures the ears, smaller earphones exist" (P8). Moreover, we noticed that when the students were wearing headphones, instructors sometimes needed to repeat questions or instructions, as the students could not hear them properly. This is consistent with other studies with PVI, in which the safety of hearing the environment was a factor in the choice of smaller earphones, such as bone-conducting headphones [9, 37]. Regarding the comfort of wearing X-Road on their head, one participant said that foam improves comfort (P12). One participant suggested extending the use of our system to guidance in the real world. However, X-Road would not be directly usable as an AR guidance system, as the visual window of the virtual world (the smartphone screen) is too small and participants' hands have to be free to handle the white cane while walking. There was no need to plug in the X-Road smartphone or to

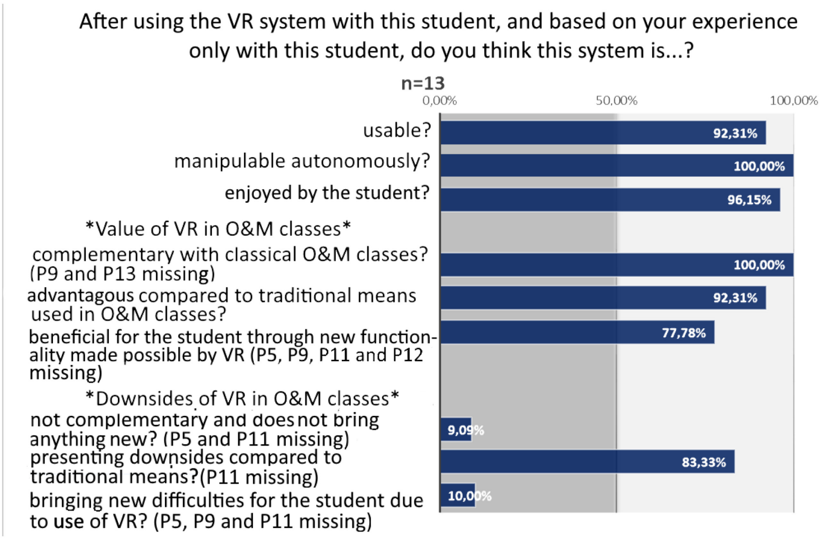


Fig. 16. Evaluation of the usability of our system by the instructors for each student (n = 13): Average results of the intructors' questionnaire about the system.

connect it to an external battery during the day (9 hours of use including 4 hours and 30 minutes using the VR and AR applications with students). Less than half of the battery was used at the end of the day.

5.4.2 Instructors' Observations. Instructors ran O&M classes with the students. Therefore, the situation in the virtual world needed to be collaboratively perceived by the students and the instructors. The instructors were able to observe the students' use of the VR world when they were holding the device in their hands by looking at the screen (see Figure 7). Thus, the instructors knew the situation of the student in the VR world and could provide pedagogical feedback and corrections. However, when students were wearing the system on their head, this was not possible.

Another limitation of the current prototype is that for instantiating the world, only visual feedback exists (see Figure 8). Therefore the instantiation cannot be done by visually impaired people themselves.

5.5 Satisfaction: Assessing Whether the System Is Compatible with O&M Classes from the Instructors' Point of View

According to the answers to the three first questions about the possible use of our system (Figure 16), and in particular its use in O&M classes, the instructors felt comfortable using our system in O&M classes. The main advantages identified by the instructors are the following:

- **Modeling and controlling the situations presented:** **(1)** X-Road provides the possibility to create scenarios with controlled complexity (ideal and not ideal situations) and dynamic events (for instance, a motorcycle overtaking when the user crosses a street because a car has stopped). This was explicitly mentioned by the instructors in the questionnaire answers for 6 of 13 participants. **(2)** X-Road could have a repository of situations (mentioned for 2 participants), **(3)** It could have a minimal situation for learning specific elements separately (3 participants), **(4)** X-Road could help to prepare a journey (1 participant).

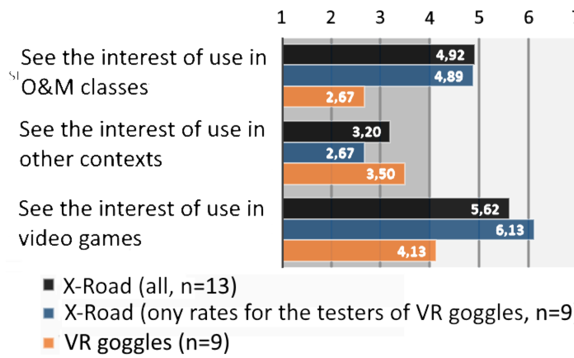


Fig. 17. Comparison of the participants' answers on a 7-point Likert scale about interest of use of X-Road and classical VR goggles in O&M training, other contexts, and video games after using both with the same virtual worlds.

- **Safe and stressless situations for the students: (1)** X-Road can provide safe and predictable environments and situations (2 participants). **(2)** It can allow students to test forbidden behaviors (2 participants). **(3)** It can reduce stress and help the student to better learn how to analyze a light crossing (3 participants).
- **Convenience of use and portability: (1)** X-Road can be used to train students on configurations without having to physically move from one place in the city to another (1 participant). **(2)** It is usable in case of bad weather or bad conditions (1 participant). **(3)** It could also be used in the context of presenting road safety and mobility for PVI to a larger public (1 participant).
- **Teaching and learning facilities: (1)** X-Road can help to explain the basics of audio analysis, orientation, and crossing techniques in lighter settings (2 participants) and can be a way to test audio analysis skills, as the skills in the VR system appear to be consistent with those used in real life (1 participant). **(2)** It is a modern approach for O&M classes, in particular for students with an interest in digital technologies (1 participant). **(3)** Feedback can be given by the system and not the instructors, so the student does not take it personally (1 participant), similarly to what has been found in previous studies [2, 70]. **(4)** X-Road can integrate play with challenges and test activities, offering engagement as in video games (2 participants).

The O&M instructors also identified a number of possible problems:

- Excitement may lead to fatigue during use (1 participant).
- The real and virtual worlds may be confused (1 participant). This was mentioned because one student struggled to establish a clear difference between the virtual and real worlds, in particular with the standard VR goggles.
- The transposition of skills has yet to be verified in the real world. In this perspective, modeling typical crossroads may be more efficient than modeling particular existing ones (2 participants). Indeed, the idea is to teach students how to cross on one virtual crossing, with the aim of them being able to cross multiple real crossings with the same configuration as the virtual one. However, we have not investigated this question in our current study.
- The student may be too keen to play (1 participant), so the pedagogical interest is uncertain.

- The sound needs to be improved (3 participants), as the artificial engine sound is short and repetitive. Moreover, some participants reported that it sounds more like a plane than a car. Also, all cars have the same engine sound, and the transitions between the sounds for different engine statuses (starting, stopping, braking, acceleration, engine idling) are not smooth.
- The visual simulation might not be realistic enough (1 participant). One student who relies heavily on visual feedback felt that he/she learned nothing in the virtual world and focused on the differences with the real world (i.e., no driver stopping when he/she crosses). Moreover, the visual simulation should provide all visual clues used by PVI, including different colors for the cars to differentiate them, signs (stop, one way, prohibited direction signs), traffic lights emitting colors, road markings, and lane width.

5.6 Role of Visual Feedback

In daily life, the smartphone screen is widely used by all students participating in the experiment and in general serves as a magnifying glass for people with residual vision. Some students with residual vision also used the screen as a magnifying glass in our study. More specifically, they looked very closely at the screen to observe distant elements not observable in reality. Considering the objective of developing spatial cognition, this is a positive aspect. As a perspective for our system, a student suggested using it to see distant exhibits in museums. Visual cues are used for environment analysis (e.g., looking at traffic signs to understand the configuration) and for situational analysis (e.g., looking at drivers' faces before deciding to initiate the crossing). During the experiment, the instructors asked the students "Do you see the car? What color is it? Do you see the pedestrian crossing?," and all low-vision students answered correctly.

The results on immersion and realism (see Figure 15) show that in our panel, the order of presentation of the virtual world has little effect on the results, while the level of vision has an impact on perceived realism. In particular, for the question on consistency, VR was rated lower by students with low vision. This might be related to the above-mentioned lack of visual realism of the VE. First, all low-vision participants mentioned that all cars had the same color, while none of the sighted users noticed this. The low-vision students use visual differences between similar elements to identify and "separate" them visually. This remark was made in the first few minutes of use, when they identified the two traffic lanes. Some participants temporally gave wrong answers when they used only visual feedback. This includes using the width of the road to conclude that there are two lanes (which is true) and inferring that it is a two-way road (which is false), because they found no one-way sign. Low-vision participants explicitly said they observed the painted white markings on the road to identify traffic patterns. They also looked for brighter light signals to infer traffic lights. All low-vision participants commented that there were no drivers in the cars, while none of the sighted users noticed this. They were looking at the drivers at the pedestrian crossing. In the real world, establishing eye-contact allows pedestrians to know that they can cross safely.

Using our system, participants with low vision could have an awareness about the real world in parallel to the virtual world. Nonetheless, most of the low-vision users held the screen very close to their eyes, limiting real-world perception. One can imagine that seeing the real world could be helpful for O&M activities; for instance, for orienting oneself parallel to a street. But seeing the real world does not help to understand the position and orientation of the virtual world, to analyze the elements inside it (street and crossing configuration), or to decide when to cross. The instructors also positioned the virtual world not to be parallel nor perpendicular to the windows. The impact of seeing the real world is limited overall, and only for low-vision students. We additionally checked participants' attention to the real world with the immersion measures, which are high. Finally, the instructors had to stop the students several times to prevent them from hitting walls, doors, and furniture. This suggests that they paid little attention to the real world.

Visual feedback was used by the instructors to find out more about the situation; in particular, about the traffic (see Figure 7). For students with low vision, the instructors took this opportunity to hide the screen, so the students needed to rely only on audio feedback for crossing the street. The instructors said that "by night, they should be able to cross the street only with audio feedback."

5.7 Case-study: VR Goggles

Out of the 13 participants, 9 voluntarily tested #VR1 and #VR2 with 3D VR goggles after the experiment with our X-Road prototype (see Figure 17). The students mostly preferred the goggles (5 out of 9) to X-Road (3 out of 9), and 1 participant liked both. However, they generally rated the goggles as less usable in O&M classes and in video games. For these nine students, the average interest for O&M classes was 4.89 (SD = 1.73) for X-Road and 2.67 for the VR goggles (SD = 1.63) on a 7-point Likert scale. The average interest for video games was 6.13 for X-Road (SD = 1.04) and 4.13 for the VR goggles (SD = 2.38). The students explained that the sensation of being immersed was stronger with the goggles, but they think the ability to move using X-Road (but not the goggles) is very important for O&M classes and video games. Use in other contexts is rated better for the VR goggles (M = 3.5, SD = 2.34) than for X-Road (M = 2.67, SD = 2.12). Ideas for other contexts included virtual visits of unknown spaces. The participants' feedback suggests that the VR goggles are preferred—in particular, by students with low vision, because they offer the impression of really being there, even if the current implementation of the VR goggles did not allow them to navigate in the virtual world. According to the participants' explanations and the interest and preferences rating, we infer that immersion strongly influences preferences, while movement is a key feature for providing more interactivity. As mentioned above, instructors also underscored that a system that is not too immersive might be advantageous, as it allows communication between students and teachers, and it avoids exposure of students to stressful situations.

6 DESIGN RECOMMENDATIONS

Through the iterative design process, we were able to gather feedback about different versions of the X-Road prototype. Here, we provide design recommendations based on these experiences.

6.1 Audio Feedback

Basic audio features are sufficient for creating accessible VR for PVI. In our X-Road prototype, we decided to use the basic audio features available on the Unity video-game engine. Nonetheless, the VR appeared to be usable by PVI for learning spatial configurations. We therefore suggest that even low-fidelity audio can be sufficient to create O&M training simulations.

Provide various types of feedback for similar objects. From the users' feedback, it appears important that cars and vehicles have differentiated engine sounds as in real life. The sound of a particular car can then be identified and followed when it approaches.

Implement a system so the user can move to avoid audio artifacts and to help locate elements in space. Walking and moving physically in space may help to spatialize the environment. This is particularly true for the discrimination of sound from the front or back. Spatial audio feedback from an object in front is the same as for an object behind a person (front-back reversals). We observed that street localization at a stationary position in the environment avoids ambiguity as long as the user moves his/her head or knows the relative position of the audio source in the world.

Set a sufficiently long audible distance of the sound. During pretests, instructors indicated that the sound of cars should be audible from far away (200 m) so the student can align with the cars' trajectories.

Ensure the Doppler effect is handled dynamically. According to the instructors, it is also necessary to implement the Doppler effect. This makes it possible to tell the difference between moving objects and static objects with audio sources in the simulation. The Doppler effect can be automatically calculated in the basic audio features of Unity from an audio source file.

Use mono sound source to ensure interactive computing of audio sources. When users move in VR, they need dynamic feedback: 3D sound depends on the user's position and head movement. To compute the audio effects (interactive stereo, 3D sound, Doppler effect) in real time, Unity uses mono files as audio sources. As described previously, binaural and ambisonic records are not appropriate, as they are only consistent in one head position. However, the sound in VR has to be computed whatever the position of the user's head and movement are.

6.2 Visual Feedback

Provide a realistic visual simulation for low-vision users. Interestingly, users with low vision noticed errors in the visual representation and proved to be very sensitive to visual clues. This is in line with previous studies [78]. Inconsistency in terms of signs (for instance, the absence of prohibited direction signs in one-way driving situations) were immediately detected. Moreover, low-vision participants also commented that the cars had no drivers, as in real life they are highly reliant on driver behavior clues to cross the street. To conclude, we suggest that a VR world has to be well modelled in terms of visual details and behavior if the VR is to be used by people with low vision.

Provide various types of feedback for similar objects to increase realism. All low-vision users noticed that all the cars were the same color. None of the sighted people noticed this.

6.3 Usability, Pedagogy, Immersion, and Preferred System

Keep the system captivating but not too immersive. In the context of O&M training, it can make sense for a VR system not to be too immersive. One advantage is that trial-and-error situations are not too stressful (such as hitting virtual cars). Moreover, communication with the O&M instructor can be maintained more easily.

Allow users to easily switch visual feedback on and off. In a teaching context, it may make sense to switch visual feedback on and off as required. Indeed, low-vision students may rely heavily on vision. But in real life, they may find themselves in situations where they cannot rely on vision (e.g., by night). Switching visual feedback on and off allows users to simulate a situation where vision is unusable.

7 DISCUSSIONS AND PERSPECTIVES

7.1 Technical Issues

Our current system is a first exploration of inclusive VR for O&M training and has some technical limitations. Movement in the world as detected by ARCore relies on movements seen by the camera. If the camera is covered for more than 2 seconds, if it is too close to a unicolored surface (no optical capture of the movement), or if there are too many successive and rapid movements, then the virtual object may disappear. This can be resolved by physically moving back to the approximate place where the application started so the system can locate the virtual object again. It is also possible to simply stop the application and start it over before instantiating a new virtual object at the same position with the same orientation.

Another issue with image recognition occurred during a cloudy day. The room was symmetric and the two sides of the room had identical lighting conditions. Consequently, the system flipped the virtual world as if it was on the other side of the room. We oriented the camera in the direction of a non-symmetric element of the room to make the virtual world come back to its correct orientation.

7.2 Augmented and Virtual Reality as Training Devices

We noticed that active learning may be helpful to understand virtual feedback. For instance, some students said that they did not understand where the cars were. But when asked "Tell me when a car is close to you. Where is it?," they were able to answer. Similarly, AR Cube Hunter proved to be helpful for understanding the position of objects in space based on audio cues and could be used to teach generic audio-spatial concepts. Indeed, some students struggled to understand why the sound was louder in one ear than the other. When the virtual object was embodied by a physical object or person (Figure 9) around which they moved, they understood the cube's position relative to their own position ("I hear it louder in my left ear, when the cube is to my left").

Finally, VR provided the opportunity to do forbidden things, such as walk on the line between two lanes and stopping in the middle of a crossing. While it is dangerous or impossible to do this in real life, in VR this was useful for some students to understand the logic of the traffic light crossings.

7.3 VR, Realism, Immersion, and Pedagogy

This work demonstrates that VR can be accessible and usable even though sound is not perfectly realistic. A parallel can be drawn with the visual feedback in existing VR systems. Sighted people can deal with non-realistic and/or low-resolution visual feedback in the same way that PVI can deal with low-fidelity audio.

Contrary to our initial idea and based on the instructors' feedback, we think that the system should not be too immersive. VR goggles were rated as more immersive than X-Road by the students. However, students were more easily able to take a step back with X-Road. The connection with the instructor is more natural using X-Road, while when participants are wearing goggles the instructor's questions come from outside the world. Moreover, we believe that more realistic and immersive solutions would exacerbate fear in stressful situations; for instance, in a situation with heavy traffic.

Regarding the results of the questionnaires for the instructors (Figure 16), the system was rated as usable with some possible improvements. We tried different solutions for improving the sharing of the screen with the instructors. These required an additional computer and manual settings (WLAN network) or an internet connection (ARCore Cloud Anchor). We observed that these solutions were too complex for autonomous use by the teachers and that the internet connection in the school was not good enough. However, from the feedback from the instructors, we learned that it would still be important to have a shared point of view. A solution could be to broadcast the virtual world or to depict the teacher in the VE, e.g., using an avatar.

Finally, instructors believe that VR supports the development of transferable skills, i.e., skills not associated with a particular crossing. However, we have not evaluated the transfer to the real world of skills acquired during VR use.

7.4 3D Model as Inclusive Support

We have seen in the "Related Work" section how spatial representations are amodal and can be similar for people with and without visual impairments. A parallel can be drawn with VR, as it aims to represent environments with 3D representations: the equivalent of the spatial representation is a 3D Mesh, the visual feedback is the colored texture, the audio source emits sound, the normal map is the tactile texture, and so on. The simulation of the virtual world, including physics and dynamics, is included in the game engine. While in most VR applications the feedback generated is mainly created for vision, the 3D models for VR can act as an amodal support to provide appropriate and inclusive feedback to the user using vision, audition, and haptics.

7.5 VR Culture, AR Culture, and Inclusion

AR and VR are today mainstream for sighted people. However, PVI are largely excluded from them. Even for low-vision people with remaining visual capacities, these technologies are rarely accessible. We observed that students consequently had misconceptions about how these technologies work. For instance, one student asked if he would still be physically in the room when he was visiting the virtual world. During the use of the familiarization application (AR Cube Hunter), the students asked “Why can I hear the cube but not touch it?,” “Why can I only see the cube on the screen and not in the room?.” Our system allows instructors to explain and experience these concepts so the students understand them better.

Moreover, it is sometimes difficult for blind students to understand the utility of having a wearable device. For instance, one student asked “Why are you not filming the room, detecting me, and sending the appropriate audio feedback?.” Our inclusive system raises some questions for blind people, such as why there is visual feedback. This shows that a common representation of feedback for sighted people and PVI is not so easily achievable.

As AR and VR are today mainstream, with games such as Pokemon Go, the fact that PVI experience and understand them can improve their integration. AR and VR are currently used extensively for entertainment by sighted people. The students were interested in VR video games, and we intend to investigate this in our future work. However, the positive results from our study and the enthusiasm of the participants may partly be due to the novelty effect, and long-term use may lead to a drop in interest.

7.6 Perspectives for Improvements

Despite a low-fidelity audio rendering, X-Road appeared to be usable by PVI for O&M training. However, by using the Resonance Audio Unity package,[15] it would be possible to create more precise sound rendering than the one offered by default features. One major improvement would be to implement smooth audio transitions between the statuses of the same object. As described above, when there was a change in behavior of the car (driving, braking, accelerating, stopping, and starting), it was very complicated to make smooth sound transitions. Thus, participants explained that they first thought that cars popped up out of nowhere because the audio file was changing. For students with low vision, this is a less serious problem, as vision allows them to see that the cars do not appear or disappear. Transition functions exist in Unity with animation through interpolation, while this is not the case for sound. We think that it would be necessary to provide such functions. Implementing rear and front audio differences is also a perspective for improvement. The instructors noticed that there is almost no difference between the rear and front sounds, as mentioned above. One technical solution would be to model this more realistically using a Head Related Transfer Function that models the specific sound reception signature related to an individual’s body.

For more collaborative use (with O&M instructors and other students), we suggest that it would be necessary to broadcast the student’s screen or to share the virtual world. We envision implementing this feature in our future work using ARCore Cloud. To make the application

[15] https://resonance-audio.github.io/resonance-audio/.

autonomously usable by people with visual impairments, another type of feedback than visual only would need to be provided for the start of the application.

7.7 Making X-Road Accessible to the General Public

X-Road is implemented as a smartphone application using mainstream devices to make it accessible to a large proportion of the population. However, for legal and intellectual property reasons in France, we cannot share the application directly. We are currently working on writing a tutorial to allow a broader audience to create their own street simulator applications. A tutorial for creating the AR Cube Hunter application is already available online.[16].

In our study, the participants nearly collided with walls and objects when they were moving around and paying close attention to virtual objects. As for any VR or AR system, we recommend that all users should be careful when using the system and encourage them to use it in collaboration with another person for greater safety.

8 CONCLUSION

In this article, we propose X-Road, an accessible VR application using mainstream hardware: a smartphone, the video-game engine Unity with ARCore, and headphones. This application presents two VR worlds—a street crossing and a traffic light crossing—to be used in O&M training for people with visual impairments. It allows users to move in space at a scale 1:1, without any external materials or settings, thus making it usable in any space. This low-cost and mainstream system was evaluated as usable by 13 blind and low-vision students and 3 O&M instructors in O&M classes. The students successfully analyzed a 3D world, performed O&M tasks, and learned by trial and error. The O&M instructors felt capable of using the VR application and system autonomously, as it is a smartphone application, i.e., it is installed on a familiar device. Finally, based on the results of this experience, we present design recommendations for inclusive VR systems, such as providing realistic visual clues for low-vision people.

Future work will offer a better understanding of inclusive VR for PVI. While low-fidelity audio proved to be sufficient to create a usable and accessible VE for O&M training, there is still a lot of room for improvement in terms of the audio quality of X-Road—for instance, the use of mainstream tools like the Unity Resonance audio package.

A APPENDIX: DETAILED ITERATIVE PROCESS

A.0.1 #0 January 2018–March 2018: Initial Demand, Modeling, and Programming the First Virtual Street. In January 2018, three O&M instructors (the project initiator and two colleagues) from IRSA Alfred Peyrelongue, a school near Bordeaux for students with visual impairments, asked us about the possibility of creating a VR crossing simulator for O&M classes. With the instructors, we defined several elements that would be relevant in the virtual world. The first environment to be created was defined as a street with a pedestrian crossing but without any traffic lights.

Based on these steps, we then designed a virtual street as a basis of our future simulator with Unity. This version was implemented as a computer video game. As previously described, the Unity framework allows versatile code, and this was compatible for later use with HTC Vive or other VR displays. Moreover, we tested various features: (1) providing different weather conditions (sun or rain), (2) modifying the hour of the day or night with light and shadow modifications, (3) controlling the number of virtual pedestrians autonomously walking on the virtual street, (4) dealing with collisions between the user/player and the virtual vehicles, with virtual pedestrians and with a type of floor ensuring that the audio feedback was adapted.

[16]https://sites.google.com/ensc.fr/laurenthevin/scientific-mediation/accessible-treasure-hunting.

*A.0.2 #1 April 2018: HTC Vive Virtual World in 4m*5m Tracked Space.* We adapted the above street model to be executable with HTC Vive VR goggles.[17] This system works with a headset, two hand controllers, and external base-stations for tracking the orientations and positions of the headset and the controllers in the play area. The main software change was an adaptation of the user interface from mouse and keyboard to HTC Vive. The tracked space was limited (typically from 3m*3m to 6m*6m) and relatively easy to set up through a tutorial. We transcribed movement in the real world to movement in the virtual world on a 1:1 scale. For instance, the user moved one meter in one specific direction in the virtual world if he or she did so in the tracked area of the HTC Vive.

For this first step, in collaboration with the instructors, we decided to continue working on the prototype with the simulation in sunny weather (no rain), during the daytime, and without virtual pedestrians to avoid a heavy workload and to keep users within a comfortable training situation. Indeed, for the proof of concept and for this first future experiment, the O&M instructors decided to work further on scenarios that are accessible to a large number of students. This choice was made bearing in mind that a limited number of scenarios would be implemented and that these scenarios would be tested in O&M classes with students who were not selected based on their O&M level.

One of the O&M instructors tested this HTC Vive prototype and this virtual environment. The instructor validated the VR street with the pedestrian crossing as suitable for simulation in O&M scenarios. However, one result of this step was the fact that a technical and specific system like HTC Vive was not suitable for O&M classes, according to the feedback from the O&M instructors. First, the cost would be too high (around 600 euros or dollars for the HTC Vive, around 1000 euros or dollars for the HTC Vive pro starter kit) for just one specific use. Next, the system would have to be installed at a certain height; in particular, the two sensors defining the tracked space. Moreover, the instructors did not feel confident that they could deal with such a technical system alone; in particular, in the event of bugs or crashes.

In terms of use, along with the instructors, we decided that the 4m*5m tracked area was too small for a street crossing. Moreover, in this version, the virtual cars were audible from 0 meter (very loud) to 50 meters (no longer audible). The O&M instructors pointed out that while the car sound should be much louder closer to the user, it was also necessary to hear the cars from far away (around 200 m) for users to be able to align themselves with the direction of the traffic flow.

Based on the findings in this step, in the next iteration, we decided to explore how to design a portable version without external sensors.

A.0.3 #2 Beginning of May 2018: Cardboard with Various Movement Features in VR. We adapted the simulation to be compatible with the Google VR Cardboard[18] system running on a smartphone with an integrated pedometer (smartphone OnePlus 5 A5000 128 GB, 8 GB RAM). The Cardboard VR platform allows the creation of smartphone apps that are portable on Android and iOS. Cardboard computes the smartphone's position from the smartphone's internal sensors. The smartphone's coordinates in a VR headset correspond to the position and orientation of the head of the user and are used to display the virtual world. The screen is split to render two images for stereoscopic vision. The Cardboard headsets (starting at around 10 euros) are affordable and consist of boxes with lenses to hold the smartphone so the two slightly different images for stereoscopic rendering are each visible to one eye only. The main goal of this step was to develop a portable solution, without any electronic devices or sensors other than a smartphone, that would be easy

[17] https://www.vive.com/us/product/vive-virtual-reality-system/.
[18] https://developers.google.com/vr/discover/cardboard.

for the instructors to use. Instructors reported that they felt comfortable launching a smartphone VR app and restarting it in the event of bugs.

This implementation used the integrated Cardboard features. Cardboard natively allows the user to move their head but not to move in the virtual world. We added features so the user could move in the virtual world. Without vision, the user cannot update his or her position in the virtual world if the movement in the virtual world does not correspond to the movement in the real world. For instance, the user can walk and come back to the starting point in the real world, but will not come back to the starting point in the virtual world and will get no feedback about it. For this reason, we included two features: movement in VR depending on the smartphone's accelerometer movement detection, and movement depending on the smartphone's step detection to initiate a 40 cm movement in VR for each step. The main idea of these features was for the user to be able to know how he/she moves in the virtual world. In other words, we sought to compute the user's movement in the real world only with the smartphone's sensors to transpose the user's movement in the virtual world. In such a configuration, the system would remain easy to use without external sensors or installations, unlike the HTC Vive installation with its base stations. However, the lack of precision of movement detection with the accelerometer led to unexpected movements and it was therefore not suitable for PVI. For step detection, we realized that moving in the direction of the head was not very convenient in the absence of vision, as the direction of the step might be different from the head orientation, and it was not possible to update the new position without vision. PVI move their head to discriminate sounds (i.e., their ear rather than their eyes would point in the direction of the target). Moreover, step sizes may vary inter- and intra-participants. Finally, this system needs to be used in real time, while a pedometer has a two-second latency for signal processing reasons. Overall, the false detection of steps or movements induces an unexpected change of position in the virtual world, and this is very confusing without vision. Testing this system also allowed us to note sound elements that needed to be improved. All cars used the same free engine sound with a two-second sound loop. We noticed that all cars in the world started at the same time, upon the initialization of the virtual world, so the engine sounds were all perfectly synchronized. This made it difficult to isolate the sound of a single car and to follow it, thus making it harder to understand the geometry of the road. At the next iteration, we consequently added a random delay for all cars before the sound started up.

This system was first tested within the scientific team before moving to the next iteration.

A.0.4 #3 End of May 2018: Vuforia Augmented Reality Prototype. While the use of a smartphone in the previous iteration was appreciated, it did not allow us to detect movement and convert it 1:1 to the virtual world. Therefore, we used Vuforia to detect movement relative to images used as targets to instantiate the virtual objects. In practice, the smartphone camera films the environment, then targets are detected using Vuforia, which runs on the smartphone. The virtual objects are instantiated at the position of the images, and Vuforia computes the distance and orientation by image processing: The smaller the target image detected by the camera, the further it is from the smartphone. Distortion of the image enables a calculation of the orientation. With this technique, instead of computing the movement of the user in the world, we compute the distance of the object in the world from the user. Moreover, if two objects (or more) are visible at the same time in the environment, with Unity they are all positioned in the same virtual scene depending on the computed position relative to the smartphone and the relative position between objects. This makes it possible to reconstruct the world as long as at least one object is visible through the smartphone camera and provided that this object is correctly positioned. Instead of using fiducial markers such as QR codes, we decided to use images representing elements in a virtual scene or in the virtual environment. We believe that using figurative images is easier for teachers than abstract

QR codes or fiducial markers. An A3 image was used to display the street, another to generate a flow of vehicles, and others for buildings and trees. We could also display the virtual object in stereoscopic rendering by applying the split-screen features used with the Cardboard application. Next, we also tried to move around these virtual objects when the image was no longer visible to the camera. This technique would have been suitable for creating a scene dynamically by adding target images in the world (for instance, creating a road and creating a second one to define a crossing between the two). However, we observed that although this principle was appreciated, it was not robust enough for correct positioning: The elements "jumped" from one place to another or from one orientation to another. The error in positioning the virtual object was particularly clear when the position was inferred without the target image being in the field of vision of the smartphone camera, and also when objects were moving in the background (such as other people) in the field of vision of the camera.

From the first implementation of iteration #0, we used the native features of Unity for 3D sound rendering. In this iteration, we were able to verify that this level of audio feedback was satisfactory to acquire a spatial representation for moving objects such as cars. Several people tested this version of our systems: five O&M instructors including the three involved in the origin of the project, three sighted people working in a specialized school, and two adults with visual impairments. We presented the prototype, and participants attempted to localize the cars, to position themselves parallel to the street, and to identify collisions with virtual cars from audio feedback. They succeeded, and the participants with visual impairments understood the virtual scene from the audio feedback alone.

In this version, we also introduced collision detection between the avatar of the user and the vehicles. In the Unity scene, we associated a "collider," a virtual box detecting collisions in Unity, with the camera representing the detected position of the smartphone. The vehicles also have a collider, so we were able to launch a braking sound when the application detected the user colliding virtually with a vehicle. The result of this iteration step was that the braking sound during virtual collisions may have generated too much anxiety for users. Therefore, in the following iterations, we removed all specific vehicle behaviors when a collision occurred.

A.0.5 #4 Beginning of June 2018: ARCore Augmented Reality Prototype. We carried out an in-depth inventory of AR solutions other than Vuforia, which were compatible with both smartphone apps and Unity. We chose to test the ARCore framework, as it was known to be a robust system for positioning virtual objects in the environment. The first result of this iteration was that the positioning of visual and audio feedback was robust enough, as verified first by the scientific team and later by the O&M participants and users with visual impairments during the pretest. Indeed, we were able to instantiate, directly at a position in the world and without any target images, an object emitting sound at a particular place in the world (e.g., on a table), visually displayed at the right place and audible at the right place. And we were able to get 3D audio feedback in correlation with the smartphone's movement, including intensity (louder or quieter, depending on the distance) and laterality (louder in the ear closest to the object). We made sure that elements that were not displayed did not disappear from the scene and remained audible. With our ARCore test application running on a smartphone (with instantiation of a simple audible object), we noticed that ARCore was also robust with moving elements in the field of vision of the camera (other people, for instance). The ARCore framework was validated for motionless virtual objects first. Next, we generated virtual moving objects (vehicles) with engine sounds attached to a smartphone application. This version also worked, and the ARCore framework was validated for VR use.

We noted that ARCore is not directly compatible with the previous split screen Cardboard features, as ARCore virtual objects do not follow movements in the world. For instance, if we position a virtual object on a table and then move the smartphone, the virtual object should still appear on the table, but this was not the case when the ARCore and split screen features were combined. As the stereoscopic display for PVI was already an issue identified by the IRSA orthoptist, we decided to continue without a stereoscopic display. Moreover, smartphones are already widely used by PVI and are low-risk in terms of use (glare, diplopia, eye tiredness, and cybersickness).

A.0.6 #5 End of June 2018: ARCore Virtual Reality Prototype. The main idea of this iteration was to verify that we could generate a virtual object that was actually the complete VR scene. This was a way to roll out VR using ARCore, as discussed above. The problem was that the scene was rendered as a single object and the user walked inside the virtual object instead of around it as in the previous iterations. This had consequences on computation time, as the object was very large and the device was only a smartphone. Finally, we had to verify whether the robustness that was good enough for AR was also sufficient for VR. Indeed, precise positioning is required for virtual objects measuring several hundred meters, as with such sizes a slight movement in the orientation of the object appears bigger because of angular movements.

During our pretest with a blind student, we gained new insights. The student used the prototype outside, on a courtyard at our university. He moved around in this area and corresponding movements were replicated in the virtual world. As the virtual world is bound to elements of the real world, elements such as the pedestrian crossing and the road in the virtual world have a specific position in the real world that are not necessarily ideal (for instance, close to a wall). During the pretest, the student was able to understand where the virtual vehicles were driving and was able to walk on the virtual sidewalk by positioning himself parallel to the street (i.e., parallel to the flow of vehicles). He also managed to identify the right time to cross when no vehicles were driving in either lane. He was able to move around in the street and also tested the AR prototypes. This student tested the prototype with Cardboard with the same virtual world and noticed that the VR goggles were not comfortable, as they were quite heavy and his remaining light perception (only forward, not in the periphery) was disturbed. According to the orthoptist, the instructors, and our observations in later demonstrations, users with residual vision perceived the image as distorted by the glass lens in their peripheral field of vision. This can also be a problem with diverging eyes. These issues do not exist with WoW. This consolidated our decision to use ARCore without a split screen for the classroom prototype. Moreover, we confirmed that the engine sound was not realistic. The user underlined that we should provide different sounds for engines, as in reality all cars have different sounds. Other feedback concerned the use of the smartphone. A blind person can inadvertently cover the camera or touch the touchscreen and stop the execution of the virtual world. Therefore, we created a 3D printed holder in the next iteration to allow the users to hold the device more easily.

A.0.7 #6 (Final) July 2018: X-Road - ARCore VR Street and Traffic Light Crossing with 3D Printed Holder. The main point of this iteration was to finalize the virtual worlds for the use of X-Road in the classroom and to create the smartphone holder. For this last iteration, we created the two final virtual worlds used in the experiment: VR#1 (street world) and VR#2 (traffic light crossing), which are shown in Figure 2, with a total surface area of 400×200m. These environments were created based on the scenarios designed with the O&M instructors during a workshop at our institution. As described in the illustrative Use Case of O&M classes above, specific techniques were used for analyzing and crossing in these two configurations. The main features added to create the VR#2 were car behavior to avoid a collision with the front car, and alternation of the traffic lights. For this experiment, we did not include dynamic behavior for the cars depending on the user's actions. The change for the street world (VR#1) was the generation of cars with a pseudo-randomized engine sound. We found four freely available engine sounds. For each car in the VR world, we attached one of the four sounds, each one different from the previous car. We created a traffic light crossing (VR#2). Here, we created more dynamic car behavior and statuses (starting, accelerating, braking, stopping), which were synchronized with the traffic light and with the behavior of the other vehicles. We could not find an easy solution for a smooth sound transition between these statuses. However, in our pilot study, one O&M instructor and two blindfolded people found this behavior usable. The O&M instructors used O&M techniques to verify the usability of X-Road. We asked the blindfolded people the O&M questions from the experiment questionnaires, as described above ("Do you hear the cars?," "Is it a street or a crossing?," "Can you position yourself parallel to the road?," etc.), and one of them answered the full questionnaire to verify the experimental protocol.

As presented in step #5, users struggled to hold the X-Road smartphone without blocking the camera. Therefore, we created a holder to enable users to hold the X-Road smartphone more easily. The holder was designed using OnShape.[19] The 3D model is shown in Figure 6. It was printed with PLA using the Makerbot replicator 2 printer, with 30% of infill and layers of 0.1 mm. We did three iterations to adjust the comfort for users wearing X-Road on their head and to enable access to the smartphone buttons and plugs. The first version was an open box where one side was the space for holding the smartphone. In the second version, we added elements so it was possible for users to wear the holder on their head (Figure 5). Specifically, the holder is not rectangular but has a more comfortable shape; we added attachments from another VR headset to fix a rubber band to it and put foam in the part in contact with the face when the rubber band was in place. The last iteration allowed us to finalize the shape to offer full access to the buttons of the smartphone, as well as the mini-jack for headphones and the charging port so we could plug in an external battery if needed during the experiment (see Figure 1).

ACKNOWLEDGMENTS

We would like to thank all students who participated in our user study. We also thank Isabelle Campaignolle, Daniel Tay, and Nathalie Bujosa-Garbolino for working with X-Road during classes. We thank Thibault Lainé for his help in designing the 3D holder, Marine Lira and Sheryna Guidicelli for implementing a virtual street during the iteration#0, and Philippe Giraudeau for information about ARCore.

REFERENCES

[1] Dragan Ahmetovic, Cole Gleason, Chengxiong Ruan, Kris Kitani, Hironobu Takagi, and Chieko Asakawa. 2016. NavCog: A navigational cognitive assistant for the blind. In *Proceedings of the 18th International Conference on Human-computer Interaction with Mobile Devices and Services (MobileHCI'16).* ACM, New York, NY, 90–99. DOI: https://doi.org/10.1145/2935334.2935361

[2] Jérémy Albouys-Perrois, Jérémy Laviole, Carine Briant, and Anke M. Brock. 2018. Towards a multisensory augmented reality map for blind and low vision people: A participatory design approach. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI'18).* ACM, New York, NY. DOI: https://doi.org/10.1145/3173574.3174203

[3] Kazuma Aoyama, Makoto Mizukami, Taro Maeda, and Hideyuki Ando. 2016. Modeling the enhancement effect of countercurrent on acceleration perception in galvanic vestibular stimulation. In *Proceedings of the 13th International Conference on Advances in Computer Entertainment Technology.* ACM, 6.

[4] Bruno Arnaldi, Pascal Guitton, and Guillaume Moreau. 2018. *Virtual Reality and Augmented Reality: Myths and Realities.* Wiley. Retrieved from https://www.wiley.com/en-fr/Virtual+Reality+and+Augmented+Reality:+Myths+and+Realities-p-9781786301055.

[5] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre. 2001. Recent advances in augmented reality. *IEEE Comput. Graph. Applic.* 21, 6 (2001), 34–47. DOI: https://doi.org/10.1109/38.963459

[6] Durand R. Begault. 1994. *3DD Sound for Virtual Reality and Multimedia.* Academic Press Professional, Inc.

[19](https://www.onshape.com/).

[7] Mark Billinghurst, Hirokazu Kato, and Ivan Poupyrev. 2001. The magicbook-moving seamlessly between reality and virtuality. *IEEE Comput. Graph. Applic.* 21, 3 (2001), 6–8.

[8] Ian Bishop and Muhammad Rizwan Abid. 2018. Survey of locomotion systems in virtual reality. In *Proceedings of the 2nd International Conference on Information System and Data Mining (ICISDM'18).* ACM, New York, NY, 151–154. DOI: https://doi.org/10.1145/3206098.3206108

[9] Jeffrey R. Blum, Mathieu Bouchard, and Jeremy R. Cooperstock. 2013. Spatialized audio environmental awareness for blind users with a smartphone. *Mob. Netw. Applic.* 18, 3 (01 June 2013), 295–309. DOI: https://doi.org/10.1007/s11036-012-0425-8

[10] Mourad Bouzit, Grigore Burdea, George Popescu, and Rares Boian. 2002. The Rutgers Master II-new design force-feedback glove. *IEEE/ASME Trans. Mechatr.* 7, 2 (2002), 256–263.

[11] Doug A. Bowman and Ryan P. McMahan. 2007. Virtual reality: How much immersion is enough? *Computer* 40, 7 (2007).

[12] Ellen Lambert Bowman and Lei Liu. 2017. Individuals with severely impaired vision can learn useful orientation and mobility skills in virtual streets and can use them to improve real street safety. *PLoS One* 12, 4 (2017), e0176534.

[13] Evren Bozgeyikli, Andrew Raij, Srinivas Katkoori, and Rajiv Dubey. 2016. Point & teleport locomotion technique for virtual reality. In *Proceedings of the Symposium on Computer-Human Interaction in Play.* ACM, 205–216.

[14] Evren Bozgeyikli, Andrew Raij, Srinivas Katkoori, and Rajiv Dubey. 2019. Locomotion in virtual reality for room scale tracked areas. *Int. J. Hum.-comput. Stud.* 122 (2019), 38–49.

[15] Anke M. Brock, Philippe Truillet, Bernard Oriola, Delphine Picard, and Christophe Jouffrais. 2014. Interactivity improves usability of geographic maps for visually impaired people. *Hum.-computer Interact.* 30 (2014), 156–194.

[16] Abbie Brown and Tim Green. 2016. Virtual reality: Low-cost tools and resources for the classroom. *TechTrends* 60, 5 (2016), 517–519.

[17] Zaira Cattaneo and Tomaso Vecchi. 2011. *Blind Vision: The Neuroscience of Visual Impairment.* The MIT Press. 269 pages.

[18] Erin C. Connors, Elizabeth R. Chrastil, Jaime Sánchez, and Lotfi B. Merabet. 2014. Virtual environments for the transfer of navigation skills in the blind: A comparison of directed instruction vs. video game based learning approaches. *Front. Hum. Neurosci.* 8 (2014), 223.

[19] Carolina Cruz-Neira, Daniel J. Sandin, Thomas A. DeFanti, Robert V. Kenyon, and John C. Hart. 1992. The CAVE: Audio visual experience automatic virtual environment. *Commun. ACM* 35, 6 (1992), 64–73.

[20] Florian Dramas, Brian F. Katz, and Christophe Jouffrais. 2008. Auditory-guided reaching movements in the peripersonal frontal space. *J. Acoust. Soc. Amer.* 123, 5 (May 2008), 3723. DOI: https://doi.org/10.1121/1.2935195

[21] Julie Ducasse, Anke M. Brock, and Christophe Jouffrais. 2018. Accessible interactive maps for visually impaired users. In *Mobility of Visually Impaired People.* Springer, 537–584.

[22] Julie Ducasse, Marc Macé, Marcos Serrano, and Christophe Jouffrais. 2016. Tangible reels: Construction and exploration of tangible maps by visually impaired users. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.* ACM, 2186–2197.

[23] Steven Feiner, Blair MacIntyre, Marcus Haupt, and Eliot Solomon. 1993. Windows on the world: 2D windows for 3D augmented reality. In *Proceedings of the 6th ACM Symposium on User Interface Software and Technology (UIST'93).* ACM, New York, NY, 145–155. DOI: https://doi.org/10.1145/168642.168657

[24] Laura Freina and Michela Ott. 2015. A literature review on immersive virtual reality in education: State-of-the-art and perspectives. *eLearn. Softw. Educ.*1 (2015).

[25] Doron Friedman. 2015. Brain-computer interfacing and virtual reality. In *Handbook of Digital Games and Entertainment Technologies.* Springer, 1–22.

[26] Marwan A. Gharaybeh and Grigore C. Burdea. 1994. Investigation of a shape memory alloy actuator for dextrous force-feedback masters. *Adv. Robot.* 9, 3 (1994), 317–329.

[27] Michael A. Gigante. 1993. Virtual reality: Definitions, history, and applications. In *Virtual Reality Systems.* Elsevier, 3–14.

[28] Stéphanie Giraud, Anke M. Brock, Marc Macé, and Christophe Jouffrais. 2017. Map learning with a 3D printed interactive small-scale model: Improvement of space and text memorization in visually impaired students. *Front. Psychol.* 8 (2017), 930.

[29] Nicholas A. Giudice. 2018. Navigating without vision: Principles of blind spatial cognition. In *Handbook of Behavioral & Cognitive Geography.* Edward Elgar Publishing, 1–32.

[30] Reginald G. Golledge. 2003. Human wayfinding and cognitive maps. In *The Colonization of Unfamiliar Landscapes.* Routledge, 49–54.

[31] T. Gotzelmann. 2016. LucentMaps 3D printed audiovisual tactile maps for blind and visually impaired people. In *Proceedings of the International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS'16).* 81–90. DOI: https://doi.org/10.1145/2982142.2982163

[32] João Guerreiro, Dragan Ahmetovic, Kris M. Kitani, and Chieko Asakawa. 2017. Virtual navigation for blind people: Building sequential representations of the real-world. In *International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS'17)*.

[33] João Guerreiro and Daniel Gonçalves. 2014. Text-to-speeches: Evaluating the perception of concurrent speech by blind people. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility (ASSETS'14)*. ACM, New York, NY, 169–176. DOI: https://doi.org/10.1145/2661334.2661367

[34] Nancy E. Guerrón, Antonio Cobo, José J. Serrano Olmedo, and Carlos Martín. 2019. Sensitive interfaces for blind people in virtual visits inside unknown spaces. *Int. J. Hum.-comput. Stud.* (Aug. 2019). DOI: https://doi.org/10.1016/J.IJHCS.2019.08.004

[35] Wijnand A. IJsselsteijn, Yvonne A. W. de Kort, and Antal Haans. 2006. Is this my hand I see before me? The rubber hand illusion in reality, virtual reality, and mixed reality. *Pres.: Teleop. Virt. Environ.* 15, 4 (2006), 455–464.

[36] Hiroo Iwata. 1993. Pen-based haptic virtual environment. In *Proceedings of the IEEE Virtual Reality International Symposium*. 287–292.

[37] Slim Kammoun, Gaétan Parseihian, Olivier Gutierrez, Adrien Brilhault, Antonio Serpa, Mathieu Raynal, Bernard Oriola, Marc Macé, Malika Auvray, Michel Denis, et al. 2012. Navigation and space perception assistance for the visually impaired: The NAVIG project. *IRBM* 33, 2 (2012), 182–189.

[38] Julian Kreimeier and Timo Götzelmann. 2018. Real world VR proxies to support blind people in mobility training. In *Proceedings of the Mensch Comput. Workshopband*.

[39] Julian Kreimeier and Timo Götzelmann. 2019. First steps towards walk-in-place locomotion and haptic feedback in virtual reality for visually impaired. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*.

[40] Merel Krijn, Paul M. G. Emmelkamp, Roeline Biemond, Claudius de Wilde de Ligny, Martijn J. Schuemie, and Charles A. P. G. van der Mast. 2004. Treatment of acrophobia in virtual reality: The role of immersion and presence. *Behav. Res. Ther.* 42, 2 (2004), 229–239.

[41] Andreas Kunz, Klaus Miesenberger, Limin Zeng, and Gerhard Weber. 2018. Virtual navigation environment for blind and low vision people. In *Proceedings of the International Conference on Computers Helping People with Special Needs*. Springer, 114–122.

[42] Orly Lahav and David Mioduser. 2008. Construction of cognitive maps of unknown spaces using a multi-sensory virtual environment for people who are blind. *Comput. Hum. Behav.* 24, 3 (May 2008), 1139–1155. DOI: https://doi.org/10.1016/j.chb.2007.04.003

[43] Jack M. Loomis, Roberta L. Klatzky, and Nicholas A. Giudice. 2013. Representing 3D space in working memory: Spatial images from vision, hearing, touch, and language. In *Multisensory Imagery*. Springer, 131–155.

[44] Pedro Lopes, Sijing You, Lung-Pan Cheng, Sebastian Marwecki, and Patrick Baudisch. 2017. Providing haptics to walls & heavy objects in virtual reality by means of electrical muscle stimulation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI'17)*. ACM Press, New York, 1471–1482. DOI: https://doi.org/10.1145/3025453.3025600

[45] Wendy E. Mackay. 2003. Educating multi-disciplinary design teams. In *Proceedings of Tales of the Disappearing Computer*. 105–118.

[46] Jonatan Martínez, Arturo García, Miguel Oliver, José Pascual Molina, and Pascual González. 2016. Identifying virtual 3D geometric shapes with a vibrotactile glove. *IEEE Comput. Graph. Applic.* 36, 1 (2016), 42–51.

[47] Thomas H. Massie, J. Kenneth Salisbury, et al. 1994. The phantom haptic interface: A device for probing virtual objects. In *Proceedings of the ASME Winter Annual Meeting, Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, Vol. 55. Citeseer, 295–300.

[48] Oussama Metatla, Nick Bryan-Kinns, Tony Stockman, and Fiore Martin. 2015. Designing with and for people living with visual impairments: Audio-tactile mock-ups, audio diaries, and participatory prototyping. *CoDesign* 11, 1 (2015), 35–48.

[49] Joshua A. Miele, Steven Landau, and Deborah Gilden. 2006. Talking TMAP: Automated generation of audio-tactile maps using Smith-Kettlewell's TMAP software. *Brit. J. Vis. Impair.* 24, 2 (2006), 93–100.

[50] Paul Milgram, Herman Colquhoun, et al. 1999. A taxonomy of real and virtual world display integration. *Mixed Real. Merg. Real and Virt. Worlds* 1 (1999), 1–26.

[51] Paul Milgram and Fumio Kishino. 1994. A taxonomy of mixed reality visual displays. *IEICE Trans. Inf. Syst.* 77, 12 (1994), 1321–1329.

[52] Roberto A. Montano Murillo, Sriram Subramanian, and Diego Martinez Plasencia. 2017. Erg-O: Ergonomic optimization of immersive virtual environments. In *Proceedings of the 30th ACM Symposium on User Interface Software and Technology*. ACM, 759–771.

[53] Upul Obeysekare, Chas Williams, Jim Durbin, Larry Rosenblum, Robert Rosenberg, Fernando Grinstein, Ravi Ramamurti, Alexandra Landsberg, and William Sandberg. 1996. Virtual workbench—A non-immersive virtual

environment for visualizing and interacting with 3D objects for scientific visualization. In *Proceedings of the Visualization Conference.* IEEE, 345–349.

[54] Claudio Pacchierotti, Stephen Sinclair, Massimiliano Solazzi, Antonio Frisoli, Vincent Hayward, and Domenico Prattichizzo. 2017. Wearable haptic systems for the fingertip and the hand: Taxonomy, review, and perspectives. *IEEE Trans. Haptics* 10, 4 (2017), 580–600.

[55] Betsy Phillips and Hongxin Zhao. 1993. Predictors of assistive technology abandonment. *Assist. Technol.* 5, 1 (1993), 36–45.

[56] Georg Regal, Elke Mattheiss, David Sellitsch, and Manfred Tscheligi. 2018. Mobile location-based games to support orientation & mobility training for visually impaired students. In *Proceedings of the 20th International Conference on Human-computer Interaction with Mobile Devices and Services.* ACM, 47.

[57] Bernhard E. Riecke, Jörg Schulte-Pelkum, Franck Caniard, and Heinrich H. Bulthoff. 2005. Towards lean and elegant self-motion simulation in virtual reality. In *Proceedings of the Virtual Reality Conference.* IEEE, 131–138.

[58] Joan Sol Roo and Martin Hachet. 2017. Towards a hybrid space combining spatial augmented reality and virtual reality. In *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI'17).* https://hal.archives-ouvertes.fr/hal-01453385.

[59] Jaime Sánchez and Claudio Oyarzún. 2011. Mobile audio assistance in bus transportation for the blind. *Int. J. Disab. Hum. Dev.* 10, 4 (2011), 365–371.

[60] Elizabeth B.-N. Sanders. 2003. From user-centered to participatory design approaches. In *Design and the Social Sciences.* CRC Press, 18–25.

[61] Stefano Scheggi, Leonardo Meli, Claudio Pacchierotti, and Domenico Prattichizzo. 2015. Touch the virtual reality: Using the Leap Motion controller for hand tracking and wearable tactile devices for immersive haptic rendering. In *Proceedings of the ACM Special Interest Group on Computer Graphics and Interactive Techniques Conference (SIGGRAPH'15).* ACM, 31.

[62] David W. Schloerb, Orly Lahav, Joseph G. Desloge, and Mandayam A. Srinivasan. 2010. BlindAid: Virtual environment system for self-reliant trip planning and orientation and mobility training. In *Proceedings of the Haptics Symposium.* 363–370.

[63] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. 2001. The experience of presence: Factor analytic insights. *Pres. Teleop. Virt. Environ.* 10, 3 (2001), 266–281.

[64] Alexander W. Siegel and Sheldon H. White. 1975. The development of spatial representations of large-scale environments. In *Advances in Child Development and Behavior.* Vol. 10. Elsevier, 9–55.

[65] Mel Slater and Maria V. Sanchez-Vives. 2016. Enhancing our lives with immersive virtual reality. *Front. Robot. AI* 3 (2016), 74.

[66] Mel Slater and Sylvia Wilbur. 1997. A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Pres. Teleop. Virt. Environ.* 6, 6 (1997), 603–616.

[67] Frank J. J. M. Steyvers and Aart C. Kooijman. 2009. Using route and survey information to generate cognitive maps: Differences between normally sighted and visually impaired individuals. *Appl. Cogn. Psychol. Offic. J. Soc. Appl. Res. Mem. Cogn.* 23, 2 (2009), 223–235.

[68] A. F. Tatham. 1991.The design of tactile maps: Theoretical and practical considerations. In *Proceedings of the International Cartographic Association: Mapping the Nations*, K. Rybaczak and M. Blakemore (Eds.). ICA, London, UK, 157–166.

[69] Shan-Yuan Teng, Tzu-Sheng Kuo, Chi Wang, Chi-huan Chiang, Da-Yuan Huang, Liwei Chan, and Bing-Yu Chen. 2018. PuPoP: Pop-up prop on palm for virtual reality. In *Proceedings of the 31st ACM Symposium on User Interface Software and Technology.* ACM, 5–17.

[70] Lauren Thevin and Anke M. Brock. 2018. Augmented reality for people with visual impairments: Designing and creating audio-tactile content from existing objects. In *Computers Helping People with Special Needs.* Springer, Cham, 193–200. DOI: https://doi.org/10.1007/978-3-319-94274-2_26

[71] Manos Tsakiris and Patrick Haggard. 2005. The rubber hand illusion revisited: Visuotactile integration and self-attribution. *J. Exper. Psychol. Hum. Percept. Perf.* 31, 1 (2005), 80.

[72] Simon Ungar. 2000. Cognitive mapping without visual experience. In *Cognitive Mapping: Past, Present, and Future*, Rob Kitchin and Scott Freundschuh (Eds.). Routledge, Oxon, UK, Chapter 13, 221–248.

[73] Eric Whitmire, Hrvoje Benko, Christian Holz, Eyal Ofek, and Mike Sinclair. 2018. Haptic revolver: Touch, shear, texture, and shape rendering on a reconfigurable virtual reality controller. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI'18).* ACM, New York, NY, Article 86, 12 pages. DOI: https://doi.org/10.1145/3173574.3173660

[74] Betsy Williams, Gayathri Narasimham, Bjoern Rump, Timothy P. McNamara, Thomas H. Carr, John Rieser, and Bobby Bodenheimer. 2007. Exploring large virtual environments with an HMD when physical space is limited. In *Proceedings of the 4th Symposium on Applied Perception in Graphics and Visualization.* ACM, 41–48.

[75] Ye Yuan and Anthony Steed. 2010. Is the rubber hand illusion induced by immersive virtual reality? In *Proceedings of the IEEE Virtual Reality Conference (VR'10)*. 95–102.

[76] Pavel Zahorik, Philbert Bangayan, V. Sundareswaran, Kenneth Wang, and Clement Tam. 2006. Perceptual recalibration in human sound localization: Learning to remediate front-back reversals. *J. Acoust. Soc. Amer.* 120, 1 (2006), 343–359.

[77] Yuhang Zhao, Cynthia L. Bennett, Hrvoje Benko, Edward Cutrell, Christian Holz, Meredith Ringel Morris, and Mike Sinclair. 2018. Enabling people with visual impairments to navigate virtual reality with a haptic and auditory cane simulation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, 116.

[78] Yuhang Zhao, Elizabeth Kupferstein, Doron Tal, and Shiri Azenkot. 2018. "It looks beautiful but scary": How low vision people navigate stairs and other surface level changes. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS'18)*. ACM, New York, 307–320. DOI: https://doi.org/10.1145/3234695.3236359

Received February 2019; revised December 2019; accepted January 2020