Motion Reactive Sound Generation System for Immersive CAVE Environment: a Design Perspective

Stefania Palmieri
Design Department
Politecnico di Milano
Milan, Italy
e-mail: stefania.palmieri@polimi.it

Alessandro Ianniello
Design Department
Politecnico di Milano
Milan, Italy
e-mail: alessandro.ianniello@polimi.it

Mario Bisson
Design Department
Politecnico di Milano
Milan, Italy
e-mail: mario.bisson@polimi.it

Giovanni Barone
Design Department
Politecnico di Milano
Milan, Italy
e-mail: giovanni2.barone@mail.polimi.it

***Abstract*— The objective of this paper is to bring attention to the world of immersive technologies and in particular to those technologies that when put in system create the Cave Automatic Virtual Environment (CAVE) system, with the aim of including within the design of these systems, the sound component, and the interactivity connected to them. Different degrees of immersiveness and fruition systems can be identified among mixed reality and immersive reality technologies. Depending on the technologies employed, immersive digital experiences are on a spectrum ranging from Virtual Reality (VR), where the individual is totally immersed in a fictional reality, to Augmented Reality (AR), where the experience is of a hybrid reality characterized by the superimposition of digital and virtual content on the real environment. Within this spectrum, where at one extreme VR can be found, while at the other AR, CAVE systems are in a central position, definable as Augmented Virtuality. Research related to the development of such a system has, to date, privileged the visual and motion capture aspect, i.e., that which causes visual and virtual elements to be responsive to the movement of the individual, making it the main element of interaction. The purpose of the paper is to bring attention to a factor not yet central to the design of CAVE systems, namely that of the sound component, with the goal of making it a new interactive element, on par with the visual component, with the aim of increasing the degree of immersiveness of the experience. The integration of this new interaction makes it possible to create systems capable of generating sound outputs, based on the movements of the human being, and therefore motion reactive, thus relying on the same motion capture technology that underlies the visual component of CAVE, and it is believed that the inclusion of this new component can be used to increase the degree of immersiveness of the experience. the final output of the paper will be to propose design directions and tools useful to the 'integration of the sound reactive component in CAVE, which remains to date a fairly unexplored terrain.**

Keywords-audioreactive; CAVE; immersive experience; sound generation; design.

I. INTRODUCTION

In recent years, to increase the degree of immersiveness of an experience, technologies are employed and integrated that allow the user to go beyond what is tangible and visible in physical reality, adding additional layers of perception. In the world of Mixed Reality (MR), the use of different technologies can cause different types of experiences to develop, which are schematized along two different axes: immersion in the virtual environment, and perception of the real world as displayed in Figure 1.

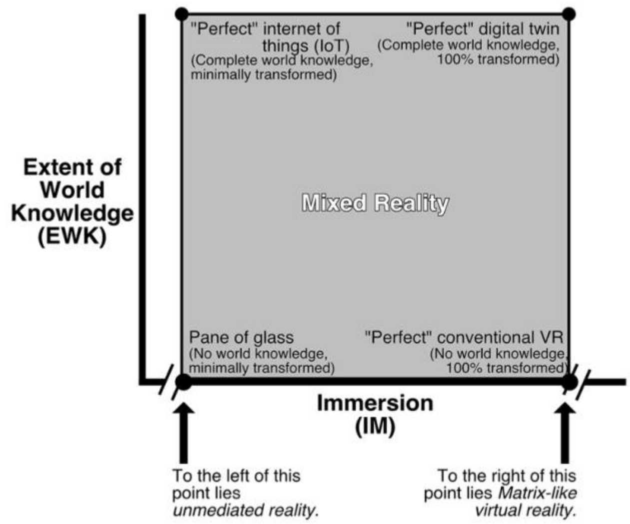

Figure 1. The graph defines Mixed Reality technologies in relation with the technological system's level of immersion, i.e., its capability of making people perceive a sense of presence, and the users's ability to perceive the real world through the technology. (Credits: Skarbez et al., 2021.)

These two dimensions create a broad spectrum of immersive experiences [1].

The different technologies available on the market determine different types of experiences: products such as

 ISBN: 978-1-68558-048-3

HoloLens enable MR based on a high perception of the external environment, which is enriched by virtual elements. Other technologies, such as Oculus, on the other hand, make a highly immersive reality experienceable, but with a minimal level of environmental perception [1].

Also found in the MR spectrum are CAVEs, integrated systems of different technologies that enable the creation of Augmented Virtuality (AV) experiences. These systems provide a high degree of spatial perception and have the potential to achieve an equally high level of immersiveness when implemented through the organic integration of new technologies [2].

The goal of the research is to increase the degree of immersiveness of these systems through the integration of sound generation machines (synths, drum machines, samplers), managed through motion tracking cams, already installed in CAVE systems. By combining these two technologies, it is possible to integrate an additional sensory component into the immersive experience, activating new types of interactions, in order to create mixed realities.

These new interactions can be starting points for thriving experience design scenarios by professionals involved in User Experience (UX) Design and Interaction Design, applicable in the world of entertainment, show business, and music production [3].

Finally, the objective of the paper is to theorize design guidelines for CAVE environments that implement Motion Reactive Sound Generation (MRSG), with the aim, following experimental phases, of creating a design framework that facilitates a systematic and extensive technological implementation of CAVE technology.

From a process of literature review and case studies analysis, information was extracted to generate the research question and consequently the thesis. Section I of the paper makes explicit the methodology used to carry out the research; Section II is devoted to defining the concept of immersiveness and evaluating technologies that can enable immersive experiences. Section III is focused on CAVE systems, highlighting their enabling technologies, structure, functionality, and potential. In section IV the area of audioreactivity is addressed, exploring the ways in which this type of systems can be integrated with CAVE environments; section V focuses on the role of design in the design of multisensory experiences based on the technologies mentioned before; section VI is devoted to the description of the MRSG system and the explication of guidelines for its design. This is followed by section VII, where the outcomes, the future opportunities and the limits of the research are summarized.

II. METHODOLOGY

This contribution is based on a literature review process, conducted on Google Scholar and Web of Science platforms searching through keywords such as audio reactive technologies, multisensory experience, music and video commutation, user experience, and audio-driven generative design. The parameters to select the papers have been the presence of experimental stage, their systemic design approaches and impacts, the exploitation of visual and audio content as fundamental features for the research. It aimed at mapping the state of the art of academic research in the area of interest, highlighting the most promising processes to exploit the goals of the research. In addition to this activity, a case studies analysis was carried out, which is essential to understand the various concepts of mixed reality, and to give an idea of which technologies are currently available and their peculiarities.

To select the case studies, different parameters have been applied: each project should be focused on music experiences, with a particular attention to the use of electronic and Musical Instrument Digital Interface (MIDI) instruments and technologies, and on the relationship between music and visual contents. To evaluate each project, their coherence, information simplicity, scopes, and application of design methodologies have been considered. A further phase of case study analysis was conducted in the area of music generation and sound synthesis and on the ways in which these can be systematized with technologies for immersive experiences, selected applying similar parameters and evaluated following the same criteria. Finally, the content of this paper focuses on the construction of a theoretical framework, which aims to integrate the above technologies, to be tested in a subsequent research phase as shown in Figure 2.

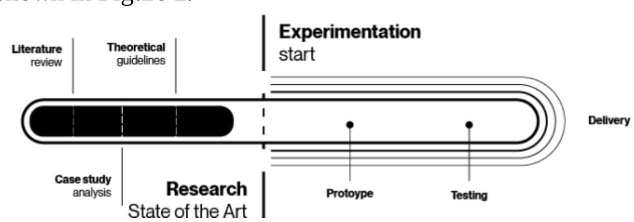

Figure 2. Review of the research process's steps highlighting its state of the art and future activities.

III. IMMERSIVE EXPERIENCE

Immersivity is a phenomenon that can be experienced when an individual is in a state of deep mental involvement [4], [5].

This mental involvement can be fostered by external stimuli, produced by technological systems; therefore, when talking about immersiveness, and especially virtual immersiveness, it is essential to also refer to the technologies that make these realities possible. Some examples of technologies are Microsoft's Hololens [6], or Meta's Oculus. These two technologies are based on AR and VR and are part of what is called Mixed Reality, as stated by Skarbez R, Smith M and Whitton MC (2021). Referring to Milgram and Kishino's [4] continuum on Mixed Reality, the different realities were placed in a diagram, to which a reference technology was assigned, as visible in Figure 3.

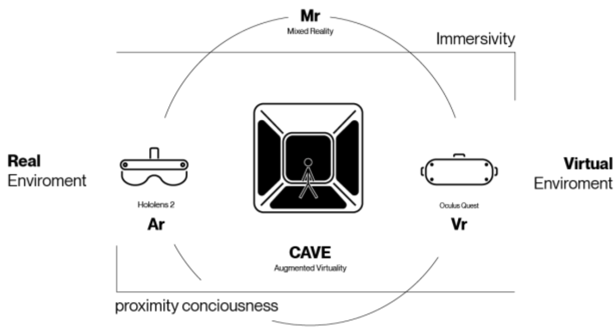

Figure 3. Review of the reality-virtuality continuum by Milgram e Kishino Milgram, P., & Kishino, F. (1994) with display's technologies examples.

As it can be observed in this scheme, the technologies that are part of mixed reality polarize on the right, where devices similar to Oculus can be found, i.e., that take advantage of immersive virtual environments, in which the user enters the virtual world and can interact with it through a control pad [6]. This technology offers a high degree of immersivity since everything can be observed at the moment of use being part of the virtual world, allowing to have a first-person Point of View (PoV) and get into the thick of the action [7].

At the opposite pole, that is, on the left side of the diagram, technologies useful for the reproduction of AR experiences can be highlighted, for example, a smartphone can be a device of AR experiences: just think of the role it plays in the application and display of Instagram filters, the positioning of digital models in space. More complex devices such as HoloLens, on the other hand, allow for greater degrees of immersiveness.

Having the visor as the only touchpoint, the user can interact with digital elements directly with their hands, making the user experience natural and intuitive. The great advantage of these technologies is the high degree of fusion with the real environment, which allows the user to have perception of his or her position in space and consequently to be able to interact with elements and individuals belonging to physical reality.

IV. CAVE SYSTEM

In a central position on the spectrum between the pole of augmented reality and that of virtual reality are CAVE systems, which belong to AV environments [8]; in these systems, the user's experience takes place within walls (the number of which varies from three to four), on which a virtual world is projected, equipped with a motion tracking system that becomes an intermediary between the virtual world and reality [8]. The advantage of these systems is the ability to move freely in space and to be able to interact with the virtual world thanks to these systems, which involve fewer constraints than those associated with wearable devices [8]. Before they were developed, and thus in earlier versions of CAVE technology, mediating devices were in fact used, which allowed interaction with virtual elements.

A CAVE has basic characterizing components and can be described, in essence, as a cube within which the user is placed. The cube shape is used to roughly simulate that of a sphere, so as to replicate the kind of view one has in the real world. The perceptual-visual component of the CAVE is made possible thanks to projectors, which can be placed either inside or outside the structure, (outside in case the walls are transparent, and the projectors point at them) thanks to which the faces of the cube are transformed into screens that envelop the viewer, allowing him to perceive himself immersed in the virtual world [9].

The interactive component, on the other hand, is made possible, as previously stated, by products, such as Microsoft's Kinect, which make motion tracking possible. These systems, placed in dialogue with a computer, make it possible to track the position of the individual in space and capture his or her movements with a level of sophistication that allows discernment of the movement of small portions of the body, such as head or even, in the most advanced devices, eyes (eye-tracking) [10]. All of this information, is subsequently translated into data that the computer uses to enable authentic interaction between the user and the virtual world, creating, for example, dynamic and responsive scenarios influenced by the user's movements.

V. AUDIOREACTIVITY

One component that often takes a back seat to the visual compartment is the world of audio. There are many articles that discuss the benefits that the sound dimension can bring within immersive systems, such as Adaptive Music Composition for Games or music mandala mindfulness [11], [12] in which they explain how the reactive audio component can increase immersiveness in virtual video game worlds in no small part.

Other articles such as Designing generative sound for responsive 3D digital environment interaction [12], on the other hand, talk about the kind of interaction that generative audio can offer within immersive 3D spaces, and the importance of the user having awareness of the spatial dimension around him or her and how generative audio, which responds in real time to movements, can be remarkably immersive [13].

By audioreactive it is mostly meant the whole world that has to do with visual elements that are generated or changed to the rhythm of music [14]. More specifically, thanks to various software, such as for example Touchdesigner, which translates sound into digital inputs, which are processed to go into the mathematical parameters with which shapes, colours, etc. are described [15], [16].

This new form of art is becoming more and more established in the world of entertainment and, in particular, in the field of live music, but it is rooted in the history of Nourathar music by Mary Hallock-Greenewalt: it evolves primarily with the goal of creating synaesthetic musical experiences, which has been pushed to the limit in experiences dedicated exclusively to this world, in which the visuals are not background of the music, but are mixed becoming one work. Examples of such installations may be the works of Jon Weinel [17] or the work of the Berlin-based artist couple Quadrature.

VI. MRSG SYSTEM

Starting from audioreactive systems and their operative functions, through this research the aim is to hypothesize a change in the process of application of such technologies: if in the field of VJing music is an input to create visualizations, through the hypothesized system the scope is to create music using visual sources, and to be precise cameras that integrates motion capture technology to control sound generating machines [18] through the MIDI communication system. Taking inspiration from research such as EyesWeb: Toward Gesture and Affect Recognition in Interactive Dance and Music Systems [19] it is intended to integrate this system within a CAVE equipped with three motion capture sensors, crossed so that a mix of synaesthetic interactions can be created in a sound mixed reality situation, so that the user can control and interact, through his or her own movement [20], [21], with both the visual part of the virtual world and the sound part, thanks to virtual musical instruments that will produce generative music.

An initial hypothesis, which is shown in Figure 4, envisioned that the person inside the cave could actually with his or her movement make the synthesizer play by producing notes with body movement, but in a multi-user scenario the risk of creating sound confusion dictated the need to investigate new possible solutions.

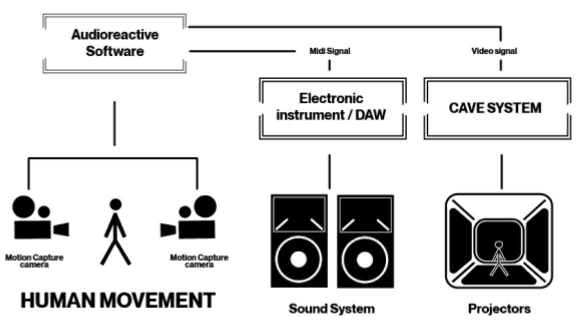

Figure 4. The scheme highlights the routing of the signals starting from the human movements, which become input for the audioreactive software. The software elaborates the input, dividing it into two different kinds of signals (MIDI and video), which finally are elaborated by the electronic instrument and the CAVE system.

Therefore, it was thought to provide a more guided and constrained experience, allowing users to manage simple parameters such as opening or closing filters, sound waveform and rhythm. The experience can be structured in two ways: using a motion sensing system given by the CAVE's cameras, with which to manage the different parameters of the virtual musical instrument, which can be assigned to each moving part of the body: limbs and head a parameter such as Low-Frequency Oscillators (LFO) speed, filter cut-off and sequencer speed, or even tuning the note itself; or, dividing the CAVE space into responsive zones dedicated to each of the instruments. In this case, as moving within the space, the user can give more or less prominence to the instrument is stepping on, while moving the hands on the Z axis will allow to modulate parameters and rhythms of melody playback. Through experimentation in the field, the interaction modes of this immersive, multisensory model will be defined in detail. The experimentation that is already being structured includes an initial phase of technical realization of the software that will act as a bridge between Mixed Reality and a Digital Audio Workstation (DAW), thus transforming the interactive system between CAVE and a human being into a true immersive MIDI controller.

VII. DISCUSSION

The contribution proposed a critical analysis and mapping of the state of the art of research in the field of audioreactive applied to immersive AV contexts, such as CAVE, with the ultimate goal of defining an implementation model replicable in other contexts with the given pre-requisites. To reach this aim a process of literature and of case study analysis were exploited.

From a design perspective, these types of immersive environments can constitute a tool that takes on several functions: on the one hand, they are spaces that have the potential to increase communication and dissemination capabilities, enabling new forms of interactive dialogue; they are functional for study and experimentation phases during the development stages of a project, facilitating the immersive visualization of certain formal elements; and finally, they allow for the application of a co-design approach to design activities, enabling the presence of multiple people interacting within the immersive environment.

Through the integration of MRSG's technologies, these potentials can be expanded to all those areas where the sound component constitutes a design element of central importance.

The research presented here involves the development of successive phases aimed at testing different configurations and applications of the MRSG system with the aim of understanding its technical limitations and systemic peculiarities; in the subsequent prototyping phase the communication between the motion capture system and the sound generation machines will have to be tested through the creation of the prototype of a dedicated platform. At this stage it will be necessary to define the interaction between the human being and the system, which will have to be tested during a UX testing phase, culminating in the delivery of a questionnaire and targeted interviews. The last step will be devoted to formalizing a systemic technology integration model, which will make replication possible.

Following this process, it will be possible to take the hypothesized and tested model to various scenarios, going to replicate the hardware and software system, contextualizing it in the target environment and according to specific requirements.

The application of these technologies and systems to immersive environments can enhance the level of multisensory and cognitive immersion that a user should be able to experience, even though at the time of paper’s writing it seems that those kinds of application are still under researched and underdeveloped, due to the complexity of the needed systems and their integration within the same space.

For sure, the main issue and challenge is going to be encountered during the development stage of the project, when the different hardware should be linked to each other through to exploitation of ad-hoc software.

REFERENCES

[1] R. Skarbez, M. Smith, and M.C. Whitton, "Revisiting Milgram and Kishino's reality-virtuality continuum", Frontiers in Virtual Reality, vol. 2, pp. 1-8, 2021, doi: https://doi.org/10.3389/frvir.2021.647997

[2] H. Regenbrecht, T. Lum, P. Kohler, C. Ott, M. Wagner, W. Wilke, and E. Mueller, "Using augmented virtuality for remote collaboration", in Presence, vol. 13(3), pp. 338-354, 2004.

[3] C. Flavián, S. Ibáñez-Sánchez, and C. Orús, "The impact of Virtual, Augmented and Mixed Reality technologies on the customer experience", in Journal of business research, vol. 100, pp. 547-560, 2019, doi: https://doi.org/10.1016/j.jbusres.2018.10.050

[4] P. Milgram and F. Kishino, "A taxonomy of Mixed Reality visual displays", Proc. IEICE TRANSACTIONS on Information and Systems, 1994, pp. 1321-1329.

[5] S. Agrewal, A.M.D. Simon, S. Bech, K.B. Bærentsen, and S. Forchammer, "Defining immersion: literature review and implications for research on audiovisual experiences", in Journal of the Audio Engineering Society, vol. 68(6), pp. 404-417, 2020, doi: https://doi.org/10.17743/jaes.2020.0039

[6] S. Park, S. Bokijonov, and Y. Choi, "Review of Microsoft Hololens applications over the past five years", in Applied sciences, vol. 11(16), pp. 1-26, 2021, doi: https://doi.org/10.3390/app11167259

[7] K. Kilteni, I. Bergstrom, I., and M. Slater, "Drumming in immersive Virtual Reality: the body shapes the way we play", in IEEE transactions on visualization and computer graphics, vol. 19(4), pp. 597-605, 2013, doi: https://doi.org/ 10.1109/TVCG.2013.29

[8] M.A. Muhanna, "Virtual reality and the CAVE: Taxonomy, interaction challenges and research directions", in Journal of King Saud University-Computer and Information Sciences, vol. 27(3), pp. 344-361, 2015, doi: https://doi.org/10.1016/j.jksuci.2014.03.023

[9] D.A. Guttentag, "Virtual Reality: Applications and implications for tourism", in Tourism management, vol. 31(5), pp. 637-651, 2010, doi: https://doi.org/10.1016/j.tourman.2009.07.003

[10] B. Burger, A. Puupponen, and T. Jantunen, "Synchronizing eye tracking and optical motion capture: How to bring them together", in Journal of Eye Movement Research, vol. 11(2), pp. 1-16, doi: https://doi.org/10.16910/jemr.11.2.5

[11] A. Adolfsson, J. Bernal, M. Ackerman, and J. Scott, "Musical mandala mindfulness: a generative biofeedback experience". [Online] Available from: https://www.semanticscholar.org/paper/Musical-Mandala-Mindfulness-Adolfsson-Bernal/7db5aeb974611536134939e436be168f7905a946

[12] P.E. Hutchings and J. McCormack, "Adaptive music composition for games", in IEEE Transactions on Games, vol. 12(3), pp. 270-280, 2019, doi: https://doi.org/10.48550/arXiv.1907.01154

[13] K. Beilharz, "Designing Generative Sound for Responsive 3D Digital Environment Interaction", Proc. Computer Aided Architecture Design Research in Asia (CAADRIA'04), 2004.

[14] N. Collins, "Generative music and laptop performance", in Contemporary Music Review, vol. 22(4), pp. 67-79, 2003, doi: https://doi.org/10.1080/0749446032000156919

[15] H. Brouwer, Self-supervised Audio-reactive Music Video Synthesis: Measuring and optimizing audiovisual correlation. Delft, NL: TuDelft, 2021.

[16] J.R. Weinel, "Visualising rave music in Virtual Reality: Symbolic and interactive approaches", Proc. Electronic Visualisation and the Arts, London, 2020, pp. 78-84, doi: https://doi.org/10.14236/ewic/eva2020.13

[17] N.N. Correia, "AVOL: Towards an integrated audio-visual expression", in Journal of Visual Art Practice, vol. 10(3), pp. 201-214, 2012, doi: http://dx.doi.org/10.1386/jvap.10.3.201_1

[18] J. Weinel, "Technoshamanic visions from the Underworld II", Proc. Internet Technologies and Applications (ITA), 2017, pp. 345-346.

[19] A. Eigenfeldt, "Exploring moment-form in generative music", Proc. Sound and Music Computing Conference, 2016, pp. 1-6, doi: https://doi.org/ 10.5281/zenodo.851207

[20] A. Camurri, S. Hashimoto, M. Ricchetti, A. Ricci, K. Suzuki, R. Trocca, and G. Volpe, G, "Eyesweb: Toward gesture and affect recognition in interactive dance and music systems", in Computer Music Journal, vol. 24(1), pp. 57-69, 2000.

[21] C. Garoufis, A. Zlatintsi, and P. Maragos, "A Collaborative System for Composing Music via Motion Using a Kinect Sensor and Skeletal Data", Proc. Sound and Music Computing Conference (SMC-2018), 2018, pp. 415-421, doi: https://doi.org/10.5281/zenodo.1341738