Syntithenai Poiēsis—Perceptual Unified Objects



André Rangel

Abstract In this text, I want to affirm that we don't look at images, that there are no such thing as external images to the human body, and that all images are mental—therefore internal to the human body because there's where they are built. Having this as a premise I argue, drawing from artistic experience, for the possibility of imagining and actualising audiovisual artworks that, originating in a fusion of mental sonic and visual images, result as perceptual objects. I affirm the equality of the sonic and visual stimuli in intermedial artistic compositional processes in which both stimuli are composed simultaneously, without precedence, hierarchies, or *paragone*. This text is about experiencing the simultaneity of visual and sonic sensations, without precedence or primacy of one over the other, both in perceptual processes and in the processes of creation and reception.

Keywords Image theory · Perceptual object · Audiovisual · Intermedia · Computational art

1 Context

The title of this book—"Sounding Images"—suggests to me the precedence of the image over sound and the possibility of images that sound. An image is a broad concept, mostly associated with the sense of sight, a visual perception. We don´t look at images. All images are mental constructions that can be triggered by sensory stimuli or result from articulations and actions of memory. All images result from sensory experiences—whether present or past—and all are mental because they arise from mental activity (Rajamanickam, 2007). Both sensory and mental images can be visual, auditory, olfactory, or related to any other module of human sensitivity. The audible and the visible are external to the human body therefore hearing and vision are internal, mental, embodied constructions. The title of this book may also refer

A. Rangel (✉)
i2ADS—Research Institute in Art, Design and Society, FBAUP—Faculty of Fine Arts of the University of Porto, Porto, Portugal
e-mail: arangel@fba.up.pt

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025

L. Correia Castilho et al. (eds.), *Perspectives on Music, Sound and Musicology II*,
https://doi.org/10.1007/978-3-031-67503-4_19

to the possibility of sonic compositions and improvisations being precisely physical materializations of mental—sonic—images, constructed in and by the bodies of musicians and composers. This implies that in an artistic process, there is a cascade of images consisting of a mental image that precedes the construction of an artistic materialization—creation, design, and construction by the artist—that will later serve as a sensory image triggering the construction of another mental image built in and by the spectator of the artistic object. In this context, and from a perspective that may seem Platonic, mental and sensory images in this cascade are not entirely correspondent. A movie—and its copies—as an example, is composed of numerous and different mental images of the same film—in the bodies of the screenwriter, director, actors, technical personnel, and viewers. The lack of fidelity of the mental image—which is at the origin of the materialization of an artistic creation process—with its own physical materialization, and the lack of fidelity of the mental image in the creator, to the mental image in the viewer—sensitized by the artistic object—is due, as the etymological origin of the term indicates, to the fact that an image is always an "imitation".[1] Since what imitates is not what is imitated, images are never what they imitate.

In this text, I aim to assert the possibility of a sonic and visual amalgam, a fusion of mental sonic and visual images, both within the body of the artist and the bodies of the spectators. That is, the experience of the simultaneity of visual and sonic sensations, without precedence or primacy of one over the other, both in perceptual processes and in the processes of creation and reception. My goal is to affirm the equality of sonic and visual stimuli as means of artistic expression in intermedial compositional processes where sonic and visual stimuli are composed simultaneously, without precedence, hierarchies, or *paragone*.

2 *Paragone* 1—Visual Perception, Sound Perception and Hierarchy of Senses

Human vision and hearing result from the action of different physical energetic flows under different conditions and different physiological structures, however, both operate with oscillation frequencies—of photons and particles respectively.

Visible light—visible electromagnetic spectrum—is electromagnetic radiation that propagates in transverse waves whose frequencies are limited to a narrow band of the entire electromagnetic spectrum. When light falls on objects it is reflected, absorbed or transmitted. The matter that constitutes objects has the ability to absorb and reflect different frequencies of the visible spectrum, meaning that the light reflected by objects has different spectral compositions than the spectral composition of the light that falls on them. When light, reflected by objects, sensitizes the human retina, it stimulates different chromatic and light sensations through a

[1] The word "image" has its etymological root in the Latin word "imitari," which means to copy or imitate.

chemical process that occurs in the photoreceptor cells—cones and rods—which constitute the sensory epithelium of the eye—retina. Human beings have the ability to construct mental images based on the interpretation of these stimuli. The greatest concentration of cones in the center of the retina—fovea—allows for great two-dimensional resolution. Depth perception is due to the difference in point of view of each eye—parallax.

Sound is the propagation of mechanical energy, in longitudinal waves, through the vibration—pressure variation—of particles in an elastic medium. Human beings interpret part of the mechanical frequency spectrum as sound. Vibrations that arrive from outside the ear vibrate the eardrum, which is in contact with a set of bones—malleus, incus, and stapes—which mechanically transmit the vibration to the cochlea, which transforms it into an electrical signal that is sent to the brain via the auditory nerve. Within the cochlea, two structural properties of the basilar membrane—increasing width and decreasing stiffness in the base-apex direction—allow, depending on the location of the membrane affected by vibration, to encode different frequencies with precision. The human ear is extremely precise regarding frequency, amplitude and time but poorly configured to locate the origin of a vibration in space. However, because we have two ears—internal and external—we can tell whether the origin of a sound vibration is in front of us, behind us or to one of our sides through bin-aural comparison and the analysis of spectral interference caused by the external ears.

In short, light and sound are energy transport phenomena that propagate through waves, longitudinal mechanical in the case of sound and transversal electromagnetic in the case of light. Mechanical ones need a material medium to propagate, but electromagnetic ones can propagate in a vacuum. Both require different periods of time between their emission and reception[2] and those that are perceptible by the human senses of hearing and vision oscillate respectively in very different frequency[3] spectrums.[4]

Among others, the senses of hearing and vision coexist in humans. According to Van Campen (2010), newborns understand all their sensory impressions as a whole, therefore they do not differentiate between light, sound, taste or smell. According to this researcher, it is around four months of age that the senses begin to modularize due to neural and social development. It makes sense to say that, despite modularization, all senses are equally important for a human being at any time in their life. However, it is easily verifiable that a hierarchy of senses was established with the following decreasing order of primacy: vision, hearing, touch, taste and smell. Horn (2015),

[2] In air, sound travels at a speed of 344 m/s and light at a speed of 300,000,000 m/s.

[3] Sound waves perceptible by the human ear oscillate approximately between 20 Hz and 20 kHz with wavelengths between 17.15 and 0.0172 m, whereas electromagnetic waves perceptible by the human eye oscillate between 430 THZ and 750 THz with lengths wavelength between 740 and 380 nm.

[4] Despite differences in magnitude and speed, both mechanical sound waves and electromagnetic waves are represented by wavelength and frequency. I am intrigued by the fact that science represents the sound spectrum in the frequency domain and the visible spectrum in the wavelength domain – considering that frequency and wavelength are inversely proportional quantities.

from the Max Planck Institute for Psycholinguistics, analyzed the study "Vision verbs dominate in conversation across cultures, but the ranking of non-visual verbs varies", which suggests that the primacy of vision is due to universal constraints and to the cultural modeling of the perception language and concluded that the hierarchy of senses is shaped by both biological predispositions and cultural influences. Without wanting to discuss the role and importance that visual images have had throughout history in the construction of humanity—from rupestrian art to digital images—particularly in their religious, political and social usefulness, it is certain that currently, there is a growing proliferation of the use of visual images as the main mediator of access to information—an interface for interaction and dissemination of information that creates opinion and consumption.

Despite the growing privilege of the sense of vision—in a society with weak visual education—the sense of hearing is equally important for human experience and survival. Our spatial perception depends on all the senses, however, the focus of this text is the amalgamation of the sound and the visual. The sense of vision allows us to look in a certain direction at a certain moment, meaning when we look forward it is physically impossible to look back. The sense of hearing allows us to perceive the direction of origin of any sound stimulus and simultaneously hear sounds originating from all directions around us. Therefore, at the level of survival and human perceptual experience, no sense takes precedence over the others. The vision/hearing *paragone* is absurd.

3 *Paragone* 2—The Complementary and Parallel Itineraries of Sound and Visual in the Arts.

In the context of artistic creation in the twentieth and twenty-first centuries, most audiovisual productions (sound and visual) also force hierarchies that give primacy to the visual over the sound or vice versa. Cinema is not a case of simultaneous composition of the visual and the sound. Cinema in its origins—silent cinema—is a succession of exclusively visual stimuli. Sound stimuli are introduced at a later[5] historical moment and even today, it is consensual that the sound itinerary—concordant or dissonant—composed of diegetic and non diegetic sounds like music and sound effects is complementary and subsequent to the visual itinerary. Despite the different techniques and technologies used asynchronously in the production of a film, despite a film being an audiovisual object, there is a consensus that the visual image is the first and privileged medium of cinema.

The music video also does not result from the simultaneous composition of sound and visual. Unlike cinema, in the music video there is a previously composed sound

[5] From orchestras that played live during screenings, through sound recordings on ceramic discs such as the vitaphone that reproduced the sound itinerary in parallel with the film projection or the recording of the sound itinerary on the film itself, to current formats that store on the same physical support the digital representation of sound and the succession of visual images.

itinerary to which a visual itinerary is juxtaposed afterwards. Even though there may exist a precise synchronism between the sound and visual itineraries, a music video is always the simultaneous diffusion of these itineraries and the composition of the visual narrative is subsequent to the sound/musical composition. Another example would be *VJing's* visual composition, which although it can be composed in real time and in synchrony with the composition made by the DJ, although there may be exact synchronism between both sound and visual compositions, it is also not about a simultaneous composition between the sound and the visual. The sonorous, the musical has priority. *VJing* is always an accompaniment to a DJ's work.

4 Simultaneous Composition of Sound and Visual Image

Today, artists who have the skills to program computers can create artistic technological systems capable of simultaneously controlling in "real time" the calculation and synthesis of visible and audible stimuli. This creative exploration of algorithms allows, in my opinion, new relationships between visual and sound images. At the end of the last century, Ox (1999) referred to the use of the computer as a medium to enter the world of interaction between different senses, relating it to the phenomenon of synesthesia—information collected by one sense that causes an experience in another sense, through the transfer of that information. Ox considered that certain programmers develop computer programs that precisely cause interactions between sound and visuals.

Personally, I consider that "real-time" programming languages like Max—which do not involve writing the code, compiling it, testing it, debugging and repeating—as they allow programming while the program is running, enhance the creation of computational compositions in which the sound and the visual are thought, calculated and executed simultaneously. This possibility of simultaneously composing sound and the visual is fundamental for maximum space–time contraction and concentration between the creation moment and the moment in which what is created is experienced.

In this context, artistic creation includes an iterative and interactive loop between the mental image of the sound and visual composition—in the artist's body—which causes/originates the physical materialization/manifestation of the composition that triggers another mental image in the artist himself, resulting from the physical stimulation of his/her visual and auditive systems. The artist's decisions in such a composition process result from the reciprocal influence of these two mental images that are in constant and mutual affectation in a process—such as the process of manual clay modeling—which can be called reflective practice. It is a dialogue between two mental images, of the same thing, in the same person, which allow the artist to make simultaneously the sound and the visual composition. Making—which should be understood here as the materialization of thought (Carter, 2004; Vaughan, 2007)—implies matter and technique as means that are therefore determinants of the process of creation and production itself. This means that the artist's relationship with materials and technologies is in itself very inspiring. There is, as stated by Noë (2006),

a reciprocity between the human and the technological: the human creates, uses and modifies technologies that affect and modify the human who with them, and in them, thinks and acts. In my doctoral research I stated that the artist does what the material allows. In the context of this essay, I update this statement to the 'artist does what technology allows', adding that the relationship—thought and action—with synchronous programming languages, in real time, is an inexhaustible source of inspiration.

The maximum spatio-temporal concentration characteristic of audiovisual intermedial practice and thought, in the simultaneous composition of sound and visual that I mentioned above, contrasts and distinguishes itself from the spatio-temporal expansion that characterizes other categories of audiovisual compositions—also previously mentioned—such as cinema, music videos, visual data sonification, visualization of sound signals, VJing, etc. In these last compositions, which do not result from the simultaneous composition of the sound and visual itineraries, the conception of the sound occurs at a different time and space than the conception of the visual itinerary, and most of the time these conceptions are not thought of and made by the same individual.

Over the last 21 years I have developed several artistic experiences that used, as a strategy, synchronism between sound and visuals, calculated by original computational algorithmic systems—computer programs of my own creation. I developed computational systems that: in real time sonified data obtained from the analysis of synthetic visuals also generated in real time[6]; in real time, calculated visualizations based on data obtained from the analysis, also in real time, of digital audio signals[7]; simultaneously calculated, in real time, visual and sound synthesis processes with different levels of interactivity and physical computing.[8] Several other artists and authors also explored different ways of articulating the sound with the visual. Ribas (2011), in the section "Strategies and principles of sound-image articulation" of her PhD dissertation, uses Golan Levin's definition "audiovisual software" to refer to various software from different authors capable of computing computational articulations of sound and visuals. Citing other authors such as Hervé Zénouda, Lev Manovich or Mark Ries, Ribas adds the mutability, transmutability, fusion, equivalence or transformability of digitized media to her discussion on audiovisual software.

---

[6] For example *Empty*, 2002 (https://3kta.net/3empty.php) co-authored with Sérgio Soares Ferreira.

[7] For example *Tonis*, 2022 (https://3kta.net/3tonis.php); *TuCaRa*, 2018 (https://3kta.net/3tucara.php); *Cage*, 2016 (https://3kta.net/3cage.php); *EPV*, 2014 (https://3kta.net/3EPV.php); or *OAV*, 2011 (https://3kta.net/3oav.php).

[8] Among others, *Um-Par*, 2020-2023 (https://3kta.net/3umpar.php); *(Des)Acceleration*, 2022 (https://3kta.net/3desacel.php); *Rechteck Fyrkant*, 2021 (https://3kta.net/3rechteckfyrkant.php); *aMar*, 2021 (https://3kta.net/3amar.php); *aSail*, 2019 (https://3kta.net/3asail.php); *NoisyWaveBeat*, 2019 (https://3kta.net/3nwb.php); *Sail*, 2019 (https://3kta.net/3sail.php); *3.11.exp*, 2017 (https://3kta.net/3311exp.php); *Dirty Dirt*, 2016 (https://3kta.net/3dirtydirt.php); *Komo*, 2017 (https://3kta.net/3komo.php); *Kephoise*, 2012 (https://3kta.net/3kephoise.php); *SynDyn*, 2011 (https://3kta.net/3syndyn.php); *SoLu*, 2010 (https://3kta.net/3solu.php); *ColMus*, 2009 (https://3kta.net/3colmus.php); *Morrell*, 2008 (https://3kta.net/3morrell.php); *Graffonic*, 2007 (https://3kta.net/3graffonicrgb.php); or *Graffonic*, 2006 (https://3kta.net/3graffonicblue.php).

The binary representation of information—digitization—about sound and visual images allows this information to be modeled—synthesized and/or processed by running computer programs—in an integrated and homogeneous way because the binary representation of sounds and images is the same thing: binary data. The integration of digital representations of different media—sound and visual for example—allows computer programming languages to be an excellent syntax for implementing intermedia dynamics, fluidizing media and reinforcing their equity. According to Friedman (2007), in a world where information and digital computer code flow, media begins to become fluid, as digital control begins to blur the boundaries between distinct forms of input, transmission and output.

5 Perceptual Objects

In our daily experience, almost every audible phenomenon has an associated visual phenomenon and vice versa. Without physical obstacles, when we see heavy rain hitting a surface we also hear it. For Bizley et al. (2016) visual and auditory objects share certain properties: both have linked characteristics that change over time and group acoustic and visual perceptual characteristics. Most objects and events in the world outside our bodies generate simultaneous stimuli in different sensory modalities (Teramoto et al., 2012).

Recent studies on multisensory perception have revealed that different sensory modalities (hearing and vision in particular) are interrelated and interact when processed simultaneously by the brain. Hearing and seeing at the same time can be termed as visual-auditory integration. According to Bizlei et al. What we hear and see takes on surprisingly different physical forms and is necessarily encoded by different sensory receptor organs, yet we effortlessly unite auditory and visual features to create a coherent perception. O'Callaghan (2016) proposed that multisensory perceptual objects are mereologically complex individuals because they are characterized by their structures being made up of different parts that exist as a whole and are treated collectively as a single perceptual unit.

Based on this brief literature review, I propose that highly synchronized audiovisual works, such as *Rechteck Fyrkant*,[9] *Um-Par*,[10] *Sail*[11] or *Asail*,[12] among others, be understood as unique perceptual objects, which simultaneously affect the senses of hearing and vision in space and in time.

Sensitive space is that which is apprehended with the senses. Spatial perception is an exploratory activity mediated by the sensorimotor apparatus of human beings. From an enactive perspective (Noë, 2006), spatial perception is enaction resulting from the structural fusion of the organism with its environment. Particularly in works

[9] *Rechteck Fyrkant*, 2021 (https://3kta.net/3rechteckfyrkant.php).

[10] *Um-Par*, 2020-2023 (https://3kta.net/3umpar.php).

[11] *Sail*, 2019 (https://3kta.net/3sail.php).

[12] *Asail*, 2019 (https://3kta.net/3asail.php).

such as the four I just mentioned, an exploratory activity precisely is provoked in which spatial properties are accessible to more than one sense (vision and hearing).

Embodiment enables us to be spatially located beings. When we are awake, we always look in any direction and listen to the sounds that surround us. So in this state, space is always present—it is constant. The way we perceive the spatial context that surrounds us is a fundamental part of our thinking and behavior. The four works I mentioned, as spatial arrangements, aim to simultaneously involve the audience's senses of sight and hearing, affecting their choices and behaviors. Reflecting a desire to experience sensual pleasure, these works provide a setting that engages the audience by making them simultaneously the fundamental element and medium of the artwork. The concept of artistic object, explored in the development of these works, implants information in an environment, constituting a perceptible space modeled for the presence of the audience.

Sail—Synchronous as in life, is a unique perceptual object that simultaneously affects the senses of hearing and vision in time and space. This object consists of a horizontal row of three hundred addressable RGB LEDs, four equally spaced loudspeakers, also aligned horizontally, a computer, a micro-controller and a sound card. An original computer program triggers synchronized audiovisual events simultaneously in the speakers and LEDs. Audiovisual events occur in different locations on the line—audiovisual spatialization caused by loudspeakers and LEDs. For example, if a sound moves from left to right through the four speakers, a synchronized light event also moves from the left to the right side of the row of LEDs.

The perceptual object *Asail—Asynchronous as in life–*, uses the same hardware as the *Sail* but new software was designed to introduce a variable latency between the visible and the audible in order to induce the illusion of depth—since that audiovisual events are actually asynchronous due to the different speeds of light and sound. When someone is three hundred meters away from a fireworks show, the sound of the fireworks has a latency of about one second. For Bizlei et al., intermodal integration refers—in perception—to the ability of one sensory modality to influence the performance of another sensory modality. This is how latency introduced in *Asail* reinforces the perception of distance.

Rechteck Fyrkant and *Um-Par* are perceptual objects realised by two intermediate, computational, algorithmic, generative compositions that generate—through stochastic processes of sound and visual synthesis—rhythmic sequences and (in)determined micro-melodic tessitures that are experienced as synchronous audiovisual events located with spatiotemporal precision. Both are materialized with custom software, computer, microcontroller, LEDs, speakers and aluminum profile. In the visual domain, in *Rechteck Fyrkant*, retinal inertia is articulated with instinctive saccades to produce retinal afterimages that are naturally mixed by the audience's individual visual experience with retinal images.

In *Um-Par*, the original computational system controls the additive—simultaneous—synthesis of color and sound, triggering consonant and dissonant chromatic sensations that are perceived as masses of sound and light located in (a)cyclic and (a)symmetric movements.

In the four works that I just described, the visual does not follow the sound and the sound does not follow the visual, they not only concur, they occur in unity, causing the concurrence of two different sensory modalities.

6 In Short

The creation processes that imply analysis of one of the itineraries—sound or visual—to produce the other (Cinema, Music Video, VJing, among others), and the processes in which the strategy of transmuting one of the itineraries into the other is used (Sonification and Visualization), imply a process of interpretation and mediation between the sound composition and the visual composition. In a different way, in the process of artistic creation that results in the works I mentioned in this text, the sound and the visual are composed simultaneously—without interpretation, mediation or transmutation. For the first time, with digital computing systems it is possible to use the same syntax for representing sound and visual signals, which makes this form of audiovisual composition viable, nullifying the vision-audition *paragone*.

References

Bizley, J. K., Maddox, R. K., & Lee, A. K. C. (2016). Defining auditory-visual objects: behavioral tests and physiological mechanisms. *Trends in Neurosciences, 39*(2), 74–85. Retrieved December 18, 2023, from https://www.sciencedirect.com/science/article/pii/S0166223615002817.

Carter, P. (2004). *Material thinking: The theory and practice of creative research.* Melbourne University Press.

Friedman, K. (2007). Intermedia, Multimedia, Media. Retrieved December 18, 2023, from https://www.yumpu.com/en/document/view/10517119/intermedia-multimedia-media-ken-friedman-intermedia-mfa

Horn, C. (2015). Look first—new light shed on the hierarchy of the senses by a study at the Max Planck Institute for Psycholinguistics in Nijmegen. In *Research news. Max Planck Institute.* Retrieved December 18, 2023, from https://www.mpg.de/8849014/hierarchy-senses

Noë, A. (2006). *Action in Perception.* The MIT Press.

O'Callaghan, C. (2016). Objects for multisensory perception. *Philosophical Studies, 173*(5), 1269–1289.

Ox, J. (1999). Introduction: Color me synesthesia. *Leonardo, 32*(1), 7–8. https://doi.org/10.1162/002409499552939. https://direct.mit.edu/leon/article-abstract/32/1/7/43730/Introduction-Color-Me-Synesthesia?redirectedFrom=fulltext

Rajamanickam, M. (2007). *Modern general psychology* (2nd ed.). Concept Publishing Company.

Ribas, L. (2011). *The nature of sound-image relations in digital interactive systems.* Faculty of Fine Arts of the University of Porto.

Teramoto, W., Hidaka, S., Sugita, Y., Sakamoto, S., Gyoba, J., Iwaya, Y., & Suzuki, Y. (2012). Sounds can alter the perceived direction of a moving visual object. *Journal of Vision, 12*, 11. https://doi.org/10.1167/12.3.11. Retrieved December 18, 2023, from https://jov.arvojournals.org/article.aspx?articleid=2121046

Van Campen, C. (2010). *The hidden sense: Synesthesia in art and science.* MIT Press
Vaughan, L. (2007). Material thinking as place making. *Studies in Material Thinking, 1*(1), 1–2. AUT University.