The Design Space of the Auditory Representation of Objects and Their Behaviours in Virtual Reality for Blind People

João Guerreiro, Yujin Kim, Rodrigo Nogueira, SeungA Chung, André Rodrigues, and Uran Oh

Abstract— As virtual reality (VR) is typically designed in terms of visual experience, it poses major challenges for blind people to understand and interact with the environment. To address this, we propose a design space to explore how to augment objects and their behaviours in VR with a nonvisual audio representation. It intends to support designers in creating accessible experiences by explicitly considering alternative representations to visual feedback. To demonstrate its potential, we recruited 16 blind users and explored the design space under two scenarios in the context of boxing: understanding the location of objects (the opponent's defensive stance) and their movement (opponent's punches). We found that the design space enables the exploration of multiple engaging approaches for the auditory representation of virtual objects. Our findings depicted shared preferences but no one-size-fits-all solution, suggesting the need to understand the consequences of each design choice and their impact on the individual user experience.

Index Terms—Inclusive Virtual Reality, Nonvisual Interaction, Blind, Auditory Feedback, Design Space

1 Introduction

Virtual reality (VR) is slowly becoming available to the masses at affordable prices, paving the way for a large number of applications in a variety of contexts, such as gaming and entertainment, education, and employee training. Despite the hype about immersive experiences that VR can offer, there are numerous accounts of their inaccessibility [51, 60, 65, 71]. VR applications rely heavily on visual feedback as an essential modality with limited or no tactile cues, neglecting people with visual impairments who are either being delivered with poor VR experiences or are excluded.

VR environments for people with visual impairments do exist. However, they have focused mostly on systems specifically designed for this population as a means to acquire either orientation and mobility skills [22] or knowledge about real-world locations [28]. These solutions often try to mimic real behaviours – e.g., using a white cane [43, 69] – with the main goal of transferring knowledge to the real world. Others often over-simplify the experience (both feedback and interaction mechanisms) for the sake of access but at the expense of functionality and engagement [3, 26, 70]. For instance, many audio games restrict navigation to a grid [53], which significantly eases the ability to move in the environment but limits its exploration to a fixed set of positions. This paradigm blocks access to the more complex virtual environments and behaviours seen in mainstream VR experiences, which remain inaccessible. This is, in part, due to a lack of awareness and standardization when designing nonvisual feedback for accessible VR applications.

In this paper, we propose a design space for augmenting objects with a nonvisual audio representation of location and behaviours in an immersive VR environment. The design space proposed is meant to guide the design of alternative audio feedback for objects' location, behaviours and interactions within a VR environment. We propose 9 categories that can be leveraged by researchers and practitioners working on creating accessible VR content. This design space includes dimensions related to how audio feedback is provided, what information is delivered, and how it can be triggered. To cite a few examples, one may convey information about objects using spatialized audio to hint at their current location with respect to the location of the user (e.g., left or right, near or far) or monaural sound if the location is not relevant (e.g., a warning sound or background music); audio feedback about multiple sound sources may be played to the user sequentially or concurrently; and the feedback may be given using speech or sonification.

To explore and validate the ability of our design space to support nonvisual representations of objects and their behaviours, we conducted a user study where 16 blind people experienced one of two virtual boxing scenarios: one focused on conveying the location of objects (S_{loc}), where participants were asked to understand the opponent's defensive stance (the location of their two hands); the other focused on conveying movement behaviour (S_{mov}), where they were asked to understand the attacking punches of their opponent (moving towards the user).

Findings show that the design space enabled the exploration of multiple solutions to support VR boxing for blind people. The user study depicted both shared and contrasting preferences for the different categories of the design space and allowed us to identify how each category affected participants' preferred design choices and their experience. Overall, and independently of their preferred configurations, participants were able to complete their tasks successfully in both scenarios. The contributions of this paper are: (1) a design space for the auditory representation of the location and behaviour of objects in virtual reality for blind people, and (2) the qualitative assessment of the potential of the proposed design space for understanding how each choice in design space categories affects the user experience.

2 Related Work

We discuss related work along three topics: first, we describe how audio feedback has been leveraged in prior research to convey or augment information (e.g. sonification to facilitate exploring a map). Second, we discuss how blind people are interacting in virtual environments. Lastly, we highlight how design spaces have been used in previous research to promote an understanding and further explorations of certain domains.

2.1 Auditory Feedback in 2D and 3D spaces

Audio feedback has been recurrently investigated in the context of wayfinding and navigation [9,30,33,67,68,72,81]. For example, Heuten *et al.* [30] presented a technique to sonify a real world map that enabled blind users to build a mental map prior to their journey. The approach attempted to allow users to easily perceive relationships between objects and detect important landmarks for navigation. On touchscreens, audio feedback was investigated to improve gesture recognition [23,44,56,57].

-

João Guerreiro, Rodrigo Nogueira, and André Rodrigues are with LASIGE, Faculdade de Ciências, Universidade de Lisboa. E-mail: jpguerreiro@fc.ul.pt, fc55721@alunos.fc.ul.pt, afrodrigues@fc.ul.pt.

Yujin Kim, SeungA Chung, and Uran Oh are with the Computer Science and Engineering, Ewha Womans University. E-mail: komgi0715@ewhain.net, ewhacsa@ewhain.net, uran.oh@ewha.ac.kr.

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

For instance, Gao *et al.* [23] designed three different types of audio feedback to support trajectory-based finger gestures. The gesture was accompanied by a discrete beep, static or gradual feedback. Their results suggest that gradual continuous feedback increases accuracy.

Other works have focused their efforts on exploring audio feedback for target acquisition and detailing target information in extended reality for 3D spaces such as VR and Augmented Reality (AR) [14, 15, 20, 24, 31, 63, 64]. For example, Ren *et al.* [63] explored audio and haptic feedback design for Mixed Reality (MR) tourism applications. They investigated the effects of spatial audio with various design factors such as volume and rhythm of audio and found that fast rhythm and adequate volume improved the user performance on target acquisition in MR applications using a head-mounted display (HMD). Similarly, Chung *et al.* [14] investigated audio and haptic feedback with spatial audio for a target acquisition task in a 3D virtual space. They found that spatial sound helps the user to understand the horizontal direction of the target, and discrete-based audio and haptic combo reduced the task completion time. The use of spatial audio is now widely used and tries to convey information about the location of objects, often replicating sounds from the real-word [61, 69], but also trying to reproduce echolocation [4]. While there are many examples of using audio to augment interactions or to add new information about the objects that populate virtual environments, there has been little focus on standardizing the process of designing audio feedback. Thus, we propose a design space for the audio representation of objects and behaviours to support systematic procedures for making design decisions.

2.2 Virtual Environments for Blind People

The accessibility of virtual environments is frequently ignored. This is especially relevant for people with visual impairments, as interaction usually depends on visual stimuli and lacks tactile cues. Research on virtual environments for visually impaired people has focused mostly on systems specifically designed for them, rather than on providing access to mainstream ones. Common approaches rely on audio and/or haptic feedback to convey information about the environment [21, 40, 78] and focus mainly on creating experiences to support mobility training and/or creating mental maps of real-world locations [19, 21, 28, 47].

Recent technological advances and the easier access to immersive VR have led to new approaches that try to improve blind people's experience through different modalities, such as more realistic audio, haptics, or locomotion. For instance, earlier approaches often used keyboards, joysticks, the smartphone, or other haptic devices to support exploring a virtual environment [17, 18, 42, 66], but more recent approaches have tried to increase immersion by supporting either walking in place (with or without a treadmill) [41] or actually walking in the real-world [34, 74] – sometimes by instrumenting a white cane for improved haptic feedback [69, 80].

The use of audio has supported multiple approaches and applications in varied contexts, such as digital games. For instance, NavStick [53] enables blind gamers to probe their surroundings by scanning a specific direction at a time with the controller joystick – in the context of a 3D adventure game – while The RAD [70] used sonification in racing games for an equitable experience for blind and sighted players. VRBubble [36] supports accessible social VR experiences by exploring peripheral awareness dividing the social space into intimate, conversation, and social bubbles, and conveying different audio feedback accordingly. Other contexts include sports and exergames where prior works have targeted, for instance, yoga [62] and tennis [49]. More recent approaches, have tried to support training or playing sports that blind people practice in the real-world – e.g., Goalball [76] and Showdown [77] – by providing an accessible VR alternative.

In order to make VR applications accessible – all, not only those specifically designed for blind people – designers should explicitly consider how to convey alternative representations to visual feedback. Audio has been fundamental to providing accessible experiences in prior work, but the literature shows a panoply of approaches to convey information through this modality. Our work aims at providing a theoretical framework for designers to create accessible VR experiences.

2.3 Design Spaces as a Theoretical Framework

Design space is a theoretical framework that describes the different possibilities for designing a type of artifact [46]. It promotes reflecting on the attributes and choices made when designing new artifacts and is based on the domain of interest, its technological possibilities, and existing artifacts used for related purposes. It consists of a number of dimensions deemed central to the domain, each with a set of proposed values that may be expanded after new design explorations. These dynamics make it possible to iteratively grow and adapt both to new technological innovations and to new research findings (e.g., related to human perception of auditory feedback). Design spaces are often used in Human-Computer Interaction as a framework to understand a domain and to explore innovative solutions in that domain (e.g., [11, 45, 50, 52, 54, 82]).

Researchers have been investigating sound design and VR together and separately. Some works have explored the design space of a specific type of feedback, such as conveying heart rate in VR [12], while others have tried to classify the different types of applications of a specific domain, such as AR/VR applications for assistive environments in manufacturing [10]. Garner [25] – with the goal of building a framework of VR sound – specifies multiple perspectives on how to approach sound. For instance, one may approach sound as an object (a spatial entity) that can be located through its sound [58]; or as an event, meaning the sound is related to a specific action about the object [55], among other perspectives. Jain et al., [35] created a taxonomy in the context of VR sound accessibility for deaf and hard of hearing (DHH) people. Still, their goal contrasts with ours as they tried to categorize the existing sound in VR, while we aim to find alternative representations to visual feedback.

In the context of accessible computing for people with visual impairments, design spaces have been proposed as a way to promote research and innovation in domains such as nonvisual word completion interfaces [54], rich representations of visual content for screen reader users [50], and accessible visualizations [38, 83]. For creating the taxonomies, prior research has relied either on a systematic review of the domain [38], or on the authors' expertise on and analysis of the domains of interest [50, 54]. The design space created by Morris et al., [50] enabled the authors to create multiple prototypes that they could then evaluate with visually impaired users. Nicolau et al., [54] engaged blind participants in exploring the different categories of the design space and creating their own preferred solution. In both cases, instantiating the design space provided valuable insights into promising solutions for their own domains.

3 Design Space for Auditory Feedback of Objects and their Behaviours

Our approach lies in the formulation of a theoretical framework that allows VR designers to instantiate a design space for the audio representation of objects and their possible interactions. We tackle the current lack of standardization when defining nonvisual feedback for VR experiences, which results in either inaccessible VR applications (the large majority) or over-simplification of the experience for blind people. In the following, we present a design space of the audio representation of objects and their behaviour in VR – summarized in Table 1 – as a means to support the design of VR experiences that are accessible to blind people.

3.1 Creating the Design Space

Defining a design space with a clear taxonomy of possible dimensions involved in this domain is an iterative process, which benefits from knowledge and experimentation. The purpose is to guide the development and research of future solutions. Prior accessibility efforts have proposed different ways to represent the location and behaviour of objects in virtual environments, but that information is scattered in different scientific articles, applications, and games. In order to build this design space, the research team was informed by prior work and relied on their expertise in HCI, accessible computing, non-visual interaction, audio-based interfaces, virtual environments, and games. The team met bi-weekly for two months in order to discuss and iteratively

Category	Description	Values
Awareness	Defines if objects have an audio representation.	*None* | *Conditional* | *Full*
Audio Field	Defines users' audio spatial awareness	*Self* | *Perimeter* | *Field of View*
Trigger	Defines how audio feedback is activated	*On: Existence* | *Proximity* | *Movement* | *Collisions* | *Request*
Representation	Indicates whether the audio feedback can change	*Static* | *Dynamic*
Sound Type	Type of audio feedback	*Speech* | *Sonification*
Time Signal	Describes the type of audio signal used	*Discrete* | *Periodic* | *Continuous*
Cardinality	Number of objects considered as a audio source	*1* | *2* | ... | *N*
Concurrency	Indicates sound sources play sequential or concurrent	*Sequential* | *Concurrent*
Spatialization	Defines the form of audio Spatialization	*Monaural* | *Dichotic* | *3D*

Table 1: The design space for auditory representation of VR objects defined in our study.

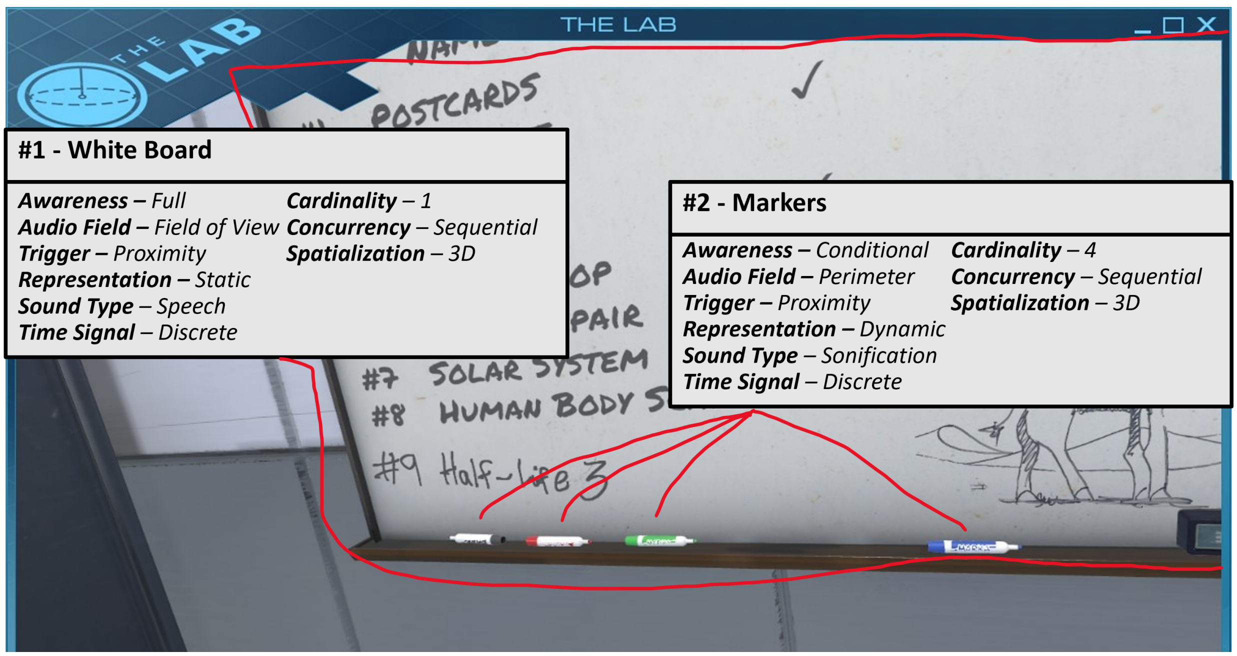

Fig. 1: Two mock-up examples of the design space applied to two interactive objects within the scene of the game named "The Lab by Valve". #1 - *White Board* is noticeable by the player when on its field of view. The board is announced through speech with the sound coming from the object location. #2 - *Markers* are only perceivable when close to the board. Each of the markers periodically produces a sound in sequence, where the colour is associated with the pitch of the sound used, originating from their location. The closer the player hand is to one, the louder it becomes. The markers only produce sound while none is in hand.

refine the design space. In addition, the team revisited and edited the design space when implementing the VR applications and preparing the user studies. Below, we present the resulting design space but we reinforce that design spaces are not final. They are meant to be iterated upon with new explorations and as new approaches and technologies come to be.

3.2 Categories and Values

We propose a taxonomy of properties that can be used as a standard for researchers and designers who seek to explore and understand what audio design choices can create an accessible VR experience and the impact on the perceived experience of each choice. Two mock-up examples of exercising this design space in an existing game are described in Figure 1. Our taxonomy includes two meta-categories (i.e., *Awareness* and *Audio Field*) which define whether or not feedback exists and under what conditions. The remaining seven categories articulate how audio can be conveyed. All representations have a value in each category.

Awareness. The awareness category indicates if users are made aware of the existence of objects of a particular type and what are the pre-conditions to represent such objects. It can be seen as a meta-category where, ultimately, a value of *None* indicates that such an object does not have an audio representation - dismissing the remaining categories. On the other end, *Full* awareness would indicate that all instances of such object have an audio representation (e.g., White board Figure 1), being their characteristics defined by the other categories. *Conditional* means there are rules which define the awareness, for example they could be Priority (e.g., only the two closest markers) or Attribute based as in the example given (e.g., only if marker is not held, Figure 1). While this category pre-defines the objects that have an audio representation, the following further detail when and how it is conveyed.

Audio Field. The Audio Field category indicates when objects can have an auditory representation based on their spatial location. When outside of the audio field, the object does not produce auditory feedback. A *Field of View* audio field corresponds to the full visual field of the user, indicating that the object would provide feedback when visible (and according to its awareness). Alternatively, feedback can be based on a *Perimeter*, meaning objects would need to be within a defined area in order to produce auditory feedback (e.g., Figure 1 - #2

Marker). A *Self* audio field would mean that feedback only occurs on self-interactions – e.g., contact with the virtual character, self footsteps.

Trigger. The Trigger category indicates what causes the outset of auditory feedback. On one end, an object can trigger auditory feedback *On Existence*, meaning that it will start producing auditory feedback whenever it enters (or starts existing) in the user's audio field (e.g., Figure 1 - #1 *White Board*). Alternatively, feedback may be triggered *On Movement* (when the object moves) or *On Proximity* (when the object gets within a distance threshold, and where zero represents on collision with self) (e.g., Figure 1 - #2 *Marker*). Other options may include audio feedback *On Collisions* or *On Request* – meaning feedback is provided when explicitly requested by the user (e.g., by clicking a button).

Representation. The Representation category indicates whether the auditory feedback has a *Static* – the feedback mapped to one object does not change (e.g., Figure 1 - #1 *White Board*) – or *Dynamic* – the feedback may change according to defined rules or characteristics (e.g., Figure 1 - #2 *Markers*, where volume is adjusted given the proximity to players' hand).

Sound Type. The Sound Type category indicates whether the audio feedback is provided through *Speech* or *Sonification*. Sonification is the "use of non-speech audio to convey information" [39], such as auditory icons – audio effects resembling real-world objects or events – earcons – artificial sounds often following musical conventions to alert users – or variations in audio characteristics to convey meaning (e.g., distance, size, or height). Specifying the possible sonification techniques [29] has its own challenges and enters a greater level of detail that is out of scope of this design space.

Time Signal. The Time Signal category indicates whether the audio feedback is provided with a *Discrete*, *Periodic*, or *Continuous* signal. Discrete indicates feedback is provided once (e.g., Figure 1 - #1 *White Board* is only announced once when it enters the field of view); Periodic represents feedback being delivered at specific time intervals (e.g., every 10 seconds); and Continuous means that feedback is provided nonstop after being triggered.

Cardinality. The Cardinality category indicates how many objects of this type are considered as a potential source of feedback at each time. This same category is used on a prior design space related to nonvisual interaction and feedback [54] and can take values from *one to N*. In the example given (Figure 1) all objects are rendered as potential audio sources, but there can be instances where (e.g. if there were 12 markers) it could be necessary to limit the cardinality.

Concurrency. The Concurrency category indicates whether the feedback of multiple objects is presented *Sequentially* or *Concurrently*. Similarly to Cardinality, we have borrowed this category from Nicolau et al.,'s design space [54]. This means that when multiple objects' feedback may overlap in time – e.g., four markers – one can either present their feedback sequentially, or concurrently while trying to ensure they are distinguishable. This category is not applicable when there is a Single object as a potential source for feedback – when cardinality equals one.

Sound Spatialization. The Sound Spatialization category indicates whether the auditory feedback provided relies on any form of spatialization (see [7] for fundamentals on spatial hearing research). *Monaural* sound means feedback is provided through a single auditory channel, which does not enable a sense of spatialization/location. *Dichotic* sound would rely on two auditory channels to convey information to the right and/or left ears (e.g., markers closer to the user's right hand would only be conveyed to the user's right ear), while *3D audio* tries to represent a specific location of a sound source relative to the user's position replicating the way people hear sound in the real world (e.g., Figure 1 #1 - *White Board* using speech to announce its position from its location).

4 Exploration of the Design Space

The design space is the means – and not the end – to build interactive VR experiences that are accessible to people with visual impairments. For that reason, we conducted a user study where blind users explored different instantiations of the design space, aiming to assess its potential to support the design (and experimentation) of nonvisual, audio representations of objects and their dynamics in VR environments. In particular, we addressed two scenarios:

- *Scenario 1 - Location (S_{loc})* : understanding the *location of stationary objects* in the virtual environment
- *Scenario 2 - Movement (S_{mov})* : understanding the *behaviour of moving objects* in the virtual environment

4.1 VR Boxing Application

In this paper, we show how the design space can be leveraged to explore nonvisual augmentative feedback. We showcase it by applying it in the context of a VR Boxing application. First, we decided to target sports applications since physical activity has a positive effect on people's health [59, 75]. In addition, we targeted a popular sport, but that is generally inaccessible to people with visual impairments as a way to afford a new – but somehow known – experience in a safe environment. Finally, we wanted an experience that is complex enough to support a variety of possible interactions and behaviours. For example, one may want to receive feedback about the opponent's (or coach's) body, head and/or hands, but such feedback may depend on the current intention of attacking or defending.

We explored two scenarios that are related to understanding either the location of objects or their movement: 1) perceiving the location of the defensive stance of the opponent (S_{loc}), and 2) punching attacks moving toward the user (S_{mov}). We explored both scenarios separately with the goal of identifying both commonalities and differences between individual explorations of the design space. Furthermore, these scenarios were explored by two different research teams – geographically distributed (Lisbon, Portugal and Seoul, South Korea) – aiming at showing the versatility of the design space in supporting the creation of accessible experiences. Both teams followed the same design space and procedure but explored, separately, the design space with different implementations, users, and a different but related task. The separate implementation of each team's VR application and exploration of the design space intended to create a greater variety of approaches and of possible experiences than what would be possible under a unique design exploration.

4.2 Scenario Exploration and the Design Space

The two scenarios allowed us to explore the design space of VR boxing both with attacking and defending in mind. We convey auditory feedback from the hands of the opponent – framed as the participant's coach in the study. In S_{loc}, we indicate the defensive stance of the opponent, by conveying feedback about the location of their hands, while in S_{mov} we indicate the attacking attempts of the opponent, by trying to convey both location and movement. In both scenarios, we fixed **Awareness** (as Full) and **Audio Field** categories (as Field of View), and **Cardinality** as two, since we wanted to always be able to convey feedback about both hands. Note that we have fixed the opponent's location to be in front of the participant's avatar. The nature of both scenarios led us to fix the **representation** category, but differently among scenarios. In S_{loc}, representation is static, meaning the location of the hands is conveyed but not its movement. On the other hand, in S_{mov}, representation is dynamic as the audio feedback changes depending on speed and proximity/distance.

All the other categories were explored in both scenarios. **Time Signal** (*discrete*, *periodic*, and *continuous*), **Concurrency** (*concurrent* or *sequential*), and **Sound Type** (*speech* or *sonification*) were explored in full. In addition, S_{mov} explored **Sound Spatialization** in full (*monaural*, *dichotic*, and *3D audio*), but S_{loc} did not explore the monaural condition given the known advantages of sound spatialization when using multiple sound sources [8]. **Trigger** category was explored differently in each scenario. In S_{loc}, it could either provide feedback on existence (meaning we would always have feedback about the hands' location), on movement (only when the hands change position), or on request (when the user explicitly requests it). In S_{mov}, audio feedback was triggered either on movement or on proximity. Note that these differences were derived from an independent exploration of the design space by the two distributed research teams, and from two related, but different scenarios.

We also highlight that combinations among these categories resulted in different-sounding approaches. For instance, in S_{loc}, combining sonification with a periodic time signal corresponded to a sonar-like sound, and a continuous beep when combined with a continuous time signal. In both scenarios, we used the sound pitch to indicate if the target was at the head (higher pitch) or body (lower pitch) height. In addition, in S_{mov}, volume indicated distance.

4.3 Apparatus

Both prototypes were implemented using Unity3D running on an Oculus Quest 2 VR system (including a headset and two controllers). The two prototypes enabled the researchers to customize the different categories and values of the design space. We used two laptops in both scenarios: one capable of running Virtual Reality (with an NVIDIA Geforce RTX 2060 in S_{loc} and a GeForce RTX 3060 in S_{mov}) and another to select the configurations and present visual feedback to the researchers. Audio customization (e.g., spatialization) was implemented with Unity3D native assets with default configurations (e.g., regarding reverb values). During the study, the researchers calibrated the VR Headset and the position of the avatar in the VR space and guided the participant to the intended location. The scale of the environment tried to represent real-size boxing gloves (both of the player's hands and of the incoming attacks), whereas the coach's protections would slightly extend his arms in width and height. Colliders were implemented to accommodate the associated shapes. In S_{loc} participants were asked to use headphones as a way to improve spatialized audio, while in S_{mov} audio was conveyed through the VR Headset speakers. This difference resulted from the two design space explorations being prepared separately. Information about configuration preferences and task success/failure was logged by the system.

4.4 Procedure

Each session started with a demographics and VR, Gaming, and Boxing experience questionnaire. Next, we presented a brief introduction to the goal of the study. Participants were informed they would be exploring a variety of audio feedback configurations to convey information about a virtual boxing opponent, their coach, in an immersive virtual environment and choose their preferred configuration.

Participants were then assisted to wear an Oculus Quest 2 and its respective hand-held controllers. Then they were given a brief overview of what they were wearing and how it enabled them to interact with an immersive environment; either hitting their coach's hand/arm protections in S_{loc} or guarding their body against their coach's punches in S_{mov}. This was followed by a practice session where participants were allowed to move their arms and hit a punching bag (S_{loc}) or shield upcoming punches (S_{mov}) to get familiar with the environment. During practice, there was no audio feedback other than the sound of collisions.

Then we started the exploration of the design space. We explored each category, one by one, and let participants try all the possible options within that category and select their preferred one. When moving to the next category, participants would keep their prior preferences but were always allowed to revisit previous selections and update their preferences. Akin to the work by Nicolau et al. [54], the procedure mimicked an optometry appointment where interface variations were experienced until the participant and researcher were confident in the selection of the participant's overall preferred configuration. When trying each configuration, participants could try out the application freely, either punching in S_{loc} or defending from punches in S_{mov}. In S_{loc}, the targets were the coach's hand/arm protections, whereas in S_{mov} the targets were the coach's incoming gloves. Sound feedback informed participants if they have missed or hit the target – e.g., the sound of hitting the hand/arm protections, head, or body was different.

After selecting their preferred configuration, participants were asked to perform a short, controlled task to assess its ability to convey the intended meaning. In S_{loc}, participants were asked to punch the coach's hand protections, by understanding their location. Participants had 5 trials, where they had a maximum of 5 attempts (per trial) to hit one of the coach's hands, without time constraints. The hands' positions would change between trials. In S_{mov}, participants were asked to guard their body or head (using their hands) to protect themselves from incoming punches that come to different locations (6 trials). Punches were thrown every 4.5 seconds (becoming audible 3 seconds prior to reaching the user in the On Proximity condition), giving enough time for users to defend themselves from punches. These values correspond to relatively slow punches and were obtained through experimentation in pilot sessions with blind novice users. In both cases, we limited the possible hand locations to a 2×3 grid for simplicity; [up(head), down(torso)] $\times$ [left, center, right].

After completing the task, participants were asked to walk the researcher through their thought and selection process. Finally, we conducted a 10-minute semi-structured interview focused on their thoughts about the experience, and possible changes they would like to see or be able to configure/change on the audio feedback provided. All participants were compensated for their time.

4.5 Participants

For S_{loc}, we recruited eight legally blind participants, seven of them with light perception at most and one with residual sight in one eye. One participant's age ranged between 18-30, five between 30-50, two were over 50. As for S_{mov}, we recruited eight participants who are legally blind. Six of them were totally blind and two had residual vision. P16 had a minor hearing impairment in the right ear. Three participants' ages ranged between 18-30, four between 30-50 and one was older than 50. Overall, two participants (P8, P10), reported having tried VR once. None of the participants had experience with boxing.

4.6 Data Analysis

We performed a thematic analysis with primarily deductive coding [73]. We created a codebook based on our concepts of interest, which included the categories and values for the design space and 17 other codes associated with how the configuration was perceived (e.g., Novelty, Comfortable, Overload). The codebook was created after the user studies were conducted, when researchers had familiarity with the data but before the coding process. The codebook was discussed and iterated by the whole research team and used to code all audio transcriptions for both scenarios. The thematic analysis result presented below encapsulates the discussions that ensued based on the coding, observations, and notes taken during the user studies. Note that researchers translated participants' quotes to English as sessions were conducted in non-English speaking countries.

5 Findings

We examine participants' preferences (Table 2) and personalization process when engaged in exploring the design space. We first present the success rate with their preferred method in each case study. Next, we discuss the established themes from reflecting, discussing and iterating over the findings by collaborative sense-checking the data resulting from both scenarios.

5.1 Success Rate

As a quantitative assessment of the effectiveness of the participants' preferred configuration, we asked participants to hit five different locations of the opponent standing still (S_{loc}) or guard themselves against moving objects coming from six different locations (S_{mov}) at the end of the design space exploration. As a result, most participants completed the task successfully on their first try regardless of their configuration choices. For S_{loc}, all participants were able to hit the target with their preferred configuration in all tasks, being that in 82.5% (33 out of 40) of the trials the target was hit on the first attempt. Similarly, for S_{mov}, six participants were able to guard themselves from the target in all trials and the other two guarded themselves successfully in half of the trials (87.5%, 42 out of 48).

5.2 The Overall Preference

The preferred configuration while exploring the design space is shown in Table 2. Regarding **Sound Type** – in both scenarios – *Speech* was

Table 2: Participants' audio configuration preference for S_{loc} (P1-P8), and S_{mov} (P9-P16).

Scenario	PID	Trigger	Sound Type	Time Signal	Concurrency	Spatialization
S_{loc}	P1	On Movement	Sonification	Discrete	Sequential	3D
	P2	On Movement	Speech	Discrete	Sequential	3D
	P3	On Movement	Speech	Discrete	Sequential	3D
	P4	On Request	Speech	Discrete	Sequential	3D
	P5	On Request	Speech	Discrete	Concurrent*	Dichotic
	P6	On Existence	Speech	Periodic	Sequential	3D
	P7	On Request	Speech	Discrete	Sequential	3D
	P8	On Existence	Speech	Periodic	Sequential	3D
S_{mov}	P9	On proximity	Speech	Discrete	Concurrent	Monaural
	P10	On proximity	Speech	Discrete	Sequential	Monaural
	P11	On proximity	Speech	Discrete	Concurrent	Monaural
	P12	On proximity	Speech	Discrete	Sequential	Monaural
	P13	On proximity	Speech	Periodic	Sequential	Dichotic
	P14	On movement	Speech	Continuous	Sequential	3D
	P15	On movement	Speech	Continuous	Concurrent	Dichotic
	P16	On movement	Speech	Periodic	Sequential	Monaural

highly preferred ($N = 15$ out of 16) when compared with *Sonification* because it conveyed greater certainty about the position of elements. As for **Concurrency**, the majority of the participants ($N = 12$) preferred the information to be delivered sequentially rather than concurrently even when there were multiple changes happening simultaneously[1] mostly due to cognitive load. For the same reason, 10 participants preferred **Feedback Mode** to be *Discrete*.

On the other hand, configurations in the remaining dimensions varied depending on the scenario. For instance, in **Spatialization**, 3D audio was highly preferred in the S_{loc} scenario, while monaural audio was more popular for the S_{mov} scenario. It is relevant to note, however, that the monaural condition was not explored in the S_{loc} scenario. Likewise, preferences for the **Trigger** category were different between scenarios. The preferred trigger type varied between participants in the S_{loc} scenario while participants in the S_{mov} scenario slightly preferred feedback triggered on proximity. We further detail the participants' preferences and their rationale in the upcoming sections.

5.3 Speech Feedback for Precise Information

All but one participant chose to use *Speech* over *Sonification* as **Sound Type**, mainly because it provided the exact location (e.g., "upper left") of the objects of interest. P8 mentioned:

> *"I think it's the assertiveness in the message because the voice tells us clearly, where it is... where the target is. Top left, down right, whatever."*

This preference contrasted with our expectations, especially when conveying information about moving objects (i.e., S_{mov}), given the ability of sonification techniques to convey feedback about dynamic changes [6, 32, 57]. One possible reason is the limited number of discrete options (a set of 6 predefined locations such as "upper right"), which make it easy and quick to describe using speech. A less restrictive environment with continuous values (e.g., allowing the location to be anywhere in the environment) could be more difficult to describe via speech and therefore put sonification as an appropriate candidate. In this scenario, the need to interpret the audio cues and convert them to a spatial location in the environment made it harder – and more cognitively demanding – to identify the exact location of objects compared to a simple speech message.

[1]Note that P5 chose *Concurrent* combined with an *On Request* trigger, meaning he could experience concurrent feedback if pressing both controllers simultaneously. Still, he never pressed both buttons simultaneously and therefore did not experience concurrent feedback

Another explanation could be the participants' little exposure to sonification. Unlike speech feedback, sonification may require some training to comprehend its meaning [57, 79]. Conversely, participants found in speech a familiar modality that they are exposed to daily on their devices. P5 stated,

> *"The voices are good, we already know them [i.e. familiar TTS voices]. The voices you got are our friends. Therefore it is very good, and they are very perceptible".*

In addition, the use of different voices (male and female) to convey information about each of the hands, helped participants to better identify the feedback source, as described by P8:

> *"The voice is the more effective of all, as it also has the detail of having a female voice on the left and a male voice on the right (...) It helps us identify much better and separate much better what is left and what is right."*

5.4 Spatialization was preferred, but not always!

The S_{loc} scenario has shown a major preference (7 out of 8) for *3D audio* in comparison to *Dichotic* presentation of audio, mainly because it transmits the real position of the target, making it more realistic than presenting sound in only one of the ears. When used together with speech, *3D audio* worked as a complement and helped participants gain confidence about the location of objects. For instance, P7 mentioned:

> *"The sound, which sounds different from left to right, informed me about the direction anyway, so I could grasp the location a little more clearly."*

In contrast to S_{loc}, the S_{mov} scenario has also explored a *monaural* condition that ended up being preferred by the majority of participants (5 out of 8). Note that in S_{mov} we used the HMD speakers, which may result in a worse perception of location when compared to the use of headphones. Still, the main reason for preferring *Monaural* was related to transmitting distance (which is not explored in S_{loc}), which made it more difficult for participants to detect the sound when the object was further away. In this scenario, participants wanted to be aware that one punch was coming and therefore preferred the condition where that information was available – and more clearly – as early as possible. Thus, despite the advantages of *3D audio* in conveying more realistic feedback and information about direction, representing distance needs careful consideration to make sure that relevant information is clearly transmitted to the users even when further away.

In addition, *Monaural* sound was preferred by P16, who has a hearing impairment in one ear, because it allowed him to rely on one ear without missing useful information, unlike *3D or Dichotic audio* which requires both ears to interpret the feedback. It suggests the need to further consider the impact of design decisions when users have multiple impairments:

> *"There is only one reason. Your right and left ears are not the same. So, I think the accuracy will be higher if listening to the right and left sides together."* (P16)

5.5 The Effect of Perceived Mental or Physical Loads on Preferred Trigger Options

Participants were split in their preference for the audio trigger. Three preferred to control when to get feedback (*On Request*) as they can expect when they will get feedback just by pressing a button on the hand-held controller and act upon it. P7 stated:

> *"I prefer to press the button (...) because this way I control the game."*

On the other hand, some participants found that pressing a button to receive feedback was an additional step that made the experience more demanding, particularly given the novelty of the experience. P2 specified that,

> *"Note that it distracts you a bit, having to press a button. It's another mechanical thing that you have to do."*

As for S_{mov}, slightly more than half of the participants chose *On proximity* over *On Movement*, suggesting that they would like to receive feedback only when the moving objects are close enough to demand action. Otherwise, it may be overwhelming (or unnecessary) to keep track of changes in the movement, which in this case was likely caused by relatively slow punches and longer times to react. P11 said,

> *"With the first option [On movement], I have to detect the sound beforehand and prepare for it. But, with this one [On proximity], it is enough to react as soon as I hear the sound. There is no need to hold my hands in advance."*

5.6 Concurrent and Continuous Feedback to Convey Urgency

Concurrency of audio feedback was important to explore in both scenarios, because the opponent may move their two hands simultaneously. Conveying feedback one hand at a time takes more time for conveying the same amount of information. However, while past research has shown people's ability to identify relevant information levering multiple audio sources [8, 27], this overwhelmed most participants when they tried to understand both audio sources concurrently. P10 stated that,

> *"The concurrent one was frantic as it told all the sounds of the punches even before I guard them."*

Still, the decision to convey multiple sounds concurrently or sequentially may be influenced by time constraints. For instance, in S_{loc} it is not time-sensitive to know the location of both hands, since the participant could hit the location of either of the coach/opponent's hands. As a consequence, participants did not see the advantage of getting that information more quickly with concurrent feedback.

Continuous feedback was also linked to a higher cognitive load due to the amount of information transmitted. We expected more participants in S_{mov} to choose *Continuous* feedback than the ones in S_{loc} as it can represent the dynamic changes during the continuous movement. The result came out as expected for S_{loc}; none of the participants preferred *Continuous* feedback. However, only two out of eight participants chose the continuous feedback in S_{mov} because others found it overwhelming and unintuitive when combined with speech feedback – note that this may be different in contexts where participants prefer sonification instead of speech. However, some participants commented that continuous feedback can be useful when giving users a sense of urgency. For example, P14 said that,

> *"I think this sound [Continuous] has the effect of being able to feel the pressuring situation and making it more immersive because it keeps telling me [like an alert]."*

5.7 The Desire for Challenging Configurations for Game-Like Experiences

Our main objective was to explore the participants' perceptions of object location and movement. However, participants' preferences were highly influenced by a tradeoff between perceiving these elements and what they thought was an interesting challenge as in a game-like environment (e.g., when hitting their coach's hands).

> *"I differentiated between gameplay and auditory perception. In terms of gameplay, it is more challenging the more obstacles we have, right?"* (P8)

Participants' approach to this experience has clearly influenced their preferred configuration. For instance, participants who were more focused on perceiving the environment would prefer to trigger feedback *On Request* in order to have more control over the experience. In contrast, those approaching it as a game-like experience associated the lack of control over the feedback as a more challenging experience and preferred *On Movement*,*On Proximity* or *On Existence* triggers.

Additionally, P8 explicitly mentioned that he selected sequential feedback because he was able to better perceive the location of both hands, but that he would feel more challenged and engaged if receiving concurrent feedback. He felt that these configurations could fit different game levels, where players could progress based on expertise and success:

> *"What I would do is have levels, for the easiest level have it discrete, for example, for the intermediate difficulty having the periodic, and then continuous (...) And when you are used to it, you move to the next level."* (P8)

These comments support that the design is highly influenced by the context and main objectives of the VR application and that designers should consider how each feedback decision contributes to their intended experience.

5.8 Feedback Realism for Immersive Experiences

Both scenarios provided feedback about the coach/opponents' hand position in addition to feedback on collision. Still, participants commented that this felt short of a fully immersive experience, which could be augmented by realistic sounds that could either provide useful information – the opponent breathing to convey their location – or just background sound (e.g., the crowd cheering). Also, many participants suggested having additional sound cues for landmarks of the surrounding space and adversary. For instance,

> *"I mean, if the opponent was talking, we could tell if they were on the left or right." (P6)*

Others suggested the "arena" could have different crowd sounds to reference when they get away from the opponent. Another frequent request was to have human voices – specifically of a coach – more excited, shouting the location the user should hit or even shout from the side during a competitive match. For instance,

> *"I think it will be more immersive if you include the sound of the coach shouting from the side or the sound of the audience." (P13)*

5.9 The Potential of VR for Embodiment

The two boxing scenarios were unembellished, having only the bare necessity for the experience. While the experience could be more immersive, participants perceived it as a reliable proxy of what boxing is actually like:

> *"I'll tell you something funny, I never payed any attention to boxing, but I came to see how it would be. I once saw a boxing bag and gloves, but I had never punched the bag. I found it super interesting, for those who can't see, using this they have a perfect notion of boxing."* (P5)

One participant suggested augmenting their own body movement in addition to the representations of other objects to better understand their position in the world. This is not surprising given that blind people, particularly congenital blind people, are at risk of having posture deficiencies and tend to have lower posture awareness [2]. P10 provided the following example:

> *"A person stretches their arm up and the sound goes up, stretches [the arm] down it goes down, to the right to the right, or left.. In other words, we could be hearing the sound corresponding to my arm movement"*

5.10 Experiencing the Unexplored with VR

While the experience could be similar to that of a boxing coach, it would hardly be fully autonomous for blind users in the real world. This sort of approach to inaccessible sports (and other activities) can enable blind people to autonomously experience a variety of contexts that would otherwise be out of – or difficult to – reach.

> *"Most blind people love baseball very much. But even though there are a lot of baseball games, we can't do it because we don't know where the ball is coming from. But if this research develops, it would be really helpful if blind people could play games like this with sighted people on an equal basis."* (P14)

6 Discussion

In this section, we present lessons learned after exploring a design space of the auditory representation of virtual objects' location and of their behaviours. In addition, we demonstrate how our design space exploration may help inform the design of accessible VR boxing experiences.

6.1 Immersive VR Environments for Supporting Accessible Experiences for Blind People

Immersive VR environments have the potential to create accessible experiences, opening new avenues for creating holistic experiences that are far more accessible than any real-world location, as everything is digitized and tracking of virtual elements is accurate and immediate.

In immersive VR environments, users' heads and hands are tracked with HMDs and hand-tracking devices (e.g., hand-held controllers). This can be cumbersome for blind people in terms of setup, compared to traditional video games and other applications with 2D input and output devices. However, it comes with significant benefits as not only users can take a first-person perspective, but also control their avatar with their own body movements – e.g., changing the viewpoint by moving their head orientation or punching by moving their hands. This contrasts with the major challenges faced by blind users when dealing with camera controls and joystick/keyboard input in traditional 3D environments. This opportunity for embodiment has put our VR Boxing applications as safe proxies for this sport, as "*for those who can't see, using this they have a perfect notion of boxing*" (P5). Moreover, because virtual environments are not bound by real-world limitations (e.g., object detection, tracking), developers can use precise information about every object in the scene (label, position, orientation) as well as user events, and have full control of the environment (e.g., changing the object properties or the feedback design) as demonstrated in prior studies [1, 5, 13, 15, 16, 37, 48].

In this study, we used such knowledge (in particular, hand location) to experiment with various configurations with blind users and understand the impact of each design decision. Still, our applications were fairly basic and participants felt the need for a more realistic and immersive experience (e.g., the audience or the coach shouting). Designers and developers should consider how to combine both realistic audio feedback (e.g., the opponent breathing or the audience cheering), and the audio representation of visual objects (supported by our design space) that do not necessarily produce sound (e.g., the opponent's hands).

6.2 The Design Space as a Framework to Create Accessible VR Experiences

The design space supported the creation of multiple configurations to represent objects and their movement in two different scenarios. Participants' variance in preferences alongside the high success rates when performing the study tasks show the viability of the different configurations implemented. Our decision to explore one scenario independently by two separate teams (who collaborated on the definition of the design space), intended to build up evidence that the design space supports the creation of accessible VR experiences. The two teams targeted two different (but related) scenarios and explored both common and different values of design space categories, coming up with different implementations. This showcases the versatility of the design space in supporting experimentation. On the other hand, it did not allow for comparisons between the explorations of the different teams. As an example, *Monaural* audio was preferred by many participants of the S_{mov} scenario, but the S_{loc} scenario did not explore such configuration. However, the main goal of the design space is to support designers and developers in exploring different solutions, and not to perform generalist comparisons to reach a preferred, one-size-fits-all solution.

6.3 No One-Size-Fits-All Solution

While we often strive for the "one" best solution, when it comes to augmenting objects with audio feedback our findings show a great variety of preferences among blind users, both within and across scenarios. This suggests that designers should consider providing users (and designers themselves) the ability to personalize their configurations.

As evident in Table 2, the preferred configuration varied depending on how each participant perceives a task and what they value. For example, participants who wished to be challenged as in game-like environments preferred the audio feedback to be given automatically. For instance, P8 envisioned a sequence of game levels where configurations get increasingly more challenging to keep the users engaged. On the other hand, participants who appreciate the sense of user control preferred feedback played on request. Yet, all of them were able to achieve high success rates implying that the effective configurations can differ depending on the user and the context. Thus, we highlight the importance of allowing: 1) designers to explore the design space to create experiences that are in line with their context and objectives – e.g., is the main goal to convey precise information about the object location or to challenge users with more demanding tasks? 2) users to try out and select different configurations to understand what are their preferred combinations.

Meanwhile, findings also imply that the preferred configuration may change over time as users would become familiar with the feedback design. For instance, novice users may prefer speech feedback for its low learning cost but they may switch to sonification for efficiency after long-term use. In addition, one should note that even when participants chose the same configuration, the reasons for their decision may vary. Thus, it is essential for designers to focus more on the participants' thought process, in particular on how each feedback decision was made rather than the decisions themselves in order to provide a more personalized experience.

6.4 Leveraging the Design Space to Create a Virtual Boxing Experience

We presented two scenarios in the context of virtual boxing to understand and exemplify how our proposed design space can be leveraged to design augmentative audio feedback for blind people. By exercising the design space, we found no one size fits all solution, with participants having a widespread selection of configurations. Still, we were able to identify trends, specific to our boxing scenarios, that inform the future creation of VR boxing experiences:

- *Trigger*: When familiarizing players with the demands of the game, trigger on request may lower the challenge and allow players to solely focus on understanding the relationship between their actions and the feedback received. *On Request* in S_{loc}, can also accommodate interaction where players are in control of the pacing of the task but this also demands action from users, which may increase their load. **On-Proximity**, in particular for moving objects (S_{mov}), appears to be the consensual choice, as feedback prompts the user only when actions are needed thus encoding additional information that *On Request* did not.
- *Sound Type*: *Speech* is to be selected due to its ability to provide precise information about location, given a limited number of (short) options (e.g., top left, bottom right).
- *Time Signal*: Since actions (punches) are relatively quick, and are a sequence of one-time events that are mostly relevant on occurrence, *Discrete* time signals are the most suitable option.
- *Concurrency*: if the time delay is minimal in sequential feedback, the disadvantages of concurrent feedback (mental load) outweigh its benefits. Still, it can be used to give a sense of urgency and challenge.
- Spatialization: use *3D audio* when conveying location where the distance of an object does not change (e.g., S_{loc}). However, for moving objects (e.g., S_{mov}), where distance is crucial, ensure that users can accurately judge it, or else it might even be optimal to rely on *Monaural* options.

Although we advocate for the ability to customize, it is equally relevant to highlight that these choices significantly affect the experience someone has with virtual environments (e.g., make it more or less perceivable or engaging). We believe the design space can provide a framework from which to explore the viable design options and let users show the consequences of the designs while simultaneously uncovering latent feedback desires. It is then up to researchers and designers to select what we want the experience to be.

6.5 Limitations and Future Work

Our user study explored the design space under two different scenarios related to boxing. Exploring these scenarios separately fits a real-world context – e.g., a training experience where the user is attacking to hit their coach's hand protections. Still, we did not explore a full, unique experience where users could get feedback both related to attack and defense. Further research is needed to understand how to combine these separate scenarios into a comprehensive boxing experience. In addition, we instantiated the design space in the context of boxing alone. Researchers and designers may explore different scenarios (e.g., in other sports or games) and as a consequence build new (and accessible) VR applications. Alternatively, one may use this design space to analyze and catalog existing applications in this domain to better understand what type of audio feedback was used and how. We also note that design spaces are not immutable and we encourage others to explore and iterate over this theoretical framework.

We have made design and implementation decisions – e.g., the sonification techniques used or the proximity perimeter chosen – that may impact user preferences. For this reason, some findings may not be generalized. For instance, we used Unity native assets for our spatial audio representation. More recent, state-of-the-art implementations could further improve this representation. In addition, having nine possible positions simplified the use of speech. Different setups without constraints on target location could make it harder to convey such information using speech. However, note that our main goal was not to find the best overall solution, but instead to use the design space as a tool to explore different options, reflect and perform conscious design decisions that augment visual information with an auditory representation.

7 Conclusion

We proposed a design space with nine categories to support designers in exploring various audio representations of virtual objects and their behaviours in order to create accessible VR experiences for blind people. We demonstrated how the design space can be used with two VR boxing example scenarios. We performed two independent design space explorations, by separate research teams, showcasing its flexibility as a theoretical framework capable of supporting the exploration of multiple viable solutions for augmenting virtual objects with a nonvisual, audio representation. In addition, it enables designers to reflect on the impact of different design decisions, supporting them in creating accessible experiences. We note that a design space is iterative and we encourage other researchers and designers to not only use our design space to support the creation of accessible VR experiences but also to iterate on it, increment, adjust or propose changes as they see fit.

Acknowledgments

The authors wish to thank the participants of the two user studies. This work was supported in part by FCT through the LASIGE Research Unit (UIDB/00408/2020, UIDP/00408/2020) and the AccessVR project (2022.08286.PTDC). It was also supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2023-2020-0-01460) supervised by the IITP (Institute for Information Communications Technology Planning Evaluation).

References

[1] K. Allain, B. Dado, M. Van Gelderen, O. Hokke, M. Oliveira, R. Bidarra, N. D. Gaubitch, R. C. Hendriks, and B. Kybartas. An audio game for training navigation skills of blind children. In 2015 IEEE 2nd VR workshop on sonic interactions for virtual environments (SIVE), pp. 1–4. IEEE, 2015. 8

[2] A. Z. Alotaibi, A. Alghadir, Z. A. Iqbal, and S. Anwer. Effect of absence of vision on posture. Journal of physical therapy science, 28(4):1374–1377, 2016. 8

[3] R. Andrade, M. J. Rogerson, J. Waycott, S. Baker, and F. Vetere. Playing blind: Revealing the world of gamers with visual impairment. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19, p. 1–14. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3290605.3300346 1

[4] R. Andrade, J. Waycott, S. Baker, and F. Vetere. Echolocation as a means for people with visual impairment (pvi) to acquire spatial knowledge of virtual space. ACM Transactions on Accessible Computing (TACCESS), 14(1):1–25, 2021. 2

[5] O. Ariza, G. Bruder, N. Katzakis, and F. Steinicke. Analysis of proximity-based multimodal feedback for 3d selection in immersive virtual environments. In 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 327–334. IEEE, 2018. 8

[6] S. Barrass and G. Kramer. Using sonification. Multimedia systems, 7(1):23–31, 1999. 6

[7] J. Blauert. Spatial hearing: the psychophysics of human sound localization. MIT press, 1997. 4

[8] A. W. Bronkhorst. The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. Acta Acustica united with Acustica, 86(1):117–128, 2000. 4, 7

[9] A. Burkins and R. Kopper. Wayfinding by auditory cues in virtual environments. In 2015 IEEE Virtual Reality (VR), pp. 155–156. IEEE, 2015. 1

[10] S. Büttner, H. Mucha, M. Funk, T. Kosch, M. Aehnelt, S. Robert, and C. Röcker. The design space of augmented and virtual reality applications for assistive environments in manufacturing: a visual approach. In Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments, pp. 433–440, 2017. 2

[11] S. K. Card, J. D. Mackinlay, and G. G. Robertson. The design space of input devices. In Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 117–124, 1990. 2

[12] H. Chen, A. Dey, M. Billinghurst, and R. W. Lindeman. Exploring the design space for multi-sensory heart rate feedback in immersive virtual reality. In Proceedings of the 29th Australian Conference on Computer-Human Interaction, OZCHI '17, p. 108–116. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3152771.3152783 2

[13] Y. F. Cheng, H. Yin, Y. Yan, J. Gugenheimer, and D. Lindlbauer. Towards understanding diminished reality. In CHI Conference on Human Factors in Computing Systems, pp. 1–16, 2022. 8

[14] S. Chung, K. Lee, and U. Oh. Investigating three-dimensional directional guidance with nonvisual feedback for target pointing task. In 2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), pp. 206–210. IEEE, 2020. 2

[15] S. Chung, K. Lee, and U. Oh. Understanding the two-step nonvisual omnidirectional guidance for target acquisition in 3d spaces. In 2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 339–346. IEEE, 2021. 2, 8

[16] S. Chung, S. Park, S. Park, K. Lee, and U. Oh. Improving mealtime experiences of people with visual impairments. In Proceedings of the 18th International Web for All Conference, pp. 1–11, 2021. 8

[17] A. Cobo, N. E. Guerrón, C. Martín, F. del Pozo, and J. J. Serrano. Differences between blind people's cognitive maps after proximity and distant exploration of virtual environments. Computers in Human Behavior, 77:294–308, 2017. 2

[18] C. Colwell, H. Petrie, D. Kornbrot, A. Hardwick, and S. Furner. Haptic virtual reality for blind computer users. In Proceedings of the third international ACM conference on Assistive technologies, pp. 92–99, 1998. 2

[19] E. C. Connors, E. R. Chrastil, J. Sánchez, and L. B. Merabet. Virtual environments for the transfer of navigation skills in the blind: a comparison of directed instruction vs. video game based learning approaches. Frontiers in human neuroscience, 8:223, 2014. 2

[20] K. Dadamis, J. Williamson, and R. Murray-Smith. 3d cues for human control of target acquisition in auditory augmented reality. 2022. 2

[21] A. R. Façanha, T. Darin, W. Viana, and J. Sánchez. O&m indoor virtual environments for people who are blind: A systematic literature review. ACM Transactions on Accessible Computing (TACCESS), 13(2):1–42, 2020. 2

[22] A. R. Façanha, T. Darin, W. Viana, and J. Sánchez. O&m indoor virtual environments for people who are blind: A systematic literature review. ACM Trans. Access. Comput., 13(2), aug 2020. doi: 10.1145/3395769 1

[23] B. Gao, H. Kim, H. Lee, J. Lee, and J.-I. Kim. Effects of continuous auditory feedback on drawing trajectory-based finger gestures. IEEE Transactions on Human-Machine Systems, 48(6):658–669, 2018. 1, 2

[24] B. Gao, Y. Lu, H. Kim, B. Kim, and J. Long. Spherical layout with proximity-based multimodal feedback for eyes-free target acquisition in virtual reality. In International Conference on Human-Computer Interaction, pp. 44–58. Springer, 2019. 2

[25] T. A. Garner. Echoes of Other Worlds: Sound in Virtual Reality: Past, Present and Future. Springer, 2017. 2

[26] D. Gonçalves, A. Rodrigues, and T. Guerreiro. Playing with others: Depicting multiplayer gaming experiences of people with visual impairments. In The 22nd International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS '20. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3373625.3418304 1

[27] J. Guerreiro and D. Gonçalves. Scanning for digital content: How blind and sighted people perceive concurrent speech. ACM Transactions on Accessible Computing (TACCESS), 8(1):1–28, 2016. 7

[28] J. Guerreiro, D. Sato, D. Ahmetovic, E. Ohn-Bar, K. M. Kitani, and C. Asakawa. Virtual navigation for blind people: Transferring route knowledge to the real-world. International Journal of Human-Computer Studies, 135:102369, 2020. 1, 2

[29] T. Hermann, A. Hunt, and J. G. Neuhoff. The sonification handbook. Logos Verlag Berlin, 2011. 4

[30] W. Heuten, D. Wichmann, and S. Boll. Interactive 3d sonification for the exploration of city maps. In Proceedings of the 4th Nordic conference on Human-Computer Interaction: Changing roles, pp. 155–164, 2006. 1

[31] E. R. Hoeg, L. J. Gerry, L. Thomsen, N. C. Nilsson, and S. Serafin. Binaural sound reduces reaction time in a virtual reality search task. In 2017 IEEE 3rd VR workshop on sonic interactions for virtual environments (SIVE), pp. 1–4. IEEE, 2017. 2

[32] O. Höner, A. Hunt, S. Pauletto, N. Röber, T. Hermann, and A. O. Effenberg. Aiding movement with sonification in "exercise, play and sport". In The sonification handbook. 2011. 6

[33] I. Hussain, L. Chen, H. T. Mirza, K. Xing, and G. Chen. A comparative study of sonification methods to represent distance and forward-direction in pedestrian navigation. International Journal of Human-Computer Interaction, 30(9):740–751, 2014. 1

[34] G. India, M. Jain, P. Karya, N. Diwakar, and M. Swaminathan. Vstroll: An audio-based virtual exploration to encourage walking among people with vision impairments. In The 23rd International ACM SIGACCESS Conference on Computers and Accessibility, pp. 1–13, 2021. 2

[35] D. Jain, S. Junuzovic, E. Ofek, M. Sinclair, J. Porter, C. Yoon, S. Machanavajhala, and M. Ringel Morris. A taxonomy of sounds in virtual reality. In Designing Interactive Systems Conference 2021, pp. 160–170, 2021. 2

[36] T. Ji, B. R. Cochran, and Y. Zhao. Vrbubble: Enhancing peripheral awareness of avatars for people with visual impairments in social virtual reality. In The International ACM SIGACCESS Conference on Computers and Accessibility. Association for Computing Machinery, New York, NY, USA, 2022. 2

[37] O. B. Kaul and M. Rohs. Haptichead: 3d guidance and target acquisition through a vibrotactile grid. In Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, pp. 2533–2539, 2016. 8

[38] N. W. Kim, S. C. Joyner, A. Riegelhuth, and Y. Kim. Accessible visualization: Design space, opportunities, and challenges. In Computer Graphics Forum, vol. 40, pp. 173–188. Wiley Online Library, 2021. 2

[39] G. Kramer, B. Walker, T. Bonebright, P. Cook, J. H. Flowers, N. Miner, and J. Neuhoff. Sonification report: Status of the field and research agenda. 2010. 4

[40] J. Kreimeier and T. Götzelmann. Two decades of touchable and walkable virtual reality for blind and visually impaired people: A high-level taxonomy. Multimodal Technologies and Interaction, 4(4), 2020. doi: 10.3390/mti4040079 2

[41] J. Kreimeier, P. Karg, and T. Götzelmann. Blindwalkvr: formative insights into blind and visually impaired people's vr locomotion using commercially available approaches. In Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments, pp. 1–8, 2020. 2

[42] O. Lahav and D. Mioduser. Haptic-feedback support for cognitive mapping of unknown spaces by people who are blind. International Journal of Human-Computer Studies, 66(1):23–35, 2008. 2

[43] A. Lécuyer, P. Mobuchon, C. Mégard, J. Perret, C. Andriot, and J.-P. Colinot. Homere: a multimodal system for visually impaired people to explore virtual environments. In IEEE Virtual Reality, 2003. Proceedings., pp. 251–258. IEEE, 2003. 1

[44] G. Leplâtre and S. A. Brewster. Designing non-speech sounds to support navigation in mobile phone menus. 2000. 1

[45] D. Lottridge, R. Weber, E.-R. McLean, H. Williams, J. Cook, and H. Bai. Exploring the design space for immersive embodiment in dance. In 2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 93–102. IEEE, 2022. 2

[46] A. MacLean, R. Young, V. Bellotti, and T. Moran. Design space analysis: Bridging from theory to practice via design rationale. Proceedings of Esprit, 91:720–730, 1991. 2

[47] S. Maidenbaum and A. Amedi. Blind in a virtual world: Mobility-training virtual reality games for users who are blind. In 2015 IEEE Virtual Reality (VR), pp. 341–342. IEEE, 2015. 2

[48] S. C. Michalski, A. Szpak, D. Saredakis, T. J. Ross, M. Billinghurst, and T. Loetscher. Getting your game on: Using virtual reality to improve real table tennis skills. PloS one, 14(9):e0222351, 2019. 8

[49] T. Morelli, J. Foley, L. Columna, L. Lieberman, and E. Folmer. Vi-tennis: a vibrotactile/audio exergame for players who are visually impaired. In Proceedings of the Fifth International Conference on the Foundations of Digital Games, pp. 147–154, 2010. 2

[50] M. R. Morris, J. Johnson, C. L. Bennett, and E. Cutrell. Rich representations of visual content for screen reader users. In Proceedings of the 2018 CHI conference on human factors in computing systems, pp. 1–11, 2018. 2

[51] M. Mott, J. Tang, S. Kane, E. Cutrell, and M. Ringel Morris. "i just went into it assuming that i wouldn't be able to have the full experience": Understanding the accessibility of virtual reality for people with limited mobility. In Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS '20. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3373625.3416998 1

[52] J. Müller, F. Alt, D. Michelis, and A. Schmidt. Requirements and design space for interactive public displays. In Proceedings of the 18th ACM international conference on Multimedia, pp. 1285–1294, 2010. 2

[53] V. Nair, J. L. Karp, S. Silverman, M. Kalra, H. Lehv, F. Jamil, and B. A. Smith. Navstick: Making video games blind-accessible via the ability

to look around. In The 34th Annual ACM Symposium on User Interface Software and Technology, UIST '21, p. 538–551. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3472749.3474768 1, 2
[54] H. Nicolau, A. Rodrigues, A. Santos, T. Guerreiro, K. Montague, and J. a. Guerreiro. The design space of nonvisual word completion. In The 21st International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS '19, p. 249–261. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3308561.3353786 2, 4, 5
[55] C. O'Callaghan. Sounds and events. 2009. 2
[56] U. Oh, S. Branham, L. Findlater, and S. K. Kane. Audio-based feedback techniques for teaching touchscreen gestures. ACM Transactions on Accessible Computing (TACCESS), 7(3):1–29, 2015. 1
[57] U. Oh, S. K. Kane, and L. Findlater. Follow that sound: using sonification and corrective verbal feedback to teach touchscreen gestures. In Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility, pp. 1–8, 2013. 1, 6
[58] R. Pasnau. What is sound? The Philosophical Quarterly, 49(196):309–324, 1999. 2
[59] F. J. Penedo and J. R. Dahn. Exercise and well-being: a review of mental and physical health benefits associated with physical activity. Current opinion in psychiatry, 18(2):189–193, 2005. 4
[60] K. U. Phillips. Virtual reality has an accessibility problem. Scientific American, January 2020. 1
[61] L. Picinali, A. Afonso, M. Denis, and B. F. Katz. Exploration of architectural spaces by blind people using auditory virtual reality for the construction of spatial knowledge. International Journal of Human-Computer Studies, 72(4):393–407, 2014. 2
[62] K. Rector, C. L. Bennett, and J. A. Kientz. Eyes-free yoga: an exergame using depth cameras for blind & low vision exercise. In Proceedings of the 15th international acm sigaccess conference on computers and accessibility, pp. 1–8, 2013. 2
[63] G. Ren, S. Wei, E. O'Neill, and F. Chen. Towards the design of effective haptic and audio displays for augmented reality and mixed reality applications. Advances in Multimedia, 2018, 2018. 2
[64] V. Reynaert, F. Berthaut, Y. Rekik, and L. Grisoni. The effect of rhythm in mid-air gestures on the user experience in virtual reality. In IFIP Conference on Human-Computer Interaction, pp. 182–191. Springer, 2021. 2
[65] A. Ryan. Thoughts on accessibility issues with vr. Ablegamers, 2020. 1
[66] J. Sánchez, A. Tadres, A. Pascual-Leone, and L. Merabet. Blind children navigation through gaming and associated brain plasticity. In 2009 virtual rehabilitation international conference, pp. 29–36. IEEE, 2009. 2
[67] T. Senan, B. Hengeveld, and B. Eggen. Sounding obstacles for social distance sonification. In Proceedings of the 17th International Audio Mostly Conference, AM '22, p. 187–194. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3561212.3561239 1
[68] S. Serafin, N. C. Nilsson, E. Sikstrom, A. De Goetzen, and R. Nordahl. Estimation of detection thresholds for acoustic based redirected walking techniques. In 2013 IEEE Virtual Reality (VR), pp. 161–162. IEEE, 2013. 1
[69] A. F. Siu, M. Sinclair, R. Kovacs, E. Ofek, C. Holz, and E. Cutrell. Virtual reality without vision: A haptic and auditory white cane to navigate complex virtual worlds. In Proceedings of the 2020 CHI conference on human factors in computing systems, pp. 1–13, 2020. 1, 2
[70] B. A. Smith and S. K. Nayar. The rad: Making racing games equivalently accessible to people who are blind. In Proceedings of the 2018 CHI conference on human factors in computing systems, pp. 1–12, 2018. 1, 2
[71] G. Stoner. Vr is here to stay. it's time to make it accessible. Wired, March 2022. 1
[72] J. Su, A. Rosenzweig, A. Goel, E. de Lara, and K. N. Truong. Timbremap: enabling the visually-impaired to use maps on touch-enabled devices. In Proceedings of the 12th international conference on Human computer interaction with mobile devices and services, pp. 17–26, 2010. 1
[73] G. Terry, N. Hayfield, V. Clarke, and V. Braun. The sage handbook of qualitative research in psychology, Sep 2017. Thematic Analysis. doi: 10.4135/9781526405555 5
[74] L. Thevin, C. Briant, and A. M. Brock. X-road: virtual reality glasses for orientation and mobility training of people with visual impairments. ACM Transactions on Accessible Computing (TACCESS), 13(2):1–47, 2020. 2
[75] D. E. Warburton, C. W. Nicol, and S. S. Bredin. Health benefits of physical activity: the evidence. Cmaj, 174(6):801–809, 2006. 4
[76] M. Watanabe, T. Miura, M. Matsuo, M. Sakajiri, and J. Onishi. Goalbaural-ii: An acoustic virtual reality training application fornbsp;goalball players tonbsp;recognize various game conditions. In Computers Helping People with Special Needs: 18th International Conference, ICCHP-AAATE 2022, Lecco, Italy, July 11–15, 2022, Proceedings, Part II, p. 79–88. Springer-Verlag, Berlin, Heidelberg, 2022. doi: 10.1007/978-3-031-08645-8_10 2
[77] R. Wedoff, L. Ball, A. Wang, Y. X. Khoo, L. Lieberman, and K. Rector. Virtual showdown: An accessible virtual reality game with scaffolds for youth with visual impairments. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19, p. 1–15. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3290605.3300371 2
[78] S. Yasmin. Virtual reality and assistive technologies: A survey. International Journal of Virtual Reality, 18(2), 2018. 2
[79] T. Yoshida, K. M. Kitani, H. Koike, S. Belongie, and K. Schlei. Edgesonic: image feature sonification for the visually impaired. In Proceedings of the 2nd Augmented Human International Conference, pp. 1–4, 2011. 6
[80] Y. Zhao, C. L. Bennett, H. Benko, E. Cutrell, C. Holz, M. R. Morris, and M. Sinclair. Enabling people with visual impairments to navigate virtual reality with a haptic and auditory cane simulation. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2018. 2
[81] Y. Zhao, E. Kupferstein, H. Rojnirun, L. Findlater, and S. Azenkot. The effectiveness of visual and audio wayfinding guidance on smartglasses for people with low vision. In Proceedings of the 2020 CHI conference on human factors in computing systems, pp. 1–14, 2020. 1
[82] D. Zielasko, J. Heib, and B. Weyers. Systematic design space exploration of discrete virtual rotations in vr. In 2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 693–702. IEEE, 2022. 2
[83] J. Zong, C. Lee, A. Lundgard, J. Jang, D. Hajas, and A. Satyanarayan. Rich Screen Reader Experiences for Accessible Data Visualization. Computer Graphics Forum (Proc. EuroVis), 2022. 2