A METHOD FOR THE ANALYSIS OF SOUND ART AND AUDIO-VISUAL PERFORMANCE

Adriana Sá
CICANT

Resumo

O termo NIME, acrónimo de "new interface for musical expression", aplica-se a uma grande diversidade de práticas criativas. Mas raramente se discute o que significa 'expressão'. Neste artigo formulamos uma noção de expressão em que a interação performer-instrumento é importante, assim como a relação entre audição, visão e espaço. Articulando prática artística e ciências da perceção, descrevemos três princípios criativos e um modelo de visualização paramétrica. O modelo inclui parâmetros relativos à interação, às dinâmicas sonora e visual, à relação áudio-visual, à configuração do espaço físico e à semântica. Estes parâmetros são aplicáveis a qualquer plataforma técnica e abordagem estética. O método de visualização aqui proposto facilita a análise do modo como a sua inter-relação conduz a experiência do público.

Palavras-chave

Expressão sonora, performance audiovisual, dominância sensorial, "spatial presence", modelo de visualização paramétrica

Abstract

The term NIME, acronym of "new interface for musical expression", applies to a great diversity of creative practices. Nevertheless, the meaning of 'expression' is *rarely discussed. In this article we formulate an understanding of expression where the reciprocal interaction between performer and instrument is important, as well as the relation between audition, vision and space. Articulating artistic practice and the science of perception, we describe three creative principles and a parametric visualisation model. The model includes parameters for interaction, sonic and visual dynamics, audio-visual relationship, physical setup and semantics. Those parameters are applicable to any technical platform and aesthetic approach. Our proposed visualisation method facilitates the analysis of how their inter-relationship drives the audience's experience.*

Keywords:

Sonic expression, audiovisual performance, sensory dominance, spatial presence, parametric visualisation model

1. INTRODUCTION

Creative practice can benefit from research in perception science, and vice-versa. This has been our experience throughout the development of our audio-visual instrument, which outputs acoustic sound, digital sound, and digital image[1]. The combination of acoustic and digital components raised questions about interaction, and we wanted to extract musical meaning from those questions. We wanted the image to function as a reactive stage scene, and that was problematic because we had the strong impression that vision tends to subordinate audition. Our creative concerns motivated a perceptual approach to instrument design, composition and performance. On the one hand, the research clarified insights derived from artistic practice. On the other, it exposed areas of scientific uncertainty, which became further creative material.

Bridging creative practice, neuroscience and psychology, music cognition, interaction design and audio-visual theory, we defined creative principles for the sound, the image and the audio-visual relationship. Whilst those principles informed the development of our instrument (2013; 2014; 2017), the underlying investigations expanded greatly beyond personal practice. They enabled us to identify the variables, or parameters relevant to sonic expression (2017), sensory dominance (Sá, 2013; Sá, Caramiaux and Tanaka, 2015) and spatial presence, i.e. the subjective sense of presence in a mediated space (Sá and Tanaka 2019). This article will summarise our research and propose a way to visualise the relation between those parameters, in abstract form. We will use this graphic systematisation to discuss our practical work.

1 http://adrianasa.planetaclix.pt/research/practiceOverview.htm

Indeed, instruments and performance situations can be analysed through different parameters, and these can be represented graphically, with a set of axes. Parametric visualisation models can provide a holistic view of the aspects relevant to any theoretical and practical framework. For example, the model created by Birnbaum et al. reveals relations between interaction, sound organisation, physical distribution in space, and semantics of sound (2005). The model created by Magnusson reveals how digital music devices condition interaction and sonic results (2010). And the model created by Ciciliani reveals how an electronic music performance might draw the focus to the performer or the environment (2015). Our own model reveals how the combination of sound, image, audio-visual relationship and physical setup inform the audience's experience.

2. CREATIVE MOTIVATIONS AND RELATED PARAMETERS

We mentioned that our audio-visual instrument combines acoustic sound, digital sound, and digital image. The acoustic component is a *concert zither* (multi-string instrument, see Figure 1) with custom strings and tuning. Its sound is used as an input to the audio-visual software, which processes 3D graphics (Figure 1) and recorded sounds. In performance we use two interfaces: the audio analysis of the zither, and a set of switches (from a game pad or a computer keyboard).

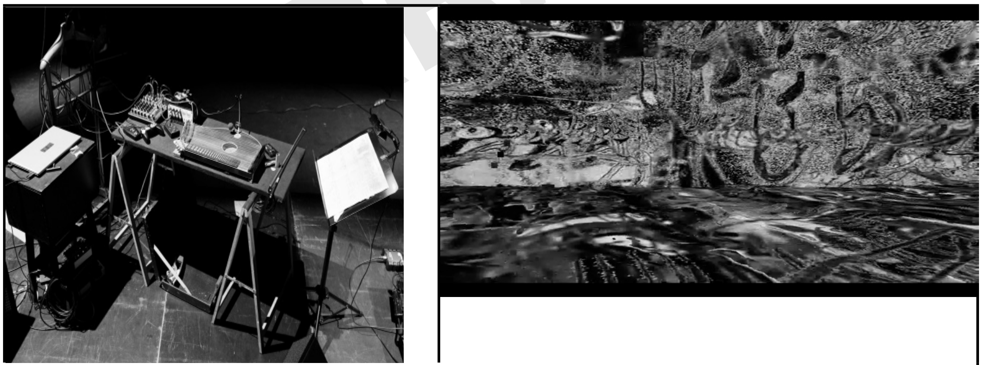

Figure 1: zither (left/ photo by Nuno Martins) and digital 3D world, version from 2021 (right)

The instrument utilises software that is originally intended for the creation of video games. Video game engines can render a digital 3D world, and make the position of elements in that world have an impact upon the spatialisation of related sounds. But as we started to work with video game engines, we needed to question certain theories embedded in the digital platform - concealed in code. For example, the software assumed that digital behaviours must be consistent, and cause-effect relationships clearly perceivable. One can adopt those theories so as to hybridise music and videogaming. This is the case with John Klima's *Glasbead*, created in the 1990s: a multi-user collaborative musical interface consisting of a rotating spherical structure with stems that resemble hammers and bells (Paul 2008). Other examples are in the work of R. Hamilton (2008) or Tarik Bari (see http://tarikbarri.nl/projects/versum). Our own creative concerns are quite different.

We were able to implement our modifications with the collaboration of John Klima, expert in video game programming, C and C++. The process entailed many discussions about motivations and feasibility. The following sections explain how we clarified our insights, and how that led to a set of parameters that can be used to discuss the convergences and divergences between different creative practices.

2.1 THE ROLE OF EFFORT IN SONIC EXPRESSION

We consider music to embrace vision and space, beyond audition. But there is also a kind of hierarchy, as in our performances the primary focus is on the sonic construction. We are driven to explore a particular understanding of expression, where the performer's interaction with the instrument is reciprocal: an instrument is simultaneously a controlled prolongation of the body, and a means of expanding action beyond intention. 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. This kind of musical motion relates to that which Jeff Pressing called dynamic complexity: a rich range of behaviours over time, an adaptation to unpredictable conditions, a monitoring of results in relation to a reference source, and an anticipation of changes in oneself or the environment (1987). In summary, we draw expression from behavioural deviations and reactions to those deviations.

These considerations lead to the creative principle of sonic complexity:

> To threshold the performer's control over the instrument, and the unpredictability of sonic outcomes - so that the instrument affords sonic complexity, in a way that suits the performer's idiosyncratic expression.

This principle can be explored in many different ways, but it does not claim to be consensual. Discussing interaction designs is particularly important because the extent to which new interfaces for musical expression (NIME) allow for agency is quite variable. For example, Tod Machover and the MIT Lab developed interfaces intended to compensate for "people's limitations" (2008). This implies that the software prescribes which output results are desirable, and which are not. In contrast, Joel Ryan (who pioneered the digital signal processing of acoustic instruments) affirms that it is interesting to make control as difficult as possible, because effort is closely related to musical expression (Ryan 1991; Sá et al. 2015). The point is, interaction designs can be governed by different notions of expression, and effort is a distinguishing parameter; it can be quantified according to the amount of cognitive processing, conscious and unconscious. As a parameter, effort is applicable to both digital and acoustic instruments. Also, it summarises parameters from Birnbaum's model (2005) as well as Magnusson's model (2010).

In our own parametric model, *little effort* can mean that the relation between deliberate human agency and sonic results is linear and clearly perceivable. Alternatively, it means that the music does not depend much on the performer's interaction. *Medium effort* indicates a convergence with our principle of sonic complexity. It denotes that the instrument requires particular skills, yet a sense of immediacy conveys musical timing, and/ or technical configurations rule out undesired outcomes. Finally, *high effort* implies high cognitive demand; the interaction with the system does not feel immediate, and/or the system does not rule out any outcomes.

2.2 THE QUESTION OF SENSORY DOMINANCE

Regardless of interaction, the experience of music exceeds audition, because perception is a process of multisensory synthesis. And the mutual impact of vision and audition leads to a perceptual surplus, as Michel Chion described (1994). However, the strength of each sensory modality tends to be unbalanced. Pierre Schaeffer noted that sounds must be detached from their originating cause to be fully experienced (1966), and Jeff Pressing described how this applies to digital 3D environments (1997). Yet in neuroscience, Sinnett et al. found that attention can be manipulated so that vision does not dominate over audition (2007). He didn't say how, but we wanted a moving image to allow for the experience of complex sonic constructions. Extrapolating from creative practice, audio-visual theory and the science of perception, we endeavoured to identify the variables that influence sensory dominance.

2.3 THE ROLE OF VISUAL DYNAMICS IN SENSORY DOMINANCE

So far, the artistic debate on sensory dominance focused on causation, but the dynamics of the image is equally influential. Sudden changes attract automatic attention (e.g. Knudsen 2007), and sudden visual changes such as light flashes cause vision to subordinate audition in multisensory integration (e.g. Kobayashi 2007). One can avoid visual disruptions by drawing from the Gestalt psychology, where researchers study how we simplify the perceptual field according to a set of cognitive principles, applicable in both the auditory and the visual domain (e.g Sá, 2013).

This understanding led us to formulate the creative principle of visual continuity:

> To dispense with sudden visual discontinuities. One can explore a wealth of visual changes, as long as Gestalt principles enable the perceptual simplification of visual dynamics.

In fact, even when the visual dynamics are not disruptive, if they are more discontinuous than the sonic construction, attention is likely to prioritise discontinuity. But that will not lead to unconscious multisensory integration processes. For example, the ventriloquist effect (i.e. the perceptual dislocation of the sound source toward the visual targe) occurs with automatic attention, but not with deliberate attention (Bertelson et al. 2000).

At this point, we must stress that our creative decisions are not driven by demonstrative aims. The initial version of our audio-visual instrument assured perceptual simplification, but further iterations explore a grey area that respects to how the audience's attention might change over time. Indeed, perception optimises the sensory information related to the attention target (Knudsen 2007), which means that directing the eyes to the screen increases visual resolution. Hence, if over time one keeps paying attention to the visual details, they become gradually more intense. Therefore, one also becomes more susceptible to automatic visual attention.

2.4 CONTINUITIES AND DISCONTINUITIES RELATED WITH INTENSITY AND ATTENTION

To analyse how any time-based work drives perception, it is useful to consider the combined effects of automatic and deliberate attention. As such, we defined intensity as the neural impact of any change in the chain of stimuli (2013). In this way, intensity depends on the event itself, on the stimuli panorama, and on a person's current perceptual resolution. We further created a taxonomy that relates intensity with different types of continuities and discontinuities. Its terminology 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.

Steady continuity is of lowest intensity; it dispenses with attention. Conscious awareness is likely to deviate and focus upon any simultaneous stimuli, or upon internal states. *Progressive continuity* occurs when successive, non-abrupt events display a similar interval of motion. It 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. *Ambivalent discontinuity* is simultaneously continuous and discontinuous. Sensing discontinuity depends on deliberate attention, as attention causes us to optimise perceptual resolution. At lower resolution, the foreseeable logic is shifted without disruption. At high resolution, the discontinuity becomes more intense. *Radical discontinuity* is disruptive, violating psychophysical expectations. It is of highest intensity, prompting automatic attention. 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, and then create continuity by repeating consistently.

2.5 THE ROLE OF PERCEIVED CAUSATION IN VISUAL DOMINANCE

Perception science can help to clarify the artistic debate on how perceived causation influences sensory dominance. Many scientific experiments revealed that people tend to respond more accurately to visual stimuli than auditory stimuli - the phenomenon is known as the *Colavita effect* (Colavita 1974). The Colavita effect might reflect a response bias, rather than an attenuation of auditory encoding due to the visual input (Spence 2009). In other words, visual dominance might happen at a level of conceptualisation, rather than at a detection level. In fact, the auditory input slows down the visual response, and not the inverse (Robinson et al.2015). Whilst audition dominates with infants, adults strategically bias their responses in favour of visual input, possibly to compensate for the poor alerting abilities of visual stimuli.

In experimental psychology, Kubovy & Schutz coined the term *ecological fit* to describe how automatic interactions between the senses draw from concepts of causation - the aural discounts the visual and the visual discounts the aural based on those concepts (2009; 2010). The greater is the ecological fit, the more we ignore any diverging sensory information. This can be explained in terms of “cognitive efficiency”: a high level of fit leads to integrated perceptual encodings and representations, which require less cognitive processing than separated encodings and representations (Brown and Boltz 2002; Boltz 2004). And, high fit leads vision to dominate over audition (ibid).

In our performances we desire the image to create a reactive stage scene, and this requires the audio-visual relationship to produce a sense of causation. Simultaneously, we want the audience to focus on the sonic construction, which requires the concepts of causation to be inconclusive. Indeed, between congruence and incongruence, there must be a medium level of audio-visual fit, corresponding to ambivalent concepts. For example, the Gestaltist term *multistability* refers to when perception pops back and forth between different interpretations (Rubin 1921; Kubovy and Yu 2012).

This led us to formulate the creative principle of audio-visual fungibility:

> To create a fungible audio-visual relationship, which produces causal percepts, but also throttles the fit between the sonic and the visual events. The fungible relationship conveys perceptual binding, but it also creates a level of interpretative discontinuity that loosens the perceptual hierarchy, so that attention embraces fitting and non-fitting information with inconclusive concepts.

We conducted a study to demonstrate how different types of audio-visual mappings inform perceptual binding and prioritisation (Sá, Caramiaux and Tanaka 2015).

The study showed that the fungible mapping combines synchronised and non-synchronised components, exhibiting complexity so as to appear confusing. This is equally applicable to the audio-visual relationship in space, involving physical gesture and sound source location.

2.6 AUDIO-VISUAL RELATIONSHIPS AND PERCEPTUAL EFFECTS

Beyond personal practice, our study (2015) mentioned in the previous section enables the parameterisation of any audio-visual relationship:

High fit means that the audio-visual relationship conveys conclusions about causes and effects; that can happen even when inconsistent information reaches conscious awareness (as shown in Gestalt psychology). Perception prioritises information that converges with those conclusions, producing integrated mental representations. High fit is of low intensity, as it requires little cognitive processing. ***Medium fit*** means that one senses causation without understanding the base cause and effect relationships. It is of medium intensity and requires a medium level of cognitive processing. It conveys perceptual chunking, but the process of audio-visual binding remains ambivalent: one can form integrated as well as separated representations of the sounds and the images. These are the effects of a fungible audio-visual mapping, but perception also binds sounds and images when there is no technological connection between them, and also, one can confound the cause and effect relationships in spite of consistent synchrony; it often happens in live coding performances where the sound unfolds according to the code projected on a screen. Furthermore, ***Low fit*** means that perceptual binding is weak because the sound/ image pairing does not activate previous memories of causation. Discerning an audio-visual relationship requires perception to create new chunks of memory, with a large amount of cognitive processing. That means high intensity.

2.7 VISUAL RELATION BETWEEN PERFORMER AND INSTRUMENT

Our creative principles for the sound, the image and audio-visual relationship informed the development of our instrument, and the physical setup in performance is equally important to the audience's experience. Accordingly, our model includes a parameter for the visual relation between the performer and the instrument/ system:

Integrated means that the instrument and the performer's physical body form a single visual scene. This is the case in our performances, where the image is projected upon the performer. ***Separated*** means that the instrument is physically separated from the performer, who is nevertheless visible. This type of arrangement tends to divide attention, or deviate attention from the performer. ***Hidden*** means that the performer is not visible. The audience does not see their agency, but knowing what type of interface is being used can influence how the work is perceived.

2.8. A WAY TO PARAMETERISE SEMANTICS

Semantics can be discussed with respect to causes and meanings, but every experience has meaning, including when we focus attention on perceptual motion itself (Dewey 1989, Johnson 2007). Jeff Pressing proposed a useful semantic characterisation of sounds (1997), which can be adapted so as to account for how the different aspects of an audio-visual performance influence attention. He distinguished *expressive, informational* and *environmental* sounds, stressing that these typologies are normally overlapped. Expressive sounds would include all kinds of music and song. Examples of informational sounds would be speech, alarms, and sonified data. Examples of environmental sounds would include animal calls, wind sounds, and machine noises.

In adapting Pressing's terminology, we consider that the semantics of an audio-visual performance result from the combination of interaction, sound, image, audio-visual relationship and physical setup. One can assess the semantics of each parameter, and then estimate their relative weight in the global meaning of the work.

Expressive semantics convey a focus upon a central target. For example, ambivalent discontinuities and radical discontinuities attract attention, and so does a performer's central position in physical space, or a spotlight. This notion of expressive semantics is applicable to any sensory domain, but it also narrows Pressing's notion of "expressive sounds", i.e. all kinds of music and song. Indeed, many musical works do not drive attention to the performer. ***Environmental*** semantics convey a focus upon space and context. This extends what Pressing termed "environmental sounds", e.g. animal calls, wind sounds and the noises of machinery. For example, a sound system distributed around the audience creates a sense of environment. Steady and progressive continuities also create a sense of environment because they fulfil expectations; attention can draw to context and space. Finally, ***informational*** semantics prompt causal percepts, shifting attention to a meaning - we can quantify the informational load of a stimulus by assessing our conclusiveness about its cause or meaning. This understanding expands beyond Pressing s "informational sounds", e.g. speech, alarms and sonified data. It embraces all situations where the sound or the image evoke something beyond themselves. It also embraces unconscious informational load; for example, audio-visual synchrony means a causal relationship.

Concluding, the informational dimension of a work can support its expressive and environmental dimensions, but the latter also exist when the informational load is very little. Regardless of concepts and interpretations, the more attention focuses upon a specific target, the less it spreads through the environment, and vice-versa.

2.9 THE PERFORMATIVE ARENA

In our performances the moving image is projected over the performer, extending the psychological space of the work to a digital 3D world that morphs with sound (Figure 2). The sonic construction unfolds creating environmental soundscapes, or drawing attention to the performer, or evoking other spaces and situations.

Figure 2: Visual projection of the 3D world, versions from 2014 and 2019.

'Spatial presence' - the subjective feeling of presence in a mediated space - has been discussed in sound art (e.g. Lopez 2004, Voegelin 2010) and media theory (e.g. Hartmann et al. 2015); from those investigations we can infer that it is influenced by the characteristics of sound and image, as well as by the audio-visual relationship. Other researches discuss how the psychological space of the work is informed by speaker placement (e.g. Chion 1997, Emmerson 2007, Ciciliani 2015). Our parametric model articulates and expands those investigations. It includes a final parameter, which summarises how the different semantic dimensions of a creative work intertwine so as to shape the performative arena.

Simon Emmerson uses the term 'performative arena' to describe the relation between performer, audience, space, sound sources and events (2007). He also defines local functions and field functions: whilst the former extend the perceived relation of human performer to the sonic result, the latter create a context, a landscape or an environment where local activity may be found. Other researches reflect similar ideas. Birnbaum et al. consider "the total physical area inhabited by the instrument/ system" as a variable in performance (2005). And Ciciliani distinguishes between *centripetal* and *centrifugal* performances (2015). In centripetal performances, the performer is visible and the centre of attention. The relation between his physical action and the sonic results is clearly perceivable, and the sound sources are placed near to him. In centrifugal performances, the performer is in a hidden position. He functions as a controlling rather than enacting entity; there is little or no correspondence between his physical actions and the sonic results, and the sound sources are spread in space.

One's sense of presence can also be expanded beyond the physical space, by means of perceptual cues. For example, film and digital 3D environments reproduce how we naturally perceive the world when the visual position of a sound emitter reflects in sound spatialisation. Also, we can be mentally transported when listening to recognisable sounds in darkness, or watching a landscape in silence. Here again, events activate unconscious memories from similar real-world events.

The performative arena can contract and expand, inextricably related to attentional processes. Those processes might depend on the characteristics of the sound, the image and the audio-visual relationship, on the performer's interaction with the system, the speakers' placement, the spatial relation between performer and visual projection, if any, the lighting, the physical architecture and the audience location. Individual predisposition might be influential as well, but it does not depend on the work, and we do not intend to parameterise experience itself. We rather consider verifiable variables, and provide methods to interpret their relationships.

We distinguish three types of performative arena, which are not mutually exclusive: ***Local arena*** means a focus upon the performer. Expressive semantics are dominant. ***Distributed arena*** means a focus upon the environment. Environmental semantics are dominant. ***Extended arena*** means a subjective sense of presence beyond the physical performance space. It requires perceptual cues, which imply informational semantics.

The three types of arena reflect the three semantic typologies but the inverse is not necessarily true. For example, there is no local arena when the performer is not visible, yet the visual discontinuities and sonic deviations might create expressive semantics. Similarly, there can be environmental semantics without the arena being distributed or extended; a sonic mass of drones emitted through a single loudspeaker illustrates the idea. Also, the semantics of an audio-visual performance can be informational without the arena being extended, as happens in many live coding performances.

3. PARAMETRIC VISUALISATION

An audio-visual performance is a construction of experienced time, which depends on multiple, intertwining parameters. By systemising those parameters in a graphical way, we obtain a useful tool for instrument design, composition and performance. Indeed, we can draw a separation between those activities, but in practice that separation might not be so obvious: ultimately, the iterative creation process must always consider the final, global experience.

3.1 AXES RELEVANT TO SONIC EXPRESSION, SENSORY DOMINANCE AND SPATIAL PRESENCE

To facilitate any discussion about sonic expression, our parametric visualisation model needs an axis for interaction effort, another for sonic dynamics, and two more for semantics – one for the informational load, and another for the environmental and the expressive dimensions of the work (a single axis suffices because expressive and environmental semantics are inversely proportional).

Time has an influence upon the perceived discontinuities, but our model does not need a dedicated axis. If one mentions the time frame of experience, two axes suffice to represent the sonic and the visual dynamics, according to our taxonomy. In this way, the model can be used to visualise moment-by-moment situations or summarise large compositional structures.

To reveal sensory dominance, the model requires an axis for the visual dynamics and two axes for the ecological fit; one for the relation between sound and image, the other for the relation between physical gesture and system output. We can infer that a work conveys visual dominance when the model shows radical visual discontinuities, high ecological fit, or both. Audition can be command in any other case.

The parameters relevant to sonic expression and sensory dominance also influence the audience's sense of presence. In addition, an axis for the performer's visibility can give indications regarding the physical set-up. And an axis for the performative arena facilitates the disambiguation of certain aspects. It can provide cues about elements that have no direct representation, such as the placement of speakers and the lighting.

3.2 USING THE PARAMETRIC VISUALISATION MODEL TO DISCUSS PRACTICAL WORK

Our solo performances are never quite equal, but their graphic parameterisation should be the same, given the model's level of abstraction (Figure 3). The digital 3D world, the visual dynamics, the audio-visual relationship and the physical setup are set prior to performance – in a way that assures compliance with the creative principles of visual continuity and audio-visual fungibility. During performance we solely focus on the sonic construction; as the creative principle of sonic complexity requires us to address the unexpected resourcefully, we pay full attention to the music.

The *Sound Organisation* axis indicates that the interaction with our instrument entails medium *Real-time Effort*, reflecting the principle of sonic complexity: control and unpredictability are thresholded so as to convey expression. Our instrument requires skills to play the zither, and integrate unexpected digital sounds in the musical logics. But our direct control over the zither sound makes the instrument extremely responsive, and digital constraints rule out undesired outcomes.

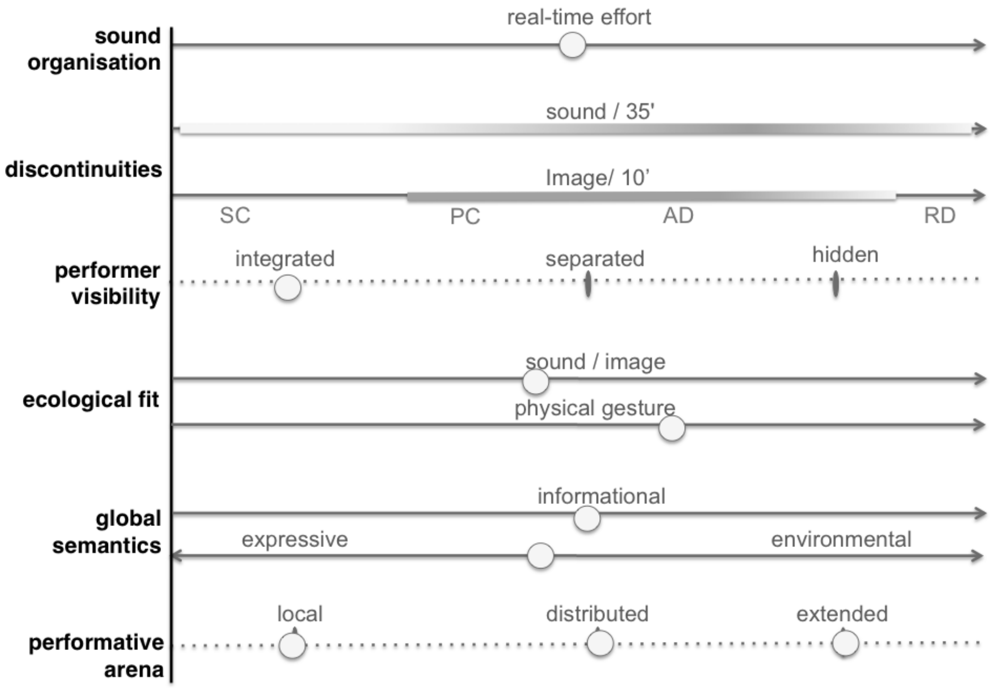

Figure 3: Parametric visualisation of a solo performance, about 40' long

The two *Discontinuities* axes represent the sonic and visual dynamics. Our compositions combine a few steady continuities, with no intrinsic motion; a few progressive continuities, where successive events display a similar interval of motion; many ambivalent discontinuities, which shift the foreseeable logic of the music without disruption; and occasional radical discontinuities, which are disruptive, prompting automatic attention. The orange gradient on *Sonic Discontinuities* shows that the music is more discontinuous than continuous: the strongest orange is closer to radical discontinuities than to progressive discontinuities. The yellow gradient reaches the left extreme and the right extreme of the axis, showing that the music combines all types of continuities and discontinuities. Ambivalent discontinuities and progressive continuities are constant in the visual dynamics, hence the stripe on *Visual Discontinuities* is mostly orange; the yellow gradient to the right represents the occasional moments where the visual changes become more intense. The axis reveals compliance with the principle of visual continuity: the image exhibits no radical discontinuities, which would attract automatic attention and subordinate audition.

The *Performer Visibility* axis represents our location relatively to the visual projection. The yellow dot on *Integrated* reflects that the image is projected over us, functioning as a stage scene. There is no dot on *Hidden* because we are always visible, and no dot on *Separated* because we never exit the frame of the visual projection.

The two *Ecological Fit* axes indicate medium fit, corresponding to a fungible audio-visual relationship. The dot on *Sound/ Image* represents the fungible audio-visual mapping, as well as the fungible relation between 3D sound (i.e. digital sound spatialisation dependent on visual dynamics) and inverted stereo diffusion. The mapping combines synchronised components, partially related components, and unrelated components. It creates a sense of causation, and confounds the base cause and effect relationships. The same happens with the audio-visual relationship in space. The combination of 3D sound and inverted stereo diffusion creates a sense of causation because the speed at which the sounds move through the speakers corresponds to the speed of the virtual camera. At the same time, it confounds the cause-effect relationships because one can hear each sound from two opposite directions, but only one corresponds to the sound emitter on the screen.

The dot on *Physical Gesture* is more to the right than the dot on Sound/ Image, due to the direct correlation between our gestures and the acoustic zither sound. It still indicates medium fit because that correlation is not clearly perceivable, for three reasons. First, our visible gestures are often not proportional to the loudness of the zither sound. Second, at times we "disappear" within the visual projection: the shadow over the projection outlines our physical body, yet our gestures are not clearly perceivable because the body is also a projection surface. And third, the digital sounds are at times purposefully undistinguishable from the acoustic sounds.

The two axes from *Global Semantics* indicate that the work has *Informational*, *Expressive* and *Environmental* semantic dimensions. The yellow dot on *Informational* is in a middle position, and that results from assessing several aspects. The fungible audio-visual relationship has informational load, because it explores perceptual cues so as to create a sense of causation; but not much, because the base cause-effect relationships are confounded. Recognisable, pre-recorded sounds (e.g. instrument sounds, wind and water sounds, etc) have informational load; but not much, because one is also led to strip perception from conclusive causes and meanings. The image is more loaded, as it represents an imaginary 3D world.

The yellow dot on the axis for *Expressive* and *Environmental* semantics shows that the expressive dimension of the work is slightly stronger than the environmental: the focus upon the performer is stronger than the focus on space. The expressive dimension of the work derives from a) the audience's sitting position, directed to the performer; b) our central position relatively to the visual projection; c) the pathos of our playing techniques; and d) a sound source placed next to us on stage (the zither amplifier). The environmental dimension comes from a) the digital 3D world (an environment per se) and the overall continuity of visual dynamics; b) the size of the projection and our integrated position in front of the screen; c) the environmental aspects of the sonic construction; and d) the sound system placed around the audience.

The *Performative Arena* axis shows a yellow dot on each discrete point, indicating that the work drives attention to shift between the performer, the environment, and imaginary spaces beyond the physical performance space. The arena is *Local* because there is an amplifier on stage, and the music often leads attention to focus on the performer's expressiveness in physical space. The arena is *Distributed* because a double inverted stereo system is placed around the audience, and attention often drifts from the performer, so as to focus upon the environment, or internal states. The arena is *Extended* because there is a large projection of a digital 3D world that morphs with sound; the audio-visual relationship explores perceptual cues, conveying spatial presence in the virtual space beyond the screen; and also, certain digital sounds, such as water and wind, convey a mental transportation to natural environments.

The *Performative Arena* axis provides information that the other parameters do not specify. The dot on *Local* shows that at least sometimes the performer has protagonism over the environment; this specifies *Performer Visibility*, with its dot on *Integrated*. The dot on *Distributed* specifies that a sound system is placed around the audience – indeed, the environmental semantics of the work could derive solely from the qualities of sound and/ or image. Also, the dot on *Extended* indicates that the image has figurative qualities – indeed, bi-dimensional abstract graphics would not extend the arena. The arena could be extended through the sonic construction alone, yet the dot on medium *Ecological Fit* permits disambiguation, as it denotes that the audio-visual relationship applies perceptual cues.

4. CONCLUSION

Our parametric visualisation model provides an operational means to glean how the different aspects of a performance inform perceptual experience. Its simple, orthogonal graphics can reveal a great amount of information, facilitating the analysis of sonic expression, sensory dominance and spatial presence. As an example, we used the model to discuss our practical work. Beyond personal practice, the model is applicable to any technical platform and aesthetical approach. It can be used to analyse existing instruments, create new audio-visual performances, and develop new audio-visual systems. One can also use only a part of the parameters so as to analyse recorded sound and film.

5. BIBLIOGRAPHIC REFERENCES

Bertelsen, O.; Breinbjerg, M. and Pold, S. (2009). Emerging Materiality: Reflections on Creative Use of Software in Electronic Music Composition. *Leonardo* 42*(3*) pp. 97–202.

Birnbaum, D.; Fiebrink, R.; Malloch, J. and Wanderley, M.(2005). Towards a Dimension Space for Musical Devices. *Proceedings of the International Conference on New Interfaces for Musical Expression,* Vancouver, BC, Canada.

Boltz, Merilyn (2004). The Cognitive Processing of Film and Musical Soundtracks. In *Memory & Cognition 32,* 1194-1205.

Braudy, Leo (1977). *The World in a Frame*. Garden City: Anchor Books.

Brown, Scott and Boltz, Merilyn (2002). Attentional Processes in Time Perception: Effects of Mental Workload and Event Structure. *Journal of Experimental Psychology: Human Perception and Performance* 28, pp. 600-615.

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

Ciciliani, Marko and Mojzysz, Zenon (2015). Evaluating a Method for the Analysis of Performance Practices in Electronic Music. *INTER-FACE: International Conference on Live Interfaces 2014,* pp. 45-61.

Colavita, Francis (1974). Human Sensory Dominance. *Perception & Psychophysics* 16(2), 409-412.

Dewey, John (1980) *Art as Experience.* New York: Penguin Putnam, Harvard (18th ed.).

Emmerson, Simon (2007). *Living Electronic Music*, Hampshire, U.K.: Ashgate

Publishing.

Hamilton, R. (2008). Maps and Legends: Designing Fps-based Interfaces for Multi-user Composition, Improvisation and Immersive Performance. *CMMR 2007 Proceedings,* Copenhagen (pp. 478-486).

Hartmann, T.; Werner, W.; Vorderer, P.; Klimmt, C.; Schramm, H. and Böcking, S. (2015). Spatial Presence Theory: State of the Art and Challenges Ahead. In *Immersed in Media: Telepresence Theory, Measurement & Technology*. Springer International Publishing, pp.115-135.

Johnson, Mark (2007). *The Meaning of the Body*: *Aesthetics of human understanding*. University of Chicago Press.

Knudsen, Eric (2007). Fundamental Components of Attention. *Annual Review of Neuroscience 30(1)*, 57-78.

Kobayashi, M.; Osada, Y.; Kashino, M. (2007). The Effect of a Flashing Visual Stimulus on the Auditory Continuity Illusion. *Perception & Psychophysics* 69(3), pp. 393-399.

Kubovy, Michael and Schutz, Michael (2010). Audio-Visual Objects. In *Review of Philosophy and Psychology 1(1)*, pp. 41–61. Springer Science and Business Media B.V.

Kubovy, Michael and Yu, Minhong (2012). Multistability, Cross-Modal Binding and the Additivity of Conjoined Grouping Principles. *Philosophical Transactions of the Royal Society Vol. B(367)*, 954–964.

López, Francisco, official Web Site, "Against the Stage" (February 2004), http://www.franciscolopez.net/stage.html (accessed May 6, 2021).

Machover, Tod (2009). Beyond Guitar Hero - Towards a New Musical Ecology. *RSA Journal*, London.

Magnusson, Thor (2010). An Epistemic Dimension Space for Musical Devices. *Proceedings of the Conference on New Interfaces for Musical Expression,* Sydney, Australia.

Paul, Christiane (2008). *Digital Art*. Thames & Hudson.

Pressing, Jeff (1987). Cognitive Complexity and the Structure of Musical Patterns. In J. Sloboda (Ed), *Generative Processes in music: The Psychology of Performance, Improvisation, and Composition*, pp. 129-178. New York, Oxford University Press.

Pressing, Jeff (1997). Some Perspectives on Performed Sound and Music in Virtual Environments. *Presence 6(4)*.

Robinson, C.;· Barnhart, W. and 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, California.

Rubin, Edgar (1921). *Visuell Wahrgenommene Figuren - Studien in Psychologischer Analyse*. Copenhagen, Gyldendalske Boghandel.

Ryan, Joel (1991). Some Remarks on Musical Instrument Design at STEIM. *Contemporary Music Review*, 6(1), 3-17.

Sá, Adriana (2013). How an Audio-Visual Instrument Can Foster the Sonic Experience. *Live Visuals/ Leonardo Almanac* 19(3), pp. 279-299. MIT Press.

Sá, Adriana (2014). Repurposing Video Game Software for Musical Expression: a perceptual approach. *Proceedings of New Interfaces for Musical Expression,* pp. 331-334.

Sá, A.; Caramiaux, B.; Tanaka, A. (2015). The Fungible Audio-Visual Mapping and its Experience. *Journal Of Science And Technology Of The Arts* 6(1), 85-96.

Sá, A.; Ryan, J.; Van Der Heide, E.; Tanaka, A.; McPherson, A., Magnusson, T. McLean, A. Carvalhais, M. and Gierson, M. (2015). Live Interfaces: Seeds of Debate. In Sá, Adriana; Carvalhais, Miguel & McLean, Alex (Ed.), *INTER-FACE: International Conference on Live Interfaces 2014,* pp. 14-28. Porto University, CECL & CESEM (NOVA University) and MITPL (University of Sussex).

Sá, Adriana (2017). Designing Musical Expression. *Proceedings of xCoAx 2017.*

Sá, Adriana and Tanaka, Atau (2019). The Variables of Spatial Presence: a parametric model. *Proceedings of xCoAx.*

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

Schutz, Michael and Kubovy, Michael (2009). Causality and Cross-Modal Integration. *Journal of Experimental Psychology: Human Perception and Performance* 35(6), 1791–1810.

Sinnett, S.; Spence, C. and Soto-Faraco, S. (2007). Visual dominance and attention: The Colavita Effect Revisited. *Perception & Psychophysics 69 (5*), 673-686.

Spence, Charles (2009). Explaining the Colavita Effect. In Srinivasan, N. (Ed.), *Progress in Brain Research 176,* pp. 245-58. Elsevier.

Voegelin, Salomé (2010). *Listening to Noise and Silence: Towards A Philosophy Of Sound Art.* Continuum.