DOI: 10.61782/fa.2023.0430

forum acusticum 2023

PERSONALIZATION IN AUDIO STORYTELLING WITHIN VIRTUAL AND AUGMENTED REALITY: STATE OF THE ART AND INSIGHTS

Alessandro Giuseppe Privitera[1*] **Federico Fontana**[1] **Michele Geronazzo**[2]

[1] Dept. Mathematics, Computer Science and Physics, University of Udine, Italy
[2] Department of Management and Engineering, University of Padua, Italy

ABSTRACT

Storytelling and narrative are essential components in diverse contexts ranging from entertainment to culture and knowledge. With the progressive evolution of communication, stories also started to be transmitted through digital media. In particular, the emerging and promising field of sonic interactions in virtual environments (SIVE) fosters the creation of immersive experiences in which personalization of user interaction and narrative content can help create engaging and interactive storytelling. The insights presented here aim to define a state of the art of available interactive audio technologies for storytelling and to illustrate the role of audio in augmented and virtual reality (AR/VR) personalized experiences by reviewing a selection of published works. Moreover, we will try to disclose the most important elements enabling an immersive experience and to explain how interactivity, emotional and personalized content help convey messages in the context of serious storytelling. Finally, we will investigate the limitations of the available methods by also outlining some promising research directions.

Keywords: *sonic interaction design, virtual/augmented reality, immersive audio, storytelling*

1. INTRODUCTION

With the growth of digital communication, storytelling evolved in different formats and supports until the advent of immersive media, such as 360° cinema and augmented and virtual realities (AR/VR). In this context, audio, soundscapes, and the acoustic environment enhance the experience's perceived immersion, especially in AR/VR [1]. Together, they form a common ground called mixed reality, MR hereafter. With MR we refer to an environment that combines physical reality and digital content enabling interaction with and among real-world and virtual objects. Personalization is a key aspect that can be applied to narrative and content design and the technological platforms that must be carefully implemented to deliver effective storytelling.

In this work, we focus on some meaningful applications and insights of immersive and interactive personalized audio storytelling designed explicitly for MR. In particular, this paper focuses on *serious storytelling*, namely storytelling designed with a specific purpose different from entertainment, where the narrative refers to a serious context, soliciting a thoughtful process. Various domains can benefit from an incisive serious storytelling design, including cultural heritage, health and medicine, education, and ethics.

While evaluating storytelling, [2] introduces four essential components with a specific focus on digital storytelling:

- **Perspective**: the perspective that an author wishes to convey with the story, including emotions, presentation, and the encoding/decoding process.
- **Narrative**: the story content by the storytelling.
- **Interactivity**: a peculiarity of digital media that can be implemented, e.g., with multiple storylines design.
- **Medium**: the influence of the type of media in the user's comprehension of the storytelling message.

*__Corresponding author__: privitera.alessandrogiuseppe@spes.uniud.it.

To explore the personalization issue, we rely on the theoretical framework offered by *Sonic Interaction in Virtual Environments* (SIVE), a new field of study introduced by [3], which refers to "the study and exploitation of sound being one of the principal channels conveying information, meaning, aesthetic and emotional qualities in immersive and interactive contexts". In particular, SIVE proposes three top-level categories that must be addressed: Immersion, Coherence, and Entanglement. To Slater's definition [4], immersion refers to the overall subjective credibility of a virtual environment (VE) and the quality of a simulation in providing the sensation of "being in a real place" and measures the technological level and its enactive potential, describing the degree of influence of the system on the user. Coherence measures the effectiveness of the sonic interaction design in terms of the plausibility of the rendering, the interactions, and possible behaviors in the VE. Finally, Entanglement describes the level of mutual interaction of the user, the technology, and the content in the *locus of agency*: in other words, the active contributions of each actor in the virtual and real environment. In Section 2 are presented works on immersive audio storytelling, and in Section 3 personalization is discussed in terms of technology and storytelling content.

2. IMMERSIVE AUDIO STORYTELLING

Immersive storytelling deals with the use of MR in storytelling, aiming at eliciting a sense of immersion and presence. In various works, content personalization enhances user experience in MR and helps create a sense of immersion. Adapting a classification proposed by [5], based on the technological platform and flexibility level of visual MRs, audio MRs experiences can be considered by using the degree of immersion, i.e. non-immersive when audio content is conveyed from a speaker and not spatialized, semi-immersive when audio content is conveyed through multichannel speakers and fully immersive when user is fully immersed in the media. While audio is considered one element of a complex systemic experience (i.e., audio/visual, audio/haptic, audio in service of system feedback), it usually has a secondary role in the overall evaluation. Additionally, very few contributions evaluating audio-specific experiences are available. In most cases, the specific contribution of the audio component is difficult to evaluate. Especially in immersive mobile applications using headphones, user tracking is crucial for conveying a personalized and immersive experience. The most common tracking method selects and personalizes audio and storytelling content based on head position and orientation in space using Inertial Measurement Units(IMU) sensors, computer vision, or a hybrid approach called Simultaneous Localisation and Mapping (SLAM) [6]. Two notable examples of tracking in audio AR for enhancing storytelling are War Children [7], a museum installation in which WWII stories told by eyewitnesses can be viewed through augmented reality. Regarding user interaction, many works use Graphical User Interface on a device screen, often relying on a smartphone as an interaction device. Other studies use tangible devices, such as commercial game controllers [8] or haptic interfaces [9]. Ec(h)o [10] uses a custom build device consisting of a wooden cube with colored faces whose position is detected by a camera for controlling an interactive audio guide. User preferences are tracked to provide meaningful audio content to the user. In an interactive installation, [11] uses a three-step staircase with pressure sensors to detect incoming visitors. Moreover, [12] implemented a custom-built pipe controller with a button and a 9-axis IMU to control movement in an audio VE.

Immersion and entanglement can also be enhanced by collaborative and social experiences in various fields, particularly when collaboration aims to improve socialization, as in the cases of inclusion, cooperative learning, and creativity. Such fields include music production and performance, cultural heritage, and psychology [13]. In Traces [14], a collaborative approach is encouraged in an artistic installation/touristic guide for St. Fagan Historical Museum in Wales. At the beginning of their visit, tourists must choose between two possible partial narratives that can only be completed by asking and conversing with other users. The Woods [15] is an AR game employing a four-channel audio spatialization that provides players with face-to-face interactions in pursuit of a shared goal, with the intention of fostering empathy and positive social impact through collaborative problem-solving.

3. PERSONALISED SONIC EXPERIENCE

To personalise audio in MR, it is crucial to coordinate the customization of the technological infrastructure and the storytelling, which should be considered a complex entangled system to achieve an effective and comprehensive evaluation [3].

3.1 Personalization in Audio MR Technologies

Personalization of the acoustic environment can significantly enhance the experience in audio MR, particularly

in terms of plausibility or the degree of congruence between the simulation and the listener's expectations [16]. Technological advancements should be closely correlated with other aspects, such as interaction design and content personalization, to elicit a coherent entanglement between humans, technological platforms, and interaction features. Considering specific contributions, an interesting approach in this direction is using individual Head Related Transfer Functions (HRTF) instead of diffused generic dummy head HRTFs that, in conjunction with reverberation and head tracking, can significantly reduce localization error and enhance externalization. Various works, including [17], discuss computationally efficient methods for HRTF synthesis in real-time, starting from a limited set of anthropometric parameters. A comprehensive discussion on personalised auralisation for SIVE can be found in the first part of the SIVE book [3]. Moreover, [18] introduces the idea of audio-augmented virtuality to evaluate the impact of various reverberation algorithms on co-immersion in VEs, highlighting a limited but significant effect of real-time digital reverberators over static spatialized sound.

3.2 Personalization in Storytelling

Interactive and non-linear narratives can enhance the sense of presence in MR, particularly in the dimensions of engagement and immersion, by affording active participation from both users and narrators [19]. Various techniques can be used to achieve nonlinearities, such as inverting the chronological order of events, creating parallel storylines, and dynamically modifying and adapting the narrative using artificial intelligence (AI). Many studies use non-linear narratives to permit users to navigate the augmented physical environment or virtual space freely. Frequently, user position tracking is used to determine the storyline. In this context, content can be predetermined or customized in various ways (point of view in space, content presentation order, dynamically generated elements, and directions) in response to a user's movement or explicit command. In [20], city tour maps and touristic information are generated dynamically based on the user's navigation path. In addition, SARIM [21] associates sound zones with audio samples corresponding to a particular artwork or historical artifact.

Though AI is an expanding and promising field of study, AI content personalization in the audio storytelling domain has very few specific contributions. A relevant example is [22], which employed AI to adapt storytelling in a VR video game based on the Baba Yaga myth with spatial audio. In the game, the user's actions in the virtual world are used to personalize the storyline and virtual avatar behaviors and emotional responses. Moreover, [23] overviews AI techniques for accessibility in cultural heritage, including interactive storytelling and an audio guide. The authors reported the lack of AI in accessibility contexts for cultural heritage and museum domains.

4. CONCLUSION

Utilizing the scientific context of SIVE, this paper introduces several specific works in order to highlight common traits and valuable examples for immersive, coherent, and entangled narrative experiences. Audio personalization can enhance storytelling through a process that combines the personalization of both technological tools and content to increase engagement and plausibility, respectively. These advancements should be considered in a tight relationship with an entanglement between humans and technology and require new methods for investigating sonic interactions as a pervasive design paradigm. In this direction, personalized, non-linear narratives and sonic interaction work together to create a rich environment that generates a sense of presence in the VE.

5. REFERENCES

[1] A. C. Kern and W. Ellermeier, "Audio in VR: Effects of a Soundscape and Movement-Triggered Step Sounds on Presence," *Frontiers in Robotics and AI*, vol. 7, Feb. 2020.

[2] M. S. Meadows, *Pause & Effect*. Upper Saddle River, NJ: New Riders Publishing, Sept. 2002.

[3] M. Geronazzo and S. Serafin, eds., *Sonic Interactions in Virtual Environments*. Springer International Publishing, 2023.

[4] M. Slater, B. Spanlang, and D. Corominas, "Simulating virtual environments within virtual environments as the basis for a psychophysics of presence," *ACM Transactions on Graphics*, vol. 29, pp. 1–9, July 2010.

[5] M. K. Bekele, R. Pierdicca, E. Frontoni, E. S. Malinverni, and J. Gain, "A survey of augmented, virtual, and mixed reality for cultural heritage," *Journal on Computing and Cultural Heritage*, vol. 11, pp. 1–36, Mar. 2018.

[6] T. A. Syed, M. S. Siddiqui, H. B. Abdullah, S. Jan, A. Namoun, A. Alzahrani, A. Nadeem, and A. B. Alkhodre, "In-Depth Review of Augmented Reality: Tracking Technologies, Development Tools, AR Displays, Collaborative AR, and Security Concerns," *Sensors*, vol. 23, p. 146, Dec. 2022.

[7] C. Zimmer, N. Ratz, M. Bertram, and C. Geiger, "War Children: Using AR in a Documentary Context," *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 390–394, 2018.

[8] W.-H. Huang, H.-M. Chiao, and W.-H. Huang, "Innovative research on the development of game-based tourism information services using component-based software engineering," *Advances in Science, Technology and Engineering Systems Journal*, vol. 3, pp. 451–459, Feb. 2018.

[9] S. Barrass, "Creative practice-based research in interaction design," *Computers in Entertainment*, vol. 6, pp. 1–17, Oct. 2008.

[10] M. Hatala and R. Wakkary, "Ontology-Based User Modeling in an Augmented Audio Reality System for Museums," *User Modeling and User-Adapted Interaction*, vol. 15, pp. 339–380, Aug. 2005.

[11] K. J. Kortbek and K. Grønbæk, "Interactive spatial multimedia for communication of art in the physical museum space," in *Proceedings of the 16th ACM international conference on Multimedia*, ACM, Oct. 2008.

[12] M. Geronazzo, A. Rosenkvist, D. S. Eriksen, C. K. Markmann-Hansen, J. Køhlert, M. Valimaa, M. B. Vittrup, and S. Serafin, "Creating an Audio Story with Interactive Binaural Rendering in Virtual Reality," *Wireless Communications and Mobile Computing*, vol. 2019, pp. 1–14, Nov. 2019.

[13] T. Deacon and M. Barthet, "Spatial Design Considerations for Interactive Audio in Virtual Reality," in *Sonic Interactions in Virtual Environments*, pp. 181–217, Springer International Publishing, Oct. 2022.

[14] S. Huws, A. John, and J. Kidd, "Evaluating the affective dimensions of Traces-Olion: a subtle mob at St Fagans National Museum of History, Wales," in *2018 3rd Digital Heritage International Congress (DigitalHERITAGE) held jointly with 2018 24th International Conference on Virtual Systems Multimedia (VSMM 2018)*, pp. 1–8, 2018.

[15] S. Swearingen, K. Swearingen, F. C. Halac, S. Ammannagari, and M. Hall, ""the woods" AR game," in *Entertainment Computing – ICEC 2021*, pp. 501–504, Springer International Publishing, 2021.

[16] A. G. Privitera, F. Fontana, and M. Geronazzo, "On the Effect of User Tracking on Perceived Source Positions in Mobile Audio Augmented Reality," in *CHItaly '23*, (Torino, Italy), Association for Computing Machinery, 2023.

[17] M. Geronazzo, J. Y. Tissieres, and S. Serafin, "A Minimal Personalization of Dynamic Binaural Synthesis with Mixed Structural Modeling and Scattering Delay Networks," in *Proc. IEEE Int. Conf. on Acoust. Speech Signal Process. (ICASSP 2020)*, (Barcelona, Spain), pp. 411–415, May 2020.

[18] D. Fantini, G. Presti, M. Geronazzo, R. Bona, A. G. Privitera, and F. Avanzini, "Co-immersion in Audio Augmented Virtuality: the Case Study of a Static and Approximated Late Reverberation Algorithm," in *IEEE Transactions on Visualization and Computer Graphics - under review*, 2023.

[19] B. Bostan and T. Marsh, "Fundamentals Of Interactive Storytelling," *Summer*, vol. 3, pp. 19–42, Aug. 2012.

[20] C. Boletsis and D. Chasanidou, "Audio Augmented Reality in Public Transport for Exploring Tourist Sites," in *Proceedings of the 10th Nordic Conference on Human-Computer Interaction*, NordiCHI '18, (New York, NY, USA), p. 721–725, Association for Computing Machinery, 2018.

[21] F. Z. Kaghat, A. Azough, and M. Fakhour, "SARIM: A gesture-based sound augmented reality interface for visiting museums," in *2018 International Conference on Intelligent Systems and Computer Vision (ISCV)*, pp. 1–9, 2018.

[22] L. Cutler, E. Darnell, N. Dirksen, A. Tucker, S. Stafford, E. Oh, A. Nagpal, E. Lee, and N. Ladd, "From Quest to Quill: Pushing the Boundaries of VR Storytelling in Baobab's Baba Yaga and Namoo," in *ACM SIGGRAPH 2021 Production Sessions*, SIGGRAPH '21, (New York, NY, USA), Association for Computing Machinery, 2022.

[23] G. Pisoni, N. Díaz-Rodríguez, H. Gijlers, and L. Tonolli, "Human-Centered Artificial Intelligence for Designing Accessible Cultural Heritage," *Applied Sciences*, vol. 11, p. 870, Jan. 2021.