EG UK Computer Graphics & Visual Computing (2025)
Y. Sheng and A. Slingsby (Editors)

Visualising Game Audio: Increasing Audience Engagement Through the Ambient Display of Video Game Soundscapes

Chris Payne [1], Mathew Dalgleish[1], Sharon Coleclough[1], Kieran Hicks[1], Billy Gray[1], and Chris Headleand[1]

[1] Staffordshire Games Institute
School of Digital, Technologies, Innovation and Business
University of Staffordshire, UK

Abstract

This paper presents an audio system originally developed by the authors as a multiplayer, audiovisual instrument-environment, here repurposed as ViGA, a real-time visualiser for spatialised game audio. ViGA translates directional and environmental audio cues into dynamic visual forms, offering players, spectators, and audiences an intuitive and ambient graphical representation of in-game sound, and, by extension, the emergent rhythms of gameplay. Drawing on influences from video game visualisation, audiovisual synthesis, historical colour organs, visual music, music visualisers, and ambient display design, the system situates itself at the intersection of creative computing and real-time telemetry rendering. In this paper we outline the design rationale and technical architecture of the system, including its data mapping, visual language, and rendering strategies. We then present early use cases of ViGA though the video game Counter-Strike II, highlighting its potential to enhance spatial comprehension, engagement, and aesthetic experience.

CCS Concepts

• ***Human-centered computing*** → ***Visualisation theory, concepts and paradigms; Interaction design;***

1. Introduction

This paper introduces ViGA, a system for the visualisation of video game soundscapes in a secondary, ambient display. Operating outside the centre of attention, ViGA provides spectators with supplementary, moment-to-moment, spatialised visual information about auditory events that are directly linked to key gameplay and its comprehension. In addition to these informational qualities, the formal properties of the visual (colour, shape, position) support the development of synchresis to help foster deeper emotional response and immersion.

Over the past two decades, video games have undergone a transformation from a largely solitary or co-located multiplayer activity to one of live spectatorship. This shift is closely intertwined with the emergence of increasingly participatory digital culture. Dedicated game streaming services such as *Twitch*[†] have enabled players to broadcast gameplay to spectators in real-time whilst also providing integrated and monetisation facilities. In parallel to casual streaming, the popularity of professional eSports has continued to grow, and numerous games have well-established and well-funded competitions attracting large audiences online and in-person. Spectator motivations, including fan interaction and experience, are surveyed by Cauteruccio and Kou [CK23]. In particular, the ViGA system aims to address the limitation that eSports events typically offer only a single perspective on the main game, thereby inhibiting the potential for richer spectator experiences that can arise from multiple perspectives or secondary experiences beyond basic social interaction and discussion.

The rest of this paper is organised as follows. After this introduction, we provide an overview of key approaches to audio visualisation and audiovisual translation in video games and wider art, technological and music experiences. This is followed by a theory overview underpinning the translation and fusion of the audiovisual relationship and how this is implemented in our foundational prior multiplayer instrument-environment, and its subsequent iteration and further development for this project, ViGA. The work concludes with a discussion of the implementation of ViGA and an outline of future research possibilities.

2. Background and Related Work

2.1. Video Game Visualisation

Most games rely on high amounts of visual data to both construct the space of the gameplay and to create useable information for the player. There are a variety of ways to present in game information with the term interface referring to hardware for implementing the game through to the user interface (UI) or player character [SN06].

[†] https://twitch.tv/

The use of heads up display (HUD) in game offers the player direct play information and can be diegetic or non-diegetic depending on UI integration and application. In terms of design, considerations of player need versus visible space in game defines the placement of this content and can be linked to genre and game style.

2.2. Voltage Control and Audiovisual Synthesis

The implications of voltage control, first presented by Robert Moog [Moo65], are that a standardised signal range can enable any module in a multi-module audio synthesizer to act upon or control any other. Similar ideas were developed contemporaneously by Donald Buchla and presented in his System 100 Electronic Music Box [Ber08]. However, while Buchla's approach separated control and audio signals, the Moog system treated these interchangeably.

After foundational technical work in the area of analogue video synthesis by Harrison III and Slater, amongst others [HHJ14], by the early 1970s, several video synthesizers carried forward the Moog-derived idea of standardised signals, meaning that any video module's output could control any other module's input via voltage. [HHJ14, VW*08]. Moreover, with both audio and image transformed into electrical signals, sound could be readily translated into image, and visuals could be generated by sound [Fur83]. Translation between the domains was explored by artists including Steina and Woody Vasulka and Nam June Paik [Fur83, YY20]. Today, similar ideas are incorporated into the design of LZX Industries' 2022 modular video synthesizer systems.

2.3. Colour Organs and Visual Music

The connection between sound and vision has long been central to both the psychology of sound and the socio-cultural associations between hue and pitch in musical contexts [PS66, Kan12]. This interplay is further explored in sound design through the concept of synchresis—the perceived fusion of auditory and visual elements into a unified experience [Chi94]. As Walker and Nees [WN11] indicate there is a taxonomy of the application of sonification and auditory displays the most recent being art and entertainment. Colour relating to frequency from low to high (paired with the established colour spectrum) and scale relating to amplitude are basic parameters which allow for easy decoding by a player when paired with synchronicity of appearance and movement.

2.4. Music Visualisers

Music visualisation is characterised by the real-time generation of images to accompany music to enhance audience experience [GCOB21]. While there is a rich history of music visualisation (including live performance) predating the computer-era, the *Atari Video Music* is acknowledged as the first commercially available electronic music visualiser. Designed by Robert Brown [Bro78], it presented a *"visual color display of objects on an unaltered TV receiver which are directly associated with the music on an audio source."*

A very substantial number of visualisation software and extensions for home consoles and personal computers have followed, but the work of Jeff Minter, founder of Llamasoft, provides several examples of music visualisers across multiple generations of home consoles. These include *Psychedelia* for the Commodore 64 and the related *Colourspace* for Atari 8-bit computers, the *Virtual Light Machine* for the Atari Jaguar CD add-on, and the *Neon Visualiser* for the Xbox 360 [Bel19, Dig24].

2.5. Ambient Displays

Building on Weiser and Browns' [WB97] vision of calm technology and labelled a sub-class of Ubicomp by Shelton et al. [SNTE21] ambient displays are described by [MDH*03] as visually appealing representations of information intended to operate on the periphery of user attention. Since 1996, a significant number of displays have attempted to present information so it may be processed in the background of awareness. For example, Shelton and Nesbitt [SN20] provide a systematic review of the high-level design features present in 459 ambient displays as reported in 410 publications over a twenty-year period. A feature Shelton and Nesbitt note is they should be *"designed to seamlessly fit into their environment and should not interrupt an individual from their primary task"* [SN16].

3. Theoretical Underpinnings of the System

3.1. Game Sound

The concerted study of game sound has a relatively brief history [Col08]. In his *Acoustic Ecology of the First Person Shooter*, Grimshaw [Gri07] observes that one early theorisation of game sound, developed by Stockburger [Sto03], has focused on the different types of sounds (categorised according to in-game sound source and organisation at a game code level) that constitute a video game soundscape. Another approach, Grimshaw [Gri07] notes, focuses on the auditory perception of the player-listener. Its origins are in the electroacoustic theories of Pierre Schaeffer, specifically acousmatic experience, whereby the source of a sound is unseen and thus appreciated for its abstract qualities.

Schaefferian ideas were subsequently applied to a film sound context and extended by Michel Chion [Chi94] in the influential *Audio-Vision: Sound on Screen.* Chion added new categories of sound such as ambient sound (territory sound), internal sound, and on-the-air sound, the latter to address some of the conceptual difficulties/instabilities introduced by technology. Chion also proposed three listening modes: causal (related to the source of a sound), semantic (related to the meaning of a sound), and reduced (related to the abstract properties of the sound itself). Both sets of ideas were further developed by Grimshaw's [Gri07] conceptual framework for the analysis of first person shooter (FPS) sound. Whereas Chion's concept of ambient sound (territory sound) concerns sound *"whose pervasive presence gives definition to a space, e.g., bird songs, churchbells"*, Grimshaw outlines a navigational mode of listening in which players leverage auditory cues within the game soundscape (a term itself derived from [Sch93]) to orient themselves and navigate effectively through the game world. At the same time, Grimshaw [Gri07] recognises that these cues are only part of the picture. More specifically, they are posited to be part of an ever-changing acoustic ecology resulting from the constantly

evolving interactions and relationships between players, and players and the soundscape.

Not only is this acoustic ecology complex and continually in flux, it is also potentially challenging for those removed from the gameplay to analyse. As a result, external observers often have an intrinsically different perspective on player intention and input gesture. This notion is also recognised in other live contexts, notably popular music and laptop music performance [Cas03, Bro78, Wai19].

3.2. Audiovisual Mapping

Mapping relates to the relationship between the input to a system and the output of that system. In the case of digital systems this relationship must be explicitly designed. The shape of this connection can be essentially any imagined by the designer, but [FGM02] identify three basic mapping types:

- One-to-One: one output parameter is driven by one input parameter
- One-to-Many: one input parameter influences multiple output parameters
- Many-to-One: one output parameter is driven by multiple input parameters

Through combination, many-to-many mappings can be formed. In an instrument context, [HK00] note that these may prove more satisfactory for player and audience alike after an initial period of learning.

Arfib et al. [ACKV02] introduce two further dichotomies. An *"explicit"* mapping clearly defines the relationship between input and output, whereas an *"implicit"* mapping functions effectively as a black box in that its behavioural rules can be broadly described, but precise values are unknown. A *"static"* mapping remains fixed and therefore will not adapt to or learn from input data, while a *"dynamic"* mapping can evolve or change over time.

Of those audiovisual compositional methods either detailed by the artist or scrutable by analysis, notable strategies include direct translation [WG16], sensory substitution [SS15], metaphorical interpretation [Cal12], the amplification of glitches or other errors [Tun12], and the use of algorithmic systems to generate sound and image concurrently [All16].

Three specific audiovisual mappings are detailed as part of work by Taylor et al. [TBT06]. These can be summarised as vocal timbre, melodic information, and vocal dynamics respectively visualised through a responsive video, the reactions of a virtual character, and the properties of an immersive virtual space. However, Callear [Cal12] notes that despite the sustained presence of sound-image practices in the twentieth century, this kind of explicit and detailed presentation of compositional methods has remained relatively rare. Perhaps this is related, at least in part, to what Chion [Chi94] terms *"added value"*, or how sound affects the perception of an image to create *"the definite [incorrect] impression that this information or expression 'naturally' comes from what is seen, and is already contained in the image itself."* A related issue is identified by Grierson [Gri18], who notes that audiovisual art continues to be *"misunderstood and poorly characterised,"* and that artists frequently approach it as a subfield. Thus, he contends, rather than explicitly consider and detail all audiovisual aspects, these artists tend to concentrate only on elements that are considered relevant within their main discipline.

There may also be more pragmatic reasons for this limited discussion. For instance, Ikeshiro [Ike13] offers a reminder that many acclaimed audiovisual works feature surprisingly simple audiovisual relationships. Carsten Nicolai's spray [Nic05] and Ryoji Ikeda's datamatics [Ike06], he notes, are based entirely on the close synchronisation of onset and panning, and onset only, respectively. This close synchronisation is central to what Chion [Chi94] terms synchresis, or the *"spontaneous and irresistible mental fusion, completely free of any logic, that happens between a sound and a visual when these occur at exactly the same time."* It is a main reason why flat, screen-based images appear to have more depth when sound is added, and, more specifically, why footfalls that land exactly on the beat of superimposed music feel more energetic and propulsive.

4. Technical Overview of the System

The precursor to this system is TYG, a local multiplayer digital musical instrument-environment [PD24, DPH24]. The system's flexible, modular architecture allows it to support a broad spectrum of game audio research applications. Additionally, this system features two modes: composition (setup) mode and a real-time performance mode. In composition (setup) mode a graphical user interface (GUI) enables players to assign fundamental musical parameters to up to five coloured spheres for subsequent use in performance mode. For this paper, TYG's coloured sphere concept has been extended into the audio-driven visualisation system ViGA.

4.1. ViGA as an Audio-Visual System

ViGA has been designed to enhance core gameplay, help make sense of a video game's complex and shifting acoustic ecology, and offer the potential for expansion into a meta-game experience. ViGA's primary visualisation component is a modified version of the TYG sphere, which the system uses to visualise sound through two main mapping aspects. This mapping is a basic spectrum alignment that maps low to high audio frequencies (specifically the spectral centre) to a corresponding colour gradient from red to violet. This approach is based on Scriabin's synesthetic perception of note-to-colour relationships and reduces the palette to seven colours for simplicity. [Pea85]. Despite some previous work, there remains no definitive relationship of colour to audio and so a simplified approach such as ROYGBIV to frequency range, low to high, offers a non-complex approach. This semiotic foundation supports the synchretic integration of audio and visual elements, enabling intuitive interpretation by the viewer.

The second mapping aspect determines the other formal (physical) properties of the visual objects and their position on screen. Spheres are spawned and their size and shape, X position, and Y position are dynamically manipulated to convey, amplitude, panning (stereo image), and spectra (coupled to colour) of discrete sound objects Figure 1. Sphere spawning is extracted from the real-time

game soundscape via a simple onset detector that blocks excessively rapid or repeated triggering. This enables the reliable handling and visual representation of polyphony, or instances where multiple sound objects occur simultaneously.

Figure 1: *Implemented mapping: sphere colour reflects frequency; size and shape reflect amplitude.*

As work in progress, the ViGA system has to date primarily used audio from the tactical shooter Counter Strike II, developed and published by Valve [Val23]. This title was selected for reasons including the complexity of its soundscape, its extensive audio options, and that it can be played without monetary cost. However, the system is generally applicable, and can be used to visualise any game audio.

5. Conclusion

With ViGA, the authors present a novel model for augmenting video game experiences through audiovisual abstraction, specifically by rendering spatial and dynamic audio cues as a secondary, ambient visualisation layer. By placing the audio information in a secondary non-invasive HUD, the information is generated by diegetic audio but visualised as non-diegetic visual representation. In removing the ViGA information from the game the approach follows traditional approaches to HUD implementations. In this case not interfering with gameplay or the visual of the playable area but augmenting for and providing secondary visual commentary. This approach has value in spectator contexts. ViGA offers an ambient and affective visual engagement device that can highlight rhythm, polyphony, tension, and directionality in a way that complements analytical data with emotional resonance. Its audiovisual translation allows for rich perceptual fusion between visual and auditory stimulus. The aesthetic qualities of the ViGA display also suggest potential uses in branding, broadcast enhancement, or installation-based performance. By adopting implicit, dynamic mappings, ViGA encourages a relationship between input and output that improves over time. This echoes practices in both musical improvisation and artistic data visualisation, where understanding is constructed through ongoing exposure and interaction. ViGA contributes to the expanding field of ambient and peripheral display technologies that aim to enrich experiences without demanding full attention.

6. Future Work

6.1. Expanded Visual Language

Further development will focus on extending the expressive capacity of ViGA's visual output. While the current system employs spectrum-based colour mappings and basic shape modulation, future iterations will explore additional parameters to increase the information bandwidth.

6.2. Dynamic Audio Broadcast Enhancement

ViGA has strong potential in scenarios where sound is present but visual information is limited or absent. One clear application is the augmentation of voice-only or audio-dominant environments such as online meetings, radio plays, or live-streamed commentary. In online meetings in particular, the absence of camera feeds is often cited as reducing engagement compared to video-enabled sessions. By introducing a generative visualisation layer, driven by tonal qualities, rhythm, and spatial dynamics in audio, ViGA can provide an ambient visual presence for otherwise invisible conversations or performances. This presents an exciting opportunity to add visual richness to communication contexts traditionally reliant on sound alone.

6.3. Asymmetric Gaming Applications

A further area of exploration is the use of ViGA in asymmetric gaming contexts. Here, some participants engage directly with the core (source) gameplay, while others interact with ViGA itself as an auxiliary game layer or creative tool. In this configuration, ViGA becomes both a spectator enhancement and a participatory system, inviting new forms of creative agency and layered gameplay.

References

[ACKV02] ARFIB D., COUTURIER J.-M., KESSOUS L., VERFAILLE V.: Strategies of mapping between gesture data and synthesis model parameters using perceptual spaces. *Organised sound 7*, 2 (2002), 127–144. 3

[All16] ALLIK A.: Evolver: An audiovisual live coding performance. In *International Conference on Live Interfaces* (2016). 3

[Bel19] BELL E.: Hacking jeff minter's virtual light machine: Unpacking the code and community behind an early software-based music visualizer. *Volume!. La revue des musiques populaires*, 16: 1 (2019), 37–59. 2

[Ber08] BERNSTEIN D. W.: *The San Francisco Tape Music Center: 1960s Counterculture and the Avant-Garde*. Univ of California Press, 2008. 2

[Bro78] BROWN R. J.: Audio activated video display, Mar. 28 1978. US Patent 4,081,829. 2, 3

[Cal12] CALLEAR S.: *Audiovisual particles: parameter mapping as a framework for audiovisual composition*. PhD thesis, Bath Spa University, 2012. 3

[Cas03] CASCONE K.: Grain, sequence, system: Three levels of reception in the performance of laptop music. *Contemporary Music Review 22*, 4 (2003), 101–104. 3

[Chi94] CHION M.: Audio-vision. sound on screen, new york: Columbia university press. 2, 3

[CK23] CAUTERUCCIO F., KOU Y.: Investigating the emotional experiences in esports spectatorship: The case of league of legends. *Information Processing & Management 60*, 6 (2023), 103516. 1

[Col08] COLLINS K.: *Game sound: an introduction to the history, theory, and practice of video game music and sound design*. Mit Press, 2008. 2

[Dig24] DIGITAL ECLIPSE: Llamasoft: The jeff minter story. https://www.digitaleclipse.com/games/llamasoft, Mar. 2024. Interactive documentary and game compilation. 2

[DPH24] DALGLEISH M., PAYNE C., HEADLEAND C.: Tyg: Collaborative music creation in an unreal engine-based local multiplayer instrument. 3

[FGM02] FELS S., GADD A., MULDER A.: Mapping transparency through metaphor: towards more expressive musical instruments. *Organised Sound 7*, 2 (2002), 109–126. 3

[Fur83] FURLONG L.: Notes toward a history of image-processed video. *Afterimage 11*, 5 (1983), 12–19. 2

[GCOB21] GRAF M., CHIJIOKE OPARA H., BARTHET M.: An audio-driven system for real-time music visualisation. In *Audio Engineering Society Convention 150* (2021), Audio Engineering Society. 2

[Gri07] GRIMSHAW M. N.: *The acoustic ecology of the first-person shooter*. PhD thesis, The University of Waikato, 2007. 2

[Gri18] GRIERSON M.: Creative coding for audiovisual art: The codecircle platform. In *The Routledge Research Companion to Electronic Music: Reaching out with Technology*. Routledge, 2018, pp. 312–335. 3

[HHJ14] HIGH K., HOCKING S. M., JIMENEZ M.: The emergence of video processing tools: Television becoming unglued. *(No Title)* (2014). 2

[HK00] HUNT A., KIRK R.: Mapping strategies for musical performance. *Trends in gestural control of music 21*, 2000 (2000), 231–258. 3

[Ike06] IKEDA R.: datamatics. https://www.ryojiikeda.com/project/datamatics/, 2006. Audiovisual art project exploring the materialisation of pure data through installations, concerts, and publications. 3

[Ike13] IKESHIRO R.: *Studio Composition: Live audiovisualisation using emergent generative systems*. PhD thesis, Goldsmiths, University of London, 2013. 3

[Kan12] KANDINSKY W.: *Concerning the Spiritual in Art*. Courier Corporation, 2012. URL: https://books.google.com/books/about/Concerning_the_Spiritual_in_Art.html?id=URvGFXf2vXgC. 2

[MDH*03] MANKOFF J., DEY A. K., HSIEH G., KIENTZ J. A., LEDERER S., AMES M.: Heuristic evaluation of ambient displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2003), ACM, pp. 169–176. URL: https://dl.acm.org/doi/10.1145/642611.642642, doi:10.1145/642611.642642. 2

[Moo65] MOOG R. A.: Voltage controlled electronic music modules. *Journal of the Audio Engineering Society 13*, 3 (1965), 200–206. 2

[Nic05] NICOLAI C.: spray. Installation exhibited at NTT InterCommunication Center [ICC], Tokyo, 2005. Part of the exhibition "Open Nature", April 29–July 3, 2005. URL: https://www.ntticc.or.jp/en/archive/works/spray/. 3

[PD24] PAYNE C., DALGLEISH M.: Tyg: A multiplayer instrument-environment in ue 5. CHIME Annual One-day Music and HCI Conference 2024. 3

[Pea85] PEACOCK K.: Synesthetic perception: Alexander scriabin's color hearing. *Music perception 2*, 4 (1985), 483–505. 3

[PS66] PANEK W., STEVENS S.: Saturation of red: A prothetic continuum. *Perception & Psychophysics 1*, 1 (1966), 59–66. 2

[Sch93] SCHAFER R. M.: *The Soundscape: Our Sonic Environment and the Tuning of the World*. Destiny Books, Rochester, Vermont, 1993. URL: https://www.simonandschuster.com/books/The-Soundscape/R-Murray-Schafer/9780892814558. 2

[SN06] SAUNDERS K., NOVAK J.: *Game development essentials: Game interface design*. Thomson Delmar Learning, 2006. 1

[SN16] SHELTON B., NESBITT K.: The aesthetic awareness display: a new design pattern for ambient information systems. In *Proceedings of the australasian computer science week multiconference* (2016), pp. 1–10. 2

[SN20] SHELTON B., NESBITT K.: A systematic review of ambient display modalities, physical forms and interactivity. In *AMBIENT 2020, The Tenth International Conference on Ambient Computing, Applications, Services and Technologies* (2020), pp. 9–15. 2

[SNTE21] SHELTON B., NESBITT K., THORPE A., EIDELS A.: Gauging the utility of ambient displays by measuring cognitive load. *Cognition, Technology & Work 23*, 3 (2021), 459–480. 2

[SS15] STILES N. R., SHIMOJO S.: Auditory sensory substitution is intuitive and automatic with texture stimuli. *Scientific reports 5*, 1 (2015), 15628. 3

[Sto03] STOCKBURGER A.: The game environment from an auditive perspective. *Level Up* (2003), 4–6. 2

[TBT06] TAYLOR R., BOULANGER P., TORRES D.: Real-time music visualization using responsive imagery. In *8th International Conference on Virtual Reality* (2006), Citeseer, pp. 26–30. 3

[Tun12] TUNALI F. S.: Glint: Audiovisual glitches. *Leonardo 45*, 3 (2012), 296–297. 3

[Val23] VALVE CORPORATION: Counter-strike 2. https://www.counter-strike.net/cs2, Sept. 2023. Video game software. 4

[VW*08] VASULKA W., WEIBEL P., ET AL.: Buffalo heads: media study, media practice, media pioneers, 1973-1990. *(No Title)* (2008). 2

[Wai19] WAITE S.: Liveness and interactivity in popular music. In *Innovation in Music*. Routledge, 2019, pp. 113–128. 3

[WB97] WEISER M., BROWN J. S.: The coming age of calm technology. In *Beyond Calculation: The Next Fifty Years of Computing*, Denning P. J., Metcalfe R. M., (Eds.). Springer, 1997, pp. 75–85. URL: https://link.springer.com/chapter/10.1007/978-1-4612-0685-9_6, doi:10.1007/978-1-4612-0685-9_6. 2

[WG16] WANG J., GAYDECKI P.: From light to sound: Signal modality translation. In *Proceedings of Acoustics 2016* (2016), Proceedings.com, p. 335. Conference paper presented at Acoustics 2016. URL: https://www.proceedings.com/content/031/031848webtoc.pdf. 3

[WN11] WALKER B. N., NEES M. A.: Theory of sonification. *The sonification handbook 1* (2011), 9–39. 2

[YY20] YOON J. W., YEO W. S.: A study on the audiovisual art of nam june paik with focus on musical synesthesia. *Journal of Korea Multimedia Society 23*, 4 (April 2020), 603–614. URL: https://journal.dcs.or.kr/_PR/view/?aidx=35764&bidx=3205. 2