The Egocentric Audio Perspective in Virtual Environments

Michele Geronazzo[1,2]

[1] Dept. of Engineering and Management - University of Padova, Italy, michele.geronazzo@unipd.it
[2] Dyson School of Design Engineering - Imperial College London, Country, United Kingdom, m.geronazzo@imperial.ac.uk

ABSTRACT

Although sound is an essential component of the grammar of digital immersion, relatively little compared to the visual domain has to be done to investigate the role of auditory space and environments. Nowadays, there is increasing consensus on the importance of spatial sound, especially in virtual reality (VR). Technologies for spatial audio rendering are now able to convey perceptually plausible simulations with stimuli that are reconstructed from real-life measurements or historical sites getting closer to a virtual experience indistinguishable from natural reality. This is made possible by a high level of personalization in acoustic transformations caused by the human body interacting with the sound field generated in room acoustic computer simulations. I believe it is time to tightly relate the real to the virtual listening experience from an egocentric audio perspective. The term audio identifies an auditory sensory component, implicitly recalling those technologies capable of immersive and interactive rendering. The term egocentric refers to the perceptual reference system for the acquisition of multi-sensory information in immersive VR technologies as well as the sense of subjectivity and perceptual/cognitive individuality.

Keywords: egocentric audio, virtual reality, sonic interaction design

1. INTRODUCTION

[1] A large body of research in computational acoustics focused on the technical challenges of quantitative accuracy characterizing engineering applications, simulations for acoustic design, and treatment in concert halls. Such simulations are very expensive in terms of computational resources and memory, so it is not surprising that the central role of perception in interactive and immersive rendering has gradually come into play. Nowadays, there is increasing consensus towards the essential contribution of spatial sound, also in virtual reality (VR) simulations. Technologies for spatial audio rendering are now able to convey perceptually plausible simulations with stimuli that are reconstructed from real-life recordings or historical archives, as for the Cathédrale Notre-Dame de Paris before and after the 2019 fire [2], getting closer to a virtual version indistin- guishable from the natural reality. This is made possible by a high level of personalization in modeling human body acoustics interacting with room acoustic computer simulations and non-acoustic factors such as familiarity and adaptability of listening.

At the terminological level, I would like to introduce a new perspective that relates the two listening experiences (i.e. real and virtual), called the egocentric audio perspective. In particular, I refer to the term audio to identify an auditory sensory component, implicitly recalling immersive audio technologies and auralization.

The term egocentric refers to the perceptual reference system for the acquisition of multi-sensory information in immersive VR technologies as well as the sense of subjectivity and perceptual/cognitive individuality that shape the self, identity, or consciousness. In such a context, immersiveness is a dynamic relationship between physical and meaningful actions by the listener in virtual environments (VEs). Performing bodily practices such as walking, sitting, talking, grasping, etc. provide meaning to virtual places, objects, and avatars.

2. THE EGOCENTRIC AUDIO PERSPECTIVE

The search for lower bounds such as the "perceptually authentic" audio-visual renderings is an ongoing process. Continuous knowledge exchange between psychophysical research and interactive algorithms development allows to test new hypotheses and propose responsive VR solutions.

2.1 Spatial Centrality

Starting from an egocentric spatial perspective of immersive VR, the learning and transformation processes of the listeners occur when their attention is guided towards external virtual sounds, e.g. the "out-of-the-head" and externalized stimuli. The three-dimensionality of the action space is one of the founding characteristics of immersive VE. Considering such space of transmission, propagation, and reception of virtually

---

[1] Extracted from Ch.1 of [1], that covers the topic.

simulated sounds, sonic experiences can assume different meanings and open up to many challenges [1].

Events in multisensory VEs are echoic, i.e., they produce auditory delays and resonances imprinted by the spatial arrangements of the avatar-VE configurations depending on the acoustical characteristics of the simulated space. Since the sound is received from the first-person point of view (generally referred to as 1PP), auralization has to take into account contextual information relating to spatial positions of sound events and self-produced sounds within the avatar's virtual body, creating a sense of proximity and meaningful relations for the listener. The concept of enaction, i.e. it is impossible to separate perception from action in a systematic way, shapes the VR experience by considering an embodied, environmentally situated perceiver with sensorimotor processes tightly connected with the exploratory action.

2.2 Binaural Hearing

Externalization can be considered a necessary condition for the place illusion, being immersed in that virtual acoustic space. From [3], one can learn that ambient reverberation and sensorimotor contingencies are key indicators for eliciting a sense of externalization, whereas head-related transfer function (HRTF) personalization and consistent visual information may reinforce the illusion. Accordingly, it is important to explore dynamic relations depending on specific links between evolving states of the listener-VE system during the VR experiences. Moreover, huge individual-based differences in the perception of externalization require an in-depth exploration of several individual factors such as monaural and binaural HRTF spectral features, and temporal processes of adaptation.

2.3 Quality of the Mediated Experience

There is still no adequately in-depth knowledge of the technical-psychological-cognitive relationship regarding spatial hearing and multisensory integration processes linked to plausibility and technological mediation. In particular, the sense of presence in VR is negatively affected if the experience is irrelevant to the listener. This means that the mediating action of the immersive technology might result in a break in presence that can hardly be restored after a pause [4]. These cognitive illusions depend, for example, on the level of hearing training, and familiarity with a stimulus/sound environment. All these aspects reinforce the term egocentric, grounding auditory information to a reference system that is naturally processed and interpreted in 1PP.

3. HUMAN-TECNOLOGY RELATIONS

Following the recent proposal of a post-phenomenological framework by Verbeek's concept of technological mediation and its extensions [5], one can identify mixed intentionality between humans and technology within VEs. We can describe the cooperation between human and technological intentionality to reveal a (virtual) reality that can only be experienced by technologies, by making accessible technological intentionality to human intentionality. The sonic information from intentional active listening is anchored to an egocentric perspective of spatiality that allows the understanding of an acoustic scene transformed by the listener's actions/movements. This process can be mathematically formalized with the active inference approach and its recent enactive interpretation [6]. Immersive audio technologies are quantitatively integrating prediction through probability and generative models as a function of the listener's beliefs and expectations to contribute to the listener's internal representation in both spatial and semantic terms, eliciting a strong sense of presence in VR.

4. CONCLUSIONS

In the entanglement within the relational network of listener-reality-simulation, configurations and actors are dynamically defined in a situated and embodied manner. Exploring the evolution of such a configuration network might enable an active search for the egocentrically meaningful experience. However, the communication between the avatar and the listener, the virtual and the physical is challenging. Considering the avatar as part of a VE configuration, I can formulate one of the fundamental questions: if we might be able to handle mediation, where/who is in charge of that? I introduced a performative perspective questioning the a priori and fixed distinctions of certain representationalism between avatar and self, technology agency and listener, physical reality, and virtuality. These boundaries have to be drawn in situated and embodied actions, which makes them dynamic and temporary. The exploration of how, when, and why such boundaries trace identity, agency, and VEs is the core challenge of the proposed theoretical framework.

ACKNOWLEDGEMENTS

5. REFERENCES

[1] M. Geronazzo and S. Serafin, Eds., *Sonic Interactions in Virtual Environments*, 1st ed. Springer International Publishing, 2022.

[2] B. F. G. Katz and A. Weber, "An Acoustic Survey of the Cathédrale Notre-Dame de Paris before and after the Fire of 2019," *Acoustics*, vol. 2, no. 4, Art. no. 4, Dec. 2020.

[3] V. Best, R. Baumgartner, M. Lavandier, P. Majdak, and N. Kopčo, "Sound Externalization: A Review of Recent Research," *Trends in Hearing*, vol. 24, Jan. 2020.

[4] M. Slater, A. Brogni, and A. Steed, "Physiological Responses to Breaks in Presence: A Pilot Study," *Presence 2003: The 6th annual international workshop on presence*, vol. 157, p. 4, 2003.

[5] J. Vindenes and B. Wasson, "A Postphenomenological Framework for Studying User Experience of Immersive Virtual Reality," *Frontiers in Virtual Reality*, vol. 2, p. 40, 2021.

[6] M. J. Ramstead, M. D. Kirchhoff, and K. J. Friston, "A tale of two densities: active inference is enactive inference," *Adaptive Behavior*, vol. 28, no. 4, pp. 225–239, Aug. 2020.