Designing an Interactive and Collaborative Experience in Audio Augmented Reality

Valentin Bauer

Advanced Placement Project at BBC Research and Development
Supervisors: Chris Baume, Josh Reiss, Chris Pike

Queen Mary University of London
School of Engineering and Computer Sciences
Media and Arts Technology Program
August 2019

Designing an Interactive and Collaborative Experience in Audio Augmented Reality

Valentin Bauer

Abstract

Audio Augmented Reality (AAR) consists of adding spatial audio entities into the real environment. Existing mobile applications combined with the affordances of current technology open questions around interactive and collaborative AAR. This study proposes an experiment to examine how can spatial audio prompt and support actions in interactive AAR experiences; how distinct auditory information influences collaborative tasks and group dynamics; and how can gamified participatory AAR enhance storytelling. We are developing an AAR interactive multiplayer game experience using the Bose Frames (BF) Audio Sunglasses. Four participants at a time go through a gamified story that attempts to interfere with group dynamics, hence prompting them to reflect on the negative effects of using digital devices. Affordances provided by BF are harnessed creatively to navigate and interact with the AAR content. We here present the AAR collaborative platform that we built and our game in terms of experience design. We then outline the user experiments that we conducted, from preliminary testing with BBC R&D staff to testing the final experience. We then detail the testing methodology that we used and the quantitative and qualitative analysis that we conducted in order to answer our research question. At last, we give insights about methodologies for participatory storytelling in AAR and future developments.

Keywords
Audio Augmented Reality · Collaboration · 3D Audio · Audio Game · Storytelling · Experience Design

Acknowledgements
Thanks to the EPSRC and AHRC Centre for Doctoral Training in Media and Arts Technology at Queen Mary University of London and BBC R&D for supporting this research. I also want to thank Anna Nagele, with whom I collaborated to conduct all this project, my supervisors at BBC R&D Chris Baume and Chris Pike, my supervisor at Queen Mary University Josh Reiss, other people from BBC R&D who helped me in the process and particularly Tim Cowlishaw, Henry Cooke and Thomas Kelly. Lastly, I want to thank all the participants who took part in the study.

Contents

1 Introduction **3**

2 Background **4**
2.1 Binaural Audio 4
2.2 Audio Augmented Reality 4
2.3 Challenges 6

3 AAR Experience Design and Storytelling **7**
3.1 Concept Development 7
3.2 Review of Existing Applications 7
3.3 Game Overview 8
3.4 Game design 8

4 Audio Augmented Reality Architecture **9**
4.1 Technical choices 9
4.2 UNet 9
4.3 Game design 11
4.4 Remarks 12

5 Methodology **12**
5.1 Testing 12
5.2 Participants 13

6 Results **14**
6.1 Main results 15
6.1.1 Research question 1: Spatial audio 15
6.1.2 Research question 2: Asymmetric information 15
6.1.3 Research question 3: Gamified AAR and storytelling 16
6.2 Other findings 18

7 Analysis **19**

8 Discussion **21**
8.1 Insights 21
8.2 Limitations 22
8.3 Future developments 22

9 Conclusion **24**

A Appendix **28**
A.1 Readme: Not-a-Robot 28
A.2 Pre-study questionnaire 33
A.3 Post-task questionnaire 37
A.4 Post-study questionnaire 40
A.5 EuroVR 2019 poster 44

1 Introduction

In the past few years, the technological development of 3D audio for headphones using binaural audio has facilitated the delivery of Audio Augmented Reality (AAR) experiences. AAR consists of adding spatial audio entities into the real environment [26]. The technology has been applied to a range of fields such as teleconferencing, accessible audio systems, location-based games or education. AAR Research has mainly been focusing on the perception of sound quality [44], realism, discrimination between real and virtual sounds [26], or adaptation to the system [21]. Yet, interaction and collaboration remain under-researched.

One of the big challenges is acoustic transparency, so that the user can stay connected to his environment as if he had no headphones. Bose Frames (BF) audio sunglasses [11] are a newly available wearable AAR consumer technology that embed the speakers and technology in the frame of the sunglasses, and are therefore perfectly acoustically transparent. This has the potential to blend real world with augmented sounds and offers new opportunities for AAR experiences or applications and multiplayer interaction.

Figure 1: BoseFrames

This project was conducted in collaboration with Anna Nagele, from my 2019 Media and Arts Technology cohort. I personally worked on building the AAR architecture, implementing and designing the game on iOS, and also the 3D audio aspects and sound design. Anna worked on the game narration and game logic. We then conducted the testing and analysis together. We developed an interactive AAR multiplayer experience, for four players at a time, that encourages human interactions. Our prototype, *Please Confirm you are not a Robot*, explores three research questions:

- RQ1: How can spatial audio prompt and support actions in interactive AAR experiences?
- RQ2: How does distinct auditory information influence collaborative tasks and group dynamics?
- RQ3: How can gamified participatory AAR enhance storytelling?

In order to answer our research questions, we first conducted a pilot study with a group of four participants, and then a user study with four groups of four participants. All participants had different levels of expertise in 3D audio and augmented reality.

Section 2 gives a theoretical overview of AAR and related studies. Section 3 outlines the concept development and design of the multiplayer game and provides a description of the game mechanics. Section 4 describes the AAR infrastructure and implementation challenges. Section 5 looks at the research methodology. Section 6 focuses on the results of the experiments. Section 7 analyses the results and section 8 discusses them. Section 9 concluded the report.

2 Background

2.1 Binaural Audio

In AAR, 3D audio is often rendered over headphones using binaural audio. This technique relies on three different cues to spatialize a sound that reaches the listener's ears [9]:

- Interaural Time Differences (ITD) between the listener's ears
- Interaural Level Differences (ILD) between the listener's ears
- Spectral Cues, that depend on the reflections, absorption, and diffraction of an incoming sound with the listener's body, and mostly the shape of the head, pinnae, and shoulders. These morphological criterion are thus individual, and modeled by the Head Related Transfert Functions (HRTF).

Because of their individual nature, choosing the HRTF pair that best suits a listener remains a challenge. Using generic HRTF can result in front-back inversions, sound timbre artifacts, or externalisation issues [7]. These issues are more noticeable in static than in dynamic binaural, which consists of the addition of a headtracking system, and also increases the user immersion and localisation accuracy. As shown by Begault [8], optimal localisation can be achieved using headtracking, a synthesis of a virtual room (reverberation), and the use of individual HRTFs. A valuable asset of wearable devices is that they can offer headtracking possibilities and thus make dynamic binaural audio more widely available.

The type of headphones used is also important. Indeed, the more acoustically transparent it is (in terms of impedance), the better the binaural rendering will be [37]. Therefore, open-headphones can achieve better results in terms of binaural rendering than closed headphones. This is very promising regarding the use of devices like Bose Frames sunglasses, which are perfectly acoustically transparent.

According to Rajar's 2015 survey, a third of the United Kingdom adult population listens to radio over headphones (see figure 2). Moreover, Rajar's 2019 survey shows that a listener spends in average 7 hours per week listening to live radio over headphones, and that half of the audiobooks, podcasts, and digital tracks are listened over headphones [40]. The connected world we live in, with the possibility to listen audio programs at anytime and anywhere via connected devices (e.g tablets, smartphones...) progressively change the listener's behaviours. Therefore, in my view, producing binaural audio content and imagining new audio formats that the audience could experience over headphones becomes more and more relevant.

2.2 Audio Augmented Reality

Representations of the AAR sound field can be either natural or pseudoacoustic. For the former, virtual audio entities are directly added to the auditory real environment. For the latter, binaural microphones are added into the listener's ears and directly routed to the earphones so that the listener perceives a synthesized version of his environment. This system is also called "hear-through" audio and has a common example with hearing aids. In all cases the aim is that the user should not be able to determine which ones of the sources are real and which ones are not. This requires using high-quality 3D audio rendering [26] but is also linked to the feeling of presence in virtual environments (VEs).

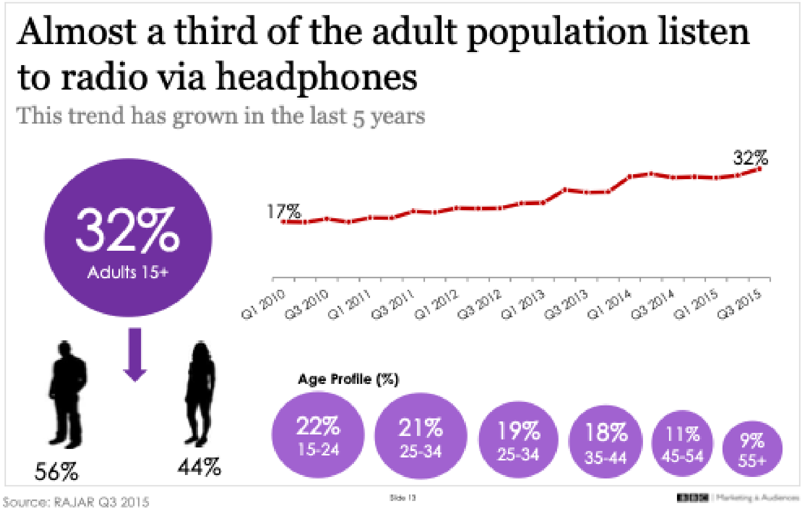


Figure 2: Profiles of people listening to radio over headphones

The feeling of presence has been studied in depth in the virtual reality (VR) literature but remains under-explored in AAR. It consists of both perceptive aspects (what the user feels in the space with his different senses) and cognitive aspects. The latter aspects are particularly linked with the concept of *affordances* [23]. This means that the virtual scene and objects have to make sense to the user (regarding his background), to match his mental model, in order to allow him to get the feeling of "being there", in the environment. The relationship between immersion and presence is complex, and, contrary to a common assumption, they differ. Immersion is more linked with the system's technology and the possibilities that the system offers to immerse the user in a VE [16]. Though, higher immersion can often be correlated with higher cumbersomeness and calibration requirements regarding the equipment used, which can hinder the feeling of presence. This leads Cummings et al. [16] to wonder "How immersive is enough?". Though, the literature does often assume that higher immersion contributes to a higher presence, and that higher presence contributes to a higher performance. But very little is known about those links. In our case, our study does not aim at evaluating the feeling of presence in AAR as a primary objective, but at giving some insights about it in AAR. Indeed, we benefit from using BF that offers a fantastic opportunity to look at these aspects. Lastly, some questionnaires can be used to evaluate the feeling of presence in VEs, such as Slater-Usoh-Steed questionnaire (SUS) or the Igroup Presence Questionnaire (IPQ). The IPQ seem to provide a good reliability within a reasonable timeframe [42].

We could imagine to complement AAR experiences with the use of haptics. So far, this field remains unexplored, except for some studies. For instance, *Kinaptic* [25] is an audio-only accessible and competitive game in stereo audio using haptics that uses a shared auditory space between sighted and blind people. It provided good results in terms of engagement and fun for all the users. Yet, to our knowledge, nothing is known about the use of haptics with AAR in 3D audio.

In the next section, we provide a list of the challenges that remain in AAR.

2.3 Challenges

- **Acoustically transparent devices** - Previous AAR studies have mainly focused on hear-through audio, but delivering AAR with transparent earphones remains under-explored. Questions arise about a seamless integration of audio entities onto the real auditory environment. BF is one of the first open-ear system that exists (see section 4.1). The mixed reality Microsoft Hololens glasses are another open-ear system that can render dynamic binaural audio and holographic 3D images, using small loudspeakers, a camera, eye tracking, and headtracking sensors integrated in the frame of the glasses [29]. Bone conduction headsets can also be used to render binaural audio. Despite some localization accuracy issues, good externalization and spatial discrimination can be achieved, even with elevation [5, 4].
- **Sound design** - Designing sounds in ARR needs more investigation, but binaural audio allows to increase the user immersion in comparison to stereo audio [14].
- **Mixing** - Mixing remains a challenge due to the dynamic nature of real sounds that change over time in both level and frequency. This can lead to an inappropriate balance between virtual and real sounds, and audio masking. To face it, some studies focus on implementing dynamic gains instead of static ones [38].
- **Real-time audio effects** - To our knowledge, the possibility to modify the listener's auditory environment in real-time is under-explored, apart from some studies that look at audio prototyping software such as PureData [43].
- **Individual vs Collaborative experiences** - Most AAR applications remain individual. Yet, some studies have focused on collaboration through location-based AAR games in stereo, using sounds triggered at specific locations [33, 20]. Others, like *Hear&There*, have tried to allow the user to leave audio imprints (music, sounds etc..) at particular locations, that can then be discovered by next users [41]. Regarding spatial AAR, the major LISTEN project (2003) allowed to wirelessly stream individualised-rendered binaural audio in an indoor environment for a maximum of eight users [19]. Then, Mariette and Katz [35] developed *SoundDelta* in 2009, a mobile multi-user AAR art/research project devoted to public events. They explored the potential of an *Ambisonic cell* approach to deliver personalized audio to a large number of users over a specific area. A server wirelessly streams ambisonic audio over WiFI to each user that depends on his position. It is then decoded in an individual binaural audio mix by the user's mobile device. *SoundDelta* authorizes the audience to move through the augmented space and to discover all the components of the sonic environment.
- **Reverberation** - Using a binaural artificial reverberation that matches the local environment where the user is raises challenges, because the place cannot be predicted in advanced. Yet, it would allow to more faithfully blend the augmented environment with the real one. Jot [31] proposed a promising solution to face this issue. He suggests a statistical reverberation model that uses a *reverberation fingerprint* of the space. The idea is to capture a single impulse response of the space, or to analyse and monitor the sounds captured in real-time by the phone's microphone of the user, in addition to compute a reverberation rendering for multiple virtual sound sources.

In the following section, we present the development of our AAR game and then the architecture the we built.

3 AAR Experience Design and Storytelling

As explained by Miller [36], every major technological advancement is followed by storytelling using that technology. In our case, binaural technology, AAR, and mobile devices possibilities have been expanding fast. AAR game design is specific because only the auditory perception is involved. This implies a range of design possibilities. On the one side, some designs aim at enabling users to carry out virtual tasks while they can be engaged in other activities with their eyes, attention or hands [28]. On the other side, some designs entirely involve the user in the virtual AAR tasks.

3.1 Concept Development

Headphones and similar devices have up until now contributed to a division of the auditory spaces into private and public. BF in contrast do not create a sound barrier to "real" auditory environment but still allow individual augmentation of sonic experiences. In addition to this asset, we wanted to use BF in order to create an AAR experience that would be different from an audio-only experience over headphones, using BF headtracking and gesture interaction possibilities (nodding, shaking the head, tapping the sides of the glasses).

To explore our research questions, we decided to imagine and create a collaborative AAR game. We were interested in fostering face-to-face interaction with a game that would be immersive and enjoyable for a group of four players at a time. Our hypothesis was that a game would make people involved in the experience, and that we could then base our methodology on it. In order to achieve a good user engagement several hypothesis were evoked: non-linearity and interactivity appeared to be important; perfect 3D audio accuracy may not be needed.

Lyons et al. [33] suggest that AAR has potential to bring people together in the same location and enhance social interactions. Considering the technological opportunities and our research interest the experience is designed around three features of AAR: Asymmetric information; Layering augmented sounds over "real life" sounds; Triggering sounds with head gestures and movements.

3.2 Review of Existing Applications

A limited amount of applications for Bose AR exist. Some apps allow users to explore a soundscape by selective listening [45], other ones use BF as a gaming device with taps and head movements as interactions [3], or make use of the technology's mobility through soundwalks [15]. Dead Drop Desperado [18] is the only known game that requires two players.

Apart from those Bose AR applications, spatial audio is used in immersive theatre to create imaginary spaces and parallel realities [17]. AAR experiences often assign a role to the user, asking him to perform. Looking at this in a multiplayer context, this is reminiscent of choreography and theatre performance. The theatre practice developed by theatre maker Augusto Boal [10] blurs the boundaries between everyday activities and performance. It is used to rehearse for desired social change [32]. Inspired by this practice, our game will result in a choreography prompting users to observe, reenact and subvert behaviours around digital devices.

3.3 Game Overview

Please Confirm you are not a Robot is a speculative fiction, constructed of four individual games. At the start, each participant meets their guide who introduces the scenario and the gesture controls: tapping, nodding and shaking head. In the first game participants are prompted to simultaneously draw a circle with one arm, and a cross with the other arm, in the air. Spatialised sounds of drawing a circle and cross will play for some players alongside the movement. We will look at whether sound cues have any effect on the participant's performance.

For the second game participants pair up and mirror each other's movements while being prompted to ask each other questions. We will look at whether this contributes to interpersonal closeness or affect between participants, or whether different layers of sound are distracting.

The third game uses BF as a gaming interface. A variety of notification sounds will appear in the sonic sphere around each participant. To turn them off they have to look at the sound and double-tap the side of the frames. Participants collect points for each sound they turn off. We will test different feedback sounds for finding sounds in space.

The last game requires the participants to tap each other's frames, following prompts of what they like about each other, to collect points. We will look at the interaction between participants. At the end, one participant will be separated from the group with a separate story-line. They will become the agent to end the whole experience by taking the other player's sunglasses off their face.

3.4 Game design

Early designs of interactive audio-only experiences highlighted the importance of sound design [33]. Sounds have to be put in context with other sounds or narration to establish a cause to effect relationship between the actions and the sounds, match the player's mental model [33], and thus present clear affordances (see section 2.2). Since varying loudness levels are a challenge in AAR, we decided on a specific room to conduct the experiments. We created the sound design with sounds gathered from personal recordings or Freesound.org (under the Creative Commons Licence), and using the software SoundParticles in combination with Reaper.

We designed the game for a persona that should not have any previous experience with 3D audio to be able to enjoy it. Though, we are aware that people who already have such experience may perceive 3D audio in a different way and have different expectations, in comparison with people who have no experience.

Then, because we wanted the user to be entirely focused on the auditory elements and gesture interactions, we minimized the interaction with the phone's screen. This required to create an ergonomic and minimal user interface (UI), to use an appropriate game logic, and to clearly communicate in the game design about the actions the user has to achieve [39].

To finish with, we are using the BF sunglasses because of the possibilities that they offer (see section 4). Though, we are aware that their design can underlie some mental models related to normal uses of sunglasses (e.g to protect from sun). Because every user will bring his background of prior models to make sense of this object and use it, we will take this phenomenon into account during the analysis to avoid some possible biases (see sections 5 and 6).

4 Audio Augmented Reality Architecture

4.1 Technical choices

In our modular AAR platform, we used BF because of their acoustic transparency, headtracking system, user interaction possibilities (nodding, shaking the head, and tapping the glasses) and ergonomics. BF have a Bluetooth low energy system. The headtracking system has an accelerometer, gyroscope and magnetometer, and a latency of around 200ms (higher than the 60ms optimal latency [12]). This may affect audio localisation but the other BF aspects made it suitable to achieve a good user engagement in the game. Indeed, one hypothesis was that since this system is less bulky in comparison with normal headphones that include a headtracking system, and since it does not cover the ears, BF may provide a good user engagement. Since our system had to be modular and support 3D audio, GPS tracking, BF API, and multiplayer possibilities, we chose to work with Unity software (version 2018.3 for compatibility). We worked with phones on iOS due to the compatibility with BF SDK, but future developments may also include Android phones. BF API gave us access to the sensor data of BF. We used Google Resonance Audio SDK for 3D audio rendering because of its high-quality with 3rd-order ambisonics [24]. The architecture also supports GPS tracking with Mapbox API [34], Audio Interactive Programming with Pure Data using the *Heavy Compiler* [2], and gives access to the phone's affordances (sensors, vibrator).

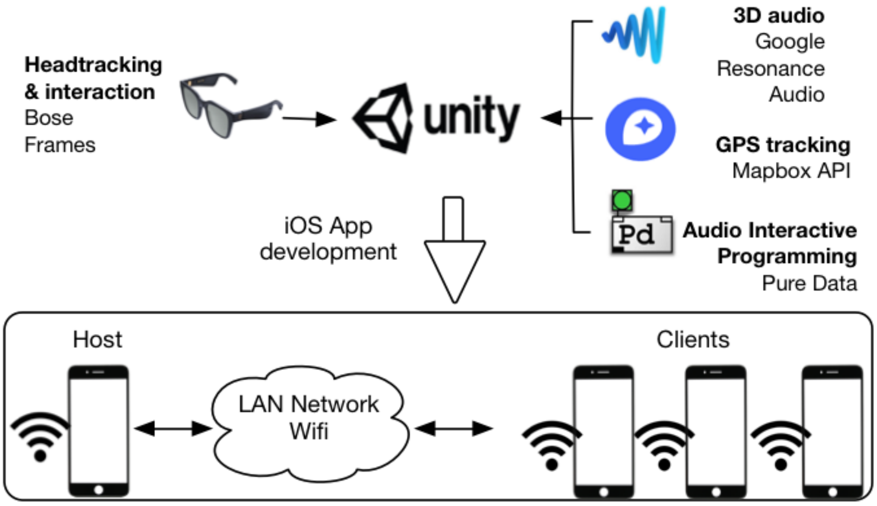


Figure 3: AAR Architecture

4.2 UNet

For multiplayer collaboration we designed a Local Area Network (LAN) over WiFi using the Unity's *UNet* system. The first player who connects to the game is the host. Being the host means that the player is a server and client at the same time. The next players (clients) who want to play have to enter the host IP address on their phone to join the game. This is only required for the first connection, since the IP address is then stored on the client's phone, using a singleton script in Unity. The collaborative logic is summurized on figure 4.

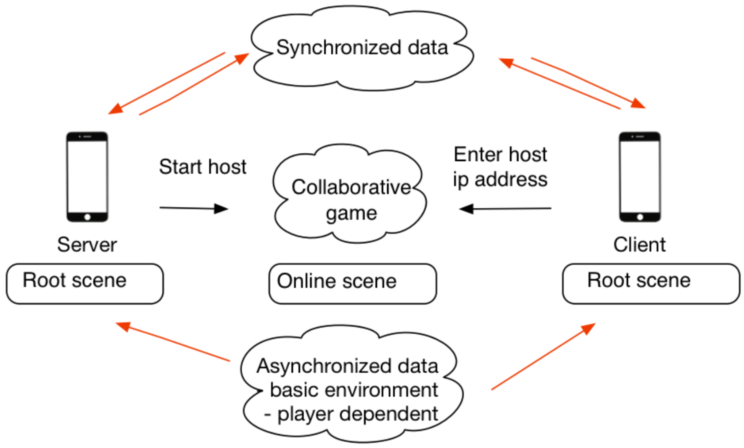


Figure 4: Collaborative Logic

The server-client communication is explained on figure 5. As in every networking system, two communication paths are necessary: one communication path from the clients to the server and a second communication path from the server to the clients. To achieve it in UNet, the following functions are used:

- Command: Function called on a client that runs on the server
- RpcClient: Function called on the server that run on the clients
- SyncVar: Variable automatically synchronized from the server on the clients

Each client "owns" a player. This is handled using the *hasAuthority* function in UNet. To give an example use of *Command* and *RpcClient* functions, imagine that client 2 wants to update the state of the player 2 that he owns (see figure 5). He first has to change it for itself, then to run a *Command* function to tell the server to also run the function. He finally has to use an *RpcClient* function (embedded in the body of the Command function) so that the server can tell all the clients to update their information for player2. Additionally, Client 2 has to take care not to update his state twice using the RpcClient. This is handled using the *hasAuthority* function.

Another example: When a client wants to update the state of every player (e.g from player 1 to player 4) we here use a non-player object (here called GO for the example) that is synchronized over the network. The logic is that the client has to modify the state of one variable on GO. This variable is then updated using Commands and RpcClient functions directly on GO so that all the clients and the host can get the changes. Then, all the players have to constantly check the state of the variables on GO in the update unity function. When they detect a change, they can then update their own variables.

To finish with, some objects are synchronised over the network, such as player dependent objects, or objects that keep track of global variables such as counters. But some asynchronised objects can also trigger events locally for each player. With this system events can be player specific, and different players can listen to different sounds synchronised over time. The networking logic is further explained in the ReadMe file (originally written for the github project) that I wrote (see appendix A.1. A template of the project will soon be publicly available on github.

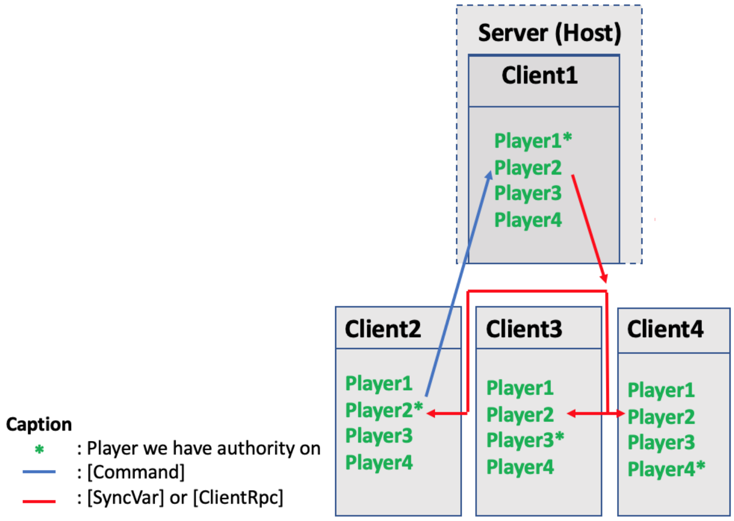


Figure 5: UNet

4.3 Game design

As explained in section 3.4, we developed the game to be ergonomic so that the user does not have to look at the screen when playing. The game starts with the introduction panel with Pi introducing the user to the experience. Then the user can go to each one of the games by clicking on the blue buttons (see figure 6 left picture). On figure 6 is displayed the first game, Circles&Crosses. The game starts when the host presses "Start Host" on the connection panel (top white button). The clients have to enter the host IP Address (bottom white button) and then press "Join Game" (middle game button). They then reach the online shared unity scene for this game, which is composed of a minimal UI, in order not to attract the user attention to much. Each player has the possibility to quit the game using the white buttons at the top (one for the host and one for the client). Otherwise, they can do the whole game and will be automatically disconnected at the end. When the user finishes a game he comes back to the main connection panel for this game. He then has to press the black button called "Main Menu" to reach the Menu panel, from which he can access the other games. All the details for the game logic and programming side are further explained in the ReadMe (see appendix A.1.

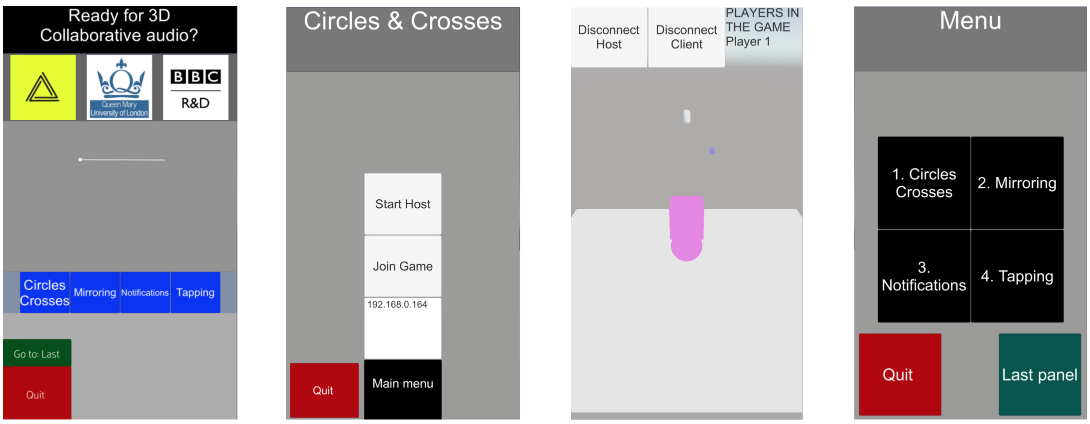


Figure 6: Game UI: main panels - from left to right: Introduction panel; Circles&Crosses Menu; Circles&Crosses Game; Main Menu

4.4 Remarks

Even if we finally did not use GPS tracking because it was not needed for *Please Confirm you are not a Robot*, we created several prototypes using Mapbox API and binaural audio during the development process. We tested two different systems: positioning sounds at specific GPS positions (1); defining GPS areas than can trigger sounds when the user reach them (2). System (1) did not work very well due to GPS inaccuracies. Indeed, the phones had a GPS precision of 6 to 10 meters. The problem was that the objects did not appear exactly where originally positioned in the space due to this inaccuracy, but also that with the GPS updates, the phone did not detect the exact user positions which led to jerky movements. This made it difficult for a user to find the sound cues in the space when spinning his head. We could argue that using the accelerometer of the phone or glasses in combination with the GPS may have helped to increase the precision, but we did not have enough time to try it. Conversely, system (2) worked very well. When a user moves in space and reaches an area that is big enough (bigger than the GPS precision), it can trigger an event (audio element). We think that this has a potential for storytelling and that it would be interesting to explore it further in future studies.

5 Methodology

5.1 Testing

First, we conducted preliminary testing with a group of four participants from BBC R&D to detect technical and narrative flaws, that allowed to refine our prototype. Second, we carried out a user study with four groups of four participants with different levels of expertise in 3D audio and AAR. The experiments lasted around one hour and a half and took place at BBC Broadcasting House, in a space where people always pass by, during two consecutive days. We used self-reported questionnaires with both quantitative and qualitative questions, along with researchers' observations of the participants' behaviours. The methodology is structured as follows:

1. Participants fill in a consent form and pre-study questionnaire
2. After each game, participants answer specific questions about the game.
3. At the end, a post-study questionnaire assess aspects of the game and is followed by a guided group discussion.

The pre-study questionnaire (see appendix A.2) asks 11 questions. Three are about demographics, then seven address the previous experience of the participant. The last one is about the feeling of connection of the participant with the group and uses the "Inclusion of Other in the Self (IOS) Scale" [1]. Regarding the previous experience of the participant, the first three questions focus on spatial audio, then two questions concern immersive theatre and participation in group performances. Finally, two questions are about the participant's use of a smartphone and social media accounts.

The post-task questionnaire (see appendix A.3) asks short questions at the end of each game about the level of engagement and participant's feelings. It allows to get direct insights that the participant could have forgotten after the experience. The questions are quite short to fill in so that it does not disturb too much the flow of the experiment. It consists of two questions for the *Circles&Crosses* game, four questions for the *Mirroring* game, four questions for the *Notification* game and three questions for the *Tapping* game.

The post-study questionnaire (see appendix A.4) asks 21 questions, about enjoyment, interactive ease, various problems, storytelling and feelings about the group. The first six questions address the general experience of the participant, regarding his overall feeling and what he liked and disliked. Then, two questions focus on the instructions given by the voice narrator and its "authority". Then, one question looks at the self-consciousness of the participant during the experiment. After that, seven questions address the feeling of presence. We took them from the IPQ presence questionnaire and chose the ones that were suitable for an AAR experience (we excluded the ones that only concerned visuals)[30]. After that, four questions address the experiment in general. They aim at getting insights to improve it for future developments. They look at the pace of the experiment, how the gesture interactions worked, and the clarity of the explanations about BF interactions and the interaction with the group. The final question is about the feeling of connection of the participant with the group and comes from the IOS scale (similar to the last question in the pre-study questionnaire).

The guided group discussion follows eight questions. The first one asks about possible contexts where such an experience could happen, and the second one about other similar multiplayer experiences that the participants could imagine. Then two questions address audio-only experiences and 3D audio. The fifth question focuses on asymmetric information. Question 6 is interested in general thoughts about the story. Question 7 concerns the reflective aspect of the experience around social media, and the last one address the overall participant's feeling about the experience.

During each experience, three to four researchers were present with the users, took notes about their behaviours and helped them if they had any questions. We filmed and recorded all the experiments and discussions to check our results if we had any doubt and to be able to conduct some behavioural analysis. We deleted all the videos and audio recordings at the end of the analysis and only kept anonymous data on spreadsheets.

Figure 7: Tapping game - Group 4

5.2 Participants

We recruited the participants from our personal and professional networks (BBC, Queen Mary University), using social media (e.g. Facebook, Linkedin) and emails. Because we were concerned about having an even number of women and men participating to the study, we used the emailing Ada's List (mailing list for women in technology) when seeing that not enough women did sign up for our study.

Regarding the pilot testing, the participants were aged from 22 to 40, counted one woman for three men. They all came from the BBC R&D staff. Three of them had a previous experience in 3D audio and binaural audio, but nobody in AAR. One participant had a previous experience in immersive theatre and another one in performance. They all used smartphones in a lot of different contexts and social media.

16 participants took part in the actual study. They were aged from 24 to 40, counted five women for eleven men, and came from a range of professions as displayed on figure 8. Research was the most represented field and counted two RD engineers; two researchers; three PhD students; one audio research scientist. The engineering field counted one engineer (for broadcast); two software engineers; one music technologist. Then the creative industries counted one interacting-creative director and one artist-art historian. Two participants had no experience in spatial audio, four no experience in binaural audio, and six no experience in AAR. Conversely, four participants were expert listeners in spatial audio and five in binaural audio (the same participants than for spatial audio plus one). Three "expert" participants for binaural audio belonged to the research field. Thus, some of the participants from the research field are familiar with audio (or work in the audio field). Regarding immersive theatre, half of the participants had never visited an immersive theatre performance, seven had experienced such an experience in the past and one was not sure. As regards the performance aspect, ten participants had already participated to group performances (music or theatre). Finally, all participants had a smartphone. Eleven participants used it in all the situations mentioned and can thus be called "heavy" users. Fifteen participants had a social media account.

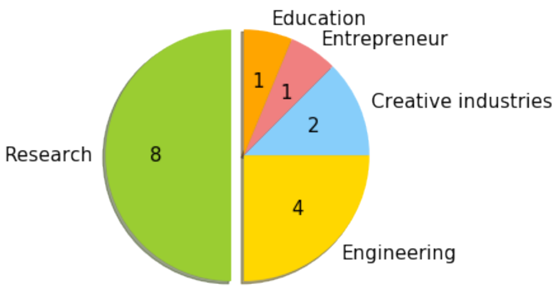


Figure 8: User experience demographics

6 Results

First, we submitted a poster that was accepted at the EuroVR-2019 Conference and describes the development process of the project (see appendix A.5). Then, regarding the analysis of the data, we used both quantitative methods (for the results of the questionnaires) and qualitative methods (questionnaires, observer's remarks, discussions, videos). To facilitate the reading, we grouped our results per finding and not per questionnaire or stage of the experiment. We first present our findings for our three research questions, namely "How can spatial audio prompt and support actions in interactive AAR experiences?" (RQ1), "How can asymmetric auditory information influence group dynamics and support or distract from collaborative tasks?" (RQ2) and "How can gamified participatory AAR enhance storytelling" (RQ3). We then summurize the main other findings that we got.

6.1 Main results

6.1.1 Research question 1: Spatial audio. 5/16 participants mention the AAR and 3D audio aspects about the elements that they liked the most about the experience (post-study questionnaire). Moreover, participants from groups 2 and 4 (G2, G4) said that they would like to see more 3D audio features in future developments (discussion). Overall, the participants enjoyed having an audio-only experiment, apart from 2 participants who would like to add visuals to it in order to feel more immersed. Lastly, participant 5 (P5) said it was "Fun to use the ears to navigate". This is overall very encouraging for future AAR experiments.

Some issues appeared about the perception of sound location. In the notification game, 11/16 participants were not satisfied of the experiment and 12/16 were a bit stressed. 11/16 participants had trouble to hear where the sound came from and 10/16 had trouble to find it when turning the head. P10 mentioned that "sound positioning was not stable", P18 said that "he got strong spatial cues but it was not consistent when he moved his head, [...], the tracking was not accurate enough, jumping from side to side...". Similar comments are said by P12, P17 and during G3 discussion. Even spatial audio expert users (P5, P6, P8, P10, P18) faced these issues. Moreover, the audio feedback used did not help them enough to find the targets, maybe because it was not distinguishable enough from the other sounds, too scarce to support the user actions (see section 6.2), or not understood by the users who were not sure about what sounds they were supposed to listen to (see section 6.2). Interestingly, most participants heard and tried to find the 3D targets on a 2D plane around them (often until one of the researchers helped them if they did not get any), or looked up to find them, but very few participants looked down. This tendency is evoked in G3 discussion with P12 telling he was constantly looking up. Last but not least, even if the participants did not have to move to find the targets in the game, looking for them often prompts the participants to move around in the space. They often stayed on their spot at the beginning of the game and then started to move (visible for Pilot group, G2, G4, G5).

Interestingly, the participants were not always sure if the sounds were real or not. This could disturb them or please them. For the former, it could make it difficult in the notification game when they did not know if they had to turn the sound off because were not sure if it was real or not. As said by P16, "It felt at times like the experience did connect to the space I was in". For the latter, during the group discussion, P20 mentioned that "I could hear someone who dropped a fork but could not tell if it was coming from the glasses. I thought it was pretty cool".

Lastly, the experience made appear the need to add an onboarding section. This would help to better explain the 3D audio sounds the participants have to focus on, and to add a short spatial audio training consisting for instance of explaining what the sounds coming from the different directions sound like (P13, discussion).

6.1.2 Research question 2: Asymmetric information. Regarding RQ2, overall, participants found interesting to listen to distinct auditory information. For instance, P12 said "When at some point I realised we got different instructions I thought it is really great", and was enthusiastic about future developments using asymmetric aspects. Some participants also liked the mystery that lies behind asynchronous information. For instance, P10 said "I got 10 and I won! Or, am

I lying?! You'll never know". Yet, other ones such as P18 mentioned that it was strange to communicate with a stranger while not knowing what the other could hear. Interestingly, participants also made choices about what to listen to between the instructions and the real persons due to asymmetric information. For instance, P9 said "Hey! She is still talking!". Lastly, some participants would have liked to know at the beginning of the experience that asymmetric information would be used and did not like to be surprised by it. This could be included in the onboarding aspect, and is further developed in section 6.2.

Asymmetric information appeared to influence group dynamics and support or distract from collaborative tasks in different ways. It first stimulated interaction and communication. The participants could tell about what they heard when it was different (P12: "Oh, for me it's different" when they are asked different things"), about what they were prompted to do (P19: "I am supposed to tell you ..."), and also often gathered at the center of the space the end of the games to talk about their respective experiences. But communication can also be non-verbal. For instance during the mirroring task, P17 adjusted her movements when she saw that they were difficult for her partner to reproduce. P13, when excluded during the tapping game, used body language to make the group aware that he was told to split from the group. Moreover, the confusion that raises from breakdowns of technology or lacks of information in the game design (see section 6.2) often prompts discussions. For instance, in the mirroring game, participants could try to mirror the others to "get back" in the game, communicated about what they heard when they did not understand the task (P9: "Did you hear that? It told me to double tap when we're ready. Are you?"), or took the role of the narrator Pi when their partner disconnected (P17). Finally, they sometimes gave accounts to explain their actions due to asymmetric information (e.g P20 "Wait, she is still talking.").

Sometimes, asynchronous information excluded one participant, for instance when P12 accidentally disconnected from the game in the Circles&Crosses activity. P12 felt excluded but thought it might have been predicted in the game. Yet, it only happened once and might be linked with the concept of the "digital authority" perceived, further developed in the analysis section. Apart from that, the participants did not feel excluded, even in the tapping game. Indeed, when one participant became the bad player and split from the group, instead of breaking communication, it often prompted it. Bad players could often feel weird at the beginning but quickly showed interest in that, because they could go into the layers of the voice assistant (P13) or "be the one" (P13). Other participants often went to give likes to the bad players so that they could still be part of the group (G3 and G4).

Lastly, it was confusing for the participants to pay attention to two similarly important sound layers at the same time. This particularly led to difficulties in the mirroring game with participants having to listen to the voice and do the tasks at the same time. All groups mentioned this difficulty during the discussion and many participants in the questionnaires.

6.1.3 Research question 3: Gamified AAR and storytelling. To start with, 13/16 participants said that they would like to try more AAR multiplayer games (post-study questionnaire). 5/16 participants liked the AAR and spatial audio aspects (as aforementioned) and 11/16 participants liked the social aspects of it (breaking social barriers, interacting with other people physically etc..). This shows the social potential of such collaborative gamified AAR experiences. Moreover, the experience was enjoyable and pleasant for most participants, since in all groups each participant was smiling or even laughing at some point (see figure 9).

During the experiment, the participants tended to perform and take the roles they were given to create their own roles. They thus added to the instructions and interpreted what to do. For instance some participant used the lexical field of theatre with "that's dramatic!' (pilot group). Other created small improvised scenarios in the mirroring game, with participants talking about what they were doing synchronously (G2, G3) (e.g making a bed "You take the pillow."). Also, the excluded participant often adopted his role at the end of the tapping game and added to the instructions and what he was prompted to say (pilot group, G4). To finish with the performance aspect, even if people were very still and focused when listening to the voice instructions to not miss anything, we noticed a great potential of the combination of movement, gesture interaction and spatial audio for social interaction and storytelling. For instance in the Circles&Crosses, Mirroring and Tapping games, participants seemed a bit reluctant to execute the actions at the beginning but quickly started to smile and laugh after around one minute (consistent over all the groups). The mirroring game also often prompted the pairs to talk between each others, apart from one in G5.

Overall, many observations and comments concerned social aspects. Firstly, some participants liked to interact with strangers and others disliked it. For instance, P17 found it weird because there was no previous small talk to break the "barriers", but P8 found it nice. In the tapping and mirroring games, interacting only with non-verbal communication was perceived as strange by some participants. Interestingly, nothing in the narration prevented them from interacting verbally. This will be further developed in the analysis section. Secondly, the experiment led to an increase of the IOS scale for all participants, which shows that everybody felt closer to the group to some degree at the end. This is emphasized by the fact that many participants considered that this experiment could be used as an ice-breaker activity. This evolution was noticed by participants who were happy to see the social evolution of the group (P7), or felt weird to interact with strangers at some point because it broke some social comfort (P8,P15,P17). Moreover, participants were more reluctant to interact with each other at the beginning than at the end (P5,P7 particularly). Thirdly, many participants judged themselves during the experiment. Some felt "silly" because they were aware of their surroundings (P15, P18, P10). Others blamed themselves when technological problems (disconnection) occurred.

Overall, the experiment captivated the attention of the participants who felt engaged, as shown by the questions about presence in the post-study questionnaire. Though, they still paid attention to their surroundings and felt that the environment did not feel "real", in comparison with real life. Therefore, they were engaged but not fully immersed in the environment. Moreover, 5 participants mentioned that the awareness of their surroundings can be increased by the feeling of being judged or seeing people around. This could sometimes lead to a decrease of the feeling of presence or increase it (when people do not know if the sources are real or not, as aforementioned). Lastly, we have to mention that the breakdowns of the technology that occurred (except for G3) may have affected the feeling of presence with people tuning in and out of the experience.

Game design	Collaboration	Digital authority	AAR Potentials	BF issues
Onboarding section needed	"Group character" emerged	Trust in technology	Design: chunky; sunglasses are strange	AAR can create expectations to see visuals
More audio feedback needed	Group synchronisation can be confusing	Voice perceived as authoritarian	Hardware: disconnection and latency issues	
Desire for more spatial sound design	Peripersonal space: positive evolution			

Table 1: Summary of main findings that differ from the research questions

6.2 Other findings

· We here present the main findings that emerged outside of the research questions (see table 1):

- **Overall game design -** Some game design insights appeared. Firstly, the experiment lacked from an onboarding section to explain the controls and how the experience works. This would have allowed to prevent the participants from being confused by explaining: how to use the device in details; how/where to tap the glasses; what sounds people have to focus on (most reported element); if the game is collaborative or not, competitive or not, uses asymmetric information or not. This last element notably confused the participants, with for instance P5 saying that he was distracted by "not knowing from the start that we might be asked different things" and could even be stressing (P15, notification game). Secondly, the experiment lacked from audio feedback, to support BF gesture interactions and show that they are recognized, and to give an auditory display of status and express the game progression. The former would allow to avoid some stress (P18, discussion), the latter some confusion, for instance due to participants not understanding their scores without looking at their screens (notification game). Finally, more spatial sound design is wanted, and the ringing sounds were perceived (notification game) as annoying and stressing.
- **Collaboration -** Participants liked AAR collaboration because 13 of them would like to try more multiplayer games. Interestingly, each group displayed a different "group character", regarding their movements and loudness. For instance, most participants from G2 and G4 smiled or laughed; G5 kept quiet and laughed little (only group composed of four complete strangers to each other). In Circles&Crosses, G2 and G4 said they were "engaged, challenged, silly, amused"; G5 felt "silly, confused, indifferent"; G3 encouraged each other. In the mirroring game, G3 moved a lot; G4 made slow movements and did not talk more than what was prompted. Then, the synchronisation of the audio elements gave consistency to the game but also led to some confusion (G2, discussion). Indeed, knowing how synchronised they were, the participants sometimes waited to be synchronised with the others to complete their actions. Lastly, some evolution of the "private sphere" of each participant, also called peripersonal space, was noticeable. For instance, in the tapping game, even if 14/16 participants felt comfortable to have their glasses touched, the interaction could be perceived as "very intimate" for strangers (G5), and the first likes as awkward because breaking some social space before becoming accepted (G3, P20). Moreover, participants accepted to have their glasses touched but were concerned about having their body possibly touched (P19).

- **Digital authority -** The concept of "digital authority", or obedience to the technology appeared. Indeed, as mentioned, P16 remained very still and confused when he was accidentally disconnected from the Circles&Crosses game. Yet, he did not say anything because he thought that it was normal. Hence, he did not criticize the system but accepted the situation in itself. Similarly, some participants followed exactly what the narration said and did nothing more. For instance G4 barely talked or laughed because it was not explicitly mentioned that they could do it. Interestingly, P15 said that she wondered "at which point the voice can tell her to do stuff and she would accept to do it". Therefore, the voice was perceived as authoritative and some participants felt that had to follow what it said. This could even lead to some surprise when the voice changed its tone (P18). One participant also got back to us by email two weeks after the experiment and said that she felt a bit in "autopilot" mode at some points.
- **Other AAR Potentials -** Some AAR potentials appeared. First, some participants expected to see the object that they were hearing because it sounded realistic in the environment. For instance, P14 was expecting to see a phone behind the couch (notification game) when hearing the ringing phone. Then, the likes given (tapping game) were perceived in a similar way as the ones given on social media. Indeed, 10/16 participants found exciting to get likes and the participant who got back to us by email said that the game had allowed her to "reflect on the consequences of [social media and digital technologies] on myself and other people".
- **BF Issues -** Some BF issues appeared due to the design and technology. In addition to the high latency aforementioned, two participants did not find BF ergonomic (a bit "chunky") and two others were annoyed by the shaded lenses that prevented them from seeing each others' eyes. Yet, P20 liked this feature because it allowed him to feel more isolated from the others. Regarding hardware problems, connection and disconnection of the glasses created frustration and impacted immersion with people tuning in and out of reality. It happened in all the groups except for G3 and we could not have control over it. Lastly, one participant found strange to wear sunglasses for that type of experiment because it is not the type of experience that we normally associate with sunglasses.
- **Game design: Comparison of the different games -** All Participants felt excited to have tried *Please Confirm you are not a Robot* and most people claimed that they would recommend it to a friend. The tapping and mirroring games were the most liked games. Results for the games are summurized on figure 9.

7 Analysis

Firstly, RQ1 results show that BF headtracking system did not improve the performances of participants in terms of audio localisation, and could even confuse them in the notification game. This is due to the system latency which is too high. Indeed, participants mentioned sounds that were unstable, moving when they were moving their head, or jumping from side to side. Secondly, and as in previous binaural audio studies, participants found it easier to detect sounds coming from left and right than from front and back (in the notification game particularly). Moreover, we can wonder if there is a cultural reason that would explain why participants were more looking up than down to find the audio targets in the notification game. Of course, the can

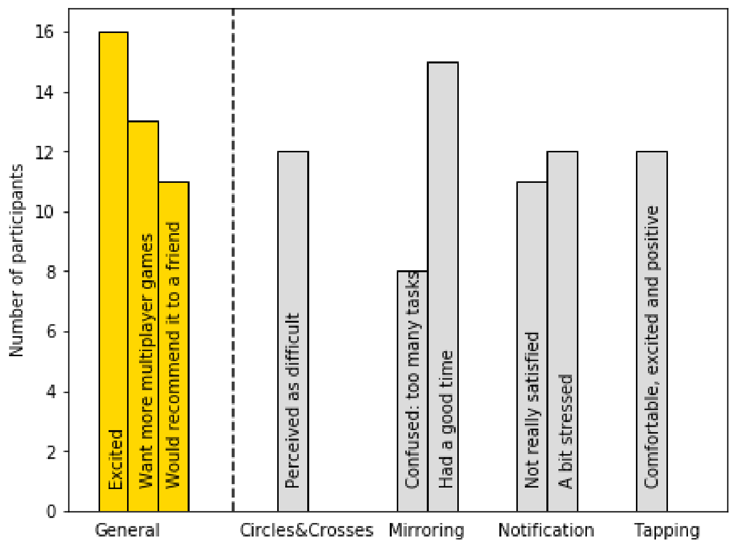


Figure 9: Main feedback for the games

be due to the generic HRTF and binaural rendering system used, that can better render the positive than negative elevation. Yet, we can argue that in our everyday life we expect ringing sounds and thus phones to be on a 2D plane or elevated, but not below us. This cultural aspect could maybe explain the tendency to look up and would need further investigation. Thirdly, the fact that the participants were not always sure if the audio sources were real or not is a great potential for AAR studies and game design. Furthermore, the perfectly acoustically transparent nature of BF may increase this possibility. Lastly, we remarked that AAR can allow or even prompt the participants to move and explore the space they are in. This offers great potential for future designs and needs further investigation.

About RQ2, interestingly, the use of asymmetric information often spurred people to give accounts about what they heard. This also happened when technology broke. Accounts are a fundamental phenomenon in human interaction. Indeed, when somebody performs an action without any apparent reason, he then has to give account for his action to the other person and explain it. This can be verbal but also non-verbal (using noddings, smalls movements of the head, shrugs etc..).

Regarding RQ3, firstly, *Please Confirm you are not a Robot* made the participant adopt the roles that they were instructed in a personal way. This is close to the theatre perspective developed by Augusto Boal [10] and detailed by Kohchar [32]. Indeed, participants progressively appropriate the script and start playing their characters. This illustrates the potential of such AAR experiments to develop experiences between theatre and AAR. Secondly, the answers to the presence questions (post-study questionnaire) show that the participants sometimes felt captivated by the experiment without being immersed in it. This implies that AAR can lead to a good feeling of presence without being fully immersed. Thirdly, the combination of movement, gesture interactions and spatial audio offers a great potential for human interaction. Indeed, even if people can seem reluctant to play at the beginning of each game, they soon start to smile and laugh when these three parameters are used together.

The experiment can lead to a reflexive process about digital technology as shown by the email that one participant sent us around two weeks after the experience.

Moreover, the likes that people gave in the tapping game were often perceived similarly as those on social media. Our take on that is that in AAR the participant who plays is instructed to complete actions and plays a role. But playing this role, the player is is not really the participant himself but also not really just a virtual player in the game. It is an "intermediary player" which lies between the real and virtual world. To our mind, this intermediary player has some resonances with the "roles" that people adopt on social media because of the kind of anonymity which surrounds it. This similarity can maybe give hints about the reason why some participants felt that the likes that they received during the tapping game felt similar in comparison with the likes that people get on social media. Therefore, we can wonder if AAR would have some potential to study the relationships we maintain on social media, and if AAR shares some similarity with social media regarding the character that we adopt.

8 Discussion

8.1 Insights

Research questions insights - To start with, RQ1 raises other questions that can be further explored, such as "Can spatial audio allow to give sound cues allowing to make the gesture easier?" We tried to get some insights about this aspect in the Circles&Crosses game, by giving some participants sound cues of a chalk for instance and other participants no sound. Though, our methodology did not allow to draw conclusions. This calls for further research in this direction. Then, RQ2 results show that the experiment could be used to dig into both verbal and non-verbal communication, with some participants exploring more non-verbal communication than they would normally do outside of the experience. Thus, AAR may allow to explore and give a novel methodological approach to what is called "embodied interaction", which means all the non-verbal communication used in conversation [6]. Hence, it could bring benefits outside of the experiment, if the experience is designed to make people more aware of the embodied interaction that they use in their everyday life. Also regarding RQ2, it appeared that participants gave accounts about what they heard in presence of asymmetric information. People always try to give logical explanations of individual situations to make sense of them (accountability), and the act of making sense of them is identical as creating it (reflexivity). This is developed by Garkinkel [22] who says that social action is constituted as reflexively accountable. Therefore, because of this ubiquitous aspect in human interaction, the participant try to avoid any misunderstanding by giving the reasons of his actions. Thus, AAR experiences would offer a novel approach to delve into the logic of giving accounts and reflexive accountability in human interaction. Lastly, RQ3 results outline the potential of movements to prompt communication and interaction between people. Moreover, according to us, one of the main AAR features is the possibility to move freely in the space while listening and interacting with audio content. To that respect, more research could be led to better understand how AAR can allow to facilitate discussion in collaborative games, and which AAR features are involved.
Other insights - Our study gave insights about the feeling of presence. It supports the idea developed by Cummings et al.[16] that calls into question the common idea which says that "better immersion leads to better presence which leads to better

performance". Even if the feeling of presence in AAR is still under researched, our study shows that the participant perception of his surroundings in AAR does not hinder the feeling of presence but prompts it in another way. Participants can be engaged, captivated, present, and have good group performances without being entirely immersed.

Some insights also concerned group dynamics. First, the group character aspect interestingly emerged. This is a phenomenon that has already been studied in the literature regarding small group dynamics, but AAR offers new opportunities to look at it. Then, the experience made appear a flexibility of the peripersonal space during the games. This notion a flexible boundaries, with a space getting bigger when people know more each other and smaller when they are strangers, has been shown by Hobeika et al. [27]. Because of the evolution of the peripersonal space that appeared, AAR collaborative games seems to be a fantastic way to dig more into this aspect. The methodology proposed by Canzoneri et al. [13] and used by Hobeika et al. [27] might be used and adapted to study it.

Lastly, what we call "digital authority", e.g the trust that some participants put into technology and other do not, is of interest. Indeed, carrying out further investigation in this direction may allow to get more results such as the participant who got back to us by email and said that the experience had allowed her to reflect about her use of technology. The game design could maybe directly integrate this aspect. A possible application could for instance be to trigger some alternative narratives, with a digital narrator prompting the user to do other actions if the player is following too much the narration.

8.2 Limitations

The project also contained some limitations. Firstly, some BF hardware issues regarding gesture recognition, or disconnection from the phones confused the participants. Secondly, some software issues emerged with the networking system, with participants sometimes disconnecting from the online scene (P16). A more robust networking solution should thus be developed. Thirdly, the lack of feedback and onboarding section mentioned in the results section could be confusing. Fourthly, we here mixed self-reported questionnaire with group discussions and behavioural analysis - with the videos and during the experiments - to avoid the inherent bias of each method. Yet, the observations of the participants that the researchers carried out brought some inevitable and obvious bias in terms of the researcher's background and what he is most interested in. Fifthly, our demographics did not contain an even gender split (despite our efforts to make it possible), and a lot of participants had a technical background. Sixthly, we tried the game with the same order each time, which can bring some bias. Lastly, because of the AAR loudness challenge (see background section), we limited our study to an indoor space and did not expand it to other spaces (indoors and outdoors).

8.3 Future developments

Our experience made emerge the following AAR design recommendations:

- Improve the audio feedback
- Make it clear when information is shared or not

- Improve the storytelling part
- Benefit from "digital authority" for storytelling and interaction patterns
- Improve onboarding and instructions during the experience
- Add 3D audio training
- Do not design for full immersion in AAR with BF
- Take the context more into account
- Design the experience for different contexts (space, time of the day etc..)
- Design for a flexible number of participants
- Dig more into the gamification aspect
- Take advantage of BF capabilities being aware of its limits (see section 6.2)
- Explore more spatial audio possibilities
- Explore human behavior and body movement beforehand and design for it

Our methodology could also be improved. Firstly, we sometimes had trouble to know what the participants were hearing during the games, and would have needed a monitoring system to achieve it. Secondly, making a game in one go would allow more people to try it without having to fill in all the questionnaires. This implies to be able to collect data from each user during the experience (with their consent), and could be done by:

- Recording BF sensor data in real time
- Using BF microphone to record some moments of the game
- Offering people the possibility to give feedback during the experience

Other developments are also considered. The loudness challenge should be addressed, with for instance a microphone (from BF or the phone) used to detect the loudness around the listener and automatically adjust the level of the audio content to it [38]. Voice recognition could also be added using the microphone. It might increase the immersion by allowing the user to orally interact with the audio content.

The group discussions also made appear future developments. The participants were keen to interact more with the voice narrator, dig into the potential offered by asymmetric information, by the collaborative experience, and by the confusion between real and virtual objects. About possible uses and applications of this kind of experiences, they first evoked a game where people have to walk that could be a guided thing, used in a museum, as a GPS memory game (leaving elements at particular locations for other people), or as an adaptation of the game *Pokemon Go*. Collaborative party games were also evoked a lot. Participants gave examples of a "Werewolf game", board game, or game to explore the space. Then, social aspects were mentioned with speed dating or ice breaking for team environments (said by all groups). After that, individual items were evoked with acting classes, accessible experiments for visually impaired, or listening to a well produced content (for instance walking in Paris and playing music from different places - P20). Lastly, regarding possible improvements, even if some participants did not want to add more technology to the experience, others mentioned adding GPS to it or haptics (P9).

Lastly, the experience showed a great potential for role taking. Thus, this would be good to carry out further work in this way and blur the boundaries between theatre and AAR as suggested by Augusto Boal [10].

9 Conclusion

We reviewed previous AAR studies and discovered that newly available technologies such as BF, which do not cover the ears, offer new opportunities for collaboration and interaction in AAR. We were inspired by previous multiplayer experiences, methods of interactive storytelling, and theatre practices to develop *Please Confirm you are not a Robot*. This game immerses a group of four players into a scene where they play and act out several scenes, guided by asymmetric information and binaural sound cues. This report details the development of the modular AAR architecture that supports our experimental game. It then presents the user experiments that we led and analyse the results that we got. We finally derive some general AAR game design recommendations and suggest future developments.

Most participant liked the spatial audio aspects despite some sound localisation issues which were mainly due to the high headtracking latency. The use of asymmetric information appeared to prompt communication, be it verbal or not. The participants were very engaged, liked the AAR collaborative aspect, and the combination of movements, gesture interaction and 3D audio showed a great potential for human interaction. The results also gave hints about the feeling of presence, by revealing that good engagement and performance in AAR do not need a full immersion. Designing experiences that lie at the boundary between theatre and AAR presents a great potential for communication and human interaction. To finish with, some interesting aspects emerged such as the "digital authority" or "group character" aspects, that would be worth to study more in future works.

References

[1] A. Aron, Aron E. N., and D. Smollan. "Inclusion of other in the self scale and the structure of interpersonal closeness". In: *Journal of Personality and Social Psychology*. Vol. 63. 1992, pp. 596–612.

[2] Enzien Audio. *Heavy Compiler Collection (hvcc)*. 2018. URL: `https://github.com/enzienaudio/hvcc`.

[3] Audiojack. *Audiojack Homepage*. 2019. URL: `https://www.audiojack.com`.

[4] A. Barde et al. "Binaural Spatialization over a Bone Conduction Headset: Elevation Perception". In: *ICAD-2016*. July 2016.

[5] A. Barde et al. "Binaural Spatialization over a Bone Conduction Headset: Minimum Discernable Angular Difference". In: *Audio Engineering Society Convention 140*. May 2016.

[6] J. B. Bavelas and J. Gerwing. "Conversational hand gestures and facial displays in face-to-face dialogue". In: *Social communication*. Ed. by K. Fiedler. New York: Psychology Press [Frontiers of Social Psychology Series], 2007. Chap. 10, pp. 283–308.

[7] D. R. Begault. *3DD Sound for Virtual Reality and Multimedia*. San Diego, CA, USA: Academic Press Professional, Inc., 1994. ISBN: 0-12-084735-3.

[8] D. R. Begault et al. "Direct Comparison of the Impact of Head Tracking, Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source". In: *Audio Engineering Society Convention 108*. 2000.

[9] J. Blauert. *Spatial hearing. The Psychophysics of Human Sound Localization*. Cambridge: MA: MIT Press, 1996. ISBN: 9780262024136.

[10] A. Boal. *Games for Actors and Non-Actors. Second Edition*. Taylor and Francis e-Library, 2005.

[11] Bose. *Bose Frames Homepage*. 2019. URL: `https://www.bose.fr/fr_fr/products/frames/bose-frames-alto.html`.

[12] D. Brungart, A. J. Kordik, and B. D. Simpson. "Effects of Headtracker Latency in Virtual Audio Displays". In: *J. Audio Eng. Soc* 54.1/2 (2006), pp. 32–44.

[13] Elisa Canzoneri, Elisa Magosso, and Andrea Serino. "Dynamic Sounds Capture the Boundaries of Peripersonal Space Representation in Humans". In: *PLOS ONE* 7.9 (Sept. 2012), pp. 1–8. DOI: `10.1371/journal.pone.0044306`. URL: `https://doi.org/10.1371/journal.pone.0044306`.

[14] T. Chatzidimitris, Gavalas D., and Michael D. "SoundPacman: Audio augmented reality in location-based games". In: *2016 18th Mediterranean Electrotechnical Conference (MELECON)*. Apr. 2016, pp. 1–6. DOI: `10.1109/MELCON.2016.7495414`.

[15] H. Cooke. *Audio AR: Sound Walk Research The Missing Voice*. 2019. URL: `https://www.bbc.co.uk/rd/blog/2019-03-audio-sound-walks-tours`.

[16] J. J. Cummings, J. N. Bailenson, and M. J. Fidler. "How immersive is enough? A meta-analysis of the effect of immersive technology on user presence". en. In: *Media Psychology* (May 2015), pp. 1–38. ISSN: 1521-3269. DOI: 10.1.1.363.6971.

[17] Darkfield. *Darkfield Homepage*. 2019. URL: http://www.darkfield.org.

[18] DeadDrop. *Dead Drop Desperados App Store Preview*. 2019. URL: https://apps.apple.com/us/app/dead-drop-desperado/id1454393037.

[19] G. Eckel et al. *Listen - augmenting everyday environments through interactive sound- scapes - Technical report*. Dec. 2003. URL: http://iem.at/~eckel/download/LISTEN-FinalReport.pdf.

[20] I. Ekman. "Sound-based Gaming for Sighted Audiences– Experiences from a Mobile Multiplayer Location Aware Game". In: *Proceedings on 2nd AudioMostly*. Germany, 2007.

[21] I. Engel and L. Picinali. "Long-Term User Adaptation to an Audio Augmented Reality System". In: *24th International Congress On Sound and Vibration*. London, July 2017.

[22] H. Garfinkel. *Studies in ethnomethodology*. Englewood Cliffs: N.J.: Prentice-Hall, 1967.

[23] J. J. Gibson. *"The theory of affordances," in Perceiving, Acting, and Knowing. Towards an Ecological Psychology.* eds Shaw R., Bransford J. Hoboken, NJ: John Wiley & Sons Inc., 1977, 127–143.

[24] M. Gorzel et al. "Efficient Encoding and Decoding of Binaural Sound with Resonance Audio". In: *Audio Engineering Society Conference: 2019 AES International Conference on Immersive and Interactive Audio*. Mar. 2019.

[25] A. Grabski et al. "Kinaptic - Techniques and insights for creating competitive accessible 3D games for sighted and visually impaired users". In: *2016 IEEE Haptics Symposium (HAPTICS)*. Apr. 2016, pp. 325–331. DOI: 10.1109/HAPTICS.2016.7463198.

[26] A. Härmä et al. "Augmented Reality Audio for Mobile and Wearable Appliances". In: *J. Audio Eng. Soc* 52.6 (2004), pp. 618–639.

[27] L. Hobeika, M. Taffou, and I. Viaud-Delmon. "Social coding of the multisensory space around us". In: *Royal Society Open Science* 6.8 (2019), p. 181878. DOI: 10.1098/rsos.181878.

[28] S. Holland, D. R. Morse, and H. Gedenryd. "AudioGPS: Spatial Audio Navigation with a Minimal Attention Interface". In: *Personal Ubiquitous Comput.* 6.4 (Jan. 2002), pp. 253–259. ISSN: 1617-4909. DOI: 10.1007/s007790200025.

[29] Hololens. *Hololens Homepage*. 2019. URL: https://www.microsoft.com/en-us/hololens/.

[30] IPQ. *Igroup presence questionnaire homepage*. 2019. URL: http://www.igroup.org/pq/ipq/download.php.

[31] J.-M. Jot and K. S. Lee. "Augmented Reality Headphone Environment Rendering". In: *Audio Engineering Society Conference: 2016 AES International Conference on Audio for Virtual and Augmented Reality*. 2016.

[32] K. Kochhar-Lindgren. "Towards A Communal Body Of Art: The Exquisite Corpse and Augusto Boal's Theatre". In: *Angelaki: Journal of Theoretical Humanities*. Vol. 7. 1. 2002, pp. 217–226. DOI: 10.1080/09697250220142137.

[33] K. Lyons, M. Gandy, and T. Starner. "Guided by Voices: An Audio Augmented Reality System". In: *Proceedings of International Conference on Auditory Display (ICAD) 2000*. 2000.

[34] Mapbox. *Mapbox API*. 2019. URL: https://www.mapbox.com.

[35] N. Mariette and B. Katz. "SoundDelta - Largescale, multi-user audio augmented reality". In: *EAA Symp. on Auralization*. Espoo, Finland, 2009, pp. 1–6.

[36] Carolyn H. Miller. *Digital Storytelling. A creator's guide to interactive entertainment*. New York, USA: Routledge, 2014. ISBN: 9780203425923.

[37] H. Møller. "Fundamentals of binaural technology". In: *Applied Acoustics* 36.3 (1992), pp. 171–218. ISSN: 0003-682X. DOI: https://doi.org/10.1016/0003-682X(92)90046-U.

[38] N. Moustakas et al. "Augmented Audio-Only Games: A New Generation of Immersive Acoustic Environments through Advanced Mixing". In: *Audio Engineering Society Convention 146*. Mar. 2019.

[39] D. A. Norman. *The Design of Everyday Things*. New York, NY, USA: Basic Books, Inc., 2002. ISBN: 9780465067107.

[40] Rajar. *MIDAS_Summer2019*. 2019. URL: https://www.rajar.co.uk/docs/news/MIDAS_Summer_2019.pdf.

[41] J. Rozier, K. Karahalios, and J. Donath. "Hear&There: An Augmented Reality System of Linked Audio". In: *ICAD-2000*. Atlanta, Georgia, Apr. 2000.

[42] V. Schwind et al. "Using Presence Questionnaires in Virtual Reality". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. Glasgow, Scotland Uk: ACM, 2019, 360:1–360:12. ISBN: 978-1-4503-5970-2. DOI: 10.1145/3290605.3300590.

[43] D. Storek, J. Stuchlik, and F. Rund. "Modifications of the surrounding auditory space by augmented reality audio: Introduction to warped acoustic reality". In: *ICAD–2015*. Graz, Austria, July 2015.

[44] M. Tikander. "Usability Issues in Listening to Natural Sounds with an Augmented Reality Audio Headset". In: *J. Audio Eng. Soc* 57.6 (2009), pp. 430–441.

[45] Traverse. *Traverse Homepage*. 2019. URL: https://www.traverse.fm.

A Appendix

A.1 Readme: Not-a-Robot

Introduction

This Audio Augmented Reality (AAR) multiplayer game is developed using ”Bose Frames” (BF) sunglasses. This is a joint project between Queen Mary's University of London and BBC RD that focuses on human computer interaction in AAR. Here we detail the AAR architecture that can be used to develop other multiplayer games.

BF Audio Sunglasses are a newly available wearable AAR consumer technology that embed the speakers and technology in the frame of the sunglasses. BF are perfectly acoustically transparent, because they are an open-ear system that do not block the user's perception of his auditory environment. They offer a headtracking system, three types of user interaction (nodding, shaking the head, tapping the side of the glasses) and a good ergonomics. They are wireless using Bluetooth low energy. The headtracking system has an accelerometer, gyroscope and magnetometer.

Since the system had to support 3D audio, GPS tracking, Bose Frames API (BF API), and multiplayer possibilities, we chose to work with Unity software (version 2018.3 for compatibility). We work with phones on iOS because of Bose SDK compatibility, but future developments may also include Android phones.

For multiplayer collaboration we designed a Local Area Network (LAN) over WiFi using the UNet system. The first player who connects to the game is the host and starts broadcasting its IP address. Following players (clients) automatically detect it and join the game. Some objects are synchronised over the network, such as player dependent objects, or non-player objects that keep track of global variables. Asynchronized objects can also trigger events locally for each player. With this system events can be player specific, and different players can listen to different sounds synchronised over time. BF API give access to the sensor data of BF. We use Google Resonance Audio SDK for 3D audio rendering because of its high-quality with 3rd-order ambisonics. The system also gives access to the phone affordances (sensors, vibrator).

Game Logic

To run the game, we start in the root scene that contains the elements we need to connect to the glasses. Then we can go to each separate game, by pressing the blue navigation buttons, or pressing a “Click to start!” Button that appears after the introductory voice narration stops. The different games that we can access correspond to the following scenes:

- CirclesCrosses_Base
- Mirroring_Base
- TargetAround_Base
- TappingGlasses_Base

In each scene, we can start the actual game by pressing the “Start Host” button if you are the first player to connect. At this moment, the host player's phone starts broadcasting its IP address. Following players (clients) automatically detect it and join the game (otherwise a failsafe "Join Game" button is available if no automatic connection happens). An automatic naming system gives names in the connection order, so that the first player to connect is Player 1, the second to connect is Player 2, etc.. Starting the game send us on the unity game scenes, respectively called:

- CirclesCrosses_Game
- Mirroring_Game
- TargetAround_Game
- TappingGlasses_Game

You can quit each game by clicking on "Disconnect Client" if you are a client, or "Disconnect Host" if you are the host. It will then bring you back to the corresponding base session. Otherwise you will be automatically disconnected at the end of the game.

When coming back to the base scene, you can access the other games by clicking on the "Main menu" button which will send you to the MainMenu Unity scene. You will have to pass by this menu scene each time you want to change the game. This scene only allows to access the different games, to quit the iOS app, or to go to the last panel of the game. This last panel corresponds displays acknowledgements to the participants, credits of the game, and allows the user to quit the iOS app.

Scenes

The project is organized with the assets for Bose API, resonance audio API in the folders respectively called Bose and ResonanceAudio. All the Prefabs mentioned below are located in the Prefabs folder, under OnBoarding for the games. Finally, the Scenes, Scripts, and Sounds folders respectively gather the elements organized by games.

Connection scenes. Root scene

WearableConnectUIPanel_VB object has two scripts: the Bose singleton WearableConnectUIPanel that enables BF connection, and the ROOT_WearableConnect_Panel_VB (described below). The object CalibrationDemoUIPanel_VB has the script ROOT_CalibrationPanel that calibrates the glasses. The calibration is basic and draws from the Bose Advanced Demo (in unity BF API). It requires the user to stay still and look in front of him for some seconds, and then pass the calibration data to the object controlled in the unity scene (the one controlled by the RotationMatcher object).

When the game starts, the WearableConnectUIPanel opens up. Then, CalibrationPanel detects when the glasses are detected. This can be over USB when using a computer, or over bluetooth when using the phone. The detection of the glasses then activates the ROOT_Calibration_bose_script (on the GlobalMenu object). This script starts the calibration of the glasses by activating the CalibrationDemoUIPanel_VB object. Then, ROOT_WearableConnect_Panel_VB indicates when the calibration finishes. At the end of the calibration, ROOT_Calibration_bose_script deactivates the CalibrationDemoUIPanel_VB and activates the GhostyVoice object, which triggers the introductory audio and narration, then deactivates itself. The player then has access to the main root panel.

When using the BF glasses over USB, ensure that the GLOBAL_GAME object contains the Bose scripts called WearableControl (where USB Provider has to be mentioned in the Editor Default Provider section), and WearableRequirements (making sure to check the sensors that are needed).

The Root Scene Panel has different options.

- Four blue buttons to go to each game scene
- A green button to go to the last scene
- A red button to quit the iOS app

The GlobalMenu object contains all the panels and buttons for this game. The scripts that enable to move to the other games/quit the iOS app are localized on the Actions game object, which is a child of the GlobalMenu. In order to go to another scene, a button has to be used and this Actions objects dragged on it as a reference. The function used is called LoadScene_single.LoadSceneOnClick, and the number of the scene you want to go to has to be mentioned (it corresponds to the build index of the scene in build settings).

StartTheApp_FrameRate is an object with a script attached that sets the frame rate at 30f/s. This is to decrease the refreshing rate of the screen in order to limit possible latency problems of the app when running it on a phone.

SoundManager is a singleton sound object. It allows to have a few objects (moving or not) that can be accessed in every scenes in order to change the audio environment. The object can be used using the following command: GameObject.Find("SoundManager").GetComponent(); An example can be found in the AudioRoot object with the script called RootSceneAudio.

The BOSE_FRAMES game object has to be positioned at the scene/game origin, and contains an AudioListener, ResonanceAudioListener and RotationMatcher (to use the headtracking of the glasses). All the audio (3D or not) is heard through this object.

To finish with, the Bose script that allows to use the gestures is called GestureDetector.

Menu scene

This scene simply allows to go from one game to another. All panels and buttons are located in the Canvas object. The scripts that allow to change the scene or quit the iOS app are located on the Actions object (as in the RootScene). The script on the MenuAudio object allows to grab the singleton instance of SoundManager and modify it.

Game scenes. In each base scene, three objects are disabled at the top of the hierarchy, namely WearableConnectUIPanel_VB, CalibrationDemoUIPanel_VB, and another one whose name relates to the current scene (CirclesCrosses, Mirroring, TargetAround, or Tapping). They have to stay deactivated if the game is started from the root scene. Otherwise, it is possible to directly start the game from each base scene by activating these objects, that allow to directly recognize the glasses.

UNet Logic

Each game works in a similar way. A NetworkManager game object handles the connection/disconnection of each player, according to Unity's UNet system. Each

NetworkManager object has a MyLanNetManager script on it. The base scene and game scene have to be dragged on MyLanNetManager, as well as the Player Prefab and the Registered Spawnable Prefabs. Moreover, the option "Don't Destroy on Load" has to be ticked to allow the object not to be destroyed when loading a new scene, and therefore keeping it where moving from one scene to another (this behaviour can also be activated by using the DontDestroyOnLoad script, located in the Bose asset package).

The objects that are networked need to have a NetworkIdentity script on them in order to be visible by everyone. If a NetworkIdentity is present on an object, the authority of the player over this game object can be checked by using the boolean: hasAuthority. This boolean is often used to trigger events locally on the game object owned by the player who has authority on it. Otherwise (as used in the PlayerConnectionObject), the isLocalPlayer boolean can also be used.

As in every networking system, clients and servers have to communicate together. In UNet:

- Command: functions that are called on a client and run on the server
- RpcClient: functions that are called on the server and run on the clients
- SyncVar: variables that are automatically synchronized from the server to the clients

PlayerConnectionObject

When a user presses Start Host, or joins the game by pressing Join Game, his Player Prefab is spawned in the game. This object is not visible, but has a PlayerConnectionObject on it that allows to instantiate and spawn the visible player unit onto the networked scene. This object is spawned by using NetworkServer.SpawnWithClientAuthority(). This function gives the player local authority over this unit (e.g. the player owns the unit). This function is called inside the CmdSpawnMyUnit() function.

PlayerUnit

The visible prefab that is spawned is named using the name of the scene followed by "PlayerUnit". It always owns a PlayerUnit script also named according to the name of the scene, that allows to:

Name the players automatically Activate the camera, audio listener, RotationMatcherfor the unit that the user owns, and deactivate it for the other players Handle some elements synchronised over the network (counters, variables etc..). In the children of the Player Unit prefab there are always:

One object called "Player" with an audio listener and rotation matcher One camera One narration object, with the script that contains the narration with respect to the scene Other objects that depend on the game narration and are handled by the narration object Non-Player unit networked

With the networking logic, if the second user, who owns player 2, has to update the counter of this player, he uses a Command function to speak to the server, followed by an RpcClient function. Therefore, the player 2 of every user will be updated. But what if one player (server or not) wants to change a value for all the players?

That is the role of non-player networked objects, that can be found in each game scene, under the prefab called StateSyncSpawner. This prefab uses a script called

StateSyncSpawner, that contains some public variables with SyncVar (automatically synchronized from server to clients). These object are mainly used for synchronisation purposes.

The logic is that for instance to start the game, one player hooks the StateSyncSpawner script and updates the boolean called StartStory to true. If it is the host who does it, all clients will see it. Otherwise, the update has to be used in conjunction with a Command function. After that, all users can check the value of this boolean (using GameObject.find()). Since the value is returned is true for all players, then they can all start the game together at the same time.

Game specificities

The particularities of each game (objects, narration, different scripts) are always located on the Player Unit Prefab.

Compatibility versions.

- Bose Frames API: version 13.0
- Google Resonance Audio: version 1.2.1
- Mapbox API: version 2.0.0
- UNet
- Unity version: 2018.3.12f1
- iOS version: 12+

Troubleshooting. To minimize the headtracker latency issues for Bose Frames, try to set all the audio files at 48kHz, or 44.1 kHz. Indeed the Bluetooth LE system sometimes rely on sample rate, and otherwise a sample rate conversion can occur that can be quite time consuming (around 100ms). If the headtracking system stops updating, it may be due to the version of Resonance Audio API used. In our scripts, a headtrackerFix() function is sometimes used to get around this problem (located in the narration scripts). Yet, the problem may have been fixed in newer versions of resonance audio API (see resonance-audio-forum).

Future developments. Future developments of this project focus on: GPS tracking with Mapbox API Audio Interactive Programming with Pure Data and the Heavy Compiler

Credits. This project was developed in collaboration between Queen Mary University of London and BBC Research and Development (Audio Team and IRFS Teams). People involved: Anna Nagele; Valentin Bauer (v.m.bauer@qmul.ac.uk); Thomas Kelly; Chris Baume; Tim Cowlishaw; Henry Cooke; Chris Pike.

A.2 Pre-study questionnaire

1

Pre-study questionnaire

- **Demographics**

1)
Age (you must be aged 18 or older to participate in this study):

2)
How do you describe your gender:
- ❏ Male
- ❏ Female
- ❏ None of the above
- ❏ Prefer not to say

3)
What is your profession and industry, or field of activity?

____
__

- **Some questions about previous experience**

4)
How often, if at all, have you listened to spatial audio (this may include sound VR and 3D videogames)?
- ❏ I don't know what it is
- ❏ Never
- ❏ Less than 5 times
- ❏ A few times a year
- ❏ A few times in the last month
- ❏ Often

5)
How often, if at all, have you listened to binaural audio?
- ❏ I don't know what it is
- ❏ Never
- ❏ Less than 5 times
- ❏ A few times a year
- ❏ A few times in the last month
- ❏ Often

6)
How often, if at all, have you experienced Audio Augmented Reality?
- ❏ I don't know what it is
- ❏ Never
- ❏ Less than 5 times
- ❏ A few times a year
- ❏ A few times in the last month
- ❏ Often

2

6.1)
If yes, what device have you used for this experience (eg. Headphones, VR Headset, AR Headset, Mobile phones, Bose Frames)?

____
—

____
—

6.2)
If yes, can you give examples of the content?

____
—

____
—

____
—

7)
Have you ever visited an immersive theatre play?

- ❏ Yes
- ❏ No
- ❏ Prefer not to say

7.1)
If yes, what was it / were they about?

____
—

____
—

____
—

7.2)
If yes, what types of media did they use?

____
—

____
—

____
—

8)
Have you ever participated in a group performance (eg. theatre, dance, stand-up comedy)?
- ❏ Yes
- ❏ No
- ❏ Prefer not to say

8.1)
If yes, what were they about?

____
—

____
—

____
—

9)
Do you have a smartphone?
- ❏ Yes
- ❏ No
- ❏ Prefer not to say

9.1)
If yes, when do you use it? (You can select as many options as you like)
- ❏ At home
- ❏ At work / school
- ❏ When travelling / commuting
- ❏ When I'm with friends
- ❏ When I'm waiting on something
- ❏ When engaged in other activities (eg. watching TV)

10)
Do you have any social media accounts?
- ❏ Yes
- ❏ No
- ❏ Prefer not to say

10.1)
If yes, what do you use them for?

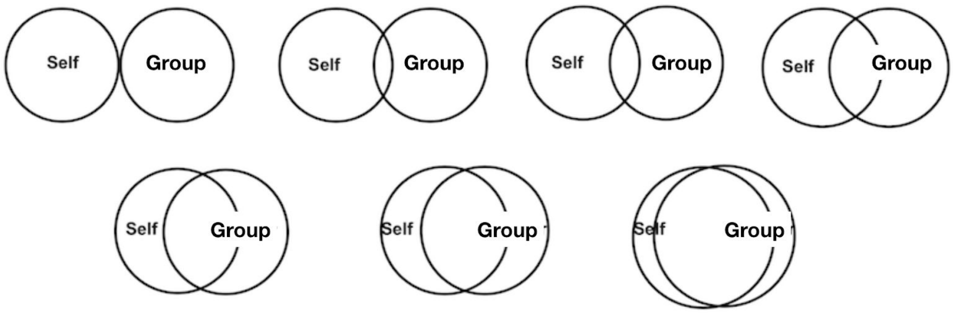
____
—

____
—

4

—

11)
This last question is about how connected you feel to the other participants of this study, who are in this room with you. Which picture best describes your relationship with the participant group?



A.3 Post-task questionnaire

Game 1 - Circles and Crosses

1)
"It was easy for me to execute the instructed movements."

Don't agree at all - Somewhat disagree - Somewhat agree - Totally agree

2)
Complete the following sentence with an adjective:

"I felt __________________ during this task."

Game 2 - Mirroring

3)
"I felt comfortable synchronising my movements with another person."

Don't agree at all - Somewhat disagree - Somewhat agree - Totally agree

3.1)
Why?

____
__

____
__

4)
"Me and my partner had a good time together during that exercise."

Don't agree at all - Somewhat disagree - Somewhat agree - Totally agree

5)
Was there anything that was distracting you from performing the tasks you were given? If so, what?

____
__

____
__

6)
Complete the following sentence with an adjective:

"I felt __________________ during this task."

Game 3 - Notifications

7)
"It was clear to hear which direction the notification sound was coming from."

Don't agree at all - Somewhat disagree - Somewhat agree - Totally agree

8)
"It was easy to find the source of the notification sound by turning my head."

Don't agree at all - Somewhat disagree - Somewhat agree - Totally agree

9)
"I felt satisfied when turning a notification sound off."

Don't agree at all - Somewhat disagree - Somewhat agree - Totally agree

9.1)
Why?

____

____

10)
How stressed were you during this task?

Not at all - not really - somewhat - totally

10.1)
Why?

____

____

Game 4 - Tapping for likes

11)
"I felt comfortable with other participants touching my frames."

Don't agree at all - Somewhat disagree - Somewhat agree - Totally agree

11.1)
Why?
____
__

____
__

12)
"It was exciting for me to receive likes."

Don't agree at all - Somewhat disagree - Somewhat agree - Totally agree

12.1)
Why?
____
__

____
__

13)
Complete the following sentence with an adjective:

"I felt ________________ during this task."

A.4 Post-study questionnaire

Post-Study Questionnaire

1)
"I'm glad I had this experience."

Don't agree at all - Somewhat disagree - Somewhat agree - Totally agree

1.1)
Why?

____
—

2)
"I would like to play more similar multiplayer games."

Don't agree at all - Somewhat disagree - Somewhat agree - Totally agree

3)
"I would recommend this experience to a friend."

Don't agree at all - Somewhat disagree - Somewhat agree - Totally agree

4)
What did you like about this experience?

____
—

____
—

5)
What did you dislike about this experience?

____
—

____
—

6)
Which of the four chapters did you find most engaging?

- ❏ 1) Circles and Crosses
- ❏ 2) Mirroring

- ❏ 3) Notifications
- ❏ 4) Tapping for likes

7)
“I followed all the instructions I was given by the narrator, Pi.”

Don’t agree at all - Somewhat disagree - Somewhat agree - Totally agree

7.1)
Why?

____
__

8)
“I questioned the authority of the narrator, Pi.”

Don’t agree at all - Somewhat disagree - Somewhat agree - Totally agree

8.1)
Why?

____
__

9)
“I felt self conscious during the experience.”

Don’t agree at all - Somewhat disagree - Somewhat agree - Totally agree

9.1)
Why?

____
__

10)
“Somehow I felt that the virtual world surrounded me.”

Don’t agree at all - Somewhat disagree - Somewhat agree - Totally agree

11)
How aware were you of the real world surrounding while navigating in the augmented world? (i.e. sounds, room temperature, other people, etc.)?

Extremely aware - Moderately aware - Not aware at all

11.1)
Why did you choose this answer?

____
__

12)
“I was not aware of my real environment.”

Don’t agree at all - Somewhat disagree - Somewhat agree - Totally agree

13)
“I still paid attention to the real environment.”

Don’t agree at all - Somewhat disagree - Somewhat agree - Totally agree

14)
“I was completely captivated by the augmented world.”

Don’t agree at all - Somewhat disagree - Somewhat agree - Totally agree

15)
How real did the augmented world seem to you?

Completely real - Moderately real - Not real at all

15.1)
Why?

____
__

16)
How much did your experience in the virtual environment seem consistent with your real world experience?

Not consistent - Moderately consistent - Very consistent

16.1)
Why?

____
__

How hurried or rushed was the pace of this experience?

Too slow - Just about right - Too fast paced

17.1)
Why? What would you change in terms of pace?

---

18)
How well did the interactions with the Bose Frames work (double tapping the side, nodding, shaking)?

Not at all - not really - somewhat - totally

18.1)
Why?

---

19)
“The interactions with the frames were well explained.”

Don’t agree at all - Somewhat disagree - Somewhat agree - Totally agree

19.1)
What could have been explained better?

---

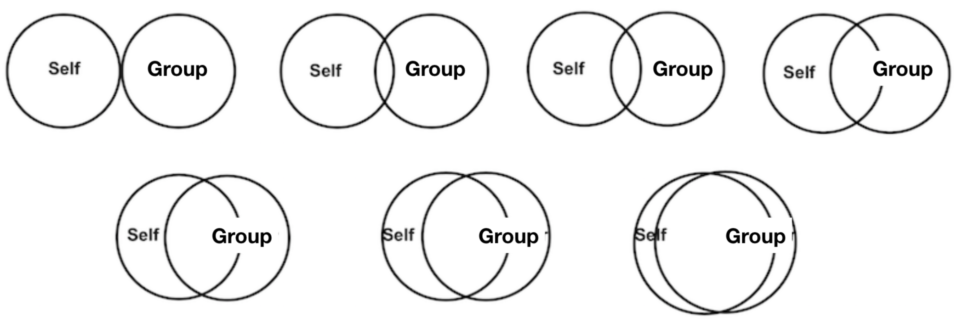
20)
“The interactions with the other participants were well explained.”

Don’t agree at all - Somewhat disagree - Somewhat agree - Totally agree

20.1)
What could have been explained better?

---

21)
How connected you feel to the other participants of this study after the experience? Which picture best describes your relationship with the participant group now?



A.5 EuroVR 2019 poster

Designing an Interactive and Collaborative Experience in Audio Augmented Reality*

Valentin Bauer*[1], Anna Nagele*[1], Chris Baume[2], Tim Cowlishaw[2], Henry Cooke[2], Chris Pike[2], and Patrick G. T. Healey[1]

[1] School of Electronic Engineering and Computer Science, Queen Mary University of London, London, UK v.m.bauer@qmul.ac.uk, a.n.nagele@qmul.ac.uk
[2] BBC Research and Development, London, UK

Abstract. Audio Augmented Reality (AAR) consists of adding spatial audio entities into the real environment. Existing mobile applications and technologies open questions around interactive and collaborative AAR. This paper proposes an experiment to examine how spatial audio can prompt and support actions in interactive AAR experiences; how distinct auditory information influence collaborative tasks and group dynamics; and how gamified AAR can enhance participatory storytelling. We are developing an interactive multiplayer experience in AAR using the Bose Frames audio sunglasses. Four participants at a time will go through a gamified story that attempts to interfere with group dynamics. In this paper we present our AAR platform and collaborative game in terms of experience design, and detail the testing methodology and analysis that we will conduct to answer our research questions.

Keywords: Audio Augmented Reality · Collaboration · 3D Audio · Audio Game · Storytelling · Experience Design.

1 Introduction

In the past few years, the technological development of 3D audio for headphones using binaural audio has facilitated the delivery of Audio Augmented Reality (AAR) experiences. AAR consists of adding spatial audio entities into the real environment [13]. The technology has been applied to a range of fields such as teleconferencing, accessible audio systems, location-based games or education. AAR Research has mainly been focusing on the perception of sound quality [18], realism, or discrimination between real and virtual sounds [13]. Yet, interaction and collaboration remain under-researched. One of the big challenges is acoustic transparency, so that the user can stay connected to his environment as if they had no headphones. Bose Frames (BF) audio sunglasses [5] are a newly available wearable AAR consumer technology that embed the speakers and technology in the frame of the sunglasses, and are therefore perfectly acoustically transparent.

*Supported by the EPSRC and AHRC Centre for Doctoral Training in Media and Arts Technology at Queen Mary University London and BBC RD.
**The first two authors have equally contributed to the paper.

We are developing an interactive AAR multiplayer experience, for four players at a time, that encourages human interactions. Our prototype, *Please Confirm you are not a Robot*, explores three research questions: How can the affordances of the technology and spatial sound prompt and support actions in interactive AAR? How can asymmetric information influence group dynamics and support or distract from collaborative tasks? How can a participatory performance create empathy and behaviour change through interactive storytelling? We will test our game in a user experience study, from which we want to derive design implications for interactive storytelling and multiplayer AAR game design.

2 Background

2.1 State of the Art

In AAR, spatial audio is often rendered over headphones. Issues have been reported regarding front-back inversions, sound timbre artifacts, or externalisation, due to the use of non-individual HRTFs [3]. These are more noticeable in static than in dynamic binaural, which consists of the addition of a headtracking system, and also increases the user immersion and localisation accuracy. A valuable asset of wearable devices is that they can offer headtracking possibilities and thus make dynamic binaural audio more widely available.

Representations of the AAR sound field can be either natural or pseudoacoustic. For the former, virtual audio entities are directly added to the auditory real environment. For the latter, binaural microphones are added into the listener's ears and routed to the earphones so that the listener perceives a synthesized version of his environment. This system, also called "hear-through" audio, is for instance common in hearing aids. In all cases, the aim is that the user should not be able to determine which sources are real and which are not. This requires using high-quality 3D audio rendering [13] and a careful mix between the virtual sources and the auditory environment.

2.2 Challenges

Previous AAR studies have mainly focussed on hear-through audio. Transparent earphones remain under-explored and questions arise about a seamless integration of audio entities onto the real auditory environment. Some open-ear systems, such as the BF system presented in section 4, exist. The mixed reality Microsoft Hololens glasses can render dynamic binaural audio and holographic 3D images, using small loudspeakers, a camera, eye tracking, and headtracking sensors integrated in the frame of the glasses [14]. Bone conduction headsets can also be used to render binaural audio. Despite some localization accuracy issues, good externalization and spatial discrimination can be achieved. [2].

Designing sounds in AAR still requires more investigation, but binaural audio has been shown to increase the user immersion in comparison to stereo audio [7]. Mixing remains a challenge due to the dynamic nature of real sounds that change over time in both level and frequency, which can lead to audio masking.

Most AAR applications remain individual. Yet, some studies have focused on collaboration through location-based AAR games in stereo, using sounds triggered at specific locations [11, 16]. Regarding spatial AAR, Mariette and Katz [17] developed *SoundDelta*, a mobile multi-user AAR architecture which uses mobile user devices and servers communicating over WiFi. They explore the potential of the *Ambisonic cell* approach to deliver personalized audio to a large number of users over a specific area.

In the following sections, we present the development of our AAR game and architecture, and introduce our research methodology and planned studies.

3 AAR Experience Design and Storytelling

Headphones and similar devices divide auditory spaces into private and public. BF, in contrast, do not create a sound barrier but allow individual augmentation of sonic experiences. Lyons et al. [16] suggest that AAR has potential to bring people together in the same location and enhance social interactions. We are designing an environment to foster face-to-face interaction, exploiting three features of AAR: Asymmetric information; Layering augmented sounds over real life sounds; Triggering sounds with head gestures and movement.

A limited amount of applications for Bose AR exist. Some apps allow users to explore a soundscape by selective listening [19], other ones use BF as a gaming device with taps and head movements as interactions [1], or make use of the technology's mobility through soundwalks [8]. Dead Drop Desperado [10] is the only known game that requires two players.

Apart from those Bose AR applications, spatial audio is used in immersive theatre to create imaginary spaces and parallel realities [9]. AAR experiences often assign a role to the user, asking them to perform. Looking at this in a multiplayer context, this is reminiscent of choreography and theatre performance. The theatre practice developed by theatre maker Augusto Boal [4] blurs the boundaries between everyday activities and performance. It is used to rehearse for desired social change [15]. Inspired by this practice, our multiplayer game will result in a choreography prompting users to observe, reenact and subvert behaviours around digital devices.

3.1 Game Overview

Please Confirm you are not a Robot is a speculative fiction, constructed of four individual games. At the start, each participant meets their guide who introduces the scenario and the gesture controls: tapping, nodding and shaking head. In the first game participants are prompted to simultaneously draw a circle with one arm, and a cross with the other arm, in the air. Spatialised sounds of drawing a circle and cross will play for some players alongside the movement. We will look at whether sound cues have any effect on the participant's performance.

For the second game participants pair up and mirror each other's movements while being prompted to ask each other questions. We will look at whether this contributes to interpersonal closeness or affect between participants, or whether different layers of sound are distracting.

The third game uses the BF as a gaming interface. A variety of notification sounds will appear in the sonic sphere around each participant. To turn them off they have to look at the sound and double-tap the side of the frames. Participants collect points for each sound they turn off. We will test different feedback sounds for finding sounds in space.

The last game requires the participants to tap each other's frames, following prompts of what they like about each other, to collect points. We will look at the interaction between participants. At the end, one participant will be separated from the group with a separate story-line. They will become the agent to end the whole experience by taking the other player's frames off their face.

3.2 Game Design

Early designs of interactive audio-only experiences highlighted the importance of sound design [16]. Sounds have to be put in context with other sounds or narration to establish a cause and effect relationship between the actions and sounds, and match the player's mental model [16]. Since varying loudness levels are a challenge in AAR, we decided on a specific room where we will conduct the experiments. We created the sound design with sounds gathered from personal recordings or Freesound.org (under the Creative Commons Licence), and using the software SoundParticles in combination with Reaper.

4 Audio Augmented Reality Architecture

In our modular AAR platform, we use BF because of their acoustic transparency, headtracking system, user interaction and ergonomics. In addition, BF have a Bluetooth low energy system and offer three interactions: nodding, shaking the head, and tapping the glasses. The headtracking system has an accelerometer, gyroscope and magnetometer, and a latency of around 200ms (higher than the 60ms optimal latency [6]). This may affect audio localisation but the other BF aspects make it suited to achieve a good user engagement in the game. Since our system had to be modular and support 3D audio, GPS tracking, BF API, and multiplayer possibilities, we chose to work with Unity software (version 2018.3 for compatibility). We work with phones on iOS due to the compatibility with BF SDK, but future developments may also include Android phones.

For multiplayer collaboration we designed a Local Area Network (LAN) over WiFi using the Unity's *UNet* system. The first player who connects to the game is the host and starts broadcasting its IP address. Following players (clients) automatically detect it and join the game. Some objects are synchronised over the network, such as player dependent objects, or objects that keep track of global variables. Asynchronised objects can also trigger events locally for each player. With this system events can be player specific, and different players can listen to different sounds synchronised over time. The BF API gives us access to the sensor data of BF. We used Google Resonance Audio SDK for 3D audio rendering because of its high-quality with 3rd-order ambisonics [12]. The architecture supports GPS tracking with Mapbox API, Audio Interactive Programming with Pure Data, and gives access to the phone's affordances (sensors, vibrator).

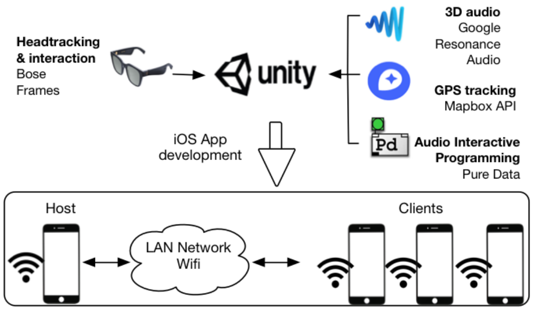


Fig. 1. Architecture of the AAR game

5 Testing methodology

We conducted preliminary testing with a group of four participants from BBC R&D to detect technical and narrative flaws that helped us to refine our prototype. We will soon conduct a user study with five groups of four participants with different levels of expertise in 3D audio and augmented reality. A pre-study questionnaire will assess previous experience with 3D audio as well as interpersonal relationships of the group. One researcher will be with each user during the experience and take notes about their behaviours. After each game, users will be asked to answer specific questions about the game. A post-study questionnaire and guided group discussion will assess aspects of the game such as enjoyment, interactive ease, problems, storytelling and feelings about the group. Experiments and discussion will both be filmed and recorded. We will conduct qualitative analysis of the recordings and participant responses, in comparison with the performance measures we set out for each game. From this analysis we will attempt to answer our research questions, derive design recommendations for AAR multiplayer games and give an indication of areas for further research.

6 Conclusion

We reviewed previous AAR studies and discovered that newly available technologies such as BF, which do not cover the ears, offer new opportunities for collaboration and interaction in AAR. We were inspired by previous multiplayer experiences, methods of interactive storytelling, and theatre practices to develop *Please Confirm you are not a Robot*. This game immerses a group of four players into a scene where they play and act out several scenes, guided by asymmetric information and binaural sound cues. This paper details the development of the modular AAR architecture that supports our experimental game, and can be extended in the future to create other multiplayer games. This is one of the first studies to our knowledge that evaluates BF AAR experiences.

References

1. Audiojack Homepage, https://www.audiojack.com. Last accessed 3 Jul 2019
2. Barde, A., Helton, W. S., Lee, G., Billinghurst, M.: Binaural spatialization over a bone conduction headset: Minimum discernable angular diference. In Proceedings of the 140th Convention of the Audio Engineering Society (2016)
3. D.-R. Begault: 3-D sound for virtual reality and multimedia. Academic Press Professional, San Diego, CA, USA (1994)
4. Boal, A.: Games for Actors and Non-Actors. Second Edition. Taylor and Francis e-Library (2005)
5. BOSE Frames Homepage, https://www.bose.co.uk/engb/products/frames.html. Last accessed 3 Jul 2019
6. Brungart, D.S., Kordik, A.J., Simpson, B.D.: Effects of Headtracker Latency in Virtual Audio Displays. In: Journal of the Audio Engineering Society, 54(1):13 (2006)
7. Chatzidimitris, T., Gavalas, D., Michael, D: SoundPacman: Audio Augmented Reality in Location-Based Games. In: 18th Mediterranean Electrotechnical Conference (MELECON), pp.1-6, Lemesos, Cyprus: (2016)
8. Cooke, H.: Audio AR: Sound Walk Research The Missing Voice (2019) https://www.bbc.co.uk/rd/blog/2019-03-audio-sound-walks-tours. Last accessed 3 Jul 2019
9. Darkfield Homepage, https://www.darkfield.org. Last accessed 3 Jul 2019
10. Dead Drop Desperados App Store Preview, https://apps.apple.com/us/app/dead-drop-desperado/id1454393037. Last accessed 3 Jul 2019
11. Ekman, I.: Sound-based Gaming for Sighted Audiences Experiences from a Mobile Multiplayer Location Aware Game. In: Proceedings on 2nd AudioMostly, Germany (2007)
12. Gorzel, M., Allen, A., Kelly, I., Kammerl, J., Gungormusler, A., Yeh, H.: Efcient Encoding and Decoding of Binaural Sound with Resonance Audio. In: Audio Eng. Soc. Convention 12 (2019)
13. Härmä, A., Jakka, J., Tikander, M., Karjalainen, M., Lokki, T., Hiipakka, J., Lorho, G.: Augmented reality audio for mobile and wearable appliances. In: Journal of the Audio Engineering Society, vol. 52, n. 6 (2004)
14. Hololens Homepage, https://www.microsoft.com/en-us/hololens/. Last accessed 28 Jul 2019
15. Kochhar-Lindgren, K.: Towards A Communal Body Of Art: The Exquisite Corpse and Augusto Boal's Theatre. In: Angelaki: Journal of Theoretical Humanities, 7:1, pp. 217-226 (2002) https://doi.org/DOI: 10.1080/09697250220142137
16. Lyons, K., Gandy, M., Starner, T.: Guided by Voices: An Audio Augmented Reality System. In: Online Proceedings of the International Conference on Auditory Display, Georgia Institute of Technology Atlanta, Georgia, USA (2000)
17. Mariette, N., Katz, B.: SoundDelta - Largescale, multiuser audio augmented reality. In: EAA Symp. on Auralization, Espoo, Finland (2009)
18. Tikander, M.: Usability issues in listening to natural sounds with an augmented reality headset. In: Journal of the Audio Engineering Society, vol. 57, no. 6 (2009)
19. Traverse Homepage, https://www.traverse.fm. Last accessed 3 Jul 2019