# Virtual Reality Audio Game for Entertainment & Sound Localization Training

Virtual Reality Audio Game for Entertainment & Sound Localization Training

LUCAS, G., GILBERTO*

Centro de Investigación y Transferencia en Acústica, CONICET, Universidad Tecnológica Nacional- Facultad Regional Córdoba, Córdoba, Argentina, lgilberto@frc.utn.edu.ar

FERNANDO, BERMEJO

Centro de Investigación y Transferencia en Acústica, CONICET, Universidad Tecnológica Nacional- Facultad Regional Córdoba, Córdoba, Argentina, fbermejo@unc.edu.ar
Facultad de Psicología, Universidad Nacional de Córdoba, Córdoba, Argentina

FABIÁN, C., TOMMASINI

Centro de Investigación y Transferencia en Acústica, CONICET, Universidad Tecnológica Nacional- Facultad Regional Córdoba, Córdoba, Argentina, ftommasini@frc.utn.edu.ar

CRISTIAN, GARCÍA BAUZA

CONICET, PLADEMA, Universidad Nacional del Centro de la Provincia de Buenos Aires, crgarcia@exa.unicen.edu.ar

Within the gaming and electronics industry, there is a continuous evolution of alternative applications. Nevertheless, accessibility to video games remains a persistent hurdle for individuals with disabilities, especially those with visual impairments due to the inherent visual-oriented design of games. Audio games (AGs) are electronic games that rely primarily on auditory cues instead of visual interfaces. This study focuses on the creation of a virtual reality AG for cell phones that integrates natural head and torso movements involved in spatial hearing. Its assessment encompasses user experience, interface usability, and sound localization performance. The study engaged eighteen sighted participants in a pre-post test with a control group. The experimental group underwent 7 training sessions with the AG. Via interviews, facets of the gaming experience were explored, while horizontal plane sound source localization was also tested before and after the training. The results enabled the characterization of sensations related to the use of the game and the interaction with the interfaces. Sound localization tests demonstrated distinct enhancements in performance among trained participants, varying with assessed stimuli. These promising results show advances for future virtual AGs, presenting prospects for auditory training. These innovations hold potential for skill development, entertainment, and the integration of visually impaired individuals.

CCS CONCEPTS • Applied computing → Psychology • Human-centered computing → Accessibility • Human-centered computing → Human computer interaction (HCI) → Interaction techniques → Auditory feedback • Software and its engineering → Contextual software domains → Virtual worlds training simulations; Interactive games

**Additional Keywords and Phrases:** audio games, spatial hearing training, natural interfaces

## 1 INTRODUCTION

Video games have become a significant part of our society's culture in recent decades. While they are predominantly used for entertainment and recreation, various studies have highlighted significant benefits when they are used as educational, training, and rehabilitation resources [1]. One of their greatest strengths lies in their motivational power; the continuous challenge and active role taken by players are highly favorable for learning and training various cognitive, social, and psychomotor skills [2]. Significant progress in the

---

ACM 1544-3965/2024/7-ART
https://doi.org/10.1145/3676557

gaming and electronic industry further enhances these alternative uses, yet video game accessibility remains an unresolved issue [3, 4, 5]. Individuals with disabilities often face restricted access to these systems, and challenges are even more complex for those with visual impairments due to the visual-centric nature of games. One alternative to overcome this barrier is the use of audio games (AGs), which are electronic games that rely primarily on auditory cues instead of visual interfaces [6]. The game mechanics, virtual environment and any necessary information (scores, instructions, time, etc.) are presented to the player mainly through loudspeakers or headphones.

Although AGs have been in development for several decades now, the progress and popularity of this type of game have been modest [7, 8]. While there are design guidelines with valuable recommendations [9, 10, 11], the field of AG design remains intricate, encompassing not only technological aspects but also crucial considerations of social, psychological and economic factors [3, 12]. For instance, previous studies [13, 14] revealed that in most cases AG interfaces used were simplistic (usually keyboard and basic stereo sounds), compromising interaction and gameplay mechanics, possibly leading to low motivation. Sometimes, the sense of immersion in AGs is degraded due to complex sonifications or audio descriptions that disrupt natural player interactions [8, 11, 15]. Furthermore, when AGs overly rely on verbal information, many players who are not fluent in the game's native language lose access to it [13, 16]. Deeper in this context, it becomes evident that the primary driving force behind AG creation stems from enthusiastic programmers, small independent studios and academic efforts that thoroughly investigate the field [17, 18, 19]. Large and established game companies do not participate due to the smaller audience compared to the global video game market [3, 12, 20].

The field of development is also characterized by a notable absence of individuals with visual disabilities who could broaden the current design scene [17]. A concomitant challenge lies in the absence of accessible and specific development tools for this type of games [8, 14]. In all these cases, the need to establish solid foundations for the development and innovation of these devices is evident. Different researchers and designers agree that knowledge about this type of game is still limited, highlighting the challenge of creating interesting AGs with an engaging gameplay experience [8, 17, 21].

Overcoming these limitations is important as AGs, beyond their entertainment value, hold significant potential for training different skills. With proper design, these developments can contribute to better adaptation to the physical and social environment for people with visual disabilities. For example, there are some AG prototypes designed to be used as educational tools in learning environments [22, 23], to train musical skills [24, 25, 26], to encourage tourism and knowledge of ancient cultural heritages [27], to practice Braille [28], to engage in physical exercise [29] and to train various aspects of hearing. The latter developments are particularly relevant considering that trained hearing allows individuals to be highly alert to their environment, enabling them to develop spatial orientation skills to perceive and navigate obstacles more effectively [30]. For example, AGs have proven to be an effective tool for echolocation training [31, 32] and for listening discrimination in noise [33, 34]. Furthermore, several studies have analyzed AGs and virtual auditory environments (VAE) designed to train spatial hearing skills in blind people. For instance, Sanchez [35] designed and evaluated the Audio Haptic Maze's impact on Orientation and Mobility (O&M) competencies, demonstrating its usability and effectiveness. Connors et al. [36] developed the Audio-based Environment Simulator, revealing a strong link between virtual environment success and real-world navigation skills. Balan et al. [37] reported improved navigation metrics after a short training session with a navigation AG. Meanwhile, Lahav et al. [38] integrated VAEs into O&M training, showing potential for competency diagnosis and exploration.

The studies mentioned above have in common that users are limited in the way they move to interact with the game. Another possible design strategy is to consider the natural movements that people make to explore the environment. In light of new approaches in Cognitive Science, encompassed under the term of Embodied [39, 40], this possibility could have a significant impact on the learning of perceptual skills. These theoretical postulations argue that perceptual acts are inseparably connected to the actions performed by individuals. Perception is structured based on the individual's exploratory activity, where they learn how sensory changes relate to their actions [41, 42, 43]. These embodied approaches provide a significant technological impact on design [44, 45, 46, 47]. The development of devices to train sensorimotor skills is meant to promote the user's engagement with the environment through active, embodied and continuous interaction [48].

There are a few studies that have tried to integrate body movements related to spatial hearing in their interfaces. For example, Bermejo et al. [30] adapted an AG called Audio Doom [49] using distributed loudspeakers, which allowed users to move their head to locate sound events. Although the training potential of this device is not further explored, this study verified the correct structuring of the virtual space by sighted and blind participants. In two independent yet similar studies, Allain et al. [50] and Cavaco et al. [51] designed audio games to train navigation skills in blind children. One study utilized the Oculus Rift headset, while the other employed a cell phone with a gyroscopic sensor. In both cases, positive feedback was collected from the players, highlighting the enjoyment and sense of immersion provided by the virtual auditory environment. However, robust indicators to evaluate the actual impact of the game on specific auditory variables were not presented. Magnusson et al. [52], using GPS data from a cell phone, present a slightly different approach. Players had to walk around to find hidden objects by interpreting monaural sounds, making it potentially useful for training mobility in blind individuals rather than spatial hearing. In its evaluation, the game was found to be enjoyable and engaging. Additionally, the rehabilitation staff participating in the study found it useful for their training programs.

Suzuki et al. [53] utilized a novel approach to conduct an integrative study, particularly advancing on the measurement of sound localization performance. The approach is known as ‘active listening’, and refers to a mode of spatial hearing that takes advantage of dynamic information induced by listeners’ movements. In one of their AGs called Hoy-Pippi [54] players must hit a virtual flying object using a handheld device (hammer). Audio is delivered through headphones and both motion sensors (for hammer and head) provide feedback to the system. The study evaluated localization skills of sighted participants through pre-post assessments in the frontal right horizontal plane, with game-trained and untrained groups. Although the reported results were limited due to the small sample size and heterogeneity of the groups, a significant improvement was observed in the trained group compared to the untrained group. In a subsequent study, the same researchers conducted an experiment with a similar AG called Bee Bee Beat [55]. In this game, the stimulus consisted of the sound of a bee's flight presented statically at a point around the player and masked by background music. Participants used a similar hammer to hit the bee and earn points. Pre-post tests with a control group were also conducted, further evaluating the use of individual or matched head-related transfer functions (HRTFs). Sound localization was evaluated using real sound sources placed in the horizontal and vertical planes. Participants used a rod to point at one of them, yielding either correct or incorrect responses. Among the main results, this study revealed a 20% improvement in correct responses in the post-test for the trained group, and this effect persisted one month later. Furthermore, the sound localization improvement using fitted HRTFs was similar to that observed when using their own HRTFs. The results of their experiments demonstrated the sound localization skills transferred by playing the AG.

In this work, we describe the development of a virtual reality AG for cell phones that incorporates spontaneous head and torso movements involved in spatial hearing. We provide a detailed description of the overall game mechanics, VAE, design of sound interaction, and how interfaces were incorporated to facilitate a flowing, motivating player experience and promote spatial hearing training. This development was subjected to, on the one hand, an examination of the user experience and usability of interfaces; and on the other hand, a pre-post test with a control group to evaluate the improvement in horizontal plane sound localization performance for sighted participants who trained with the AG. The sound localization tests were conducted under a closed-loop condition (i.e., continuous stimulus during the pointing task) [56], consistent with the active listening approach. Participant’s response was delivered with a head-pointing method, as head movement is a perceptually-driven action of particular ecological relevance [57]. Unlike previous research, our study aims to measure angular localization errors and dispersion across the entire horizontal plane allowing participants to move to locate the sound source. Through this methodology, we aim to contribute to the state of the art in AG design and further explore its use to enhance spatial hearing skills.

## 2 AUDIO GAME DESIGN

The following details the design and implementation process of the AG, entitled 'Shadows and Sounds' (S&S). Several authors highlight the essential and necessary involvement of individuals with visual impairments as integral members of the development teams [8, 17]. This inclusion in the creation process helps avoid assumptions and biases about what is expected of an AG, encouraging the development of games with a higher potential for success and relevance, thoughtfully aligned with the experiences and needs of individuals with visual impairments. In this regard, one of the recruited members of the development team was a blind university student majoring in Computer Systems Engineering.

### 2.1 Game objectives and storyline

In S&S, the player's task is to identify, locate and capture in a specific order ten farm animals (*horse, chicken, goats, cow, sheep, rooster, donkey, turkey, pig and ducks*) that are randomly distributed in a virtual square field. With no visual cues, the player must solely rely on their auditory skills and movements to accomplish this task.

When the AG begins, the player is situated in the center of a scene that describes a nocturnal soundscape. Unable to see anything, they only hear the sounds of farm and ambient animals (crickets, insects, frogs, and owls) dispersed around them. Soon, a guiding voice instructs them on the name and sound of the first animal they need to find. For example, it might say, 'Next animal: Ducks. *Quack, quack, quack.*' From there, the player must listen carefully to the sound scene to recognize where on the playing field the specified animal is located. The player can perform rotational movements with the head or use a joystick to walk freely around the playing field. The search is different each time due to the random distribution. Another complication is that the characteristic calls of the target animals are heard intermittently; at times, they make sounds, and at other times, they remain silent. The player must be patient and strategic in dealing with this situation. Also, depending on the player-target animal relative position, it may happen that crickets or some other nearby animal mask the target animal's call, complicating the search due to a poor signal-to-noise ratio perceived. In these cases, the player should move around the playing field to find more favorable listening positions. Another consideration is that as the player moves within the field, they hear the sound of their steps on the grass. This noise also hinders the signal-to-noise ratio, so the player must choose carefully when to move and when to stay still. If the player moves too far, they will hear that they are stepping on water instead of grass. This signal indicates that they are close to the boundaries of the field and should step back.

Once the target animal is located, the player should move until they are close enough to catch it pressing a specific button on the joystick. Once collected, the animal's sounds cease, and the guiding voice indicates the next animal the player must find. If the player encounters a different animal than the one indicated by the game, they cannot collect it. When the catch button is pressed in such a situation, the player will only hear the sound of their arms cutting the air, without any extra feedback. They can try to remember its position for when it's time to find it, but they should keep in mind that the animals change their positions after a certain period of time. The game concludes when the player successfully captures all ten animals.

### 2.2 Virtual Auditory Environment Design

A VAE is a digitally simulated auditory space created using binaural sound synthesis. This technique aims for a listener to obtain a three-dimensional spatial impression through the use of headphones, and to experience an immersive auditory sensation as though they were actually present in the simulated environment. The sound signals reproduced in the headphones should correspond to the signals that would be found in the listener's eardrum after being modified by the human body [58].

S&S was developed on a VAE, designed using the Unity real-time game engine (version 2019.4.26f1) and the Steam Audio plugin (version 2.0). This integration allowed the use of preconfigured filters of generic far-field HRTFs, as well as a comprehensive and customizable sound propagation simulation model. Both tools have demonstrated efficiency and flexibility in designing spatial audio for current games [59].

Unity's Audio Mixer system was used to combine all sound sources which were configured as omnidirectional. Sound propagation was modulated by the 'Physics-based attenuation' (decrease in intensity according to inverse-square law) and 'Air absorption' (air frequency-dependent, distance based attenuation) plugin options. The VAE features a free field design with no simulation of sound reflections. The ground is a square and flat surface, approximately 30 m on each side. This measurement was estimated, since the game editor software does not use standardized units of measurement. The calculation was estimated by walking with the avatar along the side of the area, measuring the travel time and considering a walking speed of 5 km/h.

The VAE recreates a natural and outdoor nocturnal soundscape with some fixed-position animals such as crickets and frogs at ground level and owls slightly elevated from the ground. Figure 1a shows a design view of the VAE implemented. The environment permits unrestricted exploration in all directions, and with each translational movement, the sound of footsteps on grass is heard. When the player approaches the lateral edges of the playing area, their footsteps sound as if they were stepping on water. Upon actually reaching one of the edges, a sound that imitates the bouncing of a ball is played, indicating that it is not possible to continue advancing in that direction.

## 2.3 Interface design

The implementation of interfaces was based on principles proposed by Froese et al. [39]. In this study, the authors propose an approach to design characterized by a natural and direct technological mediation with the environment, enriching perceptual interactions. In our design context, natural action entails users interacting with the virtual environment through their everyday hearing sensorimotor loops. Although this guideline may seem trivial, numerous AGs greatly reduce auditory interaction (e.g., discrete left-center-right panning) without articulation with the user's movements involved in listening. Direct technological mediation, on the other hand, implies that the user accesses auditory information firsthand without the need to learn any special codes to interact. The design of S&S interfaces avoided the use of audio descriptions or complex sonifications, leading the player to know the environment through their actions and obtained feedback.

Using these concepts, several possible interfaces for the design were explored. In order to work with a more socio-economically inclusive hardware, it was decided to use a low-cost cell phone mounted on virtual reality headset (VRH), headphones and a wireless joystick with analog stick and action buttons. A participant playing S&S with this hardware is shown in Figure 1b. The entire sound design was carefully crafted to ensure that all actions performed by the player have appropriate and immediate auditory feedback. The aforementioned VAE was optimized to operate efficiently on these devices by constraining the number of sound sources to accommodate the real-time simultaneous execution capabilities of this hardware type. The S&S design was tested on a cell phone with a 1.4 GHz Quad-Core processor, 3 GB of RAM, and an Android operating system (Samsung J3+).

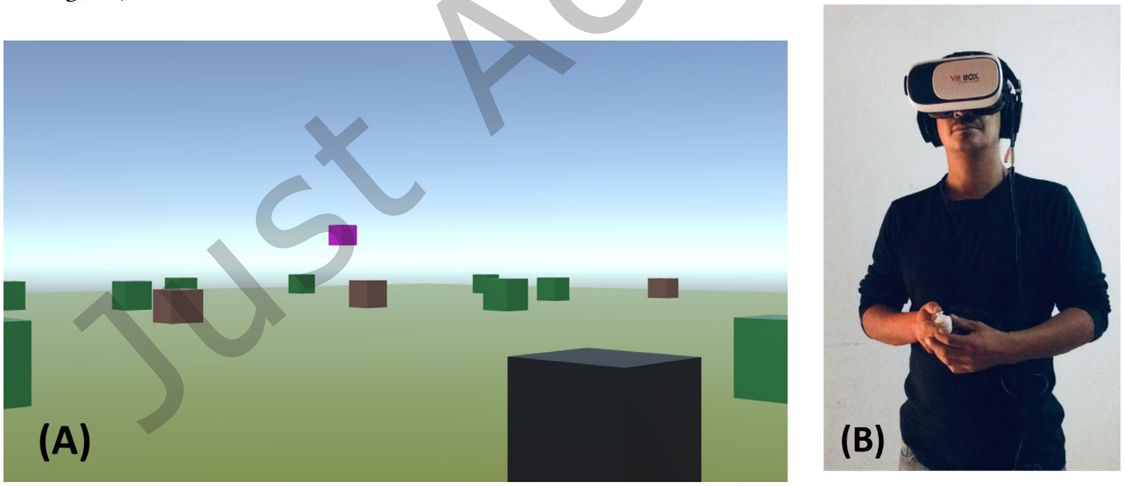

Figure 1: (A) Design view of the virtual terrain (not visible for players): crickets (green cubes), frogs (brown cubes), search animal (black cube) and owl (purple cube); (B) Participant playing S&S using the implemented interfaces (VRH, wireless joystick and headphones).

The VRH is a passive headset that allows the cell phone to be securely attached to the player's head. In this way, by utilizing the phone's gyroscope sensor, it is possible to include head and torso rotations, faithfully recreating almost the entire natural sensorimotor loop involved in spatial hearing of the horizontal plane. Additionally, the visual information on the phone was eliminated, making its screen dark throughout the game. Translational movements were controlled using the joystick stick, where the forward, backward and lateral directions consistently corresponded to the player's face orientation. Additionally, the analog stick provided variable movement speeds. Such translational control enabled S&S to be playable in a confined space, removing the requirement for a large clear area, since in virtual reality games players often lose track of their actual surroundings.

To capture an animal, the player had to press a specific button on the joystick. Each press triggered the sound of a swift, air-cutting movement (*whoosh* sound), simulating the attempt to reach out and grab the animal. This sound alternated its panning randomly between left and right with successive button presses, creating a dynamic experience that simulated different arm movements. When the player pressed the button in close proximity to the indicated animal, the whoosh sound was followed by a positive feedback (*bleep* sound), indicating a successful capture and immediately halting the animal's sound. Subsequently, the guiding voice pointed out the next animal the player needed to locate. Conversely, if the player pressed the capture button not being close enough to the right animal, only the *whoosh* feedback sound was heard. In case of confusion or forgetting, the player was able to replay the message with the name and sound of the search animal at any time by pressing another button on the joystick.

## 2.4 Game Mechanics Design

In the field of AGs, more than just computer accessibility solutions are required [9, 60]. While these recommendations are fundamental, it is essential for AGs to be enjoyable, challenging, and motivating. To achieve this goal, recommendations from different guidelines were considered in the first place [9, 10, 11, 17]. These guidelines combined several advices and suggestions based on literature reviews, prototype investigations and observations and interviews with visually impaired players. Additionally, special attention was paid to balancing the challenge of the AG with the expected user skills needed to complete it. This equilibrium, as highlighted in [9], is directly related to maintaining motivation and avoiding negative emotional states such as anxiety or disinterest. An oversimplified environment, with few interacting elements, can render the game tasks trivial. This can lead the player to feelings of boredom or apathy. Conversely, an overly cluttered soundscape elevates the complexity in identification and can overload the player's auditory channel, negatively affecting their experience [10].

The main challenge of S&S is to capture ten objects distributed in a virtual area using sound source identification and localization skills. Therefore, acoustic variables were manipulated to create a soundscape that offered a reasonable challenge, resulting in a stimulating and enjoyable gameplay experience. Achieving the right balance in sound levels and spectral profiles of simultaneously sounding sources was crucial in the soundscape composition [11]. Several pilot tests were conducted to adjust the propagation and quantity of sound sources within the scene, employing the controls of the Steam Audio plugin and Unity's Audio Mixers system.

In previous versions of the AG, five arbitrary sound elements were employed as search objects (whistle, singer, trumpeter, antique phone, and crackling bonfire). After successful informal test with these objects, it was decided to expand the number of sound sources to be searched. A selection of sounds that could provide a rich array of stimuli and contribute to the narrative cohesion of the audio game was sought. In the field of AGs, making this selection is not straightforward due to the abundance of silent objects in our surroundings. After testing several sound banks, it was decided to use animal calls. This sound category allowed for an increase in the number of search elements while maintaining variable sound characteristics that enrich the auditory challenge, including different timbres, durations, and frequency responses. Another noteworthy aspect is that, generally, some animal calls are ingrained in the common understanding of most individuals. Therefore, the audio game would not introduce search signals that could be too unfamiliar for the player.

Regarding the used audio files, they were sourced from various available sound libraries [61, 62, 63, 64], under non-commercial licenses. The selected audios had a sampling frequency of 44,100 Hz and a bit depth of 16 bits. They were processed using audio editing software (Adobe Audition) to ensure uniform sound levels. This was accomplished by adjusting the signal's amplification until the average RMS power indicated by the software was -30 dBFS +/- 1 dBFS. This level was determined to be comfortable during informal tests involving 10 randomly distributed animals emitting sounds simultaneously. Additionally, spectral profiles of sound sources were evaluated to prevent excessive concentration in any frequency band. During this adjustment, animals with greater emphasis on low frequencies and mid-high frequencies were identified. This condition is relevant since the latter imposes less difficulty for sound localization [65, 66]. It was noted that horse, pig and cow sounds have more emphasis on low frequencies, while the sounds of the other seven animals are predominantly located in mid-high frequencies. Additionally, two animals from this latter category (*sheep* and *goats*) that sound very similar were included to add an extra element of difficulty in the search.

Careful calibration was applied to the activation and deactivation intervals of the animal sounds. By alternating different voices of the same animal and periods of silence, the level of challenge proposed for each animal was adapted. Randomization of the initial position of the animals was also programmed to prevent clustering among them and to improve dispersion across the playing area. This was implemented using an algorithm that defines three zones based on the player's starting point (center of the playing area): a nearby zone, a mid-range zone, and a distant zone. The nearby zone covers a square area approximately 12 m on each side, positioned in the center of the playing field. The mid-range zone is an outer square ring surrounding the nearby zone, with a width of 3 m. The distant zone is another similar ring, located beyond the mid-range zone, also with a width of 3 m. The algorithm randomly places two animals in the nearby zone, three in the mid-range zone, and five in the distant zone, ensuring that they are separated by a minimum distance of 3 m. Figure 2 shows a particular case of this randomization. The described distribution is maintained for 90 s, following which the animals are again relocated. The new random positions vary their x and y coordinates within a margin of +/- 6 m.

Figure 2: Initial random distribution of the search animals. The red cubes represent the search animals and the green cubes represent animals from the environment. In the nearby zone (1) two animals are placed, in the mid-distance zone (2) three animals are placed and in the distant zone (3) five animals are placed. In the boundary zone of the field (4), where the player's feet step in water, no animals are placed.

It should be noted that the searched sound signal could be masked not only by the stationary animals that make up the ambient noise, but also by the other search animals present, as well as by the sound of the player's footsteps. Furthermore, the size of the playing area and the distribution of sound sources were designed so that, at times, the player could be quite far from the target animal. This distance, combined with the mentioned masking, can render the sound of the target animal imperceptible to the player. Thus, a dynamic signal-to-noise ratio is established within the game, dependent on the distances to different sound sources, as well as the player's positioning within the VAE. The latter is because discrimination and localization also rely on the spatial location of noise and signal with respect to the player [67, 68]. Ultimately, an important part of the AG mechanics depends on devising strategies to maximize this signal-to-noise ratio and thus locate the indicated animal.

### 2.5 Familiarization Scene

In line with the recommendations of García y Almeida Neris [10], a familiarization scene was constructed with a simplified sound environment that followed the principles indicated above. This has a significant impact on the appropriation of interfaces, especially for individuals with no previous experience in virtual environments. In this way, it prevents any participant from getting confused or making basic mistakes due to a lack of knowledge of this type of interface. This scene involves a brief challenge where the player must collect eight randomly distributed beacons in a playing area of the same dimensions as S&S, but with no ambient sound. The beacons are presented one by one and emit easily localizable sounds due to their sound level and high-frequency-rich spectrum.

## 3 METHODS

### 3.1 Participants

Eighteen adult participants (age: M = 31.7, SD = 7.3, 7 female) were recruited to participate in the study. All participants self-reported normal hearing, normal vision and none had previous knowledge of the experimental setup. The study was carried out in accordance with the Helsinki Declaration guidelines. All participants provided written informed consent prior to the beginning of the experiment.

### 3.2 Localization Test Setup

A psychophysical test was designed to evaluate sound localization performance using the same tools, interfaces and settings as S&S, including Unity, Steam Audio, Samsung J3+ cell phone, VRH, wireless joystick, and set of generic far-field HRTFs. For audio reproduction, Audio-Technica ATH-M50x headphones were used. The localization test incorporated 12 virtual sound sources positioned around the participant's head at 0° (frontal position), ±30°, ±60°, ±90°, ±120°, ±150° and 180° (rear position) in the horizontal plane.

Two different stimuli were used: a continuous broad-band pink noise presented in pulse trains (PN), with equal pulse and silence durations set at 500 ms; and a realistic sound loop of a clucking chicken (CC), identical to that employed in the S&S AG. Figure 3 shows temporal and frequency representation of both signals. These stimuli were selected to assess variations in sound localization performance using signals with distinct acoustic and semantic features. PN is characterized by its random frequency distribution encompassing the entire human audible spectrum (20 to 20,000 Hz) with equal energy per octave band, and is often used in sound perception experiments. Conversely, the CC stimulus has a limited frequency spectrum and a variable dynamic range. Although it has more ecological relevance, it poses a greater challenge in the localization task. The sound level of the stimuli was 60 dBA SPL, which was measured using a Brüel & Kjaer 4128-C head and torso simulator. Participants' head rotational movements were captured using the phone's gyroscope sensor and recorded in its internal memory. The phone screen remained off during the test, and participants responded using a specific button on the wireless joystick. To minimize potential ambient noise interference, the tests were conducted in acoustically treated rooms with background noise below 20 dBA SPL.

### 3.3 Experimental Design and Procedures

A pretest-posttest study was conducted, involving both an experimental group (EG) and a control group (CG). Nine participants were randomly assigned to each group (4 females in the EG and 3 in the CG). Participants in the EG underwent the pretest-training-posttest sequence, while the CG only completed the pretest-posttest procedure. The pretest and posttest were identical and consisted of two blocks of approximately 15 minutes with a 5 minute break between them. In the first block was presented the PN stimulus and in the second block the CC stimulus. Complete test sequence for both groups is shown in Figure 4.

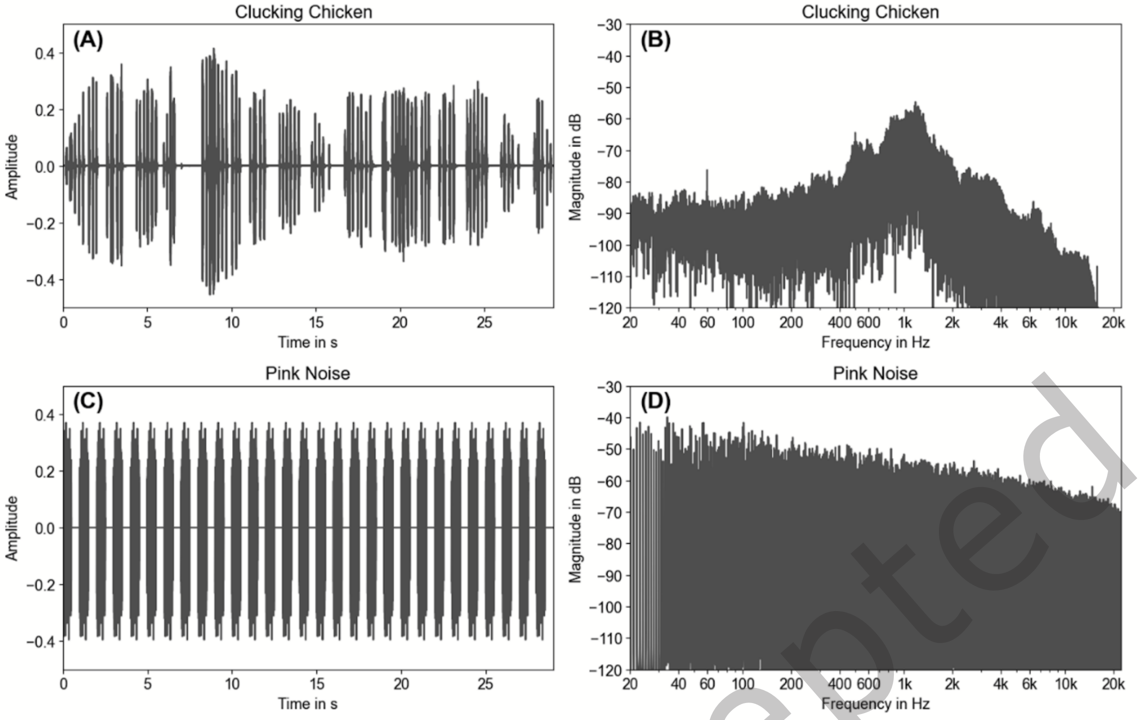

Figure 3. Test stimulus for sound localization evaluation. (A) Clucking chicken temporal representation, (B) clucking chicken power spectrum, (C) pink noise bursts temporal representation and (D) pink noise bursts power spectrum.

The participant's task was to face the sound source with their head. Once they were sure of their response, they would press a button on the joystick. Subsequently, they returned to the initial position. In order for the participant to locate this position without generating any extra training, an assistance system with non-spatialized sounds was designed. This consisted of two different sounds reproduced with the headphones. When the participant was more than 5° away from the initial position, a low-frequency (<280 Hz) pulse train (0.44 s sound, 1.70 s silence) was played, indicating that they should rotate their body to seek the initial position. If the participant entered the zone of +/- 5° from the initial position, the sound increased its pitch (<500 Hz) and pulsation (0.14 s sound, 0.30 s silence). The participant had to remain still hearing this pulse train for 2 s. When this happened, the system played a short positive feedback sound confirming the return, and after a brief silence, the next stimulus to locate was reproduced.

Sounds were presented randomly in the 12 positions, 4 times each (total of 48 trials per stimulus). Clear instructions corresponding to each phase were provided before initiation, ensuring participants had the opportunity to address any concerns. Additionally, before starting the pretest the participants solved practice trials to familiarize with the task that were not computed.

Following the pretest, EG participants proceeded to the training phase, comprising 7 sessions on different days in a maximum period of 15 days after the pretest. In each session, the participant completed 2 separate S&S games in a maximum time of 30 minutes, with a short break in between. On the first day of training, the participants also had to complete the familiarization scene. The training sessions took place in an area of 2 m x 2 m either in the participants' homes or within the aforementioned testing rooms. Throughout each trial and test, an operator ensured that the participant did not exceed the playing area, preventing any contact with real objects. The headphone sound level was adjusted to the participant's comfort.

Finally, after completing the training for the EG, or after a period of 10 to 15 days for the CG, all participants underwent the posttest under identical conditions to the pretest. At the end of the study, EG participants were interviewed and audio-recorded.

Figure 4: Test scheme for the assessment of S&S. Two groups (EG and CG) fulfilled the pretest and posttest. The experimental group underwent a 7-session training. Finally, this group participated in a semi-structured interview to gather insights into their game experience.

## 3.4 Interview

A semi-structured interview was designed to explore aspects of the participants' gaming experience. Given the absence of any validated instrument fully aligned with this research an ad hoc interview was compiled. In order to elaborate it, we take into account precedents from interviews and questionnaires used in the context of research on immersiveness in virtual environments [69, 70], sensorimotor training [71], game interfaces without visual cues [72], and engagement within computer games [73, 47]. Conversation scenarios were guided using trigger questions related to the following issues:

- Previous experiences with video games, virtual reality systems and audio games
- Emotional experience during gameplay
- Changes experienced during training with the game
- Overall immersion experience
- Experience in interacting with the virtual space

## 3.5 Data Analysis

A qualitative analysis was performed with the information collected through interviews. Six dimensions were identified to categorize participants' statements: 1) evoked emotions during gameplay, 2) sense of immersion, 3) perceived challenges within the game, 4) sensorimotor interaction, 5) interfaces usability, and 6) self-perception of learning. For each dimension, textual extracts from the interviews were compiled and their content was further analyzed.

The psychophysical test results were evaluated in terms of *accuracy* and *precision* in angular localization of presented stimuli, comparing pre- and post-tests for the EG and CG. *Accuracy* was measured by calculating the mean of unsigned errors in azimuth, while *precision* was determined by averaging the standard deviations of these signed errors. Prior to the statistical analysis, assumptions for conducting two-way mixed ANOVAs were verified, ensuring normality and homogeneity of variances using the Shapiro-Wilk and Levene tests, respectively. Subsequently, statistical significance was assessed through two-way mixed ANOVAs (one for each stimulus), with Group (EG and CG) as the between-subjects factor, and Phase (pretest and posttest) as the within-subjects factor. In the case of main effects, post-hoc analyses were conducted, including paired *t*-tests for repeated measures and unpaired *t*-tests for group comparisons. The significance level was set at $\alpha = 0.05$.

# 4 RESULTS

## 4.1 Gaming Experience

The interviews with EG participants (EPs) lasted 21.8 min on average (SD = 8.0 min). Table 1 presents verbatim excerpts of illustrative statements. In relation to prior experiences, the participants reported limited or no prior experience with virtual reality systems, and none of them had previous contact with AGs. In relation with their video games experience, responses varied: EP1, EP7, and EP8 had no prior experience, while EP3 and EP9 regularly engaged in shooter-type games, and the remaining participants had only limited or distant encounters with video games.

Table 1: Extracts from interviews with EG participants, grouped into six analyzed dimensions. Representative questions are added to illustrate the content of each dimension.

| |
|---|
| Dimension 1: Evoked emotions during gameplay<br>What emotions did you experience while playing? |
| EP4: "At first, the VR environment confused me, and I felt a certain degree of nausea or general confusion. Maybe even dizziness. But then, with more knowledge and improved skills, I started seeing it more as a game, something entertaining and fun. It also became satisfying when I started doing well, reducing the game times, and adding a competitive edge."<br>EP6: "The first day was tough. There was so much noise all around, like a cloud that didn't let me see or distinguish the sound. When I managed to focus on a specific sound and concentrate on it, everything fell into place, and the confusion went away. (...) After I got the hang of it, it even brought a sense of calm, helping me to concentrate. Disconnecting from all the other senses and solely focusing on hearing brought me a sense of relaxation."<br>EP7: "I felt frustration, especially at the beginning, and it caused moments of anxiety, but it was linked to the frustration. There was some desperation when I couldn't hear certain creatures (...) It becomes more enjoyable as you play; you start developing a connection with the game." "I didn't feel that it was inherently a fun game, but it does create a desire to improve." |
| Dimension 2: Sense of immersion<br>*To what extent did you feel that while playing, you were in a different environment from the real space?* |
| EP1: "A lot, I was in a field, but without visual representation. That seemed strange to me. It was like being in another place but only auditory."<br>EP2: "I was completely in another place. I even felt nervous stepping on the water." |
| Dimension 3: Perceived challenges within the game<br>*What did you find the most difficult to do?* |
| EP5: "I had confusions related to the similarity of some animals. I used the recall button when I got confused between ducks and chicken." "Also, the animals with lower-pitched sounds were heard less. I don't know if it's like that in general or if I heard them less, and the higher-pitched sounds were easier to hear. The turkey and the goats are heard easily, while others are heard less, like the horse."<br>EP9: "The most challenging part was when the game started, and you needed to find a not very invasive sound like the horse or the pig, and at the same time, they were far away." "It was difficult when the goats and the sheep were close to each other." |
| Dimension 4: Sensorimotor interaction<br>*How did you find the interaction with the virtual space? Did you feel coherence between your real movements and those within the virtual space?* |
| EP1: "Yes, they were coherent. I had trouble realizing when I reached the animal. It was harder to catch them than to locate them. The rule I ended up figuring out was that it would work when I felt I was right above the animal."<br>EP4: "The sound system works well, and macro-localization is straightforward. Sometimes it can be confusing when you're very close to the animal and trying to catch it (...) Probably one of the more complicated aspects, in general, is to always understand the distance to the animal, that is, the 3D dimension in which it is possible to catch the animal. After learning this internal game rule, I improved in all aspects." |
| Dimension 5: Interfaces usability<br>*How did you feel about the devices you were using?* |
| EP1: "The headset is uncomfortable; it bothered my nose, but once I started playing, I forgot about it."<br>EP3: "You feel the headset at the beginning when you put it on, but then you forget about it." |
| Dimension 6: Self-perception of learning<br>*Did you feel any changes in your gaming experience from the first session to the last?* |
| EP7: "I feel like I exercised my attention a lot. I believe my hearing improved, but it's more about attention, being able to clean up and organize the scene."<br>EP8: "Towards the end there was a point where, even with a lot of noise, I could still hear them."<br>EP9: "Yes, I felt that my hearing was more precise. Especially in terms of how close or far away I was from the animal. I knew when I would be as close as possible to the animal... As I walked towards the cow, I was making a mental map of the other animals I was hearing, so that I could react quickly when I finished grabbing the cow." |

*Dimension 1. Evoked emotions during gameplay:* most participants initially felt confused and overwhelmed by the virtual experience (EP1, EP4, EP5, EP6, EP7, EP8, EP9). They expressed feelings of anxiety, disorientation, and frustration at the beginning of their training sessions. As the sessions progressed, all participants, except for EP7, recovered from these feelings and reported experiencing fun and entertainment, associated with a sense of mastery and structuring of the virtual space. Some participants mentioned that the competitive aspect, driven by the desire to improve their completion times, heightened their engagement and induced an adrenaline rush (EP4, EP5, EP9). One participant expressed a desire for more levels to continue playing (EP4). Positive emotions also included a participant who found the game relaxing as it allowed them to focus solely on auditory input (EP6). EP7 reported that negative feelings persisted throughout the 7 training sessions, but they still maintained a desire to improve in the game and found enjoyment and connection with it. On the other hand,

EP2 declared feeling engaged with the game at all times without experiencing overwhelm, while EP3 did not report any emerging emotions.

*Dimension 2. Sense of immersion:* all participants indicated to some extent that they felt present in a space different from the physical environment where the training took place. Some expressed this with great certainty (EP1, EP2, EP3, EP5, EP6), while others described a kind of coexistence between the awareness of being in the physical place where the training took place and at the same time in the virtual field (EP4, EP7, EP8, EP9). Three participants agreed that the feeling of immersion increased as they gained more mastery within the game (EP4, EP6, EP8).

*Dimension 3. Perceived challenges within the game:* all participants acknowledged that certain animals were more difficult to find than others, especially at the beginning when all the animals were present in the scene. *Pig*, *horse* and *chicken* were frequently mentioned as "difficult" elements to identify. Some participants directly attributed the difficulty of *pig* and *horse* to their low spectral components (EP1) or because their sounds were "less invasive" than others (EP3, EP5, EP9). The *chicken* was considered difficult to identify due to confusion with another animal (EP5, EP8), its low sound level (EP2), or the perceived sense of instability (EP1). In contrast, the *turkey* was described as easy to locate due to its sharpness and high-frequency characteristics (EP3, EP5). Participants' perceptions of *goats* varied, with some finding them relatively easy to identify (EP3, EP5) and others considering them more challenging. Some participants described *goats* as having an unstable sound (EP2, EP7), while others highlighted the difficulty in differentiating their sound from *sheep*, thus making it challenging to discern between both (EP4, EP6, EP9). In challenging situations, some participants found the button that replayed the sound of the animal they were seeking to be very useful (EP5, EP6, EP9). Additionally, participants mentioned that the footstep sounds within the game contributed to the overall difficulty in identifying and locating animals. At times, participants needed to remain still to minimize ambient noise interference and improve their localization accuracy (EP4, EP9).

*Dimension 4. Sensorimotor interaction:* All participants reported feeling coherence between their rotational movements and the auditory feedback provided by the game. By employing head rotations and joystick stick navigation, participants stated they had no difficulties approaching animals when they perceived them far away. However, most participants experienced some confusion in locating the animals when they were very close. In general, the statements pointed to unclear interaction and difficulties in recognizing the distance at which the animal could be caught. Ultimately, all participants managed to overcome this issue, finding specific strategies to interact with the system effectively. Another issue mentioned by EP4 andEP5 was that they could not fully grasp the size of the playing area they were interacting with.

*Dimension 5. Interfaces usability:* regarding the comfort of the VRH, diverse opinions emerged. Some expressed that they were only aware of wearing it during the first few minutes after putting it on, and then they forgot about it (EP1, EP3, EP6, EP9). On the other hand, others declared that they felt uncomfortable most of the time they were wearing it (EP5, EP7, EP8). Notably, the discomfort was more pronounced among three female participants, who specifically cited issues related to the headset's anatomical fit, its weight and discomfort on the nasal region, as well as difficulties in achieving proper adjustment with the securing bands. Meanwhile, the majority of participants found the joystick easy and intuitive to use. Everyone successfully used this interface to navigate and interact within the virtual space.

*Dimension 6. Self-perception of learning:* participants generally perceived improvements within the game. Some mentioned that their "gameplay" (EP3) improved, as well as their "orientation" (EP6) or "attention" (EP7). Repeated expressions were found regarding how the identification of animals became simpler and faster due to familiarization with the sounds of the animals (EP1, EP2, EP4). Most participants acknowledged developing successful strategies throughout the training sessions to catch animals when they were close. Two participants expressed surprise at finding that their performances fluctuated on some occasions, without experiencing a clear learning curve (EP1, EP5). On the other hand, some mentioned that their level of improvement allowed them to create "auditory maps" that helped them orient themselves and play faster (EP1, EP9).

## 4.2 Sound Localization Performance

### *4.2.1 Data Preprocessing.*

Preprocessing of the sound localization test data was carried out to address atypical angular errors. First, front-back confusions (FBC) were identified. FBC is a perceptual error where participants struggle to distinguish whether a sound originates from a frontal or rear location. This error often arises in virtual sound systems when using generic HRTFs due to the alteration of the monaural cues, which are related to the specific anthropometric characteristics of each participant. The cause of these FBCs differs from normal localization errors [74], thus, responses registered in the hemisphere opposite to that of the presented stimulus (frontal or rear) were excluded from the analysis. The trials removed due to FBC were 44 (2.5% of the total).

Subsequently, the dataset was inspected for potential outliers, which could arise due to errors during the test. Some of these outliers were acknowledged by participants themselves, indicating erroneous responses due to accidental joystick button presses. The IQR/Tukey method with a threshold of 1.5 IQR as exclusion criteria for the unsigned localization errors was employed. A total of 81 trials (4.7%) were removed.

### *4.2.2 PN Stimulus Performance.*

The results of PN stimulus, as shown in Figure 5, demonstrate that both groups achieve an enhancement in response accuracy (reducing the localization error), although the Experimental Group (EG) accomplishes this to a greater extent. The average localization error of EG decreases by 1.5° (-27.3%), whereas CG reduces by 0.4° (-5.1%). For EG, this enhancement in accuracy is coupled with an improvement in precision. This group manages to concentrate their responses more effectively, leading to a reduction of 1.8° (-28.6%) in the average standard deviation. Conversely, CG experiences an increase in response dispersion by 0.5° (+7.6%) (see Table 2 for details).

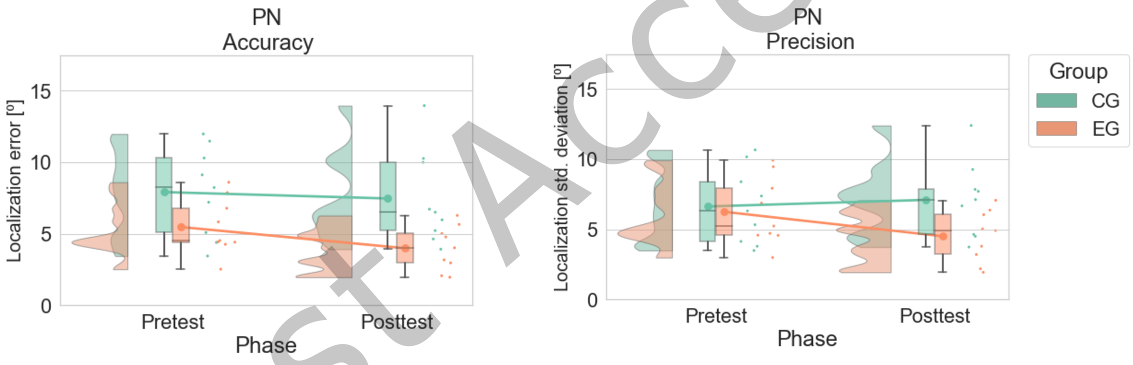

Figure 5: Distributions for accuracy measurements (localization error) and precision (localization standard deviation) for both participant groups during the pretest and posttest phases, as evaluated with the PN stimulus. Distribution curve, box plot, and scatter plot are displayed in each case. Mean calculated values are annotated for each graph, accompanied by their respective trend lines.

The outcome of the conducted ANOVA analysis for accuracy measurements reveals a main effect for Group (F(1,16) = 6.289, p = 0.019, η2: 0.260) and Phase (F(1,16) = 4.725, p = 0.045, η2: 0.028). Post hoc analysis for Group show no differences in the pretest, although they do appear in the posttest (t(16)=2,879, p =0.011), with EG showing significantly lower errors compared to CG. Furthermore, the post hoc analysis in Phase illustrates that only EG's performance exhibits significant improvements in posttest versus pretest (t(8) = 4.162, p = 0.003). CG fails to achieve statistical significance in its improvement. In summary, these analyses indicate that EG significantly enhances the accuracy of their responses, differentiating itself from CG in the posttest.

Regarding precision, the ANOVA displays a main effect in the interaction Group x Phase (F(1,16) = 10.845, p = 0.005, η2: 0.049). Post hoc analysis for Group reveals that precision in EG's responses is superior only in the posttest in comparison with CG (t(16) =2.387, p =0.030), while there are no differences between groups in the pretest. Additionally, the analysis in Phase demonstrates that posttest precision is significantly superior to the pretest only for EG (t(8) = 4.066, p = 0.004).

Taken together, the ANOVA statistical analyses demonstrate that training with the S&S AG improved the ability to localize the PN stimulus in both accuracy and precision. EG participants manage to reduce localization errors while concurrently concentrating their responses, whereas CG does not reveal any of these effects.

Table 2: Average values of sound localization error, standard deviation, and differences between pretest and posttest for the experimental group (EG) and control group (CG), across both tested sound stimuli: PN and CC.

| | | Pink Noise Burst (PN) | | | Clucking Chicken (CC) | | |
|---|---|---|---|---|---|---|---|
| | Group | Pretest | Posttest | (Δ) | Pretest | Posttest | (Δ) |
| Accuracy [°] (Mean localization error) | EG | 5.5 | 4.0 | -1.5 | 6.6 | 5.5 | -1.1 |
| | CG | 7.9 | 7.5 | -0.4 | 7.3 | 7.0 | -0.3 |
| Precision [°] (Mean localization standard deviation) | EG | 6.3 | 4.5 | -1.8 | 6.9 | 5.6 | -1.3 |
| | CG | 6.6 | 7.1 | 0.5 | 7.3 | 7.5 | 0.2 |

### *4.2.3 CC Stimulus Performance.*

For this stimulus, the tests also indicate that both groups succeed in diminishing localization error and the standard deviation of their responses. As shown in Figure 6, EG decreases the average localization error by 1.1° (-16.7%), whereas CG achieves a decrease of 0.3° (+4.1%). In terms of accuracy, EG demonstrated a reduction of 1.3° (-18.8%) in average standard deviations during the posttest compared to the pretest. In contrast, CG did not exhibit accuracy improvement; conversely, these values displayed a marginal increase in the dispersion of responses by 0.2° (+2.7%) (Table 2). Despite these trends where EG manifests better indicators than CG, the results of ANOVA statistical analyses do not present significant effects for these values.

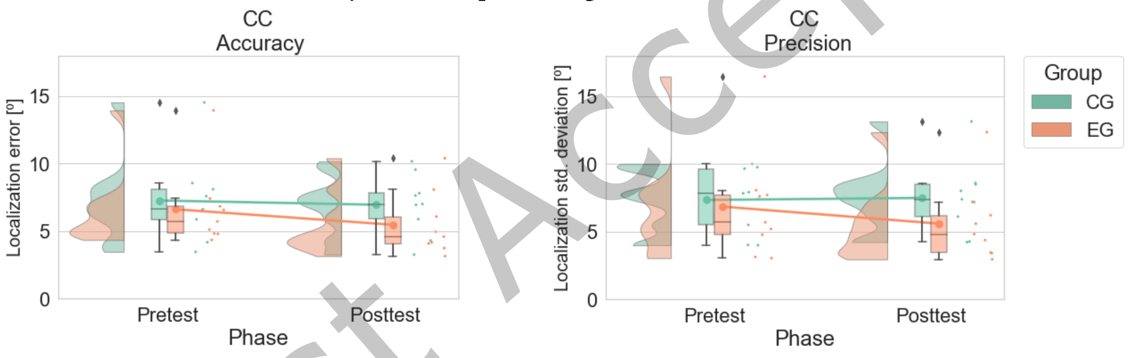

Figure 6: Distributions for accuracy measurements (localization error) and precision (localization standard deviation) for both participant groups during the pretest and posttest phases, as evaluated with the CC stimulus. Distribution curve, box plot, and scatter plot are displayed in each case. Mean calculated values are annotated for each graph, accompanied by their respective trend lines.

## 5 DISCUSSION

This study presents the design and assessment of a virtual reality AG featuring free-movement interfaces specifically developed for cell phones. The evaluation was carried out from two different approaches. Primarily, with the aim of contributing to the field of AG design, various aspects of the game experience were evaluated through interviews with nine players. Furthermore, the potential of the AG to train sound localization in the horizontal plane was studied through a test involving experimental and control groups.

The evaluation of the design revealed a positive reception from the participants, who generally found it enjoyable and challenging. The interviewees' responses suggest a ripe opportunity for developing additional levels and introducing new challenges within the existing game framework.

In the design of S&S, the decision to exclude explicit scoring mechanisms was deliberate and aligned with our intended user experience. At this stage of the project, we aimed to test the immersive and challenging nature of the gameplay designed, encouraging players to concentrate on their sensorimotor interactions without the distraction or pressure of any scoring metrics. Since interview results suggest that the game mechanics are entertaining and engaging, elements such as scoring systems, reward mechanisms, and other game feedback elements should be considered in future stages of development. These tools will be useful to enhance players' motivation and commitment across multiple playthroughs [1], keeping them engaged while also training their spatial hearing.

As for the skills needed to perform the tasks, the participants did not report difficulties in locating the position of the sound sources (animals). On the contrary, they did experience difficulties in detecting the salience of a particular animal's voice within the soundscape, disambiguating similar voices, and recognizing voices with low tonal components. Despite these challenges, most participants managed to develop strategies to complete the training games after a few minutes. Only one participant encountered complications and expressed feeling overwhelmed by the task's demands. During the interview, this participant disclosed that they had no prior experience with computer games or virtual reality interfaces. For this reason, we think that a more gradual difficulty curve is crucial especially if the user context is highly heterogeneous. This issue could be addressed from two distinct perspectives. One approach could involve designing various difficulty levels, which would be selectable based on a player's profile and learning trajectory [75]. Another option might entail employing a dynamic difficulty adjustment that continuously assesses a participant's gameplay progression through task success, timing, error counts, and other indicators. In this manner, difficulty variables could be fine-tuned to maintain an optimal balance between the presented challenge and the requisite skills for successful completion [76].

Regarding immersion, the design effectively engages participants in virtual reality without relying on visual elements. The combination of binaural sounds and head rotation tracking seems sufficient for players to perceive inhabiting a virtual space, distinct from the real physical environment. Furthermore, it was observed how this alternative reality became more plausible as they gained control over their interactions with the environment. The plausibility of a virtual reality refers to the agreement of a simulation with the user's expectation towards a corresponding real event [77, 78]. This is consistent with the embodied postulates, which indicates that the verisimilitude of a virtual environment depends on providing the user with adequate feedback to their actions rather than computing and recreating the physical reality to the maximum detail [44, 79, 80].

Concerning user interfaces, it was observed that they are appropriate and user-friendly. In most cases, participants were able to interact directly with the virtual environment without focusing their attention on the devices used. On the other hand, participants quickly learned the functioning of the joystick and the motion tracking and sound reproduction system (cell phone, VHR and headphones) demonstrated precision in recreating the natural sensorimotor loops involved in spatial hearing. This facilitated a smooth and intuitive experience in most interactions. Some problems were identified when the player and the target animal were very close to each other, leading to confusion and lack of naturalness in the interaction. We hypothesize this could be improved by adjusting acoustic parameters, such as the attenuation curve, the apparent size of the source, and distance-dependent frequency filters. The use of near-field HRTFs could also help with this issue, since they contain more adequate perceptual distance cues [81, 82, 83].

As previously mentioned, since S&S navigation is achieved with the joystick stick, it does not require large clear spaces to be played. This feature offers a significant advantage, especially for auditory training in confined spaces. However, this way of interacting may hinder the global perception of the virtual space, as some participants indicated they were unsure about the size of the playing area. The impossibility of experiencing the environment through real steps might affect how participants get distance information about the virtual space, as suggested by Grechkin [84]. The development of this type of orientation skills is outside the scope of S&S; to pursue this goal, the incorporation of translational body movement or some equivalent strategy should be designed.

A final concern arising from the utilization of interfaces was the need to improve the adaptability and ergonomic properties of headset holders. Ensuring accessibility of these implements will allow to effectively accommodate a wider spectrum of individuals.

As for the inclusion of the blind developer in the team, we believe that his participation was of great importance, as his contributions were valuable and enhanced the overall dynamic of the group. Despite having no prior experience with game programming or the development engine, their standpoint, coding skills and expertise in vision-free human-machine interaction enriched the AG design, providing ideas and solutions drawn from his personal sensory experience. However, it must be pointed out that serious accessibility issues were encountered while working with the Unity engine. A considerable portion of menus and interfaces remained inaccessible to screen readers, for example, many necessary programming tasks depend on "drag and drop" type actions, which are impossible to execute without vision. As stated in [14], there is a clear need for high-level design tools with full accessibility for individuals with visual impairments.

The sound localization tests yielded promising results. After observing pretest performances with no significant differences between groups, posttest performances with pink noise demonstrate that only the group that received training with the AG significantly improved accuracy and precision. For the clucking chicken stimulus, although the performances did not reach statistical significance, the same trend was observed. On average, the experimental group was more accurate and precise than the control group in the posttest. Probably, the lack of statistical significance is due to the more complex acoustic characteristics of this stimulus. As shown previously in Figure 3, the CC stimulus features a spectrum with reduced high-frequency content, a larger dynamic range, shorter bursts of sound information, and longer silent periods. Presumably, within our experimental context, these signal characteristics took precedence over the hypothesis that a more ecologically and familiar sound (given its appearance during the training phase) would provide certain advantages, thus potentially compensating for the signal differences during testing. It is possible that tests with this stimulus may need a special sampling, with more participants.

These results support the effectiveness of S&S as a valuable tool for training active sound localization skills for horizontal plane. Our study is consistent with previous reports that have shown auditory improvements using AG training, as conducted by [54] and [55]. It is important to highlight that this work differs from previous studies in certain methodological aspects. While Ohuchi et al. [54] focused on measuring a 90° arc with a finger pointing method, and Honda et al. [55] evaluated the entire azimuthal plane at different elevations using forced-choice responses, our study measured angular errors with a head pointing method across the full 360° of the azimuthal plane using virtual sound sources. These differences provide a novel and enriching perspective in this AG research.

Regarding error measurements, although it is difficult to establish direct comparisons with other studies due to the particularities of our approach, our tests yielded similar values to other studies. For example, Gulli et al. [85] measured sound localization of real sources using a similar stimulus and allowing participants to move their heads. They employed a manual pointing method without training or visual cues and measured a mean error of 4.1°. This performance is slightly more accurate than that achieved by our participants in the pretest (mean error for both groups: 6.7°). The small difference could be related to the type of response utilized and the use of real sound sources. However, Gröhn et al. [86], despite using a methodology more similar to [85], recorded a mean error almost equivalent to that of our participants: 6.8° in the horizontal plane. In this case, participants could move their heads freely and had to locate with a stick 3 different types of continuous stimuli (broadband noise, music and frog sounds) emitted by loudspeakers. Moreover, Bronkhorst et al. [56] also used a closed-loop head pointing task with broadband stimuli in the horizontal plane. They applied reinforcement on the responses and found mean precision errors (responses dispersion) of 5.0° for both real and virtual sources with individualized HRTF. Our results show similar values in the trained group in the post-test, of 4.5°. Note that, despite having used generic HRTFs, in our study the training was active and more extensive. This comparison with antecedents coupled with our control group design, gives us confidence that the evaluation performed was not limited by ceiling effects of performance or technical artifacts.

Regarding the type of response, we looked for an explicit participant-response involved in active listening. Head movement is a particularly ecologically relevant perceptually-driven action for experimentation on the perception of sound movement, distance, and location [57]. Head orienting is a spontaneous movement that people perform to face the sound sources [87, 88, 89]. Based on this, we believe that our results are related to everyday life orientation behaviors and are potentially more helpful for understanding the applicability of the audio game. This is also in agreement with models of sound localization involving motor components [90, 91].

In a practical sense, advances in the knowledge about the role of head movements in spatial listening are highlighting their importance in particular conditions. For example, aligned with the spirit of this study, it has been shown that they contribute to orientation in blind people [92, 93], as well as compensate for localization abilities in cases of suppression of one auditory channel [94], of pinna modification [95], of hearing impairment [96, 97], and also in cases of bilateral cochlear implants [98, 99].

A final concern about the methodology is that part of the improvement in participants' performance may have been due to procedural learning, resulting from the repetitive task required during the sound localization tests. For example, if we consider that participants were asked to locate continuous sound sources in the horizontal plane by facing them with their heads, they might have learned to complete the task by simply adjusting their perceived interaural differences to zero. If this were the case, a test with stereo sounds would have been sufficient. However, since CG did not show significant improvements in any of the indices with either stimulus, we are led to assume that the main cause of the improvement is the AJ training.

Taking together the results of sound localization performances and interviews, it is interesting to note that while participants improved their sound localization skills, they only reported improvements in their gameplay technique. This suggests that players may not be conscious of the real benefits they obtain in terms of their hearing due to an implicit learning process experienced by repeatedly solving the game [100]. Other studies have already shown the emergence of sound localization implicit learning in certain activities, such as conducting an orchestra [101] or solving daily tasks without vision [102, 103]. Therefore, it is important to emphasize the significance of continuing to research and educate players about the positive impacts that auditory training can have, both in the gaming context and in their daily lives. Furthermore, it is necessary to continue exploring and refining design strategies and feedback elements to maximize the therapeutic and entertainment benefits that such AGs can offer.

Finally, the main challenge to continue the validation of this AG is to extend the sample to blind people, who would be the most benefited users. Blind people, in addition to potentially training spatial hearing skills, will certainly be able to compare with greater expertise the engaging quality of the AG. Additionally, it would be desirable to test the effectiveness of training with the AG in more realistic scenarios that involve other skills related to O&M. It would also be important to study other dynamic variables involved in active listening, such as improvements in localization time or increased efficiency in movements made to localize. Additionally, considering that previous AGs have demonstrated efficacy in training adaptation to non-individualized HRTFs within virtual systems, it would be convenient to perform static tests under open loop conditions (with short stimuli) to evaluate the potential of S&S within this field of study [104, 105, 106]. A final idea for future work is that, considering that the sound localization testing system ran correctly and smoothly on the same hardware as the AG, we believe there is an opportunity to integrate games and testing into a single training application that would allow autonomous monitoring of the player's performance.

## 6 CONCLUSIONS

The present study introduces an innovative design in the realm of audio gaming, addressing an accessible virtual reality game that combines entertainment with auditory training through natural motion interfaces. AGs, by lacking visual information, possess the advantage of being playable by both individuals with vision and those with visual impairments. Therefore, this development focuses on an inclusive space and can contribute to a critical topic, namely the auditory training of individuals with visual impairment.

The AG design closely replicates the natural sensorimotor process individuals employ in their everyday listening, allowing players to employ rotational movements of their head within the virtual space to identify and locate sound sources. This represents a significant advance, as very few AGs extend beyond a passive auditory experience, typically confined to a seated position, using a keyboard and mouse, and receiving non-spatialized auditory signals. Moreover, this direct interaction provides a rapid and intuitive mode of engagement that benefits skill acquisition. It also enriches the sense of immersion within the virtual space, as players engage in sensorimotor tasks to know the environment through their actions, without the necessity for logical reasoning to decode sonifications methods or process verbal descriptions.

Another contribution of this AG is the design of an accessible, attractive and high quality complex soundscape. It allows free navigation, with low-cost hardware and using simple and cost-effective interfaces.

The results yielded by the conducted tests are of high value, supporting the effective design and verifying its capacity to train sound localization skills. Concurrently, these results pinpoint areas that require attention to advance the design and overall understanding of such games. While the extension of these outcomes to a broad user base is feasible, it is pertinent for future research to undertake examinations with visually impaired players to substantiate and delve further into the applicability and advantages of this inclusive approach. In this sense, to fulfill this purpose and avoid the generation of false expectations, it is important to strengthen this development and research program.

Future work is essential to refine auditory interaction as players explore proximate elements, aiming for a more natural experience. Moreover, varying difficulties for different players are necessary to enhance both entertainment and auditory training prospects. Additionally, an alternative phone holding mechanism should be investigated. While the VR headset provides adjustments for varying head sizes, it proves insufficient in certain cases. Also, given the optical components are not required, a much more optimized design could be designed. Lastly, the effectiveness and simplicity witnessed in interface usage have paved a new path for future development: integrating gameplay scenes and sound localization tests within the same application, enabling self-monitoring of skill progression.

In summary, this study lays groundwork for future research and developments in the realm of accessible virtual reality AGs, opening new opportunities for auditory training. These devices can aid in the training and acquisition of spatial hearing skills of people with impaired vision. Simultaneously, they can also offer sighted individuals an alternative to traditional video games. Our work is in line with [8, 17], where this aspect is taken as a fundamental character for the progress of this type of game. Taking into consideration the positive feedback from sighted participants, we believe that adopting a universal design approach is possible and has the potential to enhance their projection and popularity.

## ACKNOWLEDGMENTS

The authors would like to thank Emir Chesini and Juan Cruz Villanueva for their work in software development, Claudia Arias for her contributions to the development of the theoretical framework, and all the people who participated in the evaluation of S&S. This work was supported by a doctoral grant from the National Scientific and Technical Research Council – Argentina, and partially supported by Universidad Tecnológica Nacional, Argentina (PID UTN CCUTNCO0007667).

## REFERENCES

[1] Fuchs, M., Fizek, S., Ruffino, P., & Schrape, N. 2014. Rethinking gamification. Meson press.

[2] Graells, P. M. 2005. Nuevos entornos, nuevos modelos didácticos. Cuadernos de pedagogía, 363, 80-89.

[3] Pérez-Castilla Alvarez, L. 2012. Buenas prácticas de accesibilidad en videojuegos. Ed. IMSERSO. http://riberdis.cedid.es/handle/11181/3300

[4] Yuan, B., Folmer, E., & Harris, F. C. 2011. Game accessibility: a survey. Universal Access in the information Society, 10, 81-100. https://doi.org/10.1007/s10209-010-0189-5

[5] Aguado-Delgado, J., Gutierrez-Martinez, J. M., Hilera, J. R., de-Marcos, L., & Otón, S. 2020. Accessibility in video games: a systematic review. Universal Access in the Information Society, 19, 169-193. https://doi.org/10.1007/s10209-018-0628-2

[6] Friberg, J., & Gärdenfors, D. 2004. Audio games: new perspectives on game audio. In Proceedings of the 2004 ACM SIGCHI International Conference on Advances in computer entertainment technology (pp. 148-154).https://doi.org/10.1145/1067343.1067361

[7] Beksa, J. 2020. "Nobody Makes Games for Us"-An Investigation into the Independent Design of Audio Games Through the Development of the Audio Game Hub and Blind Cricket. PhD Thesis, Auckland University of Technology.

[8] Urbanek, M., & Güldenpfennig, F. 2019. Unpacking the audio game experience: Lessons learned from game veterans. In Proceedings of the Annual Symposium on Computer-Human Interaction in Play (pp. 253-264). https://doi.org/10.1145/3311350.3347182

[9] Archambault, D., & Olivier, D. 2005. How to make games for visually impaired children. In Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology (pp. 450-453).https://doi.org/10.1145/1178477.1178578

[10] Garcia, F.E. & de Almeida Neris, V.P. 2013. Design Guidelines for Audio Games. In Human-Computer Interaction. Applications and Services. HCI 2013. Lecture Notes in Computer Science, vol. 8005. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-39262-7_26

[11] Urbanek, M., Fikar, P., & Güldenpfennig, F. 2018. About the sound of bananas—Anti rules for audio game design. In 2018 IEEE 6th international conference on serious games and applications for health (SeGAH) (pp. 1-7). IEEE. https://doi.org/10.1109/SeGAH.2018.8401361

[12] Mairena, J. 2009. Videojuegos Accesibles: por qué y cómo hacerlos. Retrieved September 3, 2023 from https://biblioteca.fundaciononce.es/publicaciones/otras-editoriales/videojuegos-accesibles-por-que-y-como-hacerlos

[13] Gilberto, L.G., Arellano, P., Miloro, E., Bermejo, F. & Hüg, M. 2022. Disponibilidad y características de juegos electrónicos accesibles a personas ciegas. In Actas de Resúmenes de la XVIII Reunión Nacional y VII Encuentro Internacional de la Asociación Argentina de Ciencias del Comportamiento RACC, 2022, Suplemento (Abril) (pp. 146-147).

[14] Urbanek, M. 2020. Understanding audio game experiences: perspectives and guides for design. Phd Thesis, TU Wien. https://doi.org/10.34726/hss.2020.77261

[15] Andrade, R., Rogerson, M. J., Waycott, J., Baker, S., & Vetere, F. 2019. Playing blind: Revealing the world of gamers with visual impairment. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1-14). https://doi.org/10.1145/3290605.3300346

[16] Mangiron, C., O'Hagan, M., & Orero, P. 2014. Fun for all: translation and accessibility practices in video games. Bern: Peter Lang.

[17] Nesteriuk, S. 2018. Audiogames: Accessibility and inclusion in digital entertainment. In Digital Human Modeling. Applications in Health, Safety, Ergonomics, and Risk Management: 9th International Conference, DHM 2018, Held as Part of HCI International 2018, Proceedings 9 (pp. 338-352). Springer, Cham. https://doi.org/10.1007/978-3-319-91397-1_28

[18] Beksa, J., Fizek, S., & Carter, P. 2015. Audio games: Investigation of the potential through prototype development. A Multimodal End-2-End Approach to Accessible Computing, 211-224. https://doi.org/10.1007/978-1-4471-6708-2_11

[19] Urbanek, M., Güldenpfennig, F., & Schrempf, M. T. 2018. Building a community of audio game designers-towards an online audio game editor. In Proceedings of the 2018 ACM Conference Companion Publication on Designing Interactive Systems (pp. 171-175). https://doi.org/10.1145/3197391.3205431

[20] Wilhelmsson, U., Engstrom, H., Brusk, J., & Ostblad, P. A. 2015. Accessible game culture using inclusive game design-participating in a visual culture that you cannot see. In 2015 7th International Conference on Games and Virtual Worlds for Serious Applications (VS-Games) (pp. 1-8). IEEE. https://doi.org/10.1109/VS-GAMES.2015.7295764

[21] Archambault, D., Ossmann, R., Gaudy, T., & Miesenberger, K. 2007. Computer games and visually impaired people. Upgrade, 8(2), 43-53.

[22] Rovithis, E., Floros, A., Mniestris, A., & Grigoriou, N. 2014. Audio games as educational tools: Design principles and examples. In 2014 IEEE games media entertainment (pp. 1-8). IEEE. https://doi.org/10.1109/GEM.2014.7048083

[23] Correa, A. G., De Biase, L. C., Lotto, E. P., & Lopes, R. D. 2018. Development and usability evaluation of a configurable educational game for the visually impaired. In 2018 IEEE Games, Entertainment, Media Conference (GEM) (pp. 1-9). IEEE. https://doi.org/10.1109/GEM.2018.8516472

[24] Yuan, B., & Folmer, E. 2008. Blind hero: enabling guitar hero for the visually impaired. In Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility (pp. 169-176).https://doi.org/10.1145/1414471.1414503

[25] de Oliveira, P. A., Lotto, E. P., Correa, A. G. D., Taboada, L. G., Costa, L. C., & Lopes, R. D. 2015. Virtual stage: an immersive musical game for people with visual impairment. In 2015 14th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames) (pp. 135-141). IEEE. https://doi.org/10.1109/SBGames.2015.26

[26] Kirke, A. 2018. When the soundtrack is the game: From audio-games to gaming the music. Emotion in Video Game Soundtracking, 65-83. https://doi.org/10.1007/978-3-319-72272-6_7

[27] Rovithis, E., Moustakas, N., Floros, A., & Vogklis, K. 2019. Audio legends: Investigating sonic interaction in an augmented reality audio game. Multimodal Technologies and Interaction, 3(4), 73. https://doi.org/10.3390/mti3040073

[28] Araújo, M. C., Silva, A. R., Darin, T. G., de Castro, E. L., Andrade, R. M., de Lima, E. T., Sánchez, J:; Filho J. A. & Viana, W. 2016. Design and usability of a braille-based mobile audiogame environment. In Proceedings of the 31st Annual ACM Symposium on Applied Computing (pp. 232-238). https://doi.org/10.1145/2851613.2851701

[29] Morelli, T. 2011. Improving the lives of youth with visual impairments through exergames. Insight, 4(4), 160.

[30] Bermejo, F., Gilberto, L. G., Lunati, V., & Arias, C. 2016. Audiojuego con sonidos envolventes: una experiencia preliminar con personas ciegas y con visión normal. Investigación en Discapacidad, 5(2), 71-80.

[31] Heller, L. M., Schenker, A., Grover, P., Gardner, M., & Liu, F. 2017. Evaluating two ways to train sensitivity to echoes to improve echolocation. In Proceedings of the 23rd International Conference on Auditory Display (ICAD 2017). Georgia Institute of Technology. https://doi.org/10.21785/icad2017.053

[32] Wu, W., Morina, R., Schenker, A., Gotsis, A., Chivukula, H., Gardner, M., Liu F., Barton S., Woyach S., Sinopoli B., Grover, P. & Heller, L. M. 2017. EchoExplorer TM: A game app for understanding echolocation and learning to navigate using echo cues. In Proceedings of the 23rd International Conference on Auditory Display (ICAD 2017). https://doi.org/10.21785/icad2017.040

[33] Whitton, J. P., Hancock, K. E., & Polley, D. B. 2014. Immersive audiomotor game play enhances neural and perceptual salience of weak signals in noise. In Proceedings of the National Academy of Sciences, 111 (25), E2606-E2615. https://doi.org/10.1073/pnas.1322184111

[34] Schuchert, J. B., & Lewald, J. 2020. Training with audio and video games improves audiospatial performance in a “cocktail-party” task: A controlled intervention study in young adults. bioRxiv, 2020-11. https://doi.org/10.1101/2020.11.17.386300

[35] Sánchez, J. 2012. Development of navigation skills through audio haptic video gaming in learners who are blind. Procedia Computer Science, 14, 102-110. https://doi.org/10.1016/j.procs.2012.10.012

[36] Connors, E. C., Chrastil, E. R., Sánchez, J., & Merabet, L. B. 2014. Action video game play and transfer of navigation and spatial cognition skills in adolescents who are blind. Frontiers in human neuroscience, 8, 133. https://doi.org/10.3389/fnhum.2014.00133

[37] Bălan, O., Moldoveanu, A., Moldoveanu, F., & Dascălu, M. I. 2014. Navigational 3D audio-based game-training towards rich auditory spatial representation of the environment. In 2014 18th International Conference on System Theory, Control and Computing (ICSTCC) (pp. 682-687). IEEE. https://doi.org/10.1109/ICSTCC.2014.6982496

[38] Lahav, O., Schloerb, D. W., & Srinivasan, M. A. 2015. Virtual environments for people who are visually impaired integrated into an orientation and mobility program. Journal of Visual Impairment & Blindness, 109(1), 5-16. https://doi.org/10.1177/0145482X1510900102

[39] Newen, A., De Bruin, L., & Gallagher, S. (Eds.). 2018. The Oxford handbook of 4E cognition. Oxford University Press.

[40] Shapiro, Lawrence, and Shannon Spaulding. 2021. Embodied Cognition. The Stanford Encyclopedia of Philosophy. Winter 2021 edition. Edited by Edward N. Zalta. Available online: https://plato.stanford.edu/archives/win2021/entries/embodied-cognition/

[41] Di Paolo, E., Buhrmann, T., & Barandiaran, X. 2017. Sensorimotor life: An enactive proposal. Oxford University Press.

[42] O'regan, J. K., & Noë, A. 2001. A sensorimotor account of vision and visual consciousness. Behavioral and brain sciences, 24(5), 939-973. https://doi.org/10.1017/S0140525X01000115

[43] Varela, F., Thompson, E. y Rosch, E. The embodied mind. 1o Ed. Cambridge: MIT Press; 1991.

[44] Khatchatourov, A., Stewart, J., & Lenay, C. 2007. Towards an enactive epistemology of technics. In Proceedings of the 4th International Conference on Enactive Interfaces 2007 (Enactive / 07), 192-132.

[45] Froese, T., McGann, M., Bigge, W., Spiers, A., & Seth, A. K. 2011. The enactive torch: a new tool for the science of perception. IEEE Transactions on Haptics, 5(4), 365-375. https://doi.org/10.1109/TOH.2011.57

[46] Van Dijk, J. 2018. Designing for embodied being-in-the-world: A critical analysis of the concept of embodiment in the design of hybrids. Multimodal Technologies and Interaction, 2(1), 7. https://doi.org/10.1145/3024969.3025007

[47] Vahlo, Jukka. 2018. In gameplay: the invariant structures and varieties of the video game gameplay experience. PhD Thesis, University of Turku.

[48] Bardy, B. G., Lagarde, J., & Mottet, D. 2012. Dynamics of skill acquisition in multimodal technological environments. Skill training in multimodal virtual environments, 31-46. https://doi.org/10.1201/b12704

[49] Sánchez, J., & Lumbreras, M. 1999. Virtual environment interaction through 3D audio by blind children. CyberPsychology & Behavior, 2(2), 101-111. https://doi.org/10.1089/cpb.1999.2.101

[50] Allain, K., Dado, B., Van Gelderen, M., Hokke, O., Oliveira, M., Bidarra, R. & Kybartas, B. 2015. 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. https://doi.org/10.1109/SIVE.2015.7361292

[51] Cavaco, S., Simões, D., & Silva, T. 2015. Spatialized audio in a vision rehabilitation game for training orientation and mobility skills. In Proceedings of the International Conference on Digital Audio Effects (DAFx-15).

[52] Magnusson, C., Waern, A., Gröhn, K. R., Bjernryd, Å., Bernhardsson, H., Jakobsson, A., Salo, J., Wallon, M. & Hedvall, P. O. 2011. Navigating the world and learning to like it - mobility training through a pervasive game. In Proceedings of the 13th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI 2011) (pp. 285-294). https://doi.org/10.1145/2037373.2037416

[53] Suzuki, Y., Honda, A., Iwaya, Y., Ohuchi, M., & Sakamoto, S. 2020. Toward cognitive usage of binaural displays. In: Blauert, J., Braasch, J. (eds.). The Technology of Binaural Understanding. Modern Acoustics and Signal Processing. Springer, Cham. https://doi.org/10.1007/978-3-030-00386-9_22

[54] Ohuchi, M., Iwaya, Y., Suzuki, Y., & Munekata, T. 2005. Training effect of a virtual auditory game on sound localization ability of the visually impaired. In 11th International Conference on Auditory Display (ICAD2005).

[55] Honda, A., Shibata, H., Gyoba, J., Saitou, K., Iwaya, Y., & Suzuki, Y. 2007. Transfer effects on sound localization performances from playing a virtual three-dimensional auditory game. Applied Acoustics, 68(8), 885-896. https://doi.org/10.1016/j.apacoust.2006.08.007

[56] Bronkhorst, A. W. 1995. Localization of real and virtual sound sources. The Journal of the Acoustical Society of America, 98(5), 2542–2553. doi:10.1121/1.413219

[57] Higgins, N. C., Pupo, D. A., Ozmeral, E. J., & Eddins, D. A. 2023. Head movement and its relation to hearing. Frontiers in Psychology, 14.

[58] Paul, S. 2009. Binaural Recording Technology: A Historical Review and Possible Future Developments. Acta Acustica united with Acustica, 95(5), 767-788. https://doi.org/10.3813/AAA.918208

[59] Broderick, J., Duggan, J., & Redfern, S. 2018. The importance of spatial audio in modern games and virtual environments. In 2018 IEEE Games, Entertainment, Media Conference (GEM) (pp. 1-9). IEEE. https://doi.org/10.1109/GEM.2018.8516445

[60] Chavez Sanchez, F., Angelica Martinez de La Peña, G., Adriana Mendoza Franco, G., & Iroel Heredia Carrillo, E. 2021. Exploring Audio Game design with Visually Impaired players: A Mexican Case Study. In Proceedings of the 16th International Audio Mostly Conference (pp. 24-31). https://doi.org/10.1145/3478384.3478410

[61] Sound Ideas. 2005. Sony Pictures Sound Effects Library Volumes 1-5.

[62] Sound Ideas. 1989. BBC Sound Effects Library Second Edition.

[63] Valentino R. 2008. Valentino Production Sound Effects Library.

[64] Sound Ideas. 1990. Lucasfilm Sound Effects Library.

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

[66] Sodnik, J., & Tomažič, S. 2015. Spatial auditory human-computer interfaces. Springer.

[67] Good, M. D., & Gilkey, R. H. 1996. Sound localization in noise: The effect of signal-to-noise ratio. The Journal of the Acoustical Society of America, 99(2), 1108–1117. https://doi.org/10.1121/1.415233

[68] Shilling, R. D., & Shinn-Cunningham, B. 2002. Virtual auditory displays. Handbook of virtual environment technology, 65-92. Lawrence Erlbaum Associates Publishers.

[69] Witmer, B. G., & Singer, M. J. 1998. Measuring presence in virtual environments: A presence questionnaire. Presence, 7(3), 225-240.

[70] Gaggioli, A., Bassi, M., & Fave, A. D. 2003. Quality of experience in virtual environments. Emerging Communication, 5, 121-136.

[71] Bermejo, F., Di Paolo, E. A., Gilberto, L. G., Lunati, V., & Barrios, M. V. 2020. Learning to find spatially reversed sounds. Scientific Reports, 10(1), 4562. https://doi.org/10.1038/s41598-020-61332-4

[72] Darin, T., Andrade, R., Sánchez, J. 2019. Principles for Evaluating Usability in Multimodal Games for People Who Are Blind. In: Antona, M., Stephanidis, C. (eds) Universal Access in Human-Computer Interaction. Multimodality and Assistive Environments. HCII 2019. Lecture Notes in Computer Science(), vol 11573. Springer, Cham. https://doi.org/10.1007/978-3-030-23563-5_18

[73] Chen, M., Kolko, B. E., Cuddihy, E., & Medina, E. 2011. Modeling but NOT measuring engagement in computer games. In Proceedings of the 7th international conference on Games+ Learning+ Society Conference (pp. 55-63).

[74] Fischer, T., Caversaccio, M., & Wimmer, W. 2020. A front-back confusion metric in horizontal sound localization: The fbc score. In ACM Symposium on Applied Perception 2020 (pp. 1-5). https://doi.org/10.1145/3385955.3407928

[75] Jurado, F., & Meza, R. E. 2017. An exploratory study in the use of gamer profiles and learning styles to build educational videogames. International Journal of Engineering Education, 33(2), 797-806.

[76] Zohaib, M. 2018. Dynamic Difficulty Adjustment (DDA) in Computer Games: A Review. Advances in Human-Computer Interaction, 2018 https://doi.org/10.1155/2018/5681652

[77] Lindau, A., & Weinzierl, S. 2012. Assessing the plausibility of virtual acoustic environments. Acta Acustica united with Acustica, 98(5), 804-810. https://doi.org/10.3813/AAA.918562

[78] Oberem, J., Masiero, B., & Fels, J. 2016. Experiments on authenticity and plausibility of binaural reproduction via headphones employing different recording methods. Applied Acoustics, 114, 71-78. https://doi.org/10.1016/j.apacoust.2016.07.009

[79] van Dartel M.F., Misker J. M. V., Nigten A. M. M., van der Ster J. 2007. Virtual Reality and Augmented Reality art explained in terms of sensory-motor coordination. In Proceedings of the 4th International Conference on Enactive Interfaces 2007 (Enactive / 07), (p. 417-420).

[80] Boehner, K., DePaula, R., Dourish, P., & Sengers, P. 2005. Affect: from information to interaction. In Proceedings of the 4th decennial conference on Critical computing: between sense and sensibility (pp. 59-68). https://doi.org/10.1145/1094562.1094570

[81] Duda, R. O., & Martens, W. L. 1998. Range dependence of the response of a spherical head model. The Journal of the Acoustical Society of America, 104(5), 3048-3058. https://doi.org/10.1121/1.423886

[82] Brungart, D. S., & Rabinowitz, W. M. 1999. Auditory localization of nearby sources. Head-related transfer functions. The Journal of the Acoustical Society of America, 106(3), 1465-1479. https://doi.org/10.1121/1.427180

[83] Yu, G., Wu, R., Liu, Y., & Xie, B. 2018. Near-field head-related transfer-function measurement and database of human subjects. The Journal of the Acoustical Society of America, 143(3), EL194-EL198. https://doi.org/10.1121/1.5027019

[84] Grechkin, T. Y., Plumert, J. M., & Kearney, J. K. 2014. Dynamic affordances in embodied interactive systems: The role of display and mode of locomotion. IEEE transactions on visualization and computer graphics, 20(4), 596-605. https://doi.org/10.1109/TVCG.2014.18

[85] Gulli, A., Fontana, F., Orzan, E., Aruffo, A., & Muzzi, E. 2022. Spontaneous head movements support accurate horizontal auditory localization in a virtual visual environment. Plos one, 17(12), e0278705. https://doi.org/10.1371/journal.pone.0278705

[86] Grohn, M., Lokki, T., & Takala, T. 2002. Static and dynamic sound source localization in a virtual room. In Audio Engineering Society Conference: 22nd International Conference: Virtual, Synthetic, and Entertainment Audio. Audio Engineering Society. https://research.aalto.fi/en/publications/static-and-dynamic-sound-source-localization-in-a-virtual-room

[87] Iwaya, Y., Suzuki, Y., & Kimura, D. 2003. Effects of head movement on front-back error in sound localization. Acoustical science and technology, 24(5), 322-324. https://doi.org/10.1250/ast.24.322

[88] Thurlow, W. R., Mangels, J. W., & Runge, P. S. 1967. Head movements during sound localization. The Journal of the Acoustical society of America, 42(2), 489-493. https://doi.org/10.1121/1.1910605

[89] Valzolgher, C., Alzhaler, M., Gessa, E., Todeschini, M., Nieto, P., Verdelet, G., ... & Pavani, F. 2020. The impact of a visual spatial frame on real sound-source localization in virtual reality. Current Research in Behavioral Sciences, 1, 100003. https://doi.org/10.1016/j.crbeha.2020.100003

[90] Aytekin, M., Moss, C. F., & Simon, J. Z. 2008. A sensorimotor approach to sound localization. Neural Computation, 20(3), 603-635.

[91] McLachlan, G., Majdak, P., Reijniers, J., & Peremans, H. 2021. Towards modelling active sound localisation based on Bayesian inference in a static environment. Acta Acustica, 5, 45. https://doi.org/10.1051/aacus/2021039

[92] Mieda, T., & Kokubu, M. 2020. Blind footballers direct their head towards an approaching ball during ball trapping. Scientific Reports, 10(1), 20246. https://doi.org/10.1038/s41598-020-77049-3

[93] Miura, T., Okochi, N., Suzuki, J., & Ifukube, T. 2023. Binaural Listening with Head Rotation Helps Persons with Blindness Perceive Narrow Obstacles. International Journal of Environmental Research and Public Health, 20(8), 5573. https://doi.org/10.3390/ijerph20085573

[94] Perrott, D. R., Saberi, K., Brown, K., & Strybel, T. Z. 1990. Auditory psychomotor coordination and visual search performance. Perception & psychophysics, 48(3), 214-226. https://doi.org/10.3758/BF03211521

[95] Carlile, S., Balachandar, K., & Kelly, H. 2014. Accommodating to new ears: the effects of sensory and sensory-motor feedback. The Journal of the Acoustical Society of America, 135(4), 2002-2011. https://doi.org/10.1121/1.4868369

[96] Brimijoin, W. O., McShefferty, D., & Akeroyd, M. A. 2010. Auditory and visual orienting responses in listeners with and without hearing-impairment. The Journal of the Acoustical Society of America, 127(6), 3678-3688. https://doi.org/10.1121%2F1.3409488

[97] Gessa, E., Giovanelli, E., Spinella, D., Verdelet, G., Farnè, A., Frau, G. N., ... & Valzolgher, C. 2022. Spontaneous head-movements improve sound localization in aging adults with hearing loss. Frontiers in Human Neuroscience, 16, 1026056. https://doi.org/10.3389/fnhum.2022.1026056

[98] Mueller, M. F., Meisenbacher, K., Lai, W. K., & Dillier, N. 2014. Sound localization with bilateral cochlear implants in noise: How much do head movements contribute to localization?. Cochlear implants international, 15(1), 36-42. https://doi.org/10.1179/1754762813Y.0000000040

[99] Pastore, M. T., Natale, S. J., Yost, W. A., & Dorman, M. F. 2018. Head movements allow listeners bilaterally implanted with cochlear implants to resolve front-back confusions. Ear and hearing, 39(6), 1224-1231. https://doi.org/10.1097/AUD.0000000000000581

[100] Jackson, R. C., & Farrow, D. 2005. Implicit perceptual training: How, when, and why? Human movement science, 24(3), 308-325. https://doi.org/10.1016/j.humov.2005.06.003

[101] Münte, T. F., Kohlmetz, C., Nager, W., & Altenmüller, E. 2001. Superior auditory spatial tuning in conductors. Nature, 409(6820), 580-580. https://doi.org/10.1038/35054668

[102] Dunai, L., Lengua, I., Peris-Fajarnés, G., & Brusola, F. 2015. Virtual sound localization by blind people. Archives of Acoustics, 40(4), 561-567. https://doi.org/10.1515/aoa-2015-0055

[103] Roëder, B., Teder-Sälejärvi, W., Sterr, A., Rösler, F., Hillyard, S. A., & Neville, H. J. 1999. Improved auditory spatial tuning in blind humans. Nature, 400(6740), 162-166. https://doi.org/10.1038/22106

[104] Parseihian, G. and Katz, B. F. G. 2012. Rapid head-related transfer function adaptation using a virtual auditory environment. J Acous Soc America, 131(4):2948–2957. https://doi.org/10.1121/1.3687448

[105] Stitt, P., Picinali, L., and Katz, B. F. G. 2019. Auditory accommodation to poorly matched non-individual spectral localization cues through active learning. Scientific Reports, 9(1):1063:1–14. https://doi.org/10.1038/s41598-018-37873-0

[106] Poirier-Quinot, D. and Katz, B. F. 2021. On the improvement of accommodation to non-individual HRTFs via VR active learning and inclusion of a 3D room response. Acta Acustica, 5(25):1–17. https://doi.org/10.1051/aacus/2021019.