The 21st Century Film, TV & Media School:

TEACHING SOUND:

Aesthetics & Praxis

Editors
Michael Kowalski
Luís Cláudio Ribeiro
Samuel Larson Guerra
Elena Rusinova
J. Harry Whalley

CILECT
ASSOCIATION INTERNATIONALE DES ECOLES DE CINEMA, D'AUDIOVISUEL ET DES MEDIAS
INTERNATIONAL ASSOCIATION OF CINEMA, AUDIOVISUAL AND MEDIA SCHOOLS

The 21st Century Film, TV & Media School:

TEACHING SOUND: Aesthetics & Praxis

Edited by

Michael Kowalski
Luís Cláudio Ribeiro
Samuel Larson Guerra
Elena Rusinova
J. Harry Whalley

CILECT
The International Association of Film and Television Schools

Sofia, Bulgaria
2025

The 21st Century FILM, TV and MEDIA SCHOOL:
TEACHING SOUND: Aesthetics & Praxis

НАЦИОНАЛНА БИБЛИОТЕКА
„СВ. СВ. КИРИЛ И МЕТОДИЙ"
Каталогизация в публикацията

The 21st Century FILM, TV and MEDIA SCHOOL. Vol. 5, TEACHING SOUND: Aesthetics & Praxis/ ed. Michael Kowalski, Luís Cláudio Ribeiro ... [и др.]. – Sofia: CILECT – The International Association of Film and Television Schools, 2025

ISBN 978-619-7358-21-6 (pdf)

1. Кино – образование и просвета – сборници.

2. Телевизия – образование и просвета – сборници.

3. Висши учебни заведения – сборници

The 21st Century Film, TV & Media School: TEACHING SOUND: Aesthetics & Praxis

First Edition

Published by

www.cilect.org

ISBN 978-619-7358-21-6 (pdf)

Contents

3

PART THREE

Research

12

Using Video-Game Technologies in New Interfaces for Musical Expression

Adriana Sá

Adriana Sá is a transdisciplinary artist, performer-composer-musician, and editor of the *Live Interfaces* journal. An assistant professor at Lusófona University, Lisbon, and a member of the Centre for Research in Applied Communication, Culture and New Technologies, Sá holds a PhD in Arts and Computing from Goldsmiths, University of London (2016). In the 1990s, she began using sensor technologies to explore sound connected to light, space, movement and architecture. Twenty years later, she started developing an audio-visual instrument that combines an ancient string instrument (a zither) and 3D software that processes sound and image based on the zither sound. Her scientific publications bridge artistic practice, interaction design, audio-visual theory, neuroscience and experimental psychology.

Introduction

Three-dimensional (3D) engines are digital systems that enable the simulation of physical behaviours, conveying cinematic experiences. For example, they can transform the volume and spatialisation of sounds in real-time, depending on the image. They are used in video-gaming and interactive storytelling, as well as in audio-visual composition and performance.

Among the artists and researchers who have explored music with 3D environments, John Klima (Paul 2003), Mick Grierson (2007), Robert Hamilton (2008a, 2008b) and Florent Berthaut et al. (2011) have made explicit references to gaming while exploring participatory interaction, video-game player paradigms, and allusive iconography. I have also been using 3D technologies since 2007, but in a quite different way. My 3D software operates based on the sound from an acoustic instrument, and the image is intended to work like a reactive stage scene. The intertwining of acoustic sound, digital sound, and digital image reflects a particular understanding of musical expression and a concern with the fact that vision tends to dominate over audition. By bridging artistic work, audio-visual theory, perception sciences, and interaction theory, I could understand how visual dominance can be circumvented (Sá 2013, 2016; Sá et al. 2015), and those investigations informed the development of my NIME – new interface of musical expression (Sá 2013, 2014, 2016, 2017, 2021, 2024).

Whatever motivates a practitioner, it is useful to think of how interaction designs and audio-visual relationships drive experience.

I will begin this paper with a short description of how 3D engines render sound, and will then proceed with an overview of the diversity of approaches to music in 3D environments. Following this, I will revisit the theories of psychologist Csikszentmihalyi (1988, 1996), well known in game design, in order to analyse convergences and divergences between the notion of ‘flow’ in music and gaming. Then, I will draw from perception science and film theory (e.g., Braudy 1977; Chion 1994; Boltz 2004) to analyse how audio-visual relationships and the dynamics of sound and image drive attention. Armed with these tools, it is possible to glean how any work exploring music with 3D environments conducts perception.

3D Positional Audio and a Notion of 3D Sound

Digital 3D engines can be used to create a sense of visual depth and to spatialise digital sounds with respect to the 3D scene. They apply the principles that govern our perception of the physical world, so that the player feels immersed in the digital world beyond the screen (e.g., Dourish 2004). Positional audio effects are important in this environment, as sound spatialisation depends on visual dynamics. The creators of *FMOD Studio*, a popular 3D positional audio software, explain how this establishes the difference between what they call 2D and 3D sound (‘3D Audio’ 2014: 296). TV and CD recordings use 2D sound: each aspect of the sound environment is directed to a particular output speaker, regardless of whether there is one or many speakers.

Indeed, it seems useful to detach the term '3D sound' from the notions of 'immersion' and 'spatialisation'. With 3D sound, the sonic output depends upon the position of the source in the digital 3D world relative to the 3D camera/listener. If the camera and/or the sound-emitting object in the digital world moves, then the overall audio output reflects this shift. This is achieved by constantly altering how the signal is routed to the output channels.

The History of Music Made with Digital 3D Environments

Since the late 1990s, artists, musicians, and programmers have explored 3D technologies in audio-visual performances.

Networking through the Internet

A notable example is the *Global Visual Music Project* (Figure 1), created by Vibeke Sorensen, Miller Puckette and Rand Steiger in 1997 (Sorensen 2005). The authors developed a system for audio-visual performance, incorporating digital video, 3D graphics, acoustic musical instruments, and computer music strategies. They aimed at making connections that were clearly perceptible while avoiding direct metaphorical correspondences, such as those between pitches and colours, for example. The performances took place in spaces networked through the internet. The system used Pure Data (the visual programming language developed by Puckette), and it was intended to be played by specific musicians.

Other artists and researchers explored links between video-game technologies and music to convey the audience's networked, participatory interaction with the system. Shown for the first time in 1999, John Klima's *Glasbead* (Figure 2) is a multi-user collaborative musical interface consisting of a rotating spherical structure with stems that resemble hammers and bells (Cityarts n.d.). Sound files can be imported into the bells and triggered by flinging the hammers into the bells. Addressing the concepts of multi-user environments, gaming, and file-sharing, *Glasbead* allowed up to 20 players to remotely 'jam' with each other. Christiane Paul (2008) describes this piece as 'an instrument and "toy" that allows users to import sound files and create a myriad of soundscapes'.

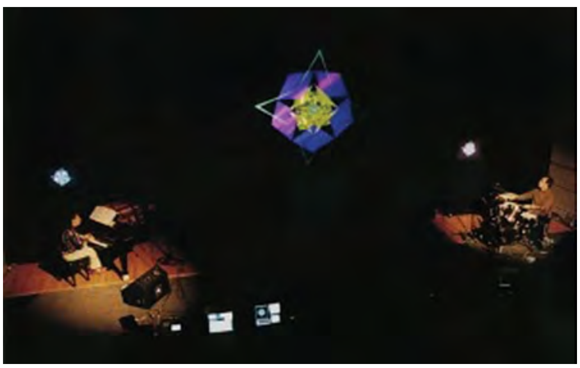

Figure 1. *Global Visual Music Project (Sorensen et al. 1997)*

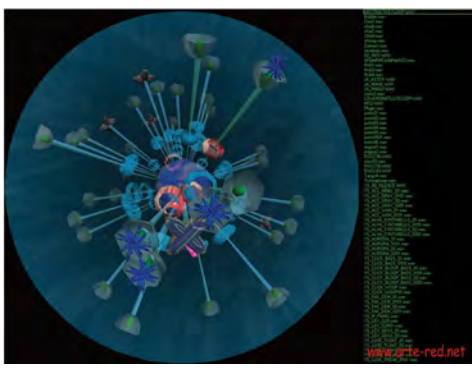

Figure 2. *John Klima's Glasbead, 1999*

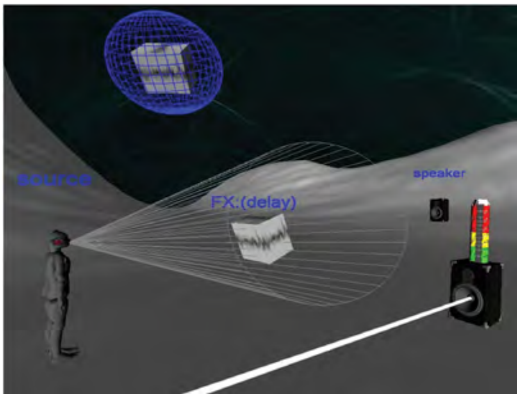

***Figure 3**. The system developed by Mike Wozniewski, Zack Settel, and Jeremy Cooperstock*

***Figure 4**. Robert Hamilton's 'q3osc', 2008 (CCRMA Wiki n.d.).*

Extending the Physical Space into the Digital Space

Mike Wozniewski, Zack Settel, and Jeremy Cooperstock (2006a, 2006b, 2006c) developed a system where the user's body can be modelled within a digital 3D world (see 'Audioscape' n.d.). The sound processing occurs at various locations in the virtual 3D space, and the transmission of sound signals from one area to another is based on physical models of sound propagation. Initially, the system was intended for the authors' performance, but subsequent versions were developed for participatory, networked interaction. The system is controlled by regular physical activity, such as moving, turning, pointing, or grabbing. Although the authors bend the rules of physics in order to explore the musical potentials of the work, the users' interaction relies on a clearly perceivable, congruent relation between gesture, sound and image (Figure 3).

Exploring the First-Person Shooter Paradigm

Other works draw upon the most popular player paradigm in video-gaming: first person shooters (FPS). In 2007, Mick Grierson created a '3D first-person composition and improvisation system' designed 'to look and behave like contemporary first-person computer games'. The user can create, adapt, and combine elements with varying physical attributes to produce musical structures. The system does not exhibit pre-determined constraints on the environment, object properties, and interactions; rather, it responds to users' behaviour through adaptive algorithms.

Robert Hamilton (2008a) describes a modification of a game engine in a paper titled 'q3osc: or how I learned to stop worrying and love the <bomb> game' (see Figure 4). The engine outputs sound-objects for real-time networked performance and spatialisation within a multi-channel audio environment. The weapon projectiles are associated with sounds triggered by collisions with the environment. In another paper, Hamilton (2008b) describes an 'interactive multi-channel multi-user networked system for real-time composition'. Using navigation and sound spatialisation, the system aims to superimpose the virtual environment and the performance space. The software draws from the topography of the

Figure 5. Three Musicians Playing the Couacs in Faders For All Mode (Berthaut et al. 2011)

Figure 6. *Tarik Barri's Versum, 2008 (Tarik Barri n.d.)*

virtual environment so that the interactor retains a level of control over the musical whole. It builds upon this visual and musical structure by leveraging the inherently flexible and indeterminate system of player motions and actions.

Musical Gaming or Musical Expression?

A few years later, Florent Berthaut et al. (2011) developed *Couacs* (Figure 5), a 'collaborative multiprocess instrument' that also uses FPS for musical interaction. It led them to ask:

> How can we use some game actions for musical control without disturbing other game actions not connected to sound, and vice-versa? Will gamers/musicians try to learn how the instrument works and how they can produce specific musical results, or will they only play without paying attention to the generated music? Should these instruments have a goal like a video game or not? (Berthaut et al. 2011)

These questions reveal a possible conflict between the notions of 'play' in gaming and music. Such conflict can become compelling creative material, whether the desire is to create a hybrid experience or to draw attention to the subtle relations between the sounds. I will further explore these issues in the chapter while looking at how perception works. For now, it is suffice to keep in mind that the meaning of 'expression' is not consensual and that the definition of NIME embraces musical instruments as well as musical games.

Tarik Barri's *Versum* (Figure 6), which originated in 2008, also uses the first-person paradigm, but 'shooting' is not involved (Tarik Barri n.d.). It combines electronic sounds and abstract images while taking advantage of 3D positional audio effects. The work is described as follows:

> The virtual world of *Versum* is seen and heard from the viewpoint of a moving virtual camera with virtual microphones attached. This camera, controlled in real time by means of a joystick (or any other kind of

Figure 7. *Left: Adriana Sá & Atau Tanaka at Iklectik, London 2022.*
Right: Zither & audio-visual software (version 2020, from a system started in 2007)

> controller), moves through space, similar to how first-person shooter games work. Within this space, I place objects that can be both seen and heard and like in reality, the closer the camera is to them, the louder you hear them. (Tarik Barri n.d.)

Circumventing Theories Embedded in Video-Game Technologies

When I started working with video-game engines in 2007, I questioned certain theories embedded in the digital platform, concealed in code. The software assumed that digital behaviours must be consistent and cause–effect relationships clearly perceivable. My creative concerns are very different.

I developed a NIME that combines acoustic sound, digital sound, and digital image (Figure 7). The acoustic component is a concert zither, a multi-string instrument, with aged strings and custom (de)tuning. Its sound is used as an input to the audio-visual 3D software, which processes graphics and recorded sounds based on frequency analysis; John Klima wrote the code (C++).

The first version of the instrument emphasised a few 'chaotic', non-linear features in the zither, exploring digital configurations and mappings that caused the software to behave inconsistently (Sá 2013). Subsequent versions implemented compositional strategies at a low level in the digital architecture rather than solely at a high level in parameterisation (Sá 2014, 2017). In performance, I use two interfaces: the audio analysis of the zither and a set of gamepad switches. I stand in front of the screen, facing the audience. The projected image is a shifting camera view over a digital 3D world, which morphs upon audio input detection.

The creative process unfolded along with a perceptual approach to instrument design, composition, and performance. This approach gave rise to a set of methods that can be used to unveil how any interaction designs, audio-visual relationships, and dynamics can inform perceptual experience.

The Meanings of 'Flow' in Gaming and Music

Repurposing video-game technologies to convey idiosyncratic expression brings the question of how the notion of 'play' might converge and diverge in gaming and music. Furthermore, how can 'flow' be understood, and how does it substantiate in interaction design?

Revisiting a Theory of Flow

Psychologist M. Csikszentmihalyi's general theory about flow (1988, 1996) became influential in research on in-game interaction (e.g., Koster 2004; Sweetser & Wyeth 2005), but interestingly, there is hardly any reference to it in music. Isn't the feeling of flow equally important whether we play a game or a musical instrument?

Digital platforms embed theories informed by specific purposes, and the history of music with 3D environments provides a wealth of examples of how theories can be adapted, expanded, and questioned. Ultimately, we can extrapolate a communication tool by debating the convergences and divergences between different interpretations of the term 'flow'.

According to Csikszentmihalyi, flow entails concentration, the merging of action and awareness, a loss of self-consciousness, the transformation of time, a balance between challenge and skills, clear goals, direct feedback, and a sense of control.

It is clear that musical flow requires interaction to enable concentration – that is, being aware only of what is relevant here and now. It also requires the merging of action and awareness: action exceeds the conscious mind. Furthermore, it implies the transformation of time – the sense of how much time passes depends on what one is doing. Beyond these convergences, the other features of flow require clarification.

Clear Goals in Gaming vs. Expanded Awareness in Music

Csikszentmihalyi states that flow entails clear goals every step of the way, meaning that one always knows what needs to be done. To some extent, this applies to both gaming and music, but it is important to define what kind of knowledge I am referring to. Indeed, a video-game player needs clear goals to act strategically. Even in sandbox games such as *The Sims,*[1] where there isn't any narrative goal, players have clear goals, such as making a bigger house to host more inhabitants or decorating characters to make them attractive for social interactions with other game players.

In contrast, many musicians feel that musical actions can be clear and deliberate without their motivations being clearly conscious. This is also implicit in that which Franziska Schroeder and Pedro Rebelo (2009) call 'the performative layer': a state that entails the wakening, extending, and refining of unused skills. As Zbigniew Karkowski (1992) wrote, 'music can heighten consciousness' and 'increase the intensity of the mind' because 'art communicates before it is understood'.

1 *For example, go to http://www.thesims3.com.*

Loss of Self-Consciousness in Gaming vs. Performative Presence in Music

According to Csikszentmihalyi, another feature of flow is the loss of self-consciousness:

> In everyday life, we are always monitoring how we appear to other people; we are on the alert to defend ourselves from potential slights and anxious to make a favourable impression. Typically, this awareness of self is a burden. In flow, we are too involved with what we are doing to care about protecting the ego. (Csikszentmihalyi 1996: page 112)

While Csikszentmihalyi speaks as a psychologist, a performer can argue that their sense of self also exists beyond any anxiety to make a 'favourable impression'. As a musician, when I am on stage, I might be too involved with performative action to care about protecting my ego, and yet I am highly aware of how I appear to the audience. It is easy to agree with Csikszentmihalyi (1996: 12–13) when he writes that 'the musician feels at one with the harmony of the cosmos'. It becomes more difficult when he states that such feeling only comes after an episode of flow is over, because 'the nervous system has definite limits on how much information it can process at a given time' (2004: 28). But a musician works with semi-conscious knowledge, and is very aware of flow as it also reflects in their bond with the audience (Sá 2017).

Balancing Challenge and Skills: Adapting Challenges or Developing New Skills?

Another feature of Csikszentmihalyi's notion of flow is the balance between challenge and skills. Such balance exists in both gaming and music, yet possibly at different levels. The game industry aims to rapidly engage the player, and difficulties are adjusted accordingly. Media researchers may find that 'to induce optimal presence, the developer of a mediated experience has to include recognition of the specific purpose of the user' (Riva et al. 2015: 90). Many designers want video games to adapt to different types of gamers while keeping all of them engaged (e.g., Hunicke & Chapman 2004).

There is also a salient, related approach in musical interface design. For example, François Pachet (2004) developed what he called 'musical mirroring effects', where the level of challenge represented by the behaviour of the system always corresponds to the level of the user.

Alternatively, one can defend that an instrument requires great investment in playing. This is often the case when the musician develops their own instrument. As Joanne Cannon and Stuart Favilla (2012) highlight, creating a new instrument must be accompanied by developing new skills to play the instrument; one does not learn to play an acoustic instrument in weeks, and that should also not be expected with digital instruments.

Sense of Control in Gaming vs. Reciprocal Interaction in Music

A sense of control is another feature of Csikszentmihalyi's flow. Again, this might be applicable to both gaming and music, but possibly at different levels. In gaming, interfaces are required to behave consistently, i.e., in clearly perceivable ways. There are different opinions in music.

For example, Cornelius Poepel and Dan Overholt (2006) suggest that in violin-related NIME design, absolute control over the instrument and clearly understandable relationships are desirable. In contrast, Tom Mudd (2015) argues that non-linear behaviours foster engagement with musical expression, as the musician focuses on the material properties of the interface itself.

For many musicians, the act of performing entails relying on the performer's skills to create unprescribed musical meanings upon unexpected events, which the performer and the audience perceive at the same time. Chaotic or entropic behaviours are often regarded as creative material (e.g., Vasulka 1996; Cascone 2000; Mudd 2015; Di Scipio 2015), entailing a reciprocal interaction between the performer and the instrument (Sá 2017). Furthermore, one may find that the unexpected must be clamped, so as to rule out undesired outcomes. As Joel Ryan observes, 'each link between performer and computer has to be invented (…) These "handles" are just as useful for the development or discovery of the piece as for the performance itself' (1991: 5).

Direct Feedback in Gaming vs. Timing in Music

Csikszentmihalyi's notion of flow also entails direct feedback, i.e., a sense of immediacy. 3D video games apply the principles of human perception mechanisms so that the player embodies the interface, and their action feels unmediated (e.g., Dourish 2004; Riva et al. 2015). The problem is more complicated in music.

'Direct' implies control over timing, and the construction of time is at the heart of music. But the term is equivocal. With certain musical interfaces that operate based on muscle energy (EMG[2]), we can say that actuation happens faster than with any acoustic instrument because the biosignal is captured when the performer initiates the physical gesture (see Tanaka's observation in Sá et al. 2015). With live coding systems, the term 'immediacy' becomes even more ambivalent: typing code is an embodied and time-based action, but it also precedes the sonic and visual results.

From a Particular Notion of Musical Flow to a Particular Notion of Musical Expression

Figure 8 summarises how the previous distinction between the notion of flow in music and gaming is reflected in interaction design.

In both music and gaming, flow requires interaction to enable concentration, the merging of action and awareness, and transformation of time. But while a game player needs clear purposes, music has no purpose beyond the perceptual experience itself; recalling Karkowski (1992), art increases the mind's intensity because it communicates before it is understood. Whereas a game player loses self-consciousness as they focus on the game, a performer is highly aware of their expanded state of perception and of how the audience is being affected. Also, a sense of control might be desired over both musical instruments and musical games, but at different levels and for different reasons.

2 Electromyography (EMG) is an electrodiagnostic medicine technique for evaluating and recording the electrical activity produced by muscles.

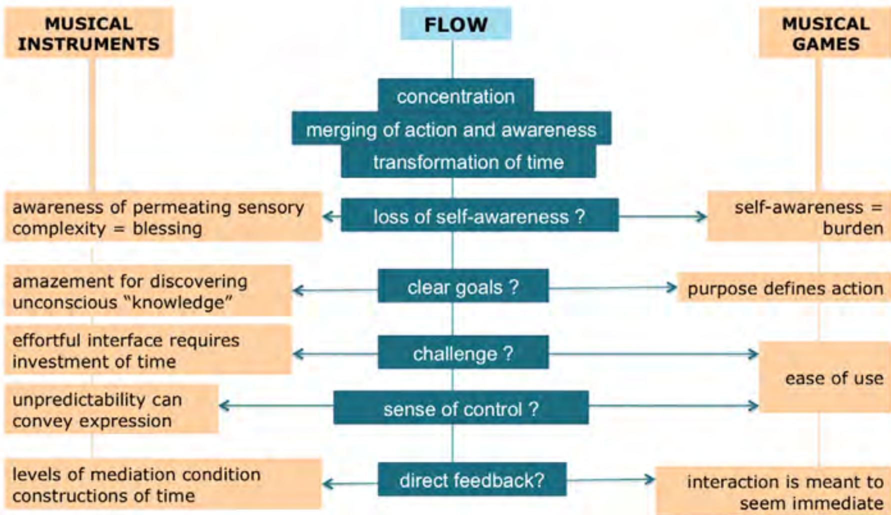

Figure 8. *Convergences and Divergences between the Sense of Flow in Music and Gaming. Diagram by the author.*

In gaming, interfaces are designed so that interaction feels as unmediated as possible. In contrast, with a musical instrument, the performer focuses on the properties and non-linear behaviours of the interface, and the unexpected can convey expression.

Differentiating Notions of Expression: Interaction Effort as a Variable

The understanding of musical expression revealed in the previous section can be explored in many different ways, but it does not claim to be consensual. The extent to which musical interfaces allow for agency is variable, and so is the amount of real-time effort required in the interaction. Different levels of real-time effort convey different understandings of flow and expression.

Little Effort

An interface requiring little effort requires a little amount of foreknowledge: the relation between deliberate human agency and sonic results is linear and clearly perceivable, and/or the music does not depend much on the people's interaction. This is the case with many installations intended to engage a broad audience.

Medium Effort

Medium effort denotes that the instrument requires particular skills, yet a sense of immediacy conveys musical timing, and/or technical configurations rule out undesired outcomes. An example is Hamilton's multi-user system for real-time composition, which captures the indeterminate movements of the interactor but draws from the topography of the virtual environment so that they retain a level of control over the musical whole (Hamilton 2008a, 2008b). Another example is Grierson's (2007) system, which does not constrain sonic and visual outcomes but responds to the user's behaviour through adaptive algorithms. My own instrument and creative approach are certainly very different, but I find it equally important to create a balance between the control over the instrument and the unpredictability of sonic outcomes. An unexpected, often minute event can produce compelling performative tension. It causes a minimal yet graspable hesitation – a moment of suspense. Resolving the musical challenge in good time then causes a sensation of release.

High Effort

Interfaces can also be designed to require high effort if they do not rule out any outcomes and/or a person needs a great amount of foreknowledge to play them. For example, live coding requires one to know a programming language. And, of course, interfaces can be simultaneously effortless and effortful. For example, with Klima's *Glasbead* (Paul 2008), the interaction of a single person is simple and clearly perceivable, but people can interact with each other remotely and play more than 20 sound files at once (Cityarts n.d.). As the mix creates a myriad of soundscapes, complexity rapidly confounds the base cause-and-effect relationships. Hence, a person's individual agency becomes challenging.

Audio-Visual Relations and Perceptual Effects

The manner in which music with 3D environments might be experienced greatly depends on how we perceive audio-visual relationships.

Modes of Perception and Sensory Interactions

Drawing from the composer Pierre Schaeffer (1966), Chion (1994) described three types of listening, or modes, which can be extended into the audio-visual domain. The first is 'causal listening', which 'consists of listening to a sound in order to gather information about its cause (or source)' (Chion 1994: 28). Causal audio-visual perception is similar; it consists of listening to the sounds and viewing the images to gather information about their cause. The second mode is 'semantic listening', which 'refers to a code or a language to interpret a message' (Chion 1994: 28). The same is applicable to semantic audio-visual perception. The third mode of listening is 'reduced listening'. Chion provides perspective over the term 'reduced' by stating that hiding the source of sounds 'intensifies causal listening in taking away the aid of sight' (1994: 32). In applying to the audio-visual domain, we consider how 'reduced' might refer to stripping the perceptual experience of conclusive causes and meanings.

Whether a person plays with an interface or simply attends a show, their senses interact with each other, producing a unified experience. Those sensory interactions, which in perception science are called multisensory integration processes, are automatic and unconscious.[3] Indeed, perception is a process of multisensory synthesis. The mutual

3 *Well-known examples are the Ventriloquist Effect, the McGurk Effect and the Double Flash Illusion.*

impact of vision and audition leads to a perceptual surplus, as Chion described in audio-visual theory; it produces effects and meanings that are not contained in the sound or the image alone. Yet, the strength of each sensory modality tends to be unbalanced.

The Problem of Visual Dominance

Since Schaeffer, acousmatic composers have argued that sounds must be detached from their visual origination to be fully experienced. Jeff Pressing also noted that, in digital 3D environments, 'where a direction of causation is discernible, it operates from visual event to sound event and not the reverse. Nevertheless, sometimes the sound for a new cinematic scene can precede the visual image' (1997: 10). Not coincidentally, 3D animators often place the sound of a footstep slightly before the foot actually hits the ground.

Chion (1994: 40) wrote:

> Each audio element enters into a simultaneous vertical relationship with narrative elements contained in the image (characters, actions) and visual elements of texture and setting. These relationships are much more direct and salient than any relations the audio element could have with other sounds.

Audio-Visual Fit, Perceptual Binding and Stimuli Prioritisation

Merilyn Boltz, a researcher in cognition applied to music and film, confirmed that a high level of audio-visual congruence produces integrated perceptual encodings and representations, which leads vision to dominate over audition (Brown & Boltz 2002; Boltz 2004). Visual dominance might happen at a level of conceptualisation rather than at a detection level, possibly to compensate for the poor alerting abilities of visual stimuli (Robinson et al. 2015). Indeed, while our hearing spans 360 degrees and apprehend all sounds at once, we see only 180 degrees, and must direct our eyes to a target at a time. To survive, we need to combine what we hear and see, and to detect causation in a timely manner.

In experimental psychology, Michael Kubovy and Michael Schutz (2010) coined the term 'ecological fit' to describe how the aural discounts the visual and the visual discounts the aural based on concepts acquired through previous experience. The greater the fit, the more we ignore diverging sensory information. A study on perceptual binding and prioritisation demonstrates the difference between high, medium, and low fit (Sá et al. 2015).

Highly Fitting Sounds and Images

'High fit' means that the audio-visual relationship conveys conclusions about causes and effects, and perception prioritises information accordingly. An example can be found in the *Global Visual Music Project* (Sorensen 2005), where clearly perceptible connections between video, 3D graphics, acoustic instruments, and computer music could draw attention to the connection between multiple performers who were not physically present in the same location. Another example is Wozniewski et al.'s (2006c) system, where both the sound and the graphics reflect a person's physical movement.

Figure 9. *Screenshot of Barri's Versum (Tarik Barri n.d.)*

From a functional perspective, Barri's *Versum* entails high fit as well – as the camera navigates through the digital environment, we see multiple colour beams (see Figure 9), and sound spatialisation provides cues about which sound is emitted by which beam (Tarik Barri n.d.). Nevertheless, there is also a level of ambiguity: the sounds are audible in areas around the beams, which have no visual representation. As we move through the digital space, we may continue to hear a sound after having ceased to see the corresponding emitter on the screen.

Expanding the Visible through the Audible

Chion coined the term 'superfield' to designate such expansion of the visible through the audible; the sound produces an environment that extends the panorama beyond the image frame. Whether we consider narrative film or music with 3D environments, the superfield does not require a high audio-visual fit; it is also compatible with a sense of openness in the overall cinematic experience. When we navigate through a digital 3D space, sound spatialisation can expand the sense of an environment without a sound emitter being visible on the screen. Similarly, in narrative film, the sound of glass breaking behind us can produce a sense of environment without requiring the event to appear on the screen; it remains open to interpretation if the sound came from a window, a mirror, or a dish. Furthermore, seeing a forest while hearing a train leads us to imagine an environment with a forest and a train – whose visual aspect remains open to imagination.

It is useful to look at how Leo Braudy spoke of the image in film a few decades before Chion's nomenclature became a foundation for all discussions relating to audio-visual works. Braudy (1977) made a notable distinction between what he called a 'closed frame' and an 'open frame'. The former provides all the information to interpret the image, leaving little room for ambiguity. Conversely, the latter does not describe a scene, leading one's imagination to build what is not seen within the frame: 'The world of the film is a momentary frame around an ongoing reality' (Braudy 1977: 47). This distinction can be extended to describe how the superfield draws from audio-visual relationships. In some cases, the sound may fit closely with the visual elements and contribute to a sense of visual coherence, similar to the notion of a 'closed frame'. In other cases, the sound does not align with the visual elements, conveying inconclusive interpretations akin to the 'open frame'.

Inconclusive Audio-Visual Relationships

Ambiguity and inconclusiveness are important if one desires attention to focus on the sonic construction; indeed, a high level of audio-visual congruence leads vision to dominate over audition (e.g., Brown & Boltz 2002; Boltz 2004). In my own performances, the image is desired to create a reactive stage scene, and this requires the audio-visual relationship to produce a sense of causation. Simultaneously, I want the audience to focus on the sonic construction, which requires the concepts of causation to be inconclusive. I call this a 'fungible' audio-visual relationship: it produces causal perceptions but also throttles the fit between the sonic and the visual events. Using methods from experimental psychology, its 'medium fit' conveys perceptual binding, but also creates a level of interpretative discontinuity that loosens the perceptual hierarchy so that attention embraces fitting and non-fitting information with inconclusive concepts (Sá et al. 2015).

The fungible relationship combines synchronised and non-synchronised components with complexity, so as to confound the base cause. This can be applied to audio-visual mapping as well as to the audio-visual relationship in space, involving physical gesture and sound source location. Moreover, perception also binds sounds and images when there is no technological connection between them, and also, one can confound cause-and-effect relationships despite consistent synchrony; it often happens in live coding performances where the sound unfolds according to the programming code projected on a screen.

Non-Fitting Sounds and Images

'Low fit' means that perceptual binding is weak because the sound/image pairing does not activate previous memories of causation. In truth, we are driven to perceive – and imagine – connections between sounds and images, even when their fit is low and perceptual binding is weak. Often, one can extrapolate meanings from images of one thing coupled with sounds from something completely different, even if there is no synchrony. Yet, discerning an audio-visual relationship requires perception to create new chunks of memory, with a large amount of cognitive processing. And making sense of it might also draw attention away from the subtle relations between the sounds.

Attention, Intensity, and Discontinuity

Attention is sometimes automatically driven through stimuli and at other times it depends on individual control. The question is when and why. So far, the artistic debate on sensory dominance has focused on the audio-visual relationship, but visual dynamics are equally influential.

Attention Dynamics and Sensory Dominance

Sudden changes attract automatic attention (Knudsen 2007), and sudden visual changes such as light flashes cause vision to subordinate audition in multisensory integration (e.g., Kobayashi 2007). One can create a wealth of non-disruptive visuals by drawing from Gestalt psychology, whose research focuses on how we simplify the perceptual field according to a set of cognitive principles (e.g., Sá 2013).

Attention causes us to optimise perceptual resolution (e.g., Knudsen 2007), and then subtle discontinuities might become prominent. In fact, even when visual dynamics are not disruptive, if they are more discontinuous than the sonic construction, attention is likely to prioritise discontinuity. But that does not necessarily lead to unconscious multisensory integration processes. For example, the perceptual dislocation of the sound source toward the visual target (the 'ventriloquist effect') occurs with automatic attention but not with deliberate attention (Bertelson et al. 2000).

A Notion of Intensity and a Taxonomy of Continuities and Discontinuities

To analyse how any time-based work drives perception, it is worth thinking of intensity as the neural impact of any change in the chain of stimuli (Sá 2013, 2016). Being proportional to discontinuity, intensity depends not only on the event itself but also on the stimuli panorama and on a person's current perceptual resolution. I created a taxonomy of continuities and discontinuities related to intensity and attention, which distinguishes whether apprehensions are primarily driven through stimuli or if they are more under individual control. The taxonomy can be used to analyse the sonic and the visual dynamics independently. It can also be used to compare their relative strength.

Continuities

'Steady continuity' is of lowest intensity; it dispenses with attention. In Klima's *Glasbead*, the virtual camera that frames the image on the screen remains still in the digital 3D environment (Cityarts n.d.) – it does not prompt attention in itself because it does not move. The graphical interface itself – a spherical structure with stems – is not 'steady' because it rotates, but those movements are never sudden. Similarly, in Barri's *Versum*, the virtual camera moves progressively through the digital world (Tarik Barri n.d.). These are examples of 'progressive continuity', which occurs whenever successive, non-abrupt events display similar intervals of motion. Progressive continuity is of low intensity, as it fulfils the expectation that once something begins to move in a certain direction, it will continue to move in that direction. Indeed, we can speak of continuity whenever expectations are fulfilled.

Discontinuities

'Ambivalent discontinuity' also leaves attention under individual control, at least to some extent. If one does not pay attention, the foreseeable logic is shifted without disruption. And if one does, the discontinuity becomes more intense – while attention causes all changes to become more salient. To convey this in my digital 3D environments, I overlap multiple images with transparencies in ways that accentuate the impression of an organic, moving painting.

'Radical discontinuity' is disruptive, violating psychophysical expectations. It is of the highest intensity, prompting automatic attention. It can occur, for example, if there is an abrupt and unexpected transition from one digital 3D environment to another or when a sound suddenly appears or disappears. Yet again, one should keep in mind that intensity depends on expectations derived from previous events: an event can be disruptive when it appears for the first time, but then create continuity if repeated consistently.

The Global Experience

Now, we are armed with tools that facilitate an understanding of how any work exploring music with digital 3D environments drives a person's global experience.

Spatial Presence and Performative Arena

That experience is often discussed while using the term 'spatial presence': the feeling of being present in a mediated space. In media theory, Draper et al. (1998) proposed that this feeling depends on attentional processes, while Schubert (2009: 15) described it as 'a feedback of unconscious processes of spatial perception that try to locate the human body in relation to its environment'.

The sense of spatial presence reflects a performative arena, which is informed by the semantics of the work we are experiencing (Sá & Tanaka 2019, 2022). The 'local arena' implies a focus on a central target, such as a performer; the 'distributed arena' implies a focus on the physical environment, and the 'extended arena' implies a subjective sense of presence beyond the actual physical context.

Extending Presence beyond the Physical Environment

Whether we speak of 3D environments, sound art, or film, extending spatial presence beyond the actual physical environment requires stimuli to be loaded with information. For example, the sound of birds can convey the sensation of a countryside environment. 3D technologies apply the study of how we naturally see and hear the physical environment to convey a sense of immersion in the digital world beyond the screen. In the physical world, we use multiple cues to detect the material environment, including the position of objects and events. Film and media can create a convincing sense of space with much fewer cues because each event activates unconscious memories of similar real-world events.

Relating Semantics and Presence

Given our three types of performative arenas, I will consider three semantic dimensions: 'expressive semantics' means that attention is directed to a target; 'environmental semantics' means a focus upon a context or environment; and 'informational semantics' prompts causal percepts, shifting attention to a meaning.

Assessing how the global semantics of a work inform spatial presence might seem ambitious, as they stem from the intertwining of multiple variables – interaction, fit, and dynamics are influential, and so is the physical set-up in performance. Yet, we can assess each variable independently and then estimate its relative weight, given our understanding of how perception and attention work.

The Meaning of Continuities and Discontinuities

Sounds and images have informational content whenever they evoke something beyond themselves. But dynamics also have an intrinsic meaning. For example, continuities are not threatening because they fulfil expectations; hence, they leave attention free to focus on the environmental context or on internal states. Ambivalent discontinuities also leave attention under individual volition, but paying attention makes the perceived discontinuities more intense, enforcing the expressive dimension of the work. Radical discontinuities make expressive semantics very strong, prompting automatic attention and monopolising conscious awareness. That is also why radical visual discontinuities lead to visual dominance.

Informational Load and Audio-Visual Binding

The binding of auditory and visual stimuli is informed by concepts of causation, which, by definition, have an informational load. The greater the audio-visual fit, the greater the load and the stronger the perceptual binding. High fit leads the brain to form integrated audio-visual representations. However, while simplifying the sensory information according to presuppositions – i.e., 'sparing' cognitive workload – perception also cancels the relations between the sounds themselves.

The Semantics of Interaction

Predictable, clearly perceivable interface behaviours provide a large amount of information about how the system should be interacted with. A system that does not depend on real-time control is effortless as well, but the information content of the interaction is low, as the audience does not perceive to what extent the performer influences system output. Furthermore, perceiving effort implies interpreting causes and meanings. And effort tends to attract attention, supporting expressive semantics.

Physical Set-up and Presence

A performer's central position, a sound source placed next to them, or a spotlight over them will have the effect of directing attention to a central target, conveying expressive semantics. Conversely, the distribution of sound sources and visual projections in space will emphasise the environment.

Conclusion

The psycho-physical space of a work intertwines multiple semantic dimensions – sometimes, attention is driven through stimuli, and at other times, it remains under individual control. The audience's sense of presence contracts and expands accordingly. There is also an interesting grey area between deliberate and automatic attention. First, certain changes may become disruptive as a person sustains attention over time. Second, because audio-visual relations and physical set-ups can be forged so as to convey inconclusive interpretations, these loosen the psycho-physical principles that usually govern attention, vitalising the mind.

Bridging artistic practice, interaction design, audio-visual theory, and perception science is useful for the analysis of existing systems/instruments as well as for instrument design, composition, and performance. It enables an understanding of how each aspect of a work drives attention, and it provides perspective on the manner in which the product exceeds the sum.

References

'3D Audio.' (2014). *FMOD Studio 101*, viewed 3 June 2023, <http://www.soundlibrarian.com/uploads/3/1/0/6/3106267/fmod_studio_101_chapter_10_3d_audio_gdc.pdf>.

'Audioscape.' (n.d.). Wikibin, viewed 13 December 2024, <http://wikibin.org/articles/audioscape.html>.

Barri, T. (2008). *Versum* project, viewed 30 June 2023, <http://tarikbarri.nl/projects/versum>.

Bertelson, P., Vroomen, J., De Gelder, B. & Driver, J. (2000). 'The Ventriloquist Effect Does Not Depend on the Direction of Deliberate Visual Attention', *Perception and Psychophysics*, vol. 62, no. 2, pp. 321–332, <https://doi.org/10.3758/BF03205552>.

Berthaut, F., Katayose, H., Wakama, H., Totani, N. & Sato, Y. (2011). 'First Person Shooters as Collaborative Multiprocess Instruments', in *Proceedings of the International Conference on New Interfaces for Musical Expression*, Oslo, Norway, 30 May–1 June, pp. 44–47, <https://www.nime.org/proceedings/2011/nime2011_044.pdf>.

Boltz, M. (2004). 'The Cognitive Processing of Film and Musical Soundtracks', *Memory & Cognition* vol. 32, pp. 1194–1205.

Braudy, L. (1976). *The World in a Frame*. Garden City, New York: Anchor Press.

Brown, S. W. & Boltz, M. (2002). 'Attentional Processes in Time Perception: Effects of Mental Workload and Event Structure', *Journal of Experimental Psychology: Human Perception and Performance*, vol. 28, no. 3, pp. 600–615.

Cannon, J. & Stuart, F. (2012). 'The Investment of Play: Expression and Affordances in Digital Musical Instrument Design', *Proceedings of the International Computer Music Conference*, Ljubljana, Slovenia, 9–14 September, pp. 459–466.

Cascone, K. (2000). 'The Esthetics of Failure: Post-Digital Tendencies in Computer Music', *Computer Music Journal*, vol. 24, no. 4, pp. 12–18.

CCRMA Wiki. (n.d.). 'q3osc: works', viewed 12 December 2024, <https://ccrma.stanford.edu/wiki/Q3osc:_works>.

Chion, M. (1994). *Audio-Vision: Sound on Screen*, trans. C. Gorbman, New York: Columbia University Press.

Cityarts. (n.d.). 'glasbead', viewed 13 December 2024, <https://www.cityarts.com/glasbeadweb/glasbead.htm>.

Csikszentmihalyi, M. (1996). *Creativity: Flow and the Psychology of Discovery and Invention*. New York: HarperCollins Publishers.

----------------------- & Csikszentmihalyi, I. S. (eds). (1988). *Optimal Experience: Psychological Studies of Flow in Consciousness*. New York: Cambridge University Press.

------------------------ (2004). *Flow: The Psychological of Optimal Experience.* New York: Harper Perennial Modern Classics.

Di Scipio, A. (2015). The Politics of Sound and the Biopolitics of Music: Weaving Together Sound-Making, Irreducible Listening, and the Physical and Cultural Environment', *Organised Sound*, vol. 20, no. 3, pp. 278–289.

Dourish, P. (2004). *Where the Action Is: The Foundations of Embodied Interaction*. Cambridge: MIT Press.

Draper, J. V., Kaber, D. B. & Usher, J. M. (1998). 'Telepresence', *Human Factors*, vol. 40, no. 3, pp. 354–375.

Grierson, M. (2007). 'Noisescape: A 3D Audiovisual Multi-User Composition Environment', in J. Birringer, T. Dumke & K. Nicolai (eds), *The World as Virtual Environment* (pp. 160–168). Dresden: Trans-Media-Akademie Hellerau.

Hamilton, R. (2008a). 'q3osc: or how i learned to stop worrying and love the game', *International Computer Music Association Conference*, Belfast, Ireland, 24–29 August. Michigan Publishing.

-----------------. (2008b). 'maps and legends: designing FPS-based interfaces for multi-user composition, improvisation and immersive performance', *Proceedings of Computer Music Modeling and Retrieval. Sense of Sounds: 4th International Symposium 2007*, Volume 4969. Copenhagen, Denmark, August 2007.

Hunicke, R. & Chapman, V. (2004). 'A.I. for Dynamic Difficulty Adjustment in Games', *Challenges in Game AAI Workshop* 2*, 19th National Conference on Artificial Intelligence*, San Jose, USA.

Karkowski, Z. (1992). 'The Method Is Science, the Aim Is Religion', viewed 3 June 2023, <https://fyours.wordpress.com/2011/01/15/the-method-is-science-the-aim-is-religion>.

'John Klima's "Glasbead".' (1999). *El Arte en la Edad del Silicio,* viewed 12 December 2024, <https://arteedadsilicio.com/2000-2/2000-john-klima-glasbead>.

Knudsen, E. I. (2007). 'Fundamental Components of Attention', *Annual Review of Neuroscience*, vol. 30, no. 1, pp. 57–78.

Kobayashi, M., Osada, Y. & Kashino, M. (2007). 'The Effect of a Flashing Visual Stimulus on the Auditory Continuity Illusion', *Perception & Psychophysics*, vol. 69, no. 3, pp. 393–399.

Koster, R. (2004). 'Theory of Fun', viewed 3 June 2023, <http://www.theoryoffun.com>.

Kubovy, M. & Schutz, M. (2010). 'Audio-Visual Objects', *Review of Philosophy and Psychology* vol. 1, no. 1, pp. 41–61.

Machover, T. (2009, January–March). 'Beyond Guitar Hero – Towards a New Musical Ecology', *RSA Journal*.

Mudd, T., Holland, S., Mulholland, P. & Dalton, N. (2015). 'Nonlinear Dynamical Systems as Enablers of Exploratory Engagement with Musical Instruments', *Inter-Face: International Conference on Live Interfaces*, Lisbon, Portugal, November 2012, pp. 172–182, <https://adrianasa.org/research/ICLI_2014-Proceedings.pdf>.

Pachet, F. (2004). 'Enhancing Individual Creativity with Interactive Musical Reflective Systems', in I. Deliège & G. Wiggins (eds), *Musical Creativity: Current Research in Theory and Practice*, Psychology Press, <https://www.francoispachet.fr/wp-content/uploads/2021/01/pachet-06-Enhancing_Individual_Creativity.pdf>.

Paul, C. (2003). *Digital Art.* New York: Thames & Hudson.

Poepel, C. & Overholt, D. (2006). 'Recent Developments in Violin-Related Digital Musical Instruments: Where Are We and Where Are We Going?' In *International Conference on New Interfaces for Musical Expression*, Paris, France, 4–8 June, pp. 390–395, <https://www.nime.org/proceedings/2006/nime2006_390.pdf>.

Pressing, J. (1997). 'Some Perspectives on Performed Sound and Music in Virtual Environments', *Presence*, vol. 6, no. 4, pp. 482–503.

Riva, G., Mantovani, F., Waterworth, E. L. & Waterworth, J. A. (2015). 'Intention, Action, Self and Other: An Evolutionary Model of Presence', in M. Lombard, F. Biocca, J. Freeman, W. IJsselsteijn & R. Schaevitz R (eds.) *Immersed in Media: Telepresence Theory, Measurement & Technology* (pp. 73–94). Cham: Springer International Publishing.

Robinson, C., Barnhart, W. & Rivera, S. (2015). 'Auditory Stimuli Slow Down Responses and First Fixations: Support for Auditory Dominance in Adults', *37th Annual Meeting of the Cognitive Science Society*, Pasadena, USA, 22–25 July 2015.

Ryan, J. (1991). 'Some Remarks on Musical Instrument Design at STEIM', *Contemporary Music Review*, vol. 6, no. 1, pp. 3–17.

Sá, A. (2013). 'How an Audio-Visual Instrument Can Foster the Sonic Experience', *Leonardo*, vol. 19, no. 3, pp. 284–305.

-------. (2014). 'Repurposing Video Game Software for Musical Expression: A Perceptual Approach', *Proceedings of the International Conference on New Interfaces for Musical Expression*, 30 June – 4 July, London, pp. 331–334.

-------. (2016). *A Perceptual Approach to Audio-Visual Instrument Design, Compostion and Performance*, PhD thesis, Goldsmiths, University of London, viewed 3 June 2023, <https://research.gold.ac.uk/id/eprint/19431>.

-------. (2017). 'Designing Musical Expression', In *xCoAx – Conference on Computation, Communication, Aesthetics & X*, Lisbon, Portugal, July, pp. 226–240.

-------. (2021). 'A Method for the Analysis of Sound Art and Audio-Visual Performance', In J. S. Sánchez & J. G. Pinto (eds), *Audiovisual e Indústrias Criativas: presente e futuro* Vol.1: McGraw Hill Education, pp. 575–589.

-------. (2024). 'Transdisciplinarity, Composition, Expression: Reflections of a Spherical Way of Thinking', *Leonardo*, vol. 57, no. 1, DOI: https://doi.org/10.1162/leon_a_02472

-------., Caramieux, B. & Tanaka, A. (2015). 'The Fungible Audio-Visual Mapping and Its Experience', *Journal of Science and Technology of the Arts*, vol. 6, no. 1, pp. 85–96.

-------., Ryan, J., Mcpherson, A., Magnusson, T., Tanaka, A., Van der Heide, E., Grierson, M., Carvalhais, M. et al. (2005). 'Live Interfaces: Seeds of Debate', in *Inter-Face: International Conference on Live Interfaces*, Lisbon, Portugal, November 2012, pp. 13–28, viewed 12 December 2024, <https://adrianasa.org/research/ICLI_2014-Proceedings.pdf>.

------- & Tanaka, A. (2019). 'The Variables of Spatial Presence', in *xCoAx – Conference on Computation, Communication, Aesthetics & X*, Milan, Italy, July, pp. 140–155.

-----------------------. (2022). 'A Parametric Model for Audio-Visual Instrument Design, Composition and Performance', in S. Gibson, S. Arisona, D. Leishman & A. Tanaka (eds), *Live Visuals: History, Theory, Practice* (pp. 240–266). London: Routledge.

Schaeffer, P. (1966). *Traité des Objets Musicaux: Essai Interdisciplines*. Paris: Éditions du Seuil.

Schroeder, F. & Rebelo, P. (2009). 'The Pontydian Performance – The Performative Layer', *Organised Sound*, vol. 14, no. 2, pp. 134–141.

Schubert, T. (2009). 'A New Conception of Spatial Presence: Once Again, with Feeling', *Communication Theory*, vol. 19, no. 2, pp. 161–187.

Sorensen, V. (2005). 'Global Visual Music Jam Project', *Leonardo*, vol. 38, no. 4 (August), p. 316.

-------------., Puckette, M. & Steiger, R. (1997). 'Global Visual Music Project', <http://vibeke.info/lemma-2>.

Sweetser, P. & Wyeth, P. (2005). 'GameFlow: A Model for Evaluating Player Enjoyment in Games', *Computers in Entertainment (CIE) – Theoretical and Practical Computer Applications in Entertainment*, vol. 3, no. 3, p. 3.

Tarik Barri. (n.d.). *Versum project*, <http://tarikbarri.nl/projects/versum>.

Vasulka, S. & Vasulka, W. (1996). *Machine Media.* New York: MoMA.

Wozniewski, M., Settel, Z. & Cooperstock, J. (2006a). 'A Framework for Immersive Spatial Audio Performance', *International Conference on New Interfaces for Musical Expression*, Paris, France, 5–7 June.

---. (2006b). 'A Paradigm for Physical Interaction with Sound in 3D Audio Space', *ICMC: International Computer Music Conference*, New Orleans, USA, 6–11 November 2006.

---. (2006c). 'A Spatial Interface for Audio and Music Production', *International Conference on Digital Audio Effects* (DAFx), Montreal, Canada, 18–20 September, pp. 271–274.