Creating an Audio Story with Interactive Binaural Rendering in Virtual Reality

Geronazzo, Michele; Leopold Hyldig Rosenkvist, Amalie ; Sebastian Eriksen, David ; Markmann-Hansen, Camilla Kirstine; Køhlert, Jeppe ; Valimaa, Miicha ; Brogaard Vittrup, Mikkel ; Serafin, Stefania

Published in:
Wireless Communications and Mobile Computing (Print)

DOI (link to publication from Publisher):
10.1155/2019/1463204

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

Hindawi
Wireless Communications and Mobile Computing
Volume 2019, Article ID 1463204, 14 pages
https://doi.org/10.1155/2019/1463204

Research Article

Creating an Audio Story with Interactive Binaural Rendering in Virtual Reality

Michele Geronazzo, Amalie Rosenkvist, David Sebastian Eriksen, Camilla Kirstine Markmann-Hansen, Jeppe Køhlert, Miicha Valimaa, Mikkel Brogaard Vittrup, and Stefania Serafin

Department of Architecture, Design, and Media Technology, Aalborg University, Copenhagen 2450, Denmark

Correspondence should be addressed to Michele Geronazzo; mge@create.aau.dk

Received 4 January 2019; Revised 21 June 2019; Accepted 4 July 2019; Published 14 November 2019

Academic Editor: Marco Picone

The process of listening to an audiobook is usually a rather passive act that does not require an active interaction. If spatial interaction is incorporated into a storytelling scenario, can open. Possibilities of a novel experience which allows an active participation might affect the user-experience. The aim of this paper is to create a portable prototype system based on an embedded hardware platform, allowing listeners to get immersed in an interactive audio storytelling experience enhanced by dynamic binaural audio rendering. For the evaluation of the experience, a short story based on the horror narrative of Stephen King's Strawberry Springs is adapted and designed in virtual environments. A comparison among three different listening experiences, namely, (i) monophonic (traditional audio story), (ii) static binaural rendering (state-of-the-art audio story), and (iii) our prototype, is conducted. We discuss the quality of the experience based on usability testing, physiological data, emotional assessments, and questionnaires for immersion and spatial presence. Results identify a clear trend for an increase in immersion with our prototype compared to traditional audiobooks, showing also an emphasis on story-specific emotions, i.e., terror and fear.

1. Introduction

Since 2015, the sale of audiobooks in the United States has expanded by almost 20 percent each year. According to a 2017 survey [1], 24% of Americans (more than 67 million people) have completed at least one audiobook in the last year (2016 data), which resulted in a 22% increase over the previous year (2015 data). There is a promising growth in the field of audiobook storytelling, which calls for a reflection upon the medium and perhaps a look for potential improvements and alterations to be made. A leading distributor and producer in the audiobook field is the Amazon-owned company Audible, which sells and produces spoken audio entertainment, information, and educational programs. However, in all their titles (more than 10000) Audible produces a year, there has been no significant development to the method of storytelling to improve the listening experience. In these traditional audiobooks, there is nothing more than a narrator or actor reading a story aloud, resulting in a passive experience without nonlinear narratives [2]. Alternatively, the state-of-the art audio drama created by BBC makes use of static binaural rendering and mixing (listen, for instance, to radio dramas of BBC Radio 3 (http://www.bbc.co.uk/programmes/articles/29L27gMX0x5YZxkSbHchstD/radio-3-in-binaural-sound).

Our main focus is the development of a technological support for audio stories which could benefit from the increasing attention and development of innovative tools in immersive virtual reality (VR). Such a framework could be extended to podcasts and radio broadcasting in general considering mobile VR, e.g., smartphone-required devices such as Google Daydream (http://vr.google.com/daydream/) or Samsung Gear VR, and all-in-one devices such as Oculus Go and Quest (http://www.oculus.com/).

One relevant aspect able to increase the pleasantness and usefulness of an audio story experience is the level of

immersion. According to Nordahl's adaptation of Mel Slater's conceptual framework for describing why individuals may respond realistically when exposed to immersive VR with sound [3], immersion is one of the main four constituents. There are many methods of improving immersion of an audio-only story. Technologies for three-dimensional sound rendering are able to create surrounding sounds, i.e., immersive soundscapes, and interactivity within virtual environments (VE) [4, 5]. Accordingly, the aim of this study is to create a prototype for interactive audio-only stories able to dynamically render spatialised soundscapes with headphones equipped with embedded movement sensors. We manipulated monoaural audio sources in accordance with the object-based audio definition of VE [6], allowing a flexible interaction design and mixing of a narrative in a full three-dimensional space around the listener. Sound propagation and occlusion in VR were kept consistent from the acoustic point of view, providing a rich and natural audio experience. Dynamic spatial audio rendering required head-tracking, allowing listeners to freely move their head localizing and exploring sound sources like in real life [7]. Two micro electromechanical system (MEMS) MPU-9250 sensors able to measure both acceleration and orientation served as the basis for the interaction part of the prototype. After various redesigns based on results from several usability tests, a movement tracker was mounted on noise-canceling headphones. Additionally, we designed a handheld controller for a basic mixing action, i.e., volume balance control between main narrator voice and auditory VE elements.

The proof-of-concept narrative was based on the short story by Stephen King called "Strawberry Springs" and was both shortened and edited in order to be considered in an scientific evaluation procedure. The story's narrator was then recorded in an anechoic room and played back monophonically, while spatialised sound sources were placed in the soundscape. The progression of the story can be described as a user moving through the storyboard on a predefined path allowing head-movements and volume adjustments. The scene was built in accordance with redesigned storyboard allowing the users to move from the room where the scene was happening, along corridors and scene connections, while triggering sound events or placing sound sources around them (see the Supplementary Material for more details about the storyboard (available here)). A similar approach was successfully adopted in [8] with the idea of *music rooms* for enhancing music genre learning with minimal visual feedback. A first evaluation was conducted comparing our final prototype with two other experiences: a passive monophonic listening (traditional audiobooks) and a static binaural listening (state-of-the art audio storytelling).

2. Background

2.1. Interactive Storytelling. Storytelling can be described as "the activity of telling or writing stories" (from Oxford Learner's Dictionary, http://www.oxfordlearnersdictionaries.com/definition/english/storytelling). It is a social and cultural act, which can include theatrics, improvisation, and the likes. As mentioned by [9], storytelling has been around for a long time and serves not only to recall past events but also to spread awe by way of fantasy, fiction, and "magic." While modern civilization continues the age-old tradition of telling stories both orally and visually, it also finds more innovative ways of telling them.

Interactive storytelling is a fairly new field from the late 1980s and 1990s, related to many different fields, like games, cinema, storytelling, programming, and mathematics [10]. When combining storytelling and interactivity in order to achieve interactive storytelling, it is important to maintain balance between the amount of control the user has over the story and the coherency of the story [11]. When considering what is required to create an interactive experience, the design of the interaction and the feedback are tightly connected [12]. Users need the ability to influence the feedback they receive before a product can be considered interactive. Influence and feedback can be achieved in several different ways. Before the digitalisation of stories, some of the ways of creating interactive stories was during public readings [13]. The reader or *teller* would actively engage the audience to prompt a response, which would either alter or further immerse the audience in the story based on the response. One can also describe this process as forward leaning or participatory storytelling, as opposed to conventional backward leaning or hypnotic storytelling achieved by a noninteractive experience.

2.2. Binaural Audio Technologies. Understanding how humans process everyday sounds is extremely relevant when designing an audio system. Simultaneous sound events in the environment can be identified, and it is possible to focus on a specific sound. The auditory system can detect the origin of a sound, elaborating its direction-of-arrival (DOA) through the head-related transfer function (HRTF), and changes in interaural level and time differences. In particular, HRTFs describe the acoustic characterization of the human body for point source around the listener, being highly individual especially for vertical sound localization [14, 15] Interestingly, sounds that are memorized through repetition are more easily identified by humans [16].

Binaural techniques usually refer to methods for recording and reproducing sound with the intent to construct an immersive auditory sensation. A common method for this is the use of so-called dummy head recording, which involves two microphones placed at blocked ear canal position of artificial head's ears, with their two isolated outputs being played to a pair of headphones worn by the listener [17]. Since the dummy's ears are built to resemble a real human ear, the sound waves are modified during the recording process, and approximate how the sound waves would be altered in a real-life scenario before reaching listener eardrums. Even though binaural recordings are not widespread within the music industry, they are being used for ambient experimental music and sound stories [18]. Furthermore, more flexible rendering techniques for sound propagation are being developed, providing sophisticated sound engine software especially for games, such as Google Resonance (https://github.com/resonance-audio), Steam

Audio (https://valvesoftware.github.io/steam-audio/), and Wwise (https://www.audiokinetic.com/products/wwise/), to name but a few.

3. Related Work

3.1. Spatial Orientation with Sound. Head and body movements are important for building cognitive maps of the real/virtual space around the listener, especially for visually impaired people [19]. Binaural audio technologies play an important role in conveying relevant spatial information via headphones in order to acquire the integrated knowledge for spatial orientation and navigation [20]. Spatial orientation also refers to how it is possible to track a user's location and orientation for rendering purposes. One can identify three common methods to manipulate user orientation in VEs: head-, body-, and device-tracking. In [21], authors tried to find differences among tracking methods when the user was moving toward a sound source in an audio-augmented reality scenario. Their results showed that there were not any statistically significance differences between the three tracking methods in terms of localization performances. On the other hand, if time to accomplish the navigation task and realism were considered, head-tracking would be the best solution. Regarding technical requirements for a head-tracking implementation, Hess [22] argued that the latency from head movement to virtual feedback can be maximum 62 ms.

Finally, it is worthwhile to notice that virtual room acoustics is important for a natural perception of reconstructed sound scenes, providing a recognizable acoustic fingerprint of specific location or event [23]. This is particularly relevant for echolocation abilities, which rely on DOA of echos in the room which could be effectively rendered with current VR technologies [24], such as those employed in this study (see Section 2.2 on binaural audio technologies). Moreover, thanks to the circular interaction between spatial presence and emotions: one can consider VR an affective medium [25] which is able to interact with user's affective states [26] and memory processes [27].

3.2. Interactive Audio Stories. In the scientific literature, there are many attempts in introducing interactivity in the passive listening of audio stories. Furini [2] developed a system architecture able to turn listener into the story director. The author proposed a script manager able to support a producer in the definition of atomic audio scenes/segments with meta data in MPEG7-DDL language and their allowed connections, i.e., a story graph. Listeners used the interaction manager in order to look at the scene transition table for possibilities in the story path. The user interface could be implemented within a touch screen or a voice detection system. In [28], Huber et al. focused on the user interface design, story sonification, and game interaction of nonlinear narrations. They defined interactive and narrative *nodes* in the story, with the opportunity to create nodes with mini-games and static 3D audio objects based on OpenAL rendering capabilities (http://www.openal.org). More recently, Marchetti and Valente [29] explored the untapped potential of audio in the context of a foreign language learning in primary and secondary schools. They were developing a prototype of mobile application which offered a multimodal experience in extending reading with social interaction: a platform for annotating, tagging, and sharing comments from written book to the correspondent audiobook.

It is worthwhile to notice that music also plays an important role to elicit the proper emotional response during a story. For example, automatically generated music scores might increase the overall quality with respect to audio stories without music [30].

Finally, a wider prospective of the use of audio narration could be in the automatic creation of daily stories based on user data from ubiquitous and wearable technologies (e.g., mobile devices/sensors) [31]. Such data presentation might help users in navigating effectively the increasingly amount of personal information in order to gain self-awareness or produce a desirable behavioral change.

3.2.1. Commercial Applications. Koob virtual reality audiobooks (https://www.koobaudio.com/) was created by the Voice Society and is a digital platform. KOOB wanted to revolutionize storytelling by using VR sound to create life-like immersive experiences in order to drag the listeners mind deep into the story, making them feel like they are living the story themselves. This was done by tricking the listeners mind into thinking that it was experiencing the binaural sound with intuition, creativity, and imagination, understanding, and reasoning.

The *Owl Field* (https://www.owlfield.com/) is a company which creates immersive binaural audio dramas. Their audiobooks are told from a first-person perspective where the events, characters, sound effects, and music surround the user. An interesting aspect to The Owl Field's storytelling is that the listener is the story's central character. The characters of the story speaks to the listener, and involves them in the story. The use of binaural audio recordings is what sets The Owl Field aside from common narrated audiobooks. Additionally, making the user an integrated part of the story gives potential for high immersiveness.

Hellblade: Senua's Sacrifice is a dark action-adventure game developed and published by the studio Ninja Theory in August 2017. The game was set in a fantasy world build on Norse and Celtic mythology, and followed the young warrior, Senua, and her journey to Hel, where she hoped to save the soul of her murdered lover, Dillion (read https://www.giantbomb.com/reviews/hellblade-senuas-sacrifice-review/1900-765/ for a review). The caveat was that Senua suffered from psychosis, meaning that she suffered from hallucinations, and as such, the world you visit was a manifestation of her mind. Furthermore, it resulted in schizophrenia and auditory hallucinations, i.e., Senua hearing voices. The general consensus between games journalists and reviewers was that the game's sound design based on binaural recordings was able to build great immersion by replicating aspects of psychosis and was a great narrative accomplishment.

4. The Interaction Design in Our Narrative

To make the user feel present in the scene, one could argue that the aspect of immersion related to spatial presence can be pursued by implementing a spatialised soundscape where the digital sound sources respond to physical movement like they would in real life. This can be achieved by implementing a dynamic audio environment played through a head-tracking pair of stereo headphones. In relation to implementing a more immersive audio experience, spatial sounds can be designed to enhance a sense of presence like in many video games [32]. Although real-time audio-processing can be computationally expensive, GPU acceleration and the fact that modern computers have become very fast helps reduce this problem [33]. Additionally, since such a kind of system does not include any graphics in the prototype, practically all computing power will be available for audio-processing.

4.1. The Narrative. Stephen King's Strawberry Spring story was written in 1978; however, it is mostly set in a flashback in 1968. In the following, we give a brief summary of the story: it revolves around a springtime murder spree that happened at the narrator's college in 1968. The weather is referred to a phenomenon known as "strawberry spring" in the story, due to the heavy fog. The narrator seems less affected than others around him and acquires a fascination for the murderer. Ten years later, in 1978, after the murders have been forgotten, the fog reappears and the narrator fears he has committed the same crimes as many years earlier.

The narrative was arbitrarily shortened and adapted for this interactive audio medium and its evaluation, trying to identify peculiar aspects of such design process. Furthermore, some words were changed in order to better adapt to the current time period, rather than the 1960s.

This specific story was chosen, based on the genre, duration, and first-person narration. In particular, this latter aspect allows the characters to directly interact with the user. Furthermore, this peculiar aspect demands the user to be active in listening, and cannot passively walk away from the narration. Spatial sound in the soundscape, such as cafeteria background sound and quiet conversations, and sound effects, such as wolf howls and crickets chirping, were actively placed around the users while they proceed in the time-line of the story. Moreover, the story segments/events and theirs audio objects were created within virtual rooms with virtual walls, corridors, and open spaces to take advantage of spacial interaction in room acoustics, e.g., sound occlusions, as design parameter (see Section 4.2 for further details). Figure 1(a) depicts the top view of the implemented story map made of such virtual elements connected in order to create the desired storyline. The adopted narrative flow was arbitrarily designed keeping in mind a feasible and short experimental evaluation for the listener (see Supplementary Material for time-stamps and durations of narration segments in the storyboard allowing a tangible time-line for testing and reproducibility).

4.2. Sonic Interactions. The listener was forced to move along a predefined track in the audio-only VE, which spatially described the narrative. The track went through several virtual rooms and corridors that acoustically resembled those described in the narrative, as this provided a realistic representation for the spatial sounds. For such purposes, the audio engine provided a set of materials which have their own unique acoustic properties, e.g., occlusion, sound absorption, and reflectance. Moreover, each recording was numbered and placed in chronological order, and each story segment was placed to take the user through the narrative in the correct order. The track system was the primary method for forcing the navigation into the story at specific movement speed, allowing the increase or decrease of speed with which the user would visit each segment/place, and thereby the speed of the story.

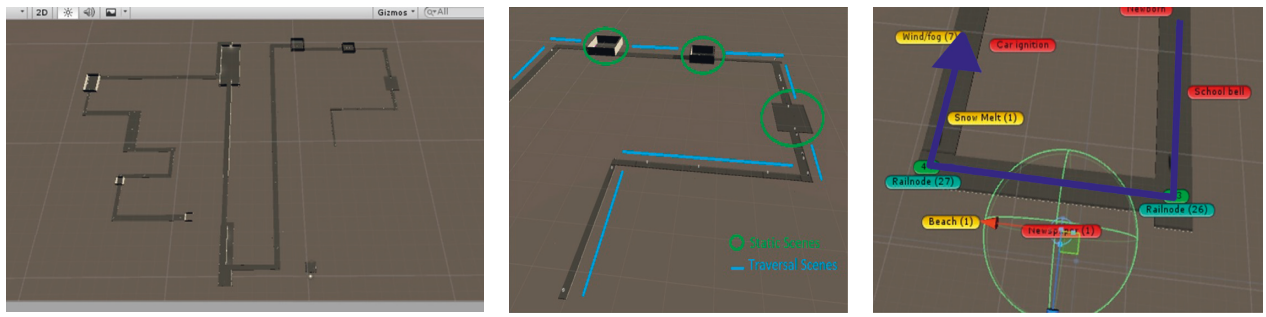
The map was built in accordance with the storyboard. Though the user did not see the visual counterpart, the story was spatially mapped visually for faster editing of source placement and organization by the experimenters.

4.2.1. Static and Traversal Scenes. The segments in the storyboard were given a "Static" or "Traversal" tag depending on whether the narrative prompted a virtual scene in a closed space/room or a scene with movements (see Figure 1(b) for an example of such distinction). The changes in speed most often occurred when users went through a change between a static and a traversal scene.

Enclosed virtual rooms correspond to static scenes, in which the user's position (not orientation) is fixed. This is meant to give a sense of being immersed in the corresponding soundscape. Static scenes were prompted by narrative phrases and conversations, with a low-speed position change. The traversal scenes were spatially described by long winding paths, through which the user would traverse through the corresponding soundscape. Accordingly, traversal scenes were commonly associated with high-speed position change.

4.2.2. Soundscape. As the user passed through the virtual scene, sounds were triggered around them. The sound samples are grouped into four categories: narrative (48 samples), ambience (25 samples), movement (6 samples), and other sounds (40 samples), some of which also were considered *reaction sounds* for special events in the story (see Supplementary Material where each category has their own colored tag for a quick overview of the complexity of the resulting short story).

The sounds were played by adding a trigger collider to the source and a box collider to the listener position in the story. Figure 1(c) depicts an example where the green sphere is the trigger radius or area of a sound source. The narration clips were separate to the ambient clips and were also triggered by collisions. The difference is that the sound source playing the narration clip was attached to the user and located slightly behind the audio-listener object. This created the effect of the sound being played behind the user. An example of such interaction could be extrapolated from



(a) (b) (c)

FIGURE 1: (a) Visual mapping of the story in VR; (b) mapping of static and traversal scenes; (c) a section of the story where the user path (blue arrow) triggers *ambient sound* (yellow labels), *narrative* (green labels), and *sample* audio (red labels).

Figure 1(c) where the traversal scenes were prompted by narrative phrases such as "Then, today's paper" and the sound of turning pages was triggered in the interactive story.

4.2.3. Additional Controls. A study by McGloin, Farrar, and Kramer aimed to investigate whether or not the controller matters when it comes to immersion in video games [34]. Their research suggests that controllers are in fact capable of influencing the perceived realism. They present the concept of controller naturalness, which refers to the overall intuitiveness a controller is perceived to have when interacting with a VE [35]. Accordingly, a remote controller was initially designed in order to provide the following basic functions:

(i) Pausing the narrative, allowing the user to explore the soundscapes

(ii) Adjusting the relative volume of the soundscape with respect to the narrative

This latter functionality should follow the gesture of turning a knob while adjusting the volume on a speaker: turning it left, or counterclockwise will turn down the volume, and turning it right, or clockwise, will turn it up. It is worthwhile to note that the pausing action was implemented but not considered in the final prototype according to a first usability test briefly described in Section 5.2.

5. Prototype Implementation

The creation of the story and its VE was based on Unity framework (https://unity3d.com/) which is an extensive game/application development kit. We used the capabilities and parameters of the Steam Audio engine (https://valvesoftware.github.io/steam-audio/), which enables a real-time rendering of realistic VE with specific audio parameters, such as audio occlusion, reflection, and propagation. Our system employed default Steam Audio rendering parameters (e.g., generic HRTFs). The narrator's voice was recorded in the anechoic room of the Multisensory Experience Lab (https://melcph.create.aau.dk/) at Aalborg University Copenhagen, with a Zoom H6 sound recorder. Such recordings composed a monophonic track resulted in *in-head* localization of the voice for the users, which guided them through the story without changing position in space. Sound effects and soundscape were selected within various collaborative databases of creative-commons licensed sounds, e.g., Freesound (https://freesound.org/).

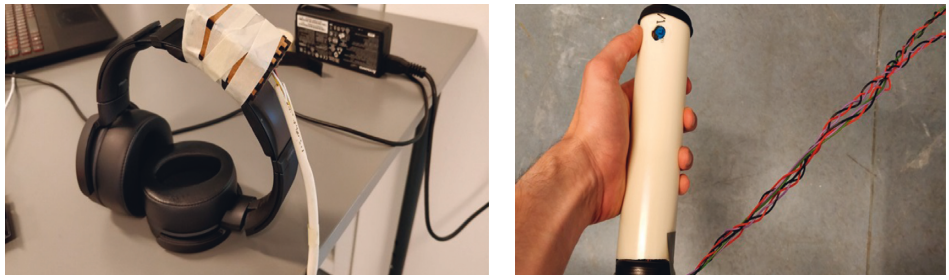
Figure 2 captures the two main components of the prototype. In Figure 2(a), one can see the head tracker that was secured in a small wooden case. In Figure 2(b), a shortened white PVC pipe contained the movement sensor and the button. Arduino Nano 2.3 was employed for a fast prototype of the devices. Specifically, two 9-axis gyroscope MPU-9250 sensors were connected to the Arduino board, which interacted with the Arduino through a I2C bus. The Arduino had both SDA and SCL pins and the platform came with built-in support functions, which made easy to implement the I2C protocol. Only the outputted angular-velocity values (Gy^X, Gy^Y, Gy^Z) were used in the computation of user's movements. Arduino was further connected to the computer by interfacing the Arduino board with a USB cable. Data were acquired in the Arduino IDE and sent to Unity through a serial port.

Finally, a custom range (−20 dB and +15 dB full scale) was set for the volume control of the soundscape in order to isolate the narrator's voice, and the lower bound effectively silences the immersive soundscape. The button might pause the linear movement in the narrative; while rotating the device horizontally, the user can adjust the volume of the ambient noises.

5.1. Remote Controller Connection. The bitrate for the serial communication between Arduino and Unity was set to 9.6 kbit/s. In the Unity environment, after instantiating and opening the serial port, the data logger was started. If an exception occurred during initialization, the script was disabled. This ensures that the program can still run without a connected Arduino, which is useful for development purposes.

An *update* method was called every frame, allowing the reading of sensors data from the Arduino. The processing steps could be summarized as follows:

(1) Reading incoming sensor data and assigning the corresponding variable



(a) (b)

Figure 2: Main components of our prototype: (a) noise-canceling headphones equipped with head tracker; (b) gestural handheld controller (remote).

(2) Parsing the raw string data to a floating-point value that can be used in Unity; a normalization factor was also applied before sending the value

(3) Comparing the value with a noise-threshold, thus not considering very small sensor responses

(4) Applying a gain to the variable controlling the volume

(5) Parsing the button state, i.e., its Boolean value

(6) Sending a pause signal to user movement in the Unity VE if the button press is true

(7) En-queuing the button state and volume for data-logging in a separate thread

The volume value was then applied to the audio-mixer of spatial sounds in the soundscape.

5.2. Usability Test and Final Prototype. The test was conducted in an ad hoc Unity scene where the user moved forward automatically, while being exposed to spatialised sound from the surroundings.

Five users were asked to perform three actions in three different navigation trials:

(1) Moving their head using the head-tracked headphones

(2) Moving the remote controller horizontally

(3) Using the button on the remote

After each trial, users were asked to explain what they thought happened with the specific action.

Table 1 showed that users interacted with the headset intuitively and in some extent also with the remote. However, some issues with the button were identified. Two out of five participants understood the concept of the button and described it as the sound in the surroundings being paused momentarily. The other three participants, when questioned about the button on the remote, replied with either confusion of that they felt no effect of clicking it, which might be because there was no immediate feedback when pressing it. When asked about the effect of the horizontal movement of the remote, almost all test participants formulated that it controlled the sound volume.

It could be concluded that the head tracker and remote were both working efficiently enough to go into final testing. The button still lacked a reliable functionality and had to be redesigned again or removed entirely. As a consequence, the button functionality was remapped for activating the volume control through gesture: when the button was held down, the user was able to direct the remote in either direction horizontally and change the volume for the spatial sounds. Left direction will turn down the volume, and the right direction will turn it up.

6. Evaluation

The primary goal of this study is to explore if interactions with auditory VR contents create a more immersive experience for the user in audiobooks. This means that the prototype has two distinct features: binaural audio rendering and sonic interactions. These two components should be tested separately in order to identify to which extent they impact the feeling of immersion, thus leading to three experimental conditions in a between-subject experimental design:

(1) Monophonic playback without interactions—common technology in audiobook (MA hereafter)

(2) Binaural audio stereo playback without interactions—state-of-the-art solutions in audio drama/book (BA hereafter)

(3) Binaural audio rendering with interactions—our proposed solution (IVE hereafter)

The experimental conditions will be tested on separate sample populations in order to maintain the novelty of experiencing the audio narrative for all participants.

However, this approach introduces a couple of issues. Firstly, there could be individual differences in the participants in each sample group in terms of personality, temperament, etc. To counter this issue, samples of similar age and ethnographic background should be used in the test and the participants were randomly assigned to an experimental condition. Thirty participants (ages M:22 SD:5) took part at our evaluation stage, all were student volunteers, self-reporting, normal hearing, and with no past experience on binaural audio.

6.1. Protocol. The evaluation was conducted in three sessions, one for each of the conditions, with 10 different

Table 1: Qualitative data from the usability testing.

Sub ID	Headset	Remote	Button
1	It switches sounds but in an unknown way. It rotates your head around in the sound space	Intensity/volume of the sound	No understanding.
2	The sound moves with the direction that I am looking. It felt like I was able to explore different scenarios. Felt natural when looking around	Adjustment of intensity/focus of the sound	Seems like it saves the sound when you click it. Heard the sound for a longer period of time.
3	When I turn my head, it is like the sound moves with me. Sound becomes more clear depending on the direction. Much like in real life	Move left the sound becomes low and muffled. Move right sound becomes louder and clearer.	I did not feel much happen.
4	It is like normal audio navigation. Able to locate sound source according to head movement	Quite intuitive volume control.	A single sound continued playing. Like the other sounds stopped moving past me.
5	Sounds are passing by and I can follow them with my headset. Feel like naturally moving through sound	Did something to the intensity of the sound but not sure how it works.	I felt no difference.

participants per condition leading to a between-subject experimental design. The experimental procedure was conducted in accordance with the Declaration of Helsinki (Edition 2013).

External uncontrolled stimuli were removed as much as possible. Test participants were placed in a dark room, ensuring that they were not influenced by external sources of light or sound that could confuse them or cause a loss of attention.

The participants were asked to sit down right away, in order to get their heart rate lowered. The participants were then introduced to the project and told about the test they were about to participate. For all of the conditions, the participants had to wear the E4 wristband from Empatica throughout the test, in order to track their heart rate. The wristband was placed on the participants' left wrist.

In conditions without interactions (MA), the participants were simply asked to wear the headphones and listen to the story. For the condition with interactions (IVE), the participants were asked to wear the headphones with the head tracker attached and hold the remote object in their dominant hand, after which they were told about the three kinds of interactions which were possible while listening to the story.

6.2. Metrics. The purpose of the proposed evaluation is to measure state immersion in a fear-inducing experience.

Immersion is a difficult term to describe. It can be divided in two separate terms; flow and spatial presence [36]. Flow is the psychological state of absorption and extreme concentration on a task at hand [37], whereas spatial presence refers to feeling physically present in a mediated environment. Measuring immersion is a challenge. Most studies use questionnaires to quantify it, but this method is sensitive to the user's opinion, mood, and other external factors that the experiment setup cannot control [38]. In an attempt to counterbalance this, a mixed-methods approach was used, as this method was effective in ensuring a broad and deep understanding of the user:

(i) Monitoring the heart rate of the user

(ii) Providing a questionnaire to assess the user's self-assessment

In particular, a physiological measurement is indeed helpful in reading the emotional state of the user, but it is unreliable in determining which type of emotion is being experienced [39]. Therefore, the physiological measurement was only used to support the self-reported data.

The nature of the project allows for a great degree of creative freedom, so the type of emotional response the prototype intends to trigger can be adapted according to what is easier to measure. The narrative of the prototype is bound to have certain themes, or a genre, that aim to produce a specific emotional response in the listener. We chose a target emotion that can be effectively measured; fear is one of many emotions looked at, and also an appropriate response in the horror genre.

Fear is the emotion associated with pain, danger, harm, etc. It bears close resemblance with anxiety, and, according to Freud (in 1924), there is a direct link between the two, and he considered the two to be the same emotion in many ways [40]. Furthermore, Freud in 1936 linked anxiety to exploratory behavior, indicating that anxiety leads to curiosity.

In this context, an important feature of fear is that the test participants were unlikely to be biased by this emotion when starting the experiment, given that they were not aware of the potentially fear-inducing experience they were about to participate. Oppositely, if happiness or sadness was measured, the participant would be biased about this emotion according to how their day was going and other trivial aspects of their life that could influence these emotions [38].

6.2.1. Immersive Response. In [41], the authors developed a quantitative measure for immersion in video games called the IMX questionnaire. Our aim is not about creating a video game, but rather an interactive story, but the two are

somehow similar that the items in the questionnaire are very relevant for us, with a few exceptions. The original IMX questionnaire consists of three parts; one that applies to all games, one that applies to games with characters, and one that applies to multiplayer games [42]. For the sake of simplicity, only the first part was used in this study, as the two others were less relevant to our purposes. Starting from them, eight questions were adapted to the context of narrative and are listed below:

Q1: The sound experience felt consistent with the real-world experience

Q2: After playing, it took me a moment to recall where I really was

Q3: During the experience, I felt at least one of the following: breathlessness, faster breathing, faster heart rate, tingling in my fingers, a fight-or-flight response

Q4: The story stimulated my reactions (panic, tension, relaxation, suspense, danger, and urgency)

Q5: Having to keep up with the speed of the story pulled me in

Q6: The story was energetic, active, and there was a sensation of movement

Q7: The story was thought-provoking for me

Q8: I felt caught up in the flow of the story

The responses were collected in 5-point Likert scale.

6.2.2. State Emotions. In the scientific literature, several theories of discrete emotions have been proposed [43]. Specific sets of discrete emotions can be extracted from research on facial or behavioral expressions, and from direct brain stimulation of animals.

In a story, multiple state emotions could be solicited. In order to measure the self-reported state emotions, the Discrete Emotions Questionnaire (DEQ) was used [44]. The questionnaire asks the participant to rate a number of emotions such as anger, dread, terror, scared, and fear using a 7-point Likert scale, according to how they felt during the experience.

6.2.3. Heart Rate Data Collection. Heart rate and pulse are two separate measurements, yet they are closely related, since the pulse stems from the heart. The heart rate is the number of heartbeats occurring in one minute. The average healthy adult heart rate is 60 to 80 bpm (beats per minute), where the normal for older adults is considered to be 60 to 100 bpm, and is affected by several conditions, e.g., emotional state, exercise, and stress [45].

The optimal method for recording heart rate usually employs ECG (electrocardiography), which is the process of recording electrical activity generated by the heart by placing electrodes on the skin. However, the Empatica E4 wristband was used for monitoring the heart rate in a more practical way for the participants.

A study on the accuracy and reliability of four different commercial heart rate monitors shows a varying accuracy when compared with electrocardiograms. As such, the wrist-worn monitors showed generally better accuracy during resting and declining with exercise [46]. Additionally, a similar study, where one of the two employed heart rate monitors is Empatica E4, showed that accuracy is once again very dependent on level and amount of movement during monitoring. The accuracy of the measured heart rate is evaluated with respect to electrocardiography, and the absolute difference of the measured heart rate and electrocardiograms was less than 10 bpm for 81–97% of the time for E4, while the percentage of accurately detected heartbeats was 68% during sitting, but only 9% during household work. The study concludes that wrist-worn devices such as the E4 are accurate and reliable for heart rate detection when hand movement is not excessive [47].

The use of a wrist-worn monitoring device, specifically the Empatica E4 at 64 Hz sample rate, is deemed sufficient for this study, since the participants are supposed to be sitting down with only moderate movement.

7. Results

The data statistics from the IMX questionnaire are plotted in Figure 3. The responses were nonnormally distributed, so a nonparametric one-way Kruskal–Wallis ANOVA test was used. In the following, results from each questionnaire items are reported: Q1 ($\chi^2(2) = 6.64$, $p < 0.05$), Q2 ($\chi^2(2) = 7.92$, $p < 0.05$), Q3 ($\chi^2(2) = 3.72$, $p = 0.15$), Q4 ($\chi^2(2) = 2.41$, $p = 0.30$), Q5 ($\chi^2(2) = 1.48$, $p = 0.47$), Q6 ($\chi^2(2) = 0.86$, $p = 0.65$), Q7 ($\chi^2(2) = 0.84$, $p = 0.65$), and Q8 ($\chi^2(2) = 0.85$, $p = 0.65$).

To find out which particular conditions differed, a Mann–Whitney test was conducted. Pairwise comparisons for statistically significant results are shown in Table 2. Questionnaire items 1 (sound consistency with reality) and 2 (spatial presence) showed differences between conditions MA and IVE; item 2 also showed a difference between condition BA and IVE while item 1 showed a value close to alpha = 0.05 between condition BA and IVE. These results suggested a differentiation of IVE towards a more immersive experience.

The nonparametric one-way Kruskal–Wallis ANOVA test is used to test for significant differences between the conditions in DEQ for each emotion in the questionnaire: anger ($\chi^2(2) = 1.21$, $p = 0.49$), sad ($\chi^2(2) = 0.3$, $p = 0.86$), grossed out ($\chi^2(2) = .19$, $p = 0.91$), happy ($\chi^2(2) = 3.0$, $p = 0.22$), terror ($\chi^2(2) = 5.8$, $p = 0.05$), rage ($\chi^2(2) = 1.8$, $p = 0.41$), grief ($\chi^2(2) = .06$, $p = 0.97$), anxiety ($\chi^2(2) = 8.4$, $p < 0.05$), nervous ($\chi^2(2) = 4.8$, $p = 0.09$), scared ($\chi^2(2) = 3.8$, $p = 0.15$), sickened ($\chi^2(2) = 0.19$, $p = 0.91$), fear ($\chi^2(2) = 6.7$, $p < 0.05$), calm ($\chi^2(2) = 2.5$, $p = 0.28$), and panic ($\chi^2(2) = 10.2$, $p < 0.01$). In order to get a deeper insight into how each condition affects the emotions, each combination of conditions was subjected to a Mann–Whitney test. The resulting significant p values are shown in Table 3, showing that the emotions terror, anxiety, nervous, fear, and panic have statistically significant differences between MA and IVE. Additionally, a difference is also detected for terror between conditions MA and BA. Moreover, Figure 4 shows a pie chart of this relevant emotions, for

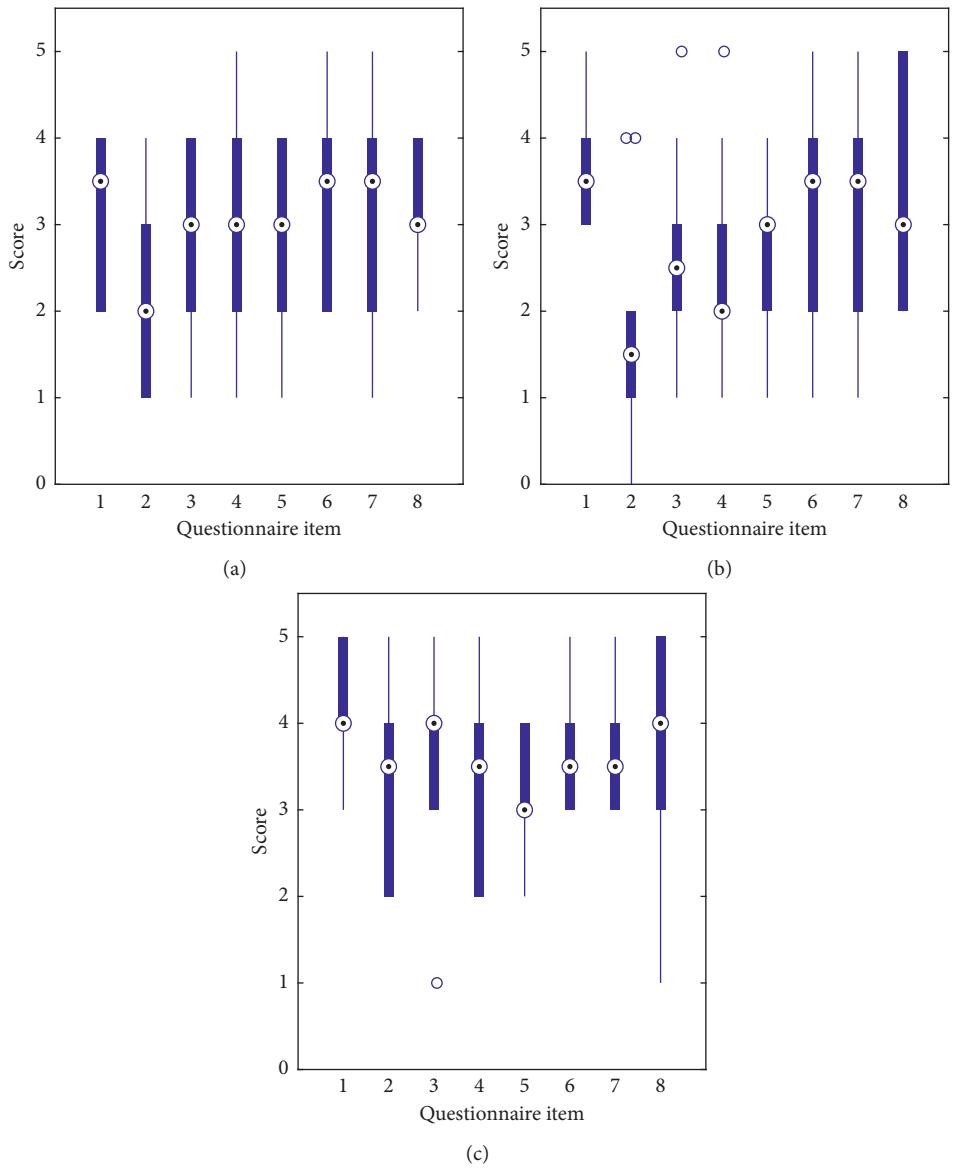


Figure 3: Boxplots of all items of the proposed IMX questionnaire grouped by listening condition. (a) Monophonic—static listening (cond. MA); (b) binaural recordings—static listening (cond. BA); (c) interactive virtual environment (cond. IVE).

Table 2: p values of pairwise comparisons in IMX questions with meaningful statistical effect.

Cond. comparison	p values	
	q1	q2
MA vs. BA	0.44	0.45
MA vs. IVE	0.02	0.03
BA vs. IVE	0.06	0.02

Table 3: p values of the pairwise comparison in emotions rating.

Cond. comparison	p values			
	Terror	Anxiety	Fear	Panic
MA vs. BA	0.68	0.02	0.38	0.17
MA vs. IVE	0.03	0.01	0.01	0.00
BA vs. IVE	0.07	0.91	0.12	0.10

each of the three test conditions. These results suggested a clear trend for an emotional enhancement for IVE experience with respect to conditions MA and BA.

From the raw data, the average heart rate and the deviation from that average were computed for each participant (see Figure 5 for a boxplot representation). The three conditions were evaluated with a one-way Kruskal–Wallis ANOVA which led to no statistically significant differences among conditions in average ($\chi^2(2) = 2.3$, $p = 0.31$), and standard deviation ($\chi^2(2) = 0.67$, $p = 0.71$). Moreover, the ratio between the first and last minute of narration was computed in order to identify a persistent increase in heart rate due to the experience. Again, no statistically significant

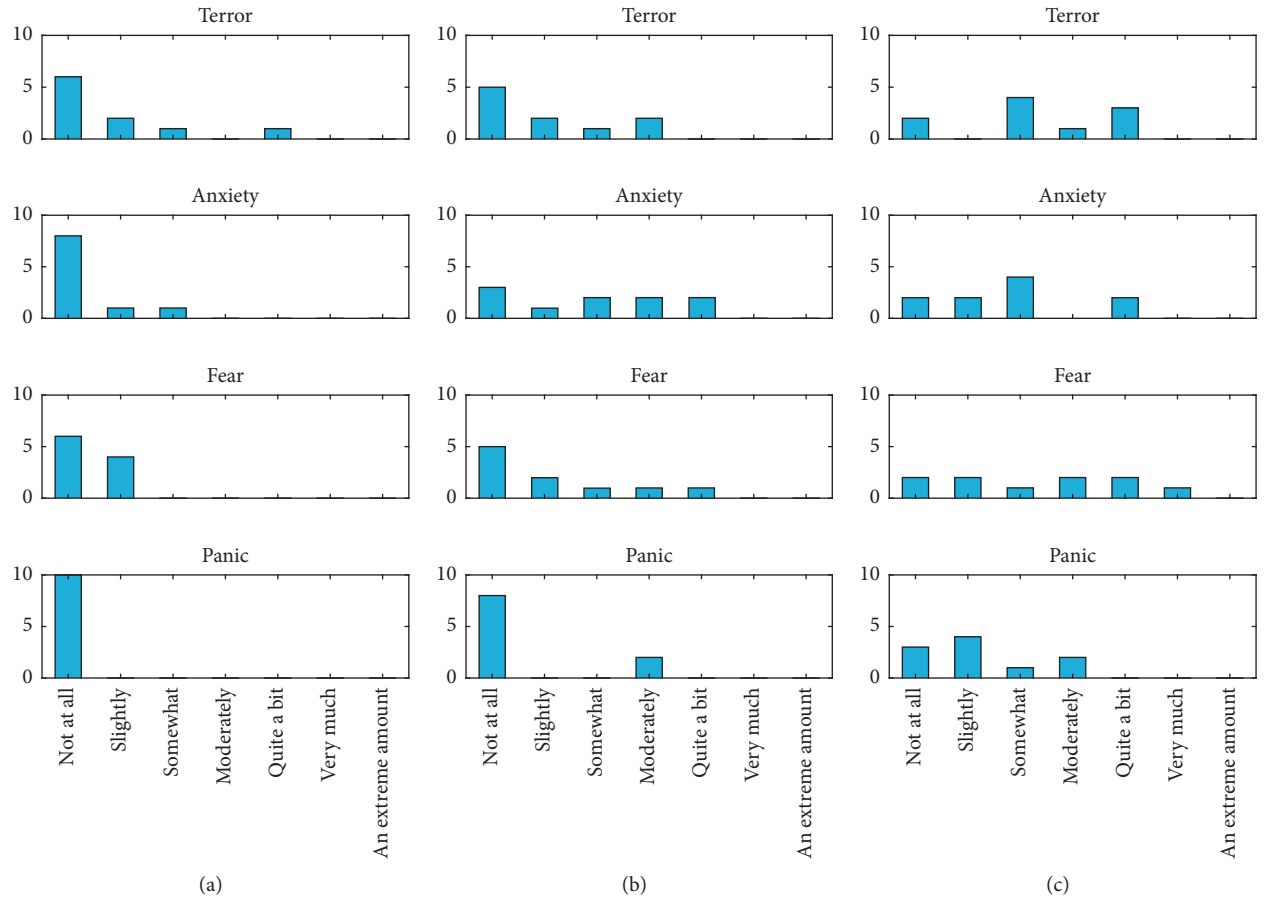


Figure 4: Data distribution of rates for the relevant emotions/items from DEQ questionnaire grouped by experimental condition. (a) MA; (b) BA; (c) IVE.

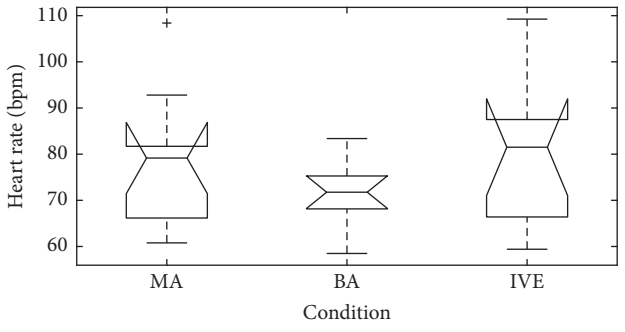


Figure 5: Notched boxplot representation of heart-rate data distributions for each condition.

differences were found among conditions ($\chi^2(2) = 1.6$, $p = 0.6$). However, the average trend suggested a higher heart rate together with a higher variability among users for the IVE experience.

8. General Discussion

By considering the IMX questionnaire with the DEQ, both the level of fear and immersion were measured, and it was interesting to evaluate these two aspects in the context of a horror-experience.

Our results showed that there was an interaction between spatial sound rendering and interactivity that enhanced the feelings of immersion with fear-induced emotions. Since analysis on heart rate data revealed no statistical difference, the questionnaires suggested that there was a difference in immersion and emotions between monophonic narration (MA) and immersive virtual environment (IVE), which was not always visible in the comparison between monophonic narration and static binaural recordings (BA). Moreover, the collected data suggested a trend of increasing level of immersion and fear-related emotions related to interactivity.

It has to be mentioned that there were several confounding variables that could have affected the heart rate data. Mainly, there was little control over the participant's initial heart rate, so a baseline resting heart rate could not be established. Instead, the average heart rate was used as the baseline, and the average deviation from that baseline was used as the main statistic. The goal was to measure how much the heart rate varied throughout the narrative. However, the development of the heart rate through the narrative could not necessarily be attributed to immersion in the narrative. If a participant had been physically active before the experiment, they would have an elevated heart rate at the beginning of the experiment, and it would decrease during the narrative as a

result of the participant not moving. To mitigate this effect, the participants could have been asked to sit still for half an hour before the narrative and then establish a baseline resting heart rate after this period of rest.

In the IMX questionnaire, items Q1 (The sound experience felt consistent with the real-world experience) and Q2 (After playing, it took me a moment to recall where I really was) showed significant differences. When looking at averages for each test condition in Figure 3, there seems to be a small, but noticeable, increase for condition IVE. Overall, all IMX answers show an increase in immersion between conditions MA and IVE on average, not detectable between MA and BA.

The first item from the IMX questionnaire is closely linked to the immersion concept of spatial presence, so detecting a difference in this question is an indicator that the level of immersion was increased. This question indicates that IVE condition provided a more realistic sound experience, and this could indicate that it was easier for the user to reach a state of spatial presence under such a condition. However, more data have to be collected before this claim can be fully asserted. The second IMX item is also an indication of increased immersion. The question can be linked to the concept of *flow*: cause a loss of self-consciousness and alter one's sense of time. This is another indicator that the experienced immersion varied with the conditions.

It is worthwhile to notice that the static binaural audio rendering (BA) led to a similar (sometimes in a worse trend) answer in IMX to MA which seems counterintuitive considering the big difference between monophony and binaural audio. However, Steam Audio engine employed generic dummy-head HRTFs which are known to cause frequent front-back confusion and inside-the-head localization without head-tracking and because of their generic solution they are not good for every listener [15, 48]. The discrete emotions questionnaire supported the result of the IMX questionnaire. The ANOVA test revealed that the emotions terror, fear, anxiety and panic were affected by the test conditions. The Mann–Whitney test revealed that the differences were primarily found between conditions MA and IVE. This is an indicator that the participants experienced a higher level of negative emotion under the interactive condition. The emotions that showed increases under IVE condition were largely the ones that were intended as the emotional response of the horror-story, so detecting a difference indicates that the emotional involvement in the narrative was affected by the test condition. The statistically significant differences were detected mostly between conditions MA and IVE, which, along with the results from the IMX, indicates that condition IVE was a more effective experience.

The fact that DEQ and IMX results both indicated a small increase in immersive and emotional responses might be viewed as an indicator that the proposed interactions had an influence on the participants' overall experience of the story and the prototype. An influence to this can be caused by several factors:

(i) An increased feeling of being present by doing something practical with your hands

(ii) The freedom of head movement

(iii) The spatial sound

(iv) Combination of previous

(v) Confounded, such as the state of mind of the participants before testing

One can eliminate, or try to restrict and contain, these confounding variables, by asking the participants to rate their current emotions before engaging in the test. For some of the participants, another confounder could be the usage of generic HRTFs, which might provide acoustic information too far from individual contribution of listener's body. Such discrepancy might lead to bad localization performances that resulted in a biased evaluation of the system like in the VR experience reported by Geronazzo et al. [49].

The results from the remote's logging data showed some different behaviors among the participants. All of them were adjusting the volume to positive and negative levels at the beginning minutes of the test, where some remained at an approximately fixed volume afterward, and others continue to use the remote continuously.

One of the primary problems with our test was the small sample size. Normally, a sample size of 28 for each group is needed to detect a large effect size in a between-subjects experimental design [50], but this experiment only had a sample size of 10. A larger sample size would be less prone to outliers and thereby reduce the risk of getting Type I errors.

The quality of the prototype was a critical aspect in providing a good representation of interaction, as a dysfunctional solution could severely affect the results of the test as opposed to the expectation. Another crucial variable of our study was the implementation of the narrative, as the chosen story was not originally designed as an interactive audio experience. The story, originally a written short story, has been adapted to fit the medium and desired duration of the experience. This means that the quality of the narrative could possibly have been decreased as the adaptation was not performed by a trained professional. Similarly, with the narrative, if the sound design is not optimal with the wanted experience, and immersiveness, the test results will have subjected to a negative effect as well. Finally, the heart rate monitor E4 Empatica provided some readings that were on unreasonable levels, as too high or too low, which had an influence on the analysis and interpretation of the data. These data points were removed from the data samples, so the overall amount of data is less than what was actually collected.

Another confounding variable was the implementation of the narrative, as the chosen story was not originally designed as an interactive audio experience. The story, originally a written short story, has been adapted to fit the medium and desired duration of the experience. This means that the quality of the narrative could possibly have been decreased as the adaptation was not performed by a trained professional.

Moreover, the heart rate monitor E4 Empatica provided some readings that were on unreasonable levels, as too high

or too low, which had an influence on the analysis and interpretation of the data. These data points were removed from the data samples, so the overall amount of data is less than what was actually collected.

In future development, the story selection could also be optimized and reviewed upon. The story can either be lengthened or shortened, depending on the goal in the evaluation. If one was to make another short 10 minute story, as done in this study, perhaps the story should be more intense and fast paced. If a longer 40 minute story was created, the story should be more intriguing in the long run and have several narrative atmospheres to keep the user actively listening.

9. Conclusion

The research on interactive storytelling revealed that both storytelling and interactivity can take several shapes. Within the story, the narrative can be linear or nonlinear and when it comes to interactivity, there are different dimensions to consider, like making sure the user has the ability to make choices. For this work, it was adopted a first-person linear narrative to make sure all the users were exposed to the same story while testing, in order to be able to use the data gathered. Our physical interface design revealed that users felt comfortable and natural in navigating a virtual world through auditory feedback with head-tracking. The quality of the experience was assessed by focusing on invoking a particular feeling in the user, as quantitative physiological measurements can be used for measuring this. Using a combination of heart rate monitoring and follow-up questionnaires was possible to determine the level of fear-induced emotion experienced, providing insight into the user's emotional involvement in the narrative. Our results suggested that dynamic binaural audio rendering allowed a greater level of immersion and more detectable fear-related emotions.

The possibility of an even further interactive story could also potentially be implemented. While interaction was incorporated in this study through head movement and a simple gesture control via hand-help controller, there might be more directions for increasing interactivity. In further studies, the user could be able to choose different paths in the storyline. This could allow the user to build the story themselves and therefore feel more involved in the process. This could be done by using the interactive remote as a selection tool, pointing in the story direction that they want to follow. This would also involve some form of crossroad areas of the story, where the narration would pause and the user would be instructed to choose between options. It is worthwhile to notice that costs and feasibility of new interactions will become more and more similar to those of computer games. The correct balance between passive and active listening should be carefully taken into consideration.

The integration of the proposed system in mobile devices for audio augmented reality (AAR) will open opportunities of new forms of spatial interactions and HRTF personalization [51]. The real deployment of such kind of audio story will require the development of story map in Unity and a definition of design guidelines for audio VE. On the other hand, the technological platform for the required interactions should take advantage from available head-mounted displays trying to integrate audio-specific featuresre into immersive VEs.

Data Availability

The data from the experimental sessions used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by the internationalization grant of the 2016–2021 strategic program "Knowledge for the World" awarded by Aalborg University to Michele Geronazzo.

Supplementary Materials

In the supplementary material, we provide the shortened narrative of Stephen King's Strawberry Springs and the storyboard built upon it for our study. Moreover, a map of all created audio objects depicts the complexity of the virtual auditory scene designed following such a storyboard. (*Supplementary Materials*)

References

[1] J. Richards, *Audiobooks Continues Double-Digit Growth—2017 Sales Survey*, Audio Publishers Association, Princeton Junction, NY, USA, 2017.

[2] M. Furini, "Beyond passive audiobook: how digital audiobooks get interactive," in *Proceedings of the 2007 4th IEEE Consumer Communications and Networking Conference*, pp. 971–975, Las Vegas, NV, USA, January 2007.

[3] R. Nordahl and N. C. Nilsson, "The sound of being there: presence and interactive audio in immersive virtual reality," *The Oxford Handbook of Interactive Audio*, Oxford University Press, Oxford, UK, 2014.

[4] D. R. Begault, *3-D Sound for Virtual Reality and Multimedia*, Academic Press Professional, Inc., San Diego, CA, USA, 1994.

[5] S. Serafin, M. Geronazzo, C. Erkut, N. C. Nilsson, and R. Nordahl, "Sonic interactions in virtual reality: state of the art, current challenges and future directions," *IEEE Computer Graphics and Applications*, vol. 38, no. 2, pp. 31–43, 2018.

[6] P. Coleman, A. Franck, J. Francombe et al., "An audio-visual system for object-based audio: from recording to listening," *IEEE Transactions on Multimedia*, vol. 20, no. 8, pp. 1919–1931, 2018.

[7] C. Kim, R. Mason, and T. Brookes, "Head movements made by listeners in experimental and real-life listening activities," *Journal of the Audio Engineering Society*, vol. 61, no. 6, pp. 425–438, 2013.

[8] E. Degli Innocenti, M. Geronazzo, D. Vescovi et al., "Mobile virtual reality for musical genre learning in primary education," *Computers & Education*, vol. 139, pp. 102–117, 2019.

[9] I. M. Konstantakos, "The magical transformation contest in the ancient storytelling tradition," *Cuadernos de filología clásica: Estudios griegos e indoeuropeos*, vol. 26, pp. 207–234, 2016.

[10] C. Crawford, *Chris Crawford on Interactive Storytelling*, New Riders, Carmel, IN, USA, 2005.

[11] M. O. Riedl and R. M. Young, "From linear story generation to branching story graphs," *IEEE Computer Graphics and Applications*, vol. 26, no. 3, pp. 23–31, 2006.

[12] J. Preece, H. Sharp, and Y. Rogers, *Interaction Design: Beyond Human-Computer Interaction*, John Wiley & Sons Inc., Chichester, UK, 2015.

[13] S. M. Lwin, "Narrativity and creativity in oral storytelling: co-constructing a story with the audience," *Language and Literature*, vol. 26, no. 1, pp. 34–53, 2017.

[14] M. Geronazzo, F. Avanzini, and F. Fontana, "Auditory navigation with a tubular acoustic model for interactive distance cues and personalized head-related transfer functions," *Journal on Multimodal User Interfaces*, vol. 10, no. 3, pp. 273–284, 2016.

[15] M. Geronazzo, E. Peruch, F. Prandoni, and F. Avanzini, "Applying a single-notch metric to image-guided head-related transfer function selection for improved vertical localization," *Journal of the Audio Engineering Society*, vol. 67, no. 6, pp. 414–428, 2019.

[16] T. R. Agus, S. J. Thorpe, and D. Pressnitzer, "Rapid formation of robust auditory memories: insights from noise," *Neuron*, vol. 66, no. 4, pp. 610–618, 2010.

[17] D. Hammershøi and H. Møller, "Methods for binaural recording and reproduction," *Acta Acustica United with Acustica*, vol. 88, no. 3, pp. 303–311, 2002.

[18] S. Paul, "Binaural recording technology: a historical review and possible future developments," *Acta Acustica United with Acustica*, vol. 95, no. 5, pp. 767–788, 2009.

[19] J. Lewald, "Exceptional ability of blind humans to hear sound motion: implications for the emergence of auditory space," *Neuropsychologia*, vol. 51, no. 1, pp. 181–186, 2013.

[20] M. Geronazzo, A. Bedin, L. Brayda, C. Campus, and F. Avanzini, "Interactive spatial sonification for non-visual exploration of virtual maps," *International Journal of Human-Computer Studies*, vol. 85, pp. 4–15, 2016.

[21] F. Heller, A. Krämer, and J. Borchers, "Simplifying orientation measurement for mobile audio augmented reality applications," in *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems-CHI'14, ACM*, pp. 615–624, New York, NY, USA, May 2014.

[22] W. Hess, "Head-tracking techniques for virtual acoustics applications," in *Proceedings of the Audio Engineering Society Convention 133*, San Francisco, CA, USA, October 2012.

[23] J. Traer and J. H. McDermott, "Statistics of natural reverberation enable perceptual separation of sound and space," *Proceedings of the National Academy of Sciences*, vol. 113, no. 48, pp. E7856–E7865, 2016.

[24] A. Andreasen, M. Geronazzo, N. C. Nilsson, J. Zovnercuka, K. Konovalov, and S. Serafin, "Auditory feedback for navigation with echoes in virtual environments: training procedure and orientation strategies," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 5, pp. 1876–1886, 2019.

[25] G. Riva, F. Mantovani, C. S. Capideville et al., "Affective interactions using virtual reality: the link between presence and emotions," *CyberPsychology & Behavior*, vol. 10, no. 1, pp. 45–56, 2007.

[26] A. Gorini and G. Riva, "Virtual reality in anxiety disorders: the past and the future," *Expert Review of Neurotherapeutics*, vol. 8, no. 2, pp. 215–233, 2008.

[27] H. Sauzéon, P. Arvind Pala, F. Larrue et al., "The use of virtual reality for episodic memory assessment: effects of active navigation," *Experimental Psychology*, vol. 59, no. 2, pp. 99–108, 2012.

[28] C. Huber, N. Rober, K. Hartmann, and M. Masuch, "Evolution of interactive audio books," in *Proceedings of the 2nd Conference on Interaction with Sound (Audio Mostly)*, pp. 166-167, Ilmenau, Germany, September 2007.

[29] E. Marchetti and A. Valente, "Interactivity and multimodality in language learning: the untapped potential of audiobooks," *Universal Access in the Information Society*, vol. 17, no. 2, pp. 257–274, 2018.

[30] S. Rubin and M. Agrawala, "Generating emotionally relevant musical scores for audio stories," in *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology-UIST'14*, pp. 439–448, ACM, Honolulu, Hawaii, USA, October 2014.

[31] D. Hilviu and A. Rapp, "Narrating the quantified self," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers-UbiComp'15*, pp. 1051–1056, ACM, Osaka, Japan, September 2015.

[32] M. Grimshaw, "Sound and immersion in the first-person shooter," *International Journal of Intelligent Games & Simulation*, vol. 5, no. 1, 2008.

[33] E. Gallo and N. Tsingos, "Efficient 3d audio processing with the GPU," in *Proceedings of the ACM Workshop on General Purpose Computing on Graphics Processor*, p. 1, ACM, Los Angeles, CA, USA, August 2004.

[34] R. McGloin, K. Farrar, and M. Krcmar, "Video games, immersion, and cognitive aggression: does the controller matter?," *Media Psychology*, vol. 16, no. 1, pp. 65–87, 2013.

[35] P. Skalski, R. Tamborini, A. Shelton, M. Buncher, and P. Lindmark, "Mapping the road to fun: natural video game controllers, presence, and game enjoyment, mapping the road to fun: natural video game controllers, presence, and game enjoyment," *New Media & Society*, vol. 13, no. 2, pp. 224–242, 2011.

[36] D. Weibel and B. Wissmath, "Immersion in computer games: the role of spatial presence and flow," *International Journal of Computer Games Technology*, vol. 2011, Article ID 282345, 14 pages.

[37] M. Csikszentmihalyi, *Flow: The Psychology of Optimal Experience Harper Perennial Modern Classics*, HarperCollins, New York, NY, USA, 1st edition, 2008.

[38] T. Bjorner, *Qualitative Methods for Consumer Research*, Hans Reitzel's Publishers, Kopenhagen, Denmark, 2015.

[39] T. Garner, M. Grimshaw, and D. A. Nabi, "A preliminary experiment to assess the fear value of preselected sound parameters in a survival horror game," in *Proceedings of the 5th Audio Mostly Conference: A Conference on Interaction with Sound, AM '10*, pp. 10:1–10:9, ACM, Piteå, Sweden, September 2010.

[40] D. Spielberger Charles and C. Reheiser Eric, "Assessment of emotions: anxiety, anger, depression, and curiosity," *Applied Psychology: Health and Well-Being*, vol. 1, no. 3, pp. 271–302, 2009.

[41] N. Curran, *The psychology of immersion and development of a quantitative measure of immersive response in games*, Ph.D. thesis, University College Cork, Cork, Ireland, 2013.

[42] L. J. Norman and L. Thaler, "Human echolocation for target detection is more accurate with emissions containing higher spectral frequencies," *i-Perception*, vol. 9, no. 3, article 2041669518776984, 2018.

[43] J. Panksepp, *Affective Neuroscience: The Foundations of Human and Animal Emotions*, Oxford University Press, Oxford, UK, 2004.

[44] C. Harmon-Jones, B. Bastian, and E. Harmon-Jones, "The discrete emotions questionnaire: a new tool for measuring state self-reported emotions," *PLoS One*, vol. 11, no. 8, Article ID e0159915, 2016.

[45] D. M. Anderson, *Mosby's Medical, Nursing, & Allied Health Dictionary*, Mosby, St. Louis, MO, USA, 2002.

[46] R. Wang, G. Blackburn, M. Desai et al., "Accuracy of wrist-worn heart rate monitors," *JAMA Cardiology*, vol. 2, no. 1, pp. 104–106, 2017.

[47] J. Pietilä, S. Mehrang, J. Tolonen et al., "Evaluation of the accuracy and reliability for photoplethysmography based heart rate and beat-to-beat detection during daily activities," in *Proceedings of the EMBEC & NBC 2017, IFMBE*, pp. 145–148, Springer, Singapore, 2017.

[48] D. R. Begault, E. M. Wenzel, and M. R. Anderson, "Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source," *Journal of the Audio Engineering Society*, vol. 49, no. 10, pp. 904–916, 2001.

[49] M. Geronazzo, E. Sikström, J. Kleimola, F. Avanzini, A. De Götzen, and S. Serafin, "The impact of an accurate vertical localization with HRTFs on short explorations of immersive virtual reality scenarios," in *Proceedings of the 17th IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 90–97, IEEE Computer Society, Munich, Germany, October 2018.

[50] A. Field and G. Hole, *How to Design and Report Experiments*, Sage Pubns Ltd., Thousand Oaks, CA, USA, 2003.

[51] M. Geronazzo, J. Fantin, G. Sorato, G. Baldovino, and F. Avanzini, "Acoustic selfies for extraction of external ear features in mobile audio augmented reality," in *Proceedings of the 22nd ACM Symposium on Virtual Reality Software and Technology (VRST 2016)*, pp. 23–26, ACM, Munich, Germany, November 2016.