PROCEEDINGS of the 2nd Symposium: The Acoustics of Ancient Theatres
6-8 July 2022 Verona, Italy

Aural Augmented and Virtual Reality Applications: Best Practices and Challenges

Andrey Borisov[1]; Anne Eiselein[2], Julien Letellier[1]; Jürgen Sieck[1], Elisabeth Thielen[1]

[1] HTW Berlin, Germany, {borisov, julien.letellier, j.sieck, thielen}@htw-berlin.de
[2] Konzerthaus Berlin, Germany a.eiselein@konzerthaus.de

ABSTRACT

Emerging technologies can help to create new kinds of interesting audio-focused applications. Especially augmented and virtual reality can allow experiencing audio in new ways since AR and VR allow new types of interactions and can also create a sense of immersion. However, there are some difficulties to consider while implementing an audio-focused AR/VR application. In this paper, we discuss these difficulties and demonstrate some best practice examples.

Keywords: Augmented Reality, Virtual Reality, Human-Computer Interaction

1. INTRODUCTION

New technologies can be used to create applications that allow experiencing the audio content in a novel manner. AR and VR applications are particularly interesting since they allow new ways for intuitive interactions which can help to create new ways of natural interactions with audio contents. From the auditive point of view are AR and VR applications also interesting because the immersion created by these technologies greatly depends on audio contents [1].

This paper describes challenges that arise while the development of audio-focused AR and VR applications. The discussed challenges are shown in already existing AR and VR applications. All presented applications were developed as part of the cooperative research project APOLLO by the research group INKA at HTW Berin and the Konzerthaus Berlin with the aim to teach the basics of classical music and to awaken an interest in classical music in general.

Four different AR/VR applications are presented and the challenges are explained using these examples. The presented AR applications are the digital "Virtual Quartet" and the "OrchestraBox" both applications are au- dio-based and the main aim is to develop and implement new interaction formats and to teach young people clas- sical music. The VR application "VR Tour and Orches- tra" shows how VR can help to create an immersive auralization of concerts and how this helps to make 3D environments more believable. The VR application "Umwelten" puts the users in a world with abstract visuals and audio and offers different interactions. After an introduction and state-of-the-art chapter, we will describe our current research and show our AR and VR applications. Then we will discuss challenges, our approach, and the development of audio-based AR/VR applications. In the end, we will summarize and discuss the future work.

2. BACKGROUND

Some definitions of VR require not only a virtual environment but also require these environments to be immersive [2]. The capability to create a sense of immersion is important for VR applications since it can make the user believe that the presented virtual environment is real. Besides the visual component of VR applications, audio plays an important role for immersion [1][3]. The sense of immersion in VR can be improved through realistic sound effects in the virtual environment and music. Music can also have a great impact on the emotions of the user in immersive environments [4].

In AR applications audio can also help to increase the sense of immersion, for instance by adding realistic sound effects to virtual objects. The main purpose of audio in AR however is to enhance the real environment of the user with audio contents. There are two different types of AR: marker-based and markerless. Marker-based AR uses image tracking algorithms such as SIFT [5], to connect virtual content with some physical objects (= markers). Markerless AR on the other hand does not need any markers and tracks the position of the camera through the extraction of the feature points from the user's surroundings. Thus, markerless AR allows the user to place virtual objects freely in his environment. For example, in some AR apps users can annotate real objects with audio [6] and other applications allow attaching music to the real objects and the user can combine the attached music by combining the physical objects [7]. There are also the so-called Auditive-AR (AAR) applications that are solely based on audio content and do not provide any form of visual input for the user. Some examples of this kind of applications are

10.58874/SAAT.2022.165

museum guides [8] and audio games [9].

An important aspect of AR and VR applications is the new way of interactions that they offer. For example, some AR and VR Head-Mounted-Displays (HMD) allow tracking of user hands which makes it possible to interact with such devices in a natural way (e.g., HoloLens or Oculus Quest). Natural interactions make it easier to create applications that allow intuitive interactions with audio content. In such applications, the users can not only hear the music but can also change it intuitively [10][7].

3. VR PROJECTS

3.1 VR Tour and Orchestra

The goal of this immersive virtual tour is to show users the impressive architecture of the Konzerthaus, the history of the building and to provide a glimpse of its orchestra in action [11]. Since audio plays a very important role in all of the activities surrounding the Konzerthaus, the VR application has to provide not only stunning visuals, but also immersive audio that makes the experience more believable.

Visually, the Konzerthaus building is presented as a high-resolution 3D model in which users are guided through the halls. Additionally, there is a short recording of a string quartet (see also Section 4.1) embedded into the Small Hall of the Konzerthaus. The second part of the VR application shows another recording of a full orchestra that can be viewed from four different viewpoints. Each viewpoint consisting of 360° spheres with 4k video and an ambisonic audio source for spatial audio.

Both additions are accompanied by spatial audio that is carefully placed within the 3D environment. Extensive user and playtesting of the application showed that the additional emphasis on the auditive components played a vital role in the overall experience.

One of the major challenges was to keep interaction techniques simple so that a wide range of users could enjoy the experience. This was achieved by implementing an intuitive interaction technique called *gaze and wait* where users simply look at a symbol and wait until the action is triggered. (This technique is now widely used within VR applications.) The app introduces the mechanic early on and constantly reminds users how to trigger certain elements. Again, extensive user testing showed that spatial audio cues are important to steer the gaze of users in the right direction.

3.2 Interactive composition in VR

Umwelten is a cooperative project of the research project APOLLO, the Konzerthaus Berlin, the visual artist Julian Bonequi and the composer Mark Barden [12].

The VR application transports users into an alien-like, eccentric 3D world where they can interact with 3D models and activate AudioSources that are attached to them.

Each of the 3D models features multiple different audio sources, that can be activated by i.e. grabbing them, drawing on them with a drawing ray or sometimes just by entering their vicinity.

At first, some of the 3D models featured up to 30 different audio sources each and would have multiple interaction concepts that could activate audio.

One of the problems with this while user testing the application was that users couldn't always discern which behaviour activated what audio and would get overwhelmed by the many different interaction possibilities. One specific interaction used a drawing ray that would activate audio upon touching an object in the scene that had an audio source. This drawing ray would however work on the object directly in front of the user as well as objects floating in the sky of the 3D world.

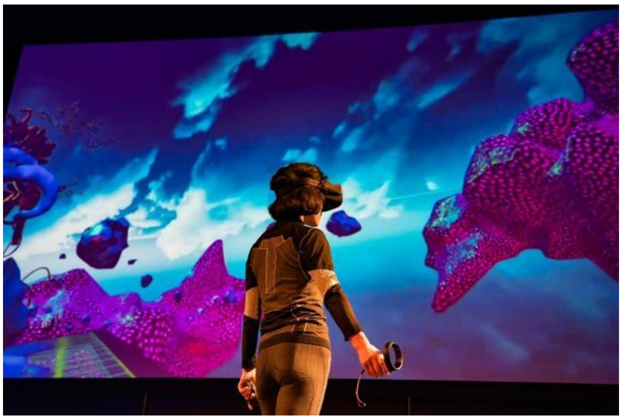

Figure 1 – VR user using Umwelten (Photo: Markus Werner)

One part of the solution to this was to use spatial audio. This meant that users could understand the correlation of the 3D models and their respective audio sources. Only the objects in the direct vicinity could be heard and it was therefore clearer what the sound belonged to. Some of the objects in the sky would still be able to be activated, as preferred by the artists working on the project.

So, a second part of the solution was to implement colour feedback when activating an object. The material of the triggered 3D model would change colour for a few seconds in order to indicate its activation and that another sound would be started.

User feedback on these changes was good, the audio activation was easier to understand and the origin of the sound easier to discern both with the spatial audio as well as the colour feedback.

4. AR PROJECTS

The next chapter will cover two AR projects of the research project APOLLO. They both dealt with the synchronisation of multiple audio tracks, both during the content creation and inside the final application.

4.1 Virtual Quartet

The virtual quartet is one of the marker-based AR applications that were developed in cooperation with the Konzerthaus Berlin [13]. The application works with four different markers, that resemble quartet cards from a playing card game. Each AR marker is connected to a musician and upon scanning the cards a visual representation of this musician is placed onto the card. In addition to this the musician's respective part from Franz Schuberts "Death and the Maiden" starts playing.

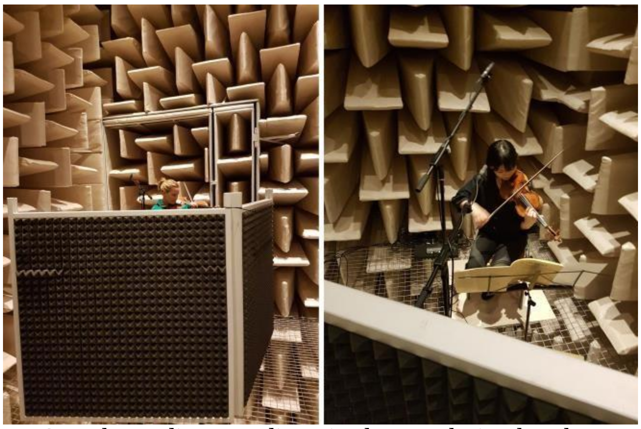

Figure 2 – The audio recordings in the anechoic chamber at TU Berlin (Photo: Annette Thoma)

The audio users can hear depends on how many markers are currently recognized in the camera view of the smartphone or tablet. Users can therefore create their own composition with i.e. only the violin, only the cello and viola, and so on. To achieve this the musicians had to be recorded separately in an anechoic chamber, resulting in four separate audio tracks, one for each musician.

The challenge for the musician was to play their piece individually. Normally, even though they have the music score, they are also playing off each other, reacting to changes in tempo etc. This could be partially solved by the musicians having earpieces with the other musician's recordings playing while they were recording their part.

The next challenge was to also have a visual recoding of their playing. This could not take place in the anechoic chamber but had to be done in front of a green screen, so that the videos could later be chroma-keyed to remove the background.

An important part when developing the application was the synchronization of the different audio tracks. Even slight inaccuracies in the playing and starting times could be heard. This was even more apparent when the video of the musicians was also introduced into the application.

In the final application, whenever a new marker is scanned, the musician is registered to a global manager class. The manager class gets the time of the video and starts the musician's audio at the respective time. This way, the audio is always in sync with the video. Fortunately, the asset sizes allow for almost no delay in starting the audio, meaning that the four quartet musician audios are also in sync with each other when they are started using the videoplayer time.

One point of discussion concerning the auditive experience was also when to first start the video-/audio playback. Early versions of the application started the audio and video whenever the first marker was recognized. However, this had to be changed since the delay in recognizing the markers meant users couldn't hear the start of the music piece with all four musicians playing.

The solution for this was to only start the audio after all the markers were scanned once. In order to give users the option to only listen to single musicians, a playbutton was introduced. While not all the musicians are scanned yet, a playbutton is attached to the first recognized musician, allowing users to start the audio and video by clicking it.

4.2 OrchestraBox

The OrchestraBox is a music box that was developed for musical education lessons at school [14]. The developed music box is playing a composition that was recorded by 18 musicians. All 18 musicians were recorded separately which allows for selecting which musicians should be audible while playing the composition. The musicians can be selected by placing their 3D printed figurines on the top surface of the music box. Since each figurine has an RFID chip attached to its bottom side, the OrchestraBox can recognize the figurines with its RFID antennas mounted on its top surface. By placing the figurines on the top surface of the OrchestraBox the user can select which musicians should be heard while playing the composition.

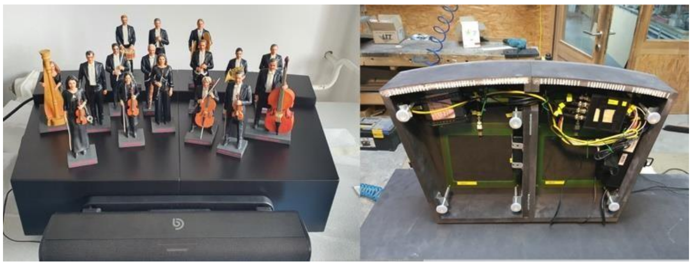

Figure 3 – left: the OrchestraBox with all musician markers, right: the underside of the box with the RFID antennas

All 18 musicians were recorded separately from each other in an anechoic chamber to provide the best audio quality. One of the challenges for the given project was the fact that the combining of 18 audio tracks in real-time requires a lot of computational power because of that a lightweight, low-power computer couldn't be used. Thus, an Intel-NUC computer was used instead of a portable and more lightweight Raspberry Pi 4B.

Also, the RFID antennas generate an electromagnetic field that affects the audio transmitted through the AUX cables, which results in noise while playing the audio. This problem was solved by rearrangement of hardware components of the OrchestraBox and better placement of the audio cable, so the RFID antennas have only a little effect on audio transmission.

5. DISCUSSION

In audio-focused AR and VR applications, it is often useful to record many different sounds that can be used to create an interesting virtual or augmented environment with interactable audio objects. However, it could become a problem if the application offers too much content (as it was the case in the Umwelten project) since it could confuse the user. Audiovisual cues can be used to direct the user's attention which could help to solve this issue.

With the effect of immersion, VR intensifies and changes the way people listen to music. Also, VR and AR offer some new ways of interacting with virtual content. Some AR and VR devices can track the hands of the users which allows them to interact with the application intuitively. This could open new ways for interaction with audio content, for example using hand gestures or activating audio content by touching it.

Overall AR and VR can help to experience audio in a new way and can make audio content more interesting and interactable.

ACKNOWLEDGEMENTS

The work described in this short paper was generously funded by the European Regional Development Fund (ERDF) and the Senate Department for Higher Education and Research in Berlin.

6. REFERENCES

[1] I. Mahalil, M. E. Rusli, A. M. Yusof, M. Z. Mohd Yusoff and A. R. Razieff Zainudin, "Study of immersion effectiveness in VR-based stress therapy," *Proceedings of the 6th International Conference on Information Technology and Multimedia*, 2014, pp. 380-384, doi: 10.1109/ICIMU.2014.7066663.

[2] Steuer, Jonathan. (2000). Defining Virtual Reality: Dimensions Determining Telepresence. Journal of Communication. 42. 10.1111/j.1460-2466.1992.tb00812.x.

[3] B, Sarah & Ladeira, Ilda & Winterbottom, Cara & Blake, Edwin. (2002). An Investigation on the Effects of Mediation in a Storytelling Virtual Environment. Lecture Notes in Computer Science. 2897. 10.1007/978-3-540-40014-1_13.

[4] Rogers, Katja & Jörg, Matthias & Weber, Michael. (2019). Effects of Background Music on Risk-Taking and General Player Experience. 213-224. 10.1145/3311350.3347158.

[5] D. G. Lowe, "Object recognition from local scale-invariant features," Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999, pp. 1150-1157 vol.2, doi: 10.1109/ICCV.1999.790410.

[6] Langlotz, Tobias & Regenbrecht, Holger & Zollmann, Stefanie & Schmalstieg, Dieter. (2013). Audio stickies: visually-guided spatial audio annotations on a mobile augmented reality platform.

[7] Yairi, Ikuko & Takeda, Takuya. (2012). A music application for visually impaired people using daily goods and stationeries on the table. 271-272. 10.1145/2384916.2384988.

[8] Fatima Zahra Kaghat, Ahmed Azough, Mohammed Fakhour, and Mohammed Meknassi. 2020. A new audio augmented reality interaction and adaptation model for museum visits. Computers & Electrical Engineering 84 (2020), 106606.

[9] Thomas Chatzidimitris, Damianos Gavalas, and Despina Michael. 2016. Sound- Pacman: Audio augmented reality in location-based games. In 2016 18th Mediterranean Electrotechnical Conference (MELECON). IEEE, 1–6.

[10] Kang, Jiyoung & Jeon, Byung-kyu & Kim, Seon-hwi & Park, Su-yong. (2021). Exposition of Music: VR Exhibition. 1-2. 10.1145/3450615.3464535.

[11] Letellier, Julien ; Sieck, Jürgen : Visualization and Interaction Techniques in Virtual Reality for Guided Tours. In: 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), S. 1041-1045, Metz, Frankreich, 2019, ISBN 978-1-7281-4068-1

[12] A. Thoma, E. Thielen, and A. Borisov. 2020. Umwelten: An immersive and interactive composition in Virtual Reality. Proceedings of the XVIII. Conference on Culture and Computer Science (2020). pp. 141-153. ISBN 978-3-86488-169-5

[13] E. Thielen, J. Letellier, J. Sieck, and A. Thoma. 2018. Bringing a virtual string quartet to life. In *Proceedings of the Second African Conference for Human Computer Interaction: Thriving Communities* (*AfriCHI '18*). Association for Computing Machinery, New York, NY, USA, Article 30, 1–4. https://doi.org/10.1145/3283458.3283477

[14] A. Borisov, J. Sieck and E. Thielen, "OrchestraBox: RFID Music Box for Musical Education at Schools," 2021 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), 2021, pp. 236-240, doi: 10.1109/IDAACS53288.2021.9660878.