Sonic Structures: 4E Visual Sound in Multi-User Mixed Reality

Krishnan Chandran
Immersive Experience Lab
TU Dresden
krishnan.chandran@tu-dresden.de

Lars Engeln
Immersive Experience Lab
TU Dresden
lars.engeln@tu-dresden.de

Daniel Zeidler
Immersive Experience Lab
TU Dresden
daniel.zeidler@tu-dresden.de

Matthew McGinity
Immersive Experience Lab
TU Dresden
matthew.mcginity@tu-dresden.de

ABSTRACT

Sonic Structures is a multi-user mixed reality experiment that explores the relationship between visual form and sound and music creation through the lens of 4E (embodied, enactive, embedded and extended) cognition. By extending a custom platform allowing large scale mixed reality experiences for up to 20 users with real-time audio processing and generative 3D graphics, Sonic Structures provides a sandbox environment for the real-time transformation of sound and music into persistent visual structures. Described here is the conceptual and architectural design of the project and the results of a workshop exploring the musical, educational and artistic potential of the system.

Author Keywords

Sound Art, 4E Cognition, Visual Music, Sonic Interaction, Mixed Reality, Enactive Musicality

CCS Concepts

•**Applied computing** → **Sound and music computing;** Performing arts; •**Information systems** → *Music retrieval;*

1. INTRODUCTION

Cybernetic artist Nicolas Schöffer proposed a synaesthetic mode of expression in which sonic and visual structures coexist symbiotically, transcending their individual spatiotemporal constraints [18]. When sound or music is manifest visually and with concrete shape and form, Schöffer argued, it is given a tangible persistence that sets it free from the roar of time. In this paper, we discuss how large-scale multi-user mixed reality can be used to realise Schöffer's vision. We propose a system for real-time composition of audiovisual structures - transformations of sound and music into 3D forms and structures that can be viewed and explored immersively. The proposed concept approaches the creation of these sonic structures through the lens of *4E* cognition. In particular, we describe a system that enables and encourages collaborative, performative experiences in which people create, view and explore virtual worlds through collective sound-making. To do this, we construct a software architecture supporting co-located large-scale mixed reality with head-mounted displays, coupled with methods for real-time analysis of sound and music and real-time generative graphics. This paper provides an overview of our motivation and conceptual framework, design process and the resulting technical architecture.

In the terminology of this paper, the terms *music* and *sound* are used interchangeably in the broadest possible sense, to refer to the full spectrum of musical and tonal possibilities, whether improvised or composed, harmonic, rhythmic or otherwise. We also use the term "sonic structure" as shorthand for the transformation of sound into persistent 3D visual forms.

2. RELATED WORK

2.1 Visual Structure from Sound

Visual structures that correspond in some way to music and sound can be found in a variety contexts, from musical notation [19] and compositional aids, to analytical and pedagogical tools. As a form or method of artistic expression, the transformation of sound into visual form has a history encompassing painting, film, video and dance [1].

The use of computers to programmatically transform sound into visual form can be traced to the beginning of computer graphics, and today is a staple of electronic music performances. Here, however, we are specifically interested in real-time transformations into 3D structures in a 3D virtual world (compare Figure 1). Precedents here can be found in the works of Lintermann and Shaw (*Utopia Triumphans*, 2002), Rokeby (*Surface Tension* 2009, *Quaver* 2010) and throughout the demo-scene. Use of real-time transformations in live performances have been demonstrated in recent years by artists Quayola and Tobias Gremmler or presented within the Ars Electronica DeepSpace theatre.

2.2 4E Cognition

The view of 4E cognition views cognition as fundamentally *embodied*, *enactive*, *embedded* and *extended*, and that it can

NIME'24, 4–6 September, Utrecht, The Netherlands.

Figure 1: Visionary image of a collaborative mixed reality sonic structure.

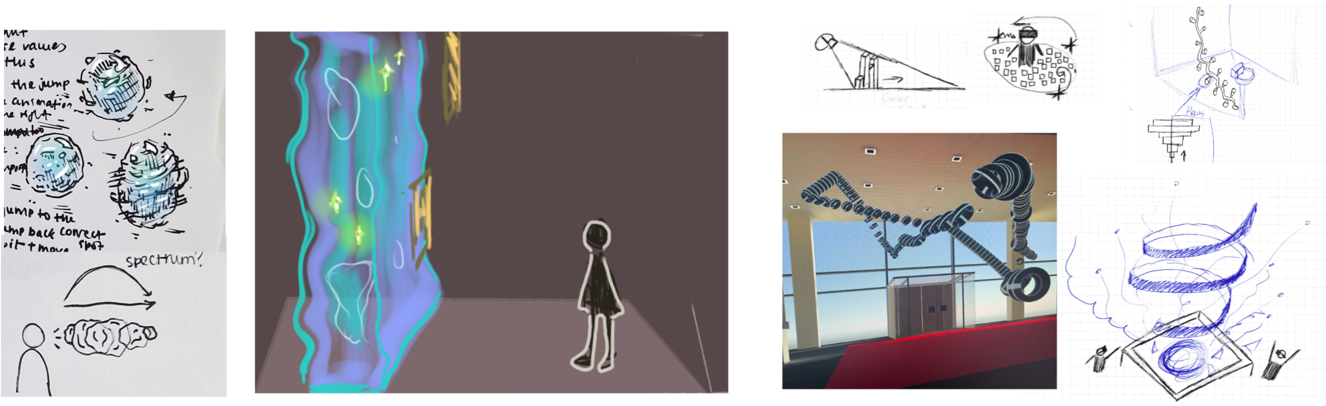

Figure 2: Sketches and Prototypes generated during the ideation workshop

only be understood by incorporating our bodies, actions and environmental interactions into the picture [14].

Visual art that emphasises a 4E mode of engagement has a rich history. Site-specific work is embedded, by nature [6]. Interactivity implies enaction, and while active participation has a history in theatre and music, it is a hallmark of computer-based media arts. And with the advent of virtual reality and motion sensors like Microsoft Kinect, embodiment has become a central feature of many recent media art. Cognitive extension, in which the cognitive processes of the viewer can only be understood as a dynamic coupling between things in the world and mental processes, could more or less be offered as a definition of art. These four traits abound in art, and as examples we offer the site-specific video walks of Janet Cardiff and George Bures Miller, the virtual reality of Char Davies, the interactive and immersive cinematic experiments of Jeffrey Shaw and the highly enactive and embodied works of David Rokeby or Stelarc. More generally, an *ecological* theory of aesthetics, in which the 4Es play primal roles is well formulated in [2].

3. CONCEPTUAL FRAMEWORK

We adopt, that musical behaviour is fundamentally embodied, enactive, embedded and extended, and so 4E cognition is an ideal framework to analyse the possibilities of the creation and viewing of sonic structures. The view that, an embodied sensorimotor engagement is essential to both the production and perception of music has found increasing theoretical and empirical grounding over the last decades [10] [9] [11]. The enactive approach to shared musical experience offers insight into the interpersonal dynamics of musical collaboration [8, 23]. The expertise of musicians in collective music-making is derived from the dynamic relationships between the patterns of perception and action that define the sonic environments [15, 16].

- *Embodied* implies that sonic structures require a corporeal engagement and relationship. They must have a physical presence and phenomenal realness [12], with a sense of persistence and continuity. They must provide affordances, presenting opportunities for physical engagement, such as pathways, openings, enclosures or supports, barriers or obstacles.

- *Embedded* describes a situated creative act influenced by the physical and social environment in which it is performed. As such, sonic structures, despite being virtual, must be embedded within the real world in a perceptually coherent and unified manner. Further, they must be responsive to their environment, respecting the spatial, and structural constraints and opportunities it affords.

- *Extended* implies a creative act that extends the mind to the outside involving instruments, objects, sound and most importantly other performers. To use the system is to couple oneself with a dynamic system strongly.

- *Enactive* implies a continuous chain of causality between performer, sound, virtual structures, audience and environment. Users being the system's engines, drive the system's behaviour.

3.1 Ideation Workshop

To explore the design space of sonic structures, a 2-week workshop (Figure 3) in collaboration with a renowned composer, Esmeralda Conde Ruiz was conducted. The focus was to explore how concepts related to emergence, complexity and realtime systems are related to sound and music. Participants were provided with a prototype implementation of our Unity-based software and were encouraged to explore the concepts that emerge when trying to use sounds/voices as interfaces to sculpt virtual worlds. During the collective music listening and reflection sessions, several concepts related to generation, emergence and dynamic coupling arose. Collaborative music-making sessions guided by the co-organiser of the workshop provided valuable insights into the collective correspondence that occurs when a large number of people are trying to improvise together. Participants generated sketches either on paper or as rapid prototypes in Unity (see Figure 2). The workshop ensured ethical participation through adherence to the recommended XR safety guidelines.

Figure 3: Workshop participants collaborating in the mixed reality system

3.2 System Properties

Here we specify some desirable properties of the system derived from the insights of the workshop:

- **Sound to structure.** The system should respond to sounds made by the voice, the body, digital or acoustic musical instruments or any other objects, and not be restricted to digital or MIDI devices.
- **Immersive**. The structures are to be created and experienced in large-scale mixed realty, embedded seamlessly in the real world (See Figure 4). These structures are to be rendered with as much perceptual realness as possible, such that they are experienced as if they were physically real and present.
- **Collaborative**. The world can be shared by many people simultaneously, allowing collaboration and group behaviour. In particular, we seek modes of interaction that encourage emergent collaborative musicality - spontaneous cooperation between users.
- **Persistent**. The visual structures must persist in time and space, and the virtual world can be progressively constructed over time. When music stops, the virtual world must remain, such that it can be explored. This invites a accumulative, constructive approach to musical and visual composition, in which the present is always contraposed against the past.
- **Causal relationships.** The causal relationships between sound and resulting structures must be unfailingly perceptible. This is not to say that they must be simple or easily understood, but the user must directly perceive themselves as the cause of events, demanding visual feedback with low latency.

3.3 Emergent Behaviours

Here we outline a handful of possible engagements that the system could be used for.

3.3.1 Musically driven creation

The visual domain is treated purely as a product of the music making process, an epiphenomenon having no causal role in the music making process. For example, when the music composition is entirely fixed, or even pre-recorded. It would also be the case when the musicians themselves are not in mixed reality and don't have access to the visual forms while making music. While an audience may see and explore the visual forms during or even after the performance, the musicians are oblivious to the forms they are creating.

3.3.2 Visually driven creation

On the other end of the spectrum is the generation and improvisation of arbitrary sounds to achieve desired visual forms. At the most extreme, no concern is given to the sonic composition, sounds are seen purely as a means of sculpting space.

3.3.3 Synaesthetic creation

In between these two extremes lies the fertile area of *synaesthetic sculpting*, in which the artist-musician simultaneously attends to both the musical and visual forms. Here we anticipate rich emergent behaviours, in which tightly coupled dynamics between musical and visual structures emerge. Indeed, mappings between sound and form can be considered potent when they yield emergent structures in both vision and sound. In particular, we seek mappings that encourage self-organisation between participants, or in other words, mappings that encourage the emergence of musical structures. For example, harmony or synchrony between two or more users can result in more prominent visual structures, creating feedback loops that can evolve noise towards music. In such loops, cause and effect between sound and vision is circular.

3.3.4 Music Skill Acquisition & Pedagogy

Joint music behaviour such as synchronisation and imitation can be visualised as shown in Figure 5. Users can synchronise in musical features pitch or phoneme to jointly sculpt a structure. Music learners can also try to imitate an expert to generate similar structures. Users can also take turns in sculpting of the structures facilitating in the learning of collaborative music behavior.

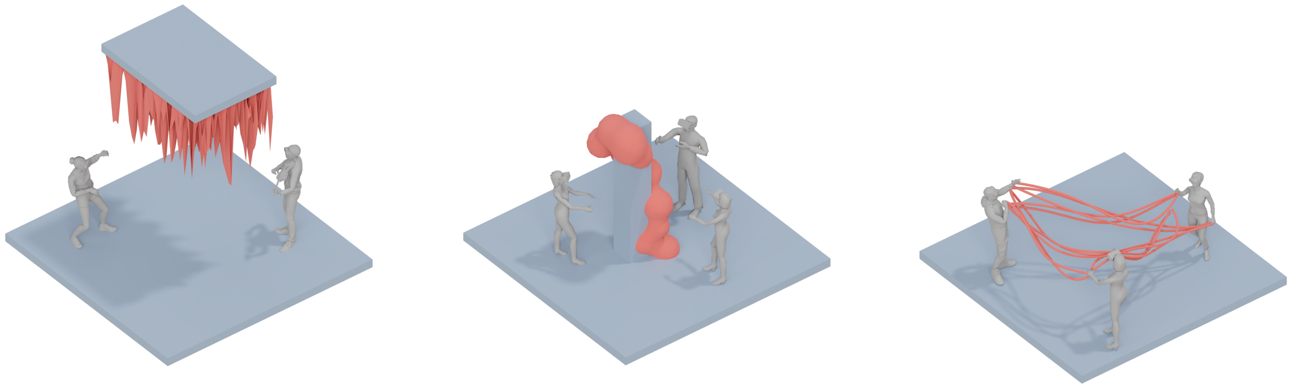

Figure 4: Users can use variations of particles, geometric forms and patterns to create virtual worlds that persists and can be actively explored. Sonic structures several limitless possibilities for audiovisual expression and it is up to the musicians to explore and experiment with different structures to create unique performances.

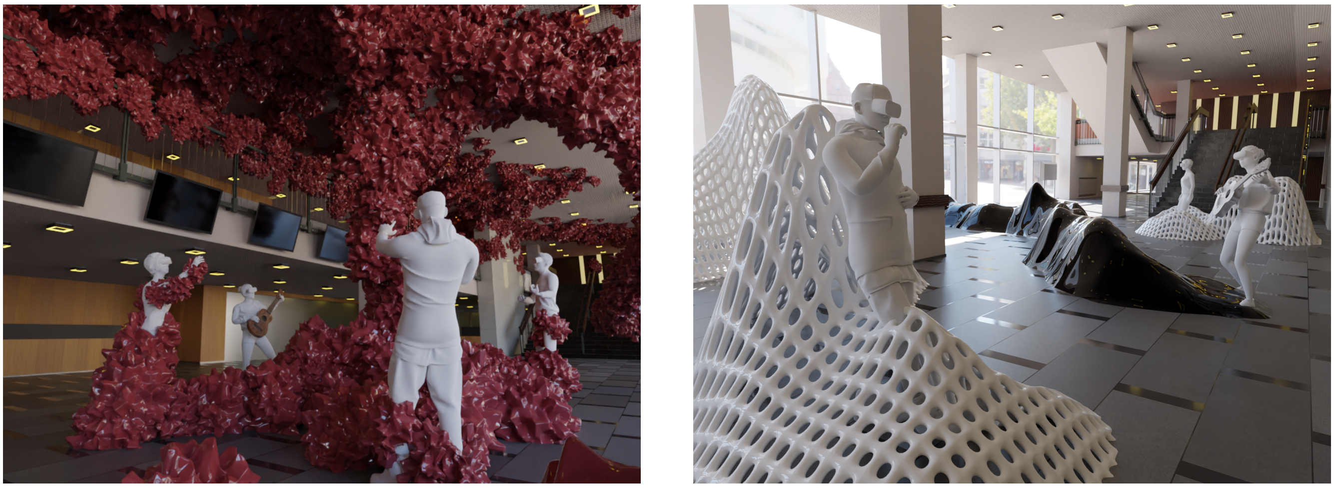

Figure 5: Musicians can use their voice or instruments to create sonic structures that correspond to their musical behaviour. The images show a few artistic renders of how sonic structures could be like.

Some other applications that we envision for sonic structures are active music interventions in therapeutic settings like in [17]. In Addition, we provide a phoneme detection module that can also be used as a visual tool to improve phonological awareness. In addition, we can also envision including elements of gamification to aid musical behaviour.

3.4 Discussion

While the basic premise of audiovisual structures is easily grasped, conceptual complexity very quickly arises when we begin to consider exactly how sound might be rendered visual, and how these transformations might be set in motion, and for what purpose, especially since there is a wide range of cross-modal mappings are possible [21]. Syntactic approaches are using the underlying structure, i.e. categorising visual forms according to their mathematical properties [24]. Semantic approaches are using cross-modal perceptual links between shape and sound, such as the bouba/kiki effect [7], sensory translation [21], or analysing the mental model [5, 4]. These include music to color [22], auditive and visual harmony [20], and roughness perception [3]. Pragmatic approaches were especially encountered as mediator, for instance by using a mood based mapping like the valance and arousal model [13].

In audiovisual performances, artists retain creative control of these mappings. Therefore we propose an architecture (Figure 6) that is flexible enough to accommodate the vast array of mapping possibilities

4. CONCLUSION

Sonic Structures is a visual sonic art experiment that aims to investigate the relationship between form and music with the 4E cognition. By leveraging cutting-edge technologies in multi-user mixed reality and real-time audio processing, the project aims to explore the musical, educational, and artistic potential of this unique approach. As the project continues to evolve and develop, we look forward to seeing the results of future experiments and artistic collaborations. A first step towards affective implementation has been done, but it needs to be extended and evaluated in cooperation with experts and in the public domain.

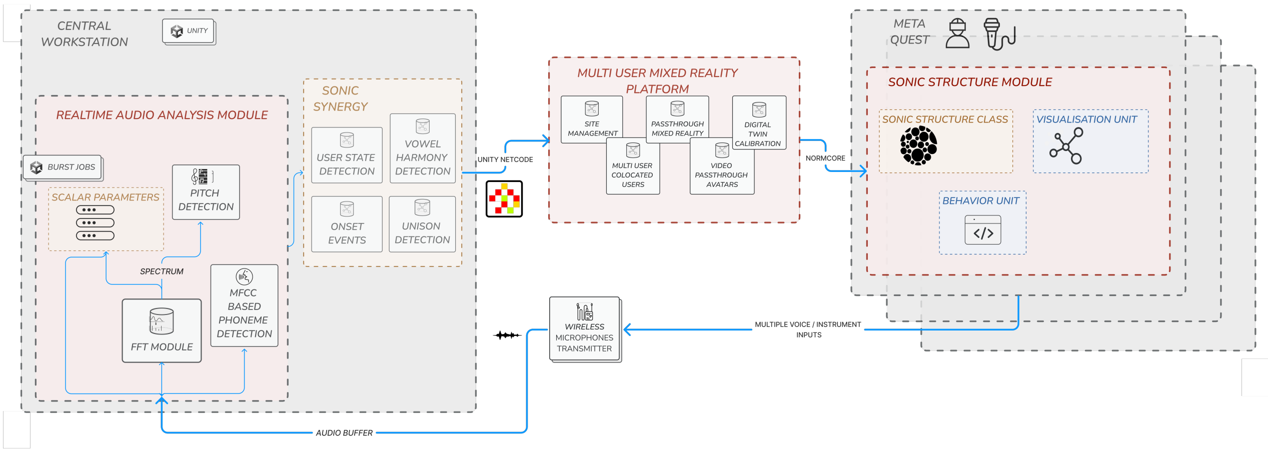

Figure 6: The proposed architecture for Sonic structure leverages multi-user mixed reality using Meta Quest devices in Unity. It allows co-located users to see each other and interact with shared sonic environments. A central workstation manages the experience, capturing audio, performing analysis and it across users.

5. REFERENCES

[1] K. Brougher and O. Mattis. *Visual music: synaesthesia in art and music since 1900*. Thames & Hudson, 2005.

[2] P. Crowther. *Art and embodiment: From aesthetics to self-consciousness*. Oxford University Press, 1993.

[3] N. Di Stefano and C. Spence. Roughness perception: A multisensory/crossmodal perspective. *Attention, Perception, & Psychophysics*, 84(7):2087–2114, 2022.

[4] L. Engeln and R. Groh. Cohearence of audible shapes—a qualitative user study for coherent visual audio design with resynthesized shapes. *Personal and Ubiquitous Computing*, 25:651–661, 2021.

[5] L. Engeln, N. L. Le, M. McGinity, and R. Groh. Similarity Analysis of Visual Sketch-based Search for Sounds. In *Proceedings of the 16th International Audio Mostly Conference*, AM '21, pages 101–108, New York, NY, USA, Oct. 2021. Association for Computing Machinery.

[6] N. Kaye. *Site-specific art: performance, place and documentation*. Routledge, 2013.

[7] V. Kovic, K. Plunkett, and G. Westermann. The shape of words in the brain. *Cognition*, 114(1):19–28, 2010.

[8] J. Krueger. Empathy, enaction, and shared musical experience. eds. cochrane, t., fantini, b., scherer, kr the emotional power of music: Multidisciplinary perspectives on musical expression, arousal and social control, 2011.

[9] M. Leman. *Embodied music cognition and mediation technology*. MIT press, 2007.

[10] P.-J. Maes. Sensorimotor grounding of musical embodiment and the role of prediction: A review. *Frontiers in psychology*, 7:308, 2016.

[11] J. Matyja and A. Schiavio. Enactive Music Cognition: Background and Research Themes. *Constructivist Foundations*, 8, July 2013.

[12] R. Mausfeld. The attribute of realness and the internal organization of perceptual reality. 2013.

[13] J. Posner, J. A. Russell, and B. S. Peterson. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and psychopathology*, 17(3):715–734, 2005.

[14] M. J. Rowlands. *The new science of the mind: From extended mind to embodied phenomenology*. Mit Press, 2010.

[15] A. Schiavio and H. De Jaegher. Participatory sense-making in joint musical practice. In *The Routledge companion to embodied music interaction*, pages 31–39. Routledge, 2017.

[16] A. Schiavio, J. Stupacher, R. Parncutt, and R. Timmers. Learning Music From Each Other: Synchronization, Turn-taking, or Imitation? *Music Perception*, June 2020.

[17] L. Schneider, L. Gossé, M. Montgomery, M. Wehmeier, A. Villringer, and T. H. Fritz. Components of active music interventions in therapeutic settings—present and future applications. *Brain Sciences*, 12(5):622, 2022.

[18] N. Schöffer. Sonic and visual structures: Theory and experiment. *Leonardo*, pages 59–68, 1985.

[19] S. Smith and S. Smith. Visual music. *Perspectives of New Music*, 20(1):75, 1981.

[20] C. Spence and N. Di Stefano. Crossmodal harmony: Looking for the meaning of harmony beyond hearing. *i-Perception*, 13(1):20416695211073817, 2022.

[21] C. Spence and N. Di Stefano. Sensory translation between audition and vision. *Psychonomic Bulletin & Review*, pages 1–28, 2023.

[22] C. Spence and N. D. Stefano. Coloured hearing, colour music, colour organs, and the search for perceptually meaningful correspondences between colour and sound. *i-Perception*, 13(3):20416695221092802, 2022. PMID: 35572076.

[23] D. van der Schyff, A. Schiavio, A. Walton, V. Velardo, and A. Chemero. Musical creativity and the embodied mind: Exploring the possibilities of 4E cognition and dynamical systems theory. *Music & Science*, 1:2059204318792319, Jan. 2018. Publisher: SAGE Publications Ltd.

[24] R. Zalaya and J. Barrallo. A Classification of Mathematical Sculpture. *Recreational Mathematics Magazine*, 5(9):71–94, Sept. 2018.