

# Resonant Aesthetics: Attention and Synaesthesia in Digital Media Art



Pedro Alves da Veiga 

**Abstract** This article posits the creative exploration of the interconnections between image and sound, as well as the intricate interplay between aesthetics and poetics within the contemporary landscape of new digital artistic formats. To facilitate this exploration, it analyses the nuanced roles attributed to visual and sonic elements and their cognitive effect as attention grabbing mechanisms, in order to discern their mutual interferences and the consequential amplification of their artistic resonance. Moreover, it extends its purview beyond the visual and auditory domains, contemplating the ramifications for the broader spectrum of human senses. Through this analysis of aesthetic dimensions, including sensory experiences that transcend the Aristotelian realms of perception, it seeks to unravel the cognitive and communicational implications inherent in the synergistic employment of these media and the transformative potential encapsulated within the convergence of digital formats, unearthing the profound sensory, cognitive, and communicational impacts that resonate within the realms of art and human perception.

**Keywords** Multisensoriality · Synaesthesia · Media convergence · Digital art

## 1 Introduction

The topic of the intersections between image and sound has a history as rich and ancient as the act of storytelling through the simultaneous use of graphic elements (images, drawings, or paintings) and sounds (the storyteller's voice, actors, or sound effects). However, because the focus of this text is contemporary art, this will be its starting point.

Lipovetsky and Serroy (2015) argue for the existence of a global aestheticization phenomenon that encompasses all areas of our daily life, reducing aesthetic appreciation to a decorative effect, a projection of status and lifestyle. Aesthetic capitalism, or *capitalisme artiste*, is based on this aestheticization that permeates everything,

---

P. A. da Veiga (✉)

Centro de Investigação em Artes e Comunicação, Universidade Aberta, Lisboa, Portugal  
e-mail: [pedro.veiga@uab.pt](mailto:pedro.veiga@uab.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\\_18](https://doi.org/10.1007/978-3-031-67503-4_18)

303

from gourmet restaurants with their delicately plated courses, to the styling and decoration of our homes, the way we dress, the photographs we take, the choice of holiday destinations, among many more. This phenomenon of the *aestheticization of the mundane* also extends to other areas such as advertising, where the conveyed message content no longer remains objective, focused on products and services (as expected), but is deeply buried beneath aestheticized layers, and is often delivered through short video pieces, bordering on video art.

News photography is also frequently analysed from an aesthetic point of view, which seems to overshadow the semantic analysis of its content. This becomes more evident when we are confronted with reports made in situations of war, migrant crises, refugees, forest fires with a high number of victims and material damage, or even during recent COVID-19 crises, involving all the situations that ensued, from deserted streets in cities in 2020, to mass graves. Many of these photographic records end up being collected and subsequently displayed in exhibitions where they are analysed—and awarded—from an aesthetic perspective, relegating the analysis of their content and context to a secondary level.

In the field of design, which combines form and function, the reflection of the aestheticization of the world is the overlap of formal aspects with functional aspects, relegating many design objects to a decorative role rather than a functional one.

In art itself, this phenomenon is also evident, as artistic objects migrate from their contexts of creation and exhibition to be incorporated into everyday use items. This is particularly noticeable in illustrations or paintings from existing museums, which are transposed into merchandise items such as t-shirts, coffee mugs, or notepads.

This phenomenon also has repercussions on the approach used by museums and galleries to promote their spaces and collections: if the previous prevailing paradigm was the prohibition of image capture, today the audience is encouraged to engage in this activity through communication and promotion actions in which events appear as *Instagrammable*. As we will see later, this paradigm shift also has significant impacts on how the public interacts with the artworks and their appreciation.

This perspective also—and necessarily—finds different echoes within academia, addressing the aestheticization of areas such as violence, war, politics or the inaeesthetic, approaches that were unthinkable two centuries ago when aesthetics focused on the search for the sublime.

An article published in *The Conversation* (2020) poses the question: “If you use social media, the chances are you see (and forward) some of the more than 3.2 billion images and 720,000 h of video shared daily. When faced with such a glut of content, how can we know what’s real and what’s not?” Through various text-to-image AI-enabled applications, such as DALL-E<sup>1</sup> and Mid Journey,<sup>2</sup> vast quantities of new images are generated—so many that requests for generation are queued, waiting for the computing capacity to respond—and are then published as if they were true.

These issues are also very relevant in the realm of music, as we are going through an era of mass-produced artistic production, not only in visual art but in all genres, thanks

---

<sup>1</sup> <https://openai.com/dall-e-2/>.

<sup>2</sup> <https://midjourney.com/>.

to the democratization of access to equipment, instruments, materials and media. In the field of music, there is a shift from old collectible items (vinyl records, cassettes, and CDs) to an increasing number of streaming services. Physical ownership of the artwork has been relegated to a secondary position, not only due to the (financial) impossibility of acquiring every music album that is released but also because there has been a decline in the enjoyment time, whereby each song is quickly replaced by a new, trendier one. This has also influenced musical composition itself, as it becomes crucial to capture the audience's attention in the first few seconds of a musical piece, or risk them switching to another song without a second thought. Just like with images, music also grapples with the issues of what's real and fake (Golsen, 2023), with Spotify being accused of encouraging the creation of fake artists (on demand) to fill playlists at a low cost. While it is mathematically easy to determine that the combinations of the seven musical notes (Freke, 2014), even considering multiple octaves, sharps, or flats, can only be combined in a finite number of distinct combinations corresponding to melodies (i.e., non-chaotic), the reality is that most of the current musical production does not seek a particular variety in composition (Serrà et al., 2012).

With an average viewing time of an artwork in a museum hovering around 27 s, there is also a movement originating from within the museum institutions that seeks to encourage lengthier aesthetic appreciation (Cascone, 2019). However, success is the enemy of that movement, as in exhibitions that attract thousands of visitors, the time spent is short, and the public itself contradicts that attempt. The popular capture of *arties* (selfies taken alongside an artwork), in which the person capturing the image turns their back to the artwork, shows that the most important thing is to capture and share the proof of being present at the location or event, associating it with some form of social or cultural prestige, overshadowing the unmediated enjoyment of the artworks.

All these aspects have considerable impacts, not only on how art is perceived but also on how it is conceived, created, and exhibited. Therefore, it is within the intersectional characteristics that can be incorporated into audio-visual digital artefacts, and the aesthetic and poetic impact they can have on the audience, that the following sections are developed.

## 2 Attention and Visual Stimulation

According to Crawford (2014), attention can be categorized based on its orientation either towards objectives, where it serves the individual's own will, or towards stimuli, which occur independently of the individual's own will. An artist immersed in the creative process dedicates their executive attention to their artwork; it is their objective to be engaged in the creative process. In contrast, if they glimpse a sudden flash in their peripheral vision, it can be asserted that attention has been stimulated—and probably diverted. Thus, the individual may act accordingly, turning their head to understand what happened, or choose to ignore it—but the claim on attention

was involuntary. Whatever the response, it required a concerted effort of executive attention.

When we use technological devices as mediators of aesthetic appreciation, whether because we want to take a photograph or because we use an augmented reality application, for instance, we become subject to having our attention fragmented due to the number of stimuli we receive through smartphones, tablets, or personal computers. A fluid, uninterrupted aesthetic appreciation results in a deeper and more intense engagement with an artefact, as opposed to engagement subject to constant interruptions from external stimuli. Therefore, it is natural for an artist to wish to introduce mechanisms into their artefact to ensure that the capture of the audience's executive attention surpasses the external stimuli to which they are exposed. It can then be asserted that the "more fluently perceivers can process an object, the more positive their aesthetic response" (Reber et al., 2004, p. 365).

Every estheticized or artistic artefact is inherently endowed with different amounts of information. For example, a black and white icon possesses very low information content, allowing for immediate apprehension. If we are faced with a photograph or an illustration, then the amount of information is greater, although our appreciation still falls within the 27 s time span mentioned earlier, even if our gaze traverses certain features of the photograph repeatedly, searching for some more noteworthy detail. However, in the case of moving images, whether the result of animation, video, or generative programming, the medium itself requires a more extended appreciation since it unfolds over time. A highly significant aspect of moving images, proposed by Christian Metz in 1965, is the indirect contribution that movement makes to the perception of reality, endowing objects with dimension. Moreover, movement is directly perceived as real.

It is movement (one of the greatest differences, doubtless the greatest, between still photography and the movies) that produces the strong impression of reality. This, of course, has often been pointed out, but the observation has perhaps never been pushed far enough. "The combination of the reality of motion and the appearance of forms gives us the feeling of concrete life and the perception of objective reality. Forms lend their objective structure to movement and movement gives body to the forms", observes Edgar Morin in *Le Cinéma ou L'homme imaginaire* (Metz, 1990, p. 7).

Thus, the aesthetic appreciation could potentially take place in a considerably more protracted manner, although here too, the current condition of excess supply and formulaic repetition is evident. As early as 2008, Tom Sherman asserted:

Displayed recordings will continue to diminish in duration, as television time, compressed by the demands of advertising, has socially engineered shorter and shorter attention spans. Videophone transmissions, initially limited by bandwidth, will radically shorten video clips. The use of canned music will prevail. Look at advertising. Short, efficient messages, post conceptual campaigns, are sold on the back of hit music. Recombinant work will be more and more common. Sampling and the repeat structures of pop music will be emulated in the repetitive 'deconstruction' of popular culture (Sherman, 2008, p. 161).

Just as in music, the audience must establish a connection within the first few seconds, under the risk of moving on without pausing for the more extended appreciation that the medium demands.

### 3 Attention and Auditory Stimulation

Auditory stimulation triggers a cognitive process capable of relating form and motion according to its duration and content. When we listen to a dog's bark, we identify the source of the sound, its producer, but we can also spatialize it through stereophonic analysis, identifying a position that may remain static or in motion (not visual but auditory). Thus, sound also contributes to the effect of making what is perceived and recognized seem real, even if it comes from a synthesized source.

However, auditory stimulation in the form of a melodic phrase, musical rhythm, or another identifiable sequence can exert a remarkable influence on attention, as demonstrated by the introduction of the ability to attach a musical theme to a visual post on all social networks, with the intention of better capturing executive attention.

Although Serrà et al.'s analysis (2012) is not overly optimistic about the evolution of popular music, it does provide some clues as to how the *déjà entendu* sensation can be overcome, particularly by delving into a hybrid field that mixes soundscapes with music. Soundscapes, named for their ability to evoke complex scenes through the overlapping use of various sound elements from different sources, introduced a form of virtual reality well before digital technologies. The aesthetic appreciation of soundscapes lead to the construction of a dematerialized, non-physical universe where various sources constituting the sound landscape were identified and superimposed. For Schafer (1993), the soundscape, or sonic environment, was composed of sound elements incorporated into musical or atmospheric compositions, derived from natural, synthesized, or processed sources. Later on, Rudi (2011, p. 187) proposed the designations of (1) keynote sound—sounds that identify places (such as a beach, field, or city), (2) soundmark—sounds specific to an area or location (like the sound of a foghorn), and (3) sound signal—to refer to sounds intended to capture an individual's attention (like a dog's bark or a smartphone notification).

In what Rudi calls reflective listening (p. 192), the listener converts the flow of sound data into objects of aesthetic appreciation through the reinterpretation of sonic information and the assignment of corresponding value. An example provided by the author is the singing of birds, where aesthetic appreciation can go beyond the interpretation of intra- and interspecies signalling proportions and positions to the appreciation of the melody as music. This adds the perception of melodic and rhythmic patterns to what is heard.

Schafer (1993, p. 152) also distinguished two types of sound from the standpoint of Gestalt psychology: figure and ground. Depending on the listening environment, the interpretation of these types of sounds can vary, including their roles: the figure becomes the ground and vice versa. Contextual listening occurs when the listener's location is considered: this location adds information to the experience. Listening to the sound of a car skidding from the comfort of a sofa at home does not determine the same experience as listening to that sound while crossing a street. When a recorded sound migrates to a new context and is remixed into a new soundscape, selection and representation issues become particularly relevant as they will determine different interpretation factors for the audience. In this context, sounds belonging to

the sound signal category become particularly relevant, because they can capture the audience's attention (through stimuli), while the other sounds (keynote and soundmark) are responsible for focusing attention on objectives (narrative or others, including conflict, intrigue, surprise, or novelty, as previously mentioned). Thus, the soundscape allows the artist to create attention-capturing stimuli, environmental and narrative contexts.

## 4 Intersections of Image and Sound

A hundred years after cinema's birth, cinematic ways of seeing the world, of structuring time, of narrating a story, of linking one experience to the next, are being extended to become the basic ways in which computer users access and interact with all cultural data (Manovich, 2001, p. 87).

When we think about conflict, intrigue, surprise, or novelty, most of the time we associate some of these concepts with something negative, and we might even consider them as serving lesser creativity. However, any of them may not be taken literally but rather as a disruption of current norms: aesthetic conflict, intrigue with the established way of life, semantic surprise, or conceptual novelty. Brandenburger proposes a term for this disruption: defamiliarization.

Working in the early twentieth century, a Russian literary theorist called Viktor Shklovsky pointed out how Tolstoy achieved heightened effect in his writing via techniques such as describing objects from a distorted perspective and refusing to use the customary names for objects, and by generally "making strange" (de-familiarizing) the otherwise familiar. Later, the great French director Jean-Luc Godard revolutionized cinema with his use of jump cuts in *Breathless*. Taken for granted today, this innovation must have seemed baffling to many people at the time. Up until then, great efforts had gone into creating a smooth continuous flow ("continuity") on the screen. After all, a continuous flow is how we experience vision, thanks to the workings of our brains. This is the familiar. But Godard decided to break up this flow to force us to step away from our usual assumptions and see his characters as, literally, jumpy and disconnected. (...) Godard lifted the technique of de-familiarization from the page to the screen (Brandenburger, 2019, n.a.).

One interesting initial effect resulting from the simple overlay of an auditory landscape, thus involving temporal development, onto a static visual element is that the aesthetic appreciation of the whole becomes governed by the duration of the sound element. This sound can contribute to influencing the interpretation of the visual. This exercise was taken to a controversial extreme in the film *Branca de Neve*<sup>3</sup> by Portuguese director João César Monteiro, where most of its duration consists of a black screen punctuated by dialogues. Marguerite Duras's *L'Homme Atlantique*<sup>4</sup> also uses a similar artifice, as does Derek Jarman's *Blue*,<sup>5</sup> in this case with a blue

---

<sup>3</sup> <https://www.imdb.com/title/tt0267155/>.

<sup>4</sup> <https://www.imdb.com/title/tt0258655/>.

<sup>5</sup> <https://www.imdb.com/title/tt0106438/>.

screen echoing his own loss of vision. Ritter's installation *Intersection*<sup>6</sup> (1993) took this exercise into the realm of interaction, by placing a battery of eight sound columns in a completely dark room, creating a three-dimensional sound traffic game. Each audience member entering the room felt like they were in the midst of intense traffic that they couldn't see, and certain positions triggered the sound of a car skidding toward them. If they didn't change position, collision sounds ensued, all directed at them. With more people in the room, the overall sensation was one of total disaster, as the auditory stimulus triggered all other perceptions.

In summary, despite living in an image-obsessed society, auditory stimuli can contribute to overcoming the absence of image, as it can evoke or suggest the same (virtual) image, maintaining a continuous flow of cognitive and aesthetic appreciation.

This is also why many of the major social networks encourage the inclusion of music in static image publications as a means to (1) increase the amount of time spent on the application and (2) increase audience engagement, capitalizing on the complementary communicational and emotional role that music can play in shared content (MacDonald, 2021).

One might be led to think that the opposite is also true, but when placing a sequence of images over a sound signal, for instance, the overall aesthetic appreciation time is not influenced, and the sound signal is easily disregarded in favour of the visual stimulus, without the previously described contamination effect.

The fact that sound can contribute to enhance visual perception has also been the focus of a study by Vroomen and Gelder (2000). The researchers concluded that giving the individual the sensation that the perception of simultaneous (visual) motion, and a sound figure or sound signal corresponding to the same real object, facilitated its identification. Synchronicity was the determining factor in that study since the synchronization of stimuli allowed for the rapid identification of the source. They conclude the findings (p. 1590) by noting that several animals have multisensory cells that respond to stimuli of various modalities but are also capable of integrating information from these different modalities by increasing the number of impulses in a multiplicative ratio when affected by multimodal/multi-channel inputs. Thus, a prominent sound in the auditory modality can trigger an equal prominence in the visual modality.

Sutherland et al. (2014) also documented this type of synesthetic multisensory experience occurring between auditory and visual stimuli, particularly analysing what they referred to as *looming sounds*, which indicate a rapid approach:

(...) we found that static visual objects paired with looming tones (but not static or receding tones) were perceived as larger and brighter than their actual physical properties, as if they appear closer to the observer. (...) We conclude that looming sounds impact visual perception through a mechanism transferring in-depth sound motion information onto the relevant in-depth visual dimensions (such as size and luminance but not shape) in a cross-modal remapping of information for a genuine, evolutionary advantage in stimulus detection (p. 312).

---

<sup>6</sup> [https://youtu.be/--5pcJB9k-k?si=-Ohi3UWUMnek\\_rqW](https://youtu.be/--5pcJB9k-k?si=-Ohi3UWUMnek_rqW).

## 5 Digital Media and Resonant Aesthetics

Current digital and wireless technologies appear to lead us towards a paradigm shift in how we perceive, in other words, in how we use the human senses to understand our surroundings. The sensory transaction of the twenty-first century is compact, immediate, mobile, and personal, materialized through laptops, tablets, and smartphones, where all types of media converge into a single binary-encoded digital format. According to our mood, we pause or fast-forward through audio and video tracks, upload or download various media, or remix images and sounds, producing and publishing our version of how reality should be. The amalgamation of sonic and graphic elements is fostered by most social media applications, allowing users to overlay their photos and videos with text and music, and creating the ephemeral pieces commonly known as “shorts” or “stories”.

However, the convergence rate of different media is not uniform within a specific culture, as those who are wealthier and possess greater technological literacy will stand out in leadership, while other segments of the population will follow in their wake—or devise divergent formats altogether. And yet, being able to use black-boxes to produce the above mentioned mainstream pieces does not imply a profound knowledge of the creative act. In fact, the ease of use tends to lessen it, to strip it of deeper meaning and consequence, at least for the majority of users.

Thus, this media convergence in the digital realm, facilitated and fuelled by social networks, has the effect, in more developed countries, of fostering the belief that one is contributing in a participatory and democratic manner (although often in service to brand influences and commercial products) to the creation of popular culture.

In reality, it portrays a very different world from the actual one, attempting to make it appear as the real world. This cultural positioning, based on *culture-jamming*, which interrupts the media flow from an external position, can be countered by artists who organically, consciously, and actively shape that flow.

The organic connection among different media, which has evolved into its current digital convergence, was possibly identified in its broadest sense at the Bauhaus, by Moholy-Nagy (1996), who theorized about its intricate and metaphorically complex organic nature:

The contemporary painting exhibits a multiplicity of colour and surface interrelationships, which gain their effect, on the one hand, from their conscious and logical statement of problems, and on the other, from the unanalysable intangibles of creative intuition. In the same way, the Theatre of Totality with its multifarious complexities of light, space, plane, form, motion, sound, man—and with all the possibilities for varying and combining these elements—must be an ORGANISM (Moholy-Nagy, 1996, p. 60).

These variations, recombinations, and intersections gain additional interest when we extend them to the entire set of human senses, beyond the limits of the Aristotelian five senses. Currently, the study of the human sensorium is significantly more advanced than in Aristotle’s time, encompassing senses such as nociception, chronoception, proprioception, thermoception, among others. Even the Aristotelian senses have undergone specialization: for example, the sense of taste can distinguish

between *sweet*, *salty*, *bitter*, *sour*, and the more recent *umami*, associated with glutamate. Each of these tastes can have varying intensity in each individual, and all taste sensations arise from the combination of these five factors (Yamaguchi & Ninomiya, 1998).

Thus, by exploring the entirety of the human sensorium through the organicity of its combinations, we also arrive at the concepts of synaesthesia and *multisensoriality*. Let us then define synaesthesia as the way external stimuli perceived by one sense trigger vivid sensations in another sense. The degree of vividness varies from individual to individual and is consciously uncontrollable. In other words, a synesthete will always be a synesthete and cannot avoid this sensory overlap.

Synaesthesia is a perceptual phenomenon that occurs when stimulation of one sensory pathway leads to experiences or perceptions in another, unrelated sensory pathway. For instance, upon seeing the colour green or the word “green,” an individual might experience a literal taste of lemon. However, this phenomenon might also manifest as a sensation of volume or various other associative perceptions. The occurrence of synaesthesia, much like any perceptual phenomenon within the human mind, is contingent upon the mental state or cerebral condition of the perceiving individual (Leote, 2014, p. 56).

Some forms of synaesthesia may be more discreet and common than others. Considering the human sensory system, as a whole, we are likely to recall occasions when our aesthetic appreciation—perceived through the senses—was influenced by the sense of temperature: if we are in an excessively cold or hot environment, the overall sensation will almost always be one of displeasure and discomfort. In contrast, in a normal-temperature environment, the overall sensation becomes more dependent on the remaining senses and perception. In this case, it is legitimate to assert that our sense of temperature has overlapped—or even overtaken—the other senses in aesthetic appreciation, and we were involved in a synesthetic experience.

When a stimulus perceived by one sense triggers (cognitive and aesthetic) interpretations associated with another sense, we encounter a phenomenon of representative synaesthesia. For example, upon hearing sounds from a field with birds singing and a breeze, we construct a mental (synesthetic) image of it. Representative synaesthesia can take two distinct forms: (1) a representation of aesthetic synthesis triggered by a multisensory artefact that activates various senses—e.g., creating mental images through the combination of auditory and olfactory stimuli; and (2) an aesthetic translation triggered by a monosensory artefact, which develops with the intention of also activating another sense—in line with the first example, the field with birds.

Thus, resuming the postulate in the preceding sections, we can establish a direct connection between the amount of information affecting various sensory processes (particularly vision and hearing, which are more detailed in this text but also touch, smell, proprioception, thermoception, chronoception, nociception, and all the other senses), and their relationship with language, attention, memory, and thought, resulting in a complex aesthetic organism, as proposed by Moholy-Nagy.

One of the attributes of the audio-visual medium is how it “align[s] the rhythmic nature of our perception with ‘the thread of the world’ and places us in a time window that offers motor action without fatigue, danger without damage, and mood alignment without self-judgment” (Antunes, 2012, p. 524). Understanding media convergence

from the perspective of media ecology reinforces the idea that audio-visual stimuli can lead to an aesthetic experience that goes beyond the senses of sight and hearing to include, among others, temperature and movement, without the risks associated with direct contact with the sensory energy of these modalities—but with an equally remarkable intensity. The idea that an audio-visual experience can give rise to a multisensory perception experience is thus proposed by Antunes (2018).

In a distinct yet convergent view, McLuhan wrote to Ezra Pound, about “The Cantos”:

Your Cantos, I now judge, to be the first and only serious use of the great technical possibilities of the cinematograph. Am I right in thinking of them as a montage of personae and sculptured images? Flash-backs providing perceptions of simultaneities? (McLuhan, 1988, p. 193).

His concept of those “perceptions of simultaneities” was illustrated as a mosaic, a methodology and technique whose efficacy is the empowering of the audience to discover in the gaps between materials a kind of generative energy and insight: a world of intervals.

The mosaic is a world of intervals in which maximal energy is transferred across the gaps. This is the “massage” effect. The Gutenberg Galaxy is a world in which energy is generated in the intervals, not by the connections. And the massage – the shaping, the twisting, the bending of the whole human environment by the technology—the reconditioning of the entire human environment by this technology—is a violent process, like all new technologies often revolting, as well as revolutionary. That is why Joyce calls them “thunders (McLuhan, 1997, p. 73).

McLuhan’s mosaic shares some conceptual similarities with the photographic collage works of David Hockney. For Hockney, his approach allowed him to overcome deficiencies of traditional photography, which limits the observer to a single frozen perspective.

He remarked that “photography is all right if you don’t mind looking at the world from the point of view of a paralysed cyclops—for a split second”. His photo collages expand the observer’s perspective because he allows us to add that very special ingredient: time. This we do by arranging the individual photographs successively into a meaningful sequence, with each providing a context for another while being itself contextualized by other photographs in the collage (Vrobel, 2011, p. 2).

As those photos used in the collages were taken at consecutive moments in time, one can argue that Hockney’s collages are, in fact, a planar and static rendering of a video, where the audience recreates the motion—by filling in the intervals of the mosaic—by looking at each photo. And since no order is pre-established, the audience has the ability to render a multitude of different videos from the same photos, only by changing the order in which they are perceived.

Ezra Pound provides the unifying field for McLuhan’s and Hockney’s stances, by stating that an image presents an intellectual and emotional complex at a given moment in time. For him, the image is not an idea but a radiant node, which he calls a vortex, “from which, and through which, and into which, ideas are constantly rushing” (Pound, 1914, p. 51).

We can, therefore, propose the concept of *resonant aesthetics*, in which the combination of various sensory stimuli interferes, like a radiant vortex, with other senses in their perception and cognitive appreciation. All these factors contribute to determining an individual's involvement in aesthetic appreciation, shaping how their attention is directed towards objectives, providing deeper engagement, leading to more detailed analysis of the artefact, and creating a more consistent relationship between the artefact and its audience and, consequently, a greater impact.

## 6 Conclusion

Sonic elements can play a pivotal role in the creative process behind digital artworks, facilitating the creation of context, the identification of sources and objects, setting a mood by means of a carefully crafted soundscape or the capture of attention. When these sonic elements are dutifully combined with visual components, they establish avenues for engagement, stimulation, and interpretation within artistic creations. Cinema and television productions have long explored these combinations, but modern-day technologies have made the production processes available to everyone with a smart device through countless free applications. As far back as 2001, Manovich posited the importance of the computer—and, therefore, digital tools—in these processes:

In this way, the computer fulfils the promise of cinema as a visual Esperanto which preoccupied many film artists and critics in the 1920s, from Griffith to Vertov. Indeed, millions of computer users communicate with each other through the same computer interface. And, in contrast to cinema where most of its “users” were able to “understand” cinematic language but not “speak” it (i.e., make films), all computer users can “speak” the language of the interface (Manovich, 2001, p. 87).

This synergy between visual and sonic elements is thus particularly relevant in the digital era, where media convergence, facilitated by—but not exclusive to—advancements in digital technologies, can heighten the intersections between aesthetics and poetics.

But rather than engaging in the careless use of black-boxed applications to those effects, this article advocates for the mindful combination of visual and sonic elements, by understanding the roles each piece may represent in the final creation. This careful construction can not only amplify the richness of artworks, but also enhance the audience's experience and perception. By using these components, artworks become multi-layered, capturing the audience's attention, prompting deeper engagement and interpretation. Moreover, the remixing and recombination of mosaics of visual elements within digital artworks fosters intricate semantic analyses. These recombined visual elements, which might encompass abstract and textual forms, or figurative representations, evoke varied interpretations, compelling the audience to scrutinize and engage more deeply with the artworks.

A careful balance must be attained, by delivering artworks that present the audience with mosaics and intervals—therefore allowing them to *fill-in the blanks*.

Furthermore, the potential for artworks to simultaneously embody both a poetic and aesthetic dimension—where the artistic expression aligns with ideological, political, or social narratives—offers digital artists powerful tools for creating impactful and attention-grabbing artistic and communicational experiences. This duality enables artists to convey multifaceted messages, triggering emotional, intellectual, and multi-sensory responses in the audience. Exploring the full range of the human sensorium and how intervals can be perceived, may contribute to deliver more complex and nuanced messages, fostering contemplation, discourse, and a deeper appreciation for the convergence of art and its societal contexts.

Ultimately, this convergence of multimedia elements in digital art exemplifies the evolving nature of artistic expression, leveraging technology and creative ingenuity to craft sensorially immersive and thought-provoking experiences. The interplay between these elements elevates the potential for art to not only captivate the senses but also act as a catalyst for dialogue, introspection, and societal reflection.

## References

- Antunes, L. R. (2012). The vestibular in film: Orientation and balance in Gus Van Sant's 'cinema of walking.' *Essays in Philosophy*, 13(2), 522–549.
- Antunes, L. R. (2018). Slow TV: The experiential and multisensory documentary. In *Cognitive theory and documentary film* (pp. 205–221). Palgrave Macmillan, Cham.
- Brandenburger, A. (2019). To change the way you think, change the way you see. *Harvard Business Review*. Retrieved from <https://hbr.org/2019/04/to-change-the-way-you-think-change-the-way-you-see>
- Cascone, S. (2019). The average person spends 27 seconds looking at a work of art. now, 166 museums are joining forces to ask you to slow down. *Artenet News*. Retrieved from <https://news.artnet.com/art-world/slow-art-day-2019-1508566>
- Crawford, M. (2014). *The world beyond your head*. Farrar.
- Freke, O. (2014). How many melodies are there? *Plus—bringing mathematics to life*. Retrieved from <https://plus.maths.org/content/how-many-melodies-are-there>
- Golsen, T. (2023). What is the deal with all the fake artists on Spotify? *Far out magazine*. Retrieved from <https://faroutmagazine.co.uk/what-is-the-deal-with-all-the-fake-artists-on-spotify/>
- Leote, R. (2014). Multisensorialidade e sinestesia: Poéticas possíveis? *ARS (São Paulo)*, 12, 42–61.
- Lipovetsky, G., & Serroy, J. (2015). *A estetização do mundo: viver na era do capitalismo artista*. Editora Companhia das Letras.
- MacDonald, R. (2021). The social functions of music: Communication, wellbeing, art, ritual, identity and social networks (C-WARIS). In A. Creech, D. A. Hodges, & S. Hallam (Eds.), *Routledge international handbook of music psychology in education and the community* (pp. 5–20). Routledge/Taylor & Francis Group. <https://doi.org/10.4324/9780429295362-3>
- Manovich, L. (2001). *The language of new media*. MIT Press.
- McLuhan, M. (1988). In M. Molinaro, C. McLuhan, W. Toye (Eds.), *Letters of Marshall McLuhan*. Oxford University Press.
- McLuhan, M. (1997). The hot and cool interview media. In M. A. Moos (Ed.), *Research: technology, art, communication*. Amsterdam: G+B Arts.
- Metz, C. (1990). *Film language*. The University of Chicago Press.
- Moholy-Nagy, L. (1996). Theater, circus, variety. In W. Gropius, & A. S. Wensinger (Eds.), *The theater of the Bauhaus* (pp. 49–70). The John Hopkins University Press.

- Pound, E. (1914). Vorticism. *The Fortnightly Review*, 96, 461–471. Retrieved from <https://fortnightlyreview.co.uk/vorticism/>
- Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review*, 8(4), 364–382.
- Ritter, D. (1993). Intersection [Obra de arte]. Retrieved from <http://aesthetic-machinery.com/intersection.html>
- Rudi, J. (2011). Soundscape and listening. In J. Rudi (Ed.), *Soundscape in the arts* (pp. 185–194). NOTAM.
- Serrà, J., Corral, A., Boguñá, M., Haro, M., Arcos, J. (2012). Measuring the evolution of contemporary western popular music. *Scientific Reports*, 2. <https://doi.org/10.1038/srep00521>
- Schafer, R. M. (1993). *The soundscape: Our sonic environment and the tuning of the world*. Simon and Schuster.
- Sherman, T. (2008). Vernacular video. Video Vortex Reader. In *Responses to YouTube* (pp. 161–172). Institute of Network Cultures.
- Sutherland, C. A. M., Thut, G., & Romei, V. (2014). Hearing brighter: Changing in-depth visual perception through looming sounds. *Cognition*, 132(3), 312–323. <https://doi.org/10.1016/J.COGNITION.2014.04.011>
- The Conversation. (2020). 3.2 billion images and 720,000 hours of video are shared online daily. Can you sort real from fake? *The Conversation*. Retrieved from <https://theconversation.com/3-2-billion-images-and-720-000-hours-of-video-are-shared-online-daily-can-you-sort-real-from-fake-148630>
- Vrobel, S. (2011). A second eye and wings for the cyclops: A fractal temporal analogy to Hockney's Collages. *Acta Systemica*, 11(1). The International Institute for Advanced Studies in Systems Research and Cybernetics. ISSN 1813-4769
- Vroomen, J., & Gelder, B. D. (2000). Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception and Performance*, 26(5), 1583.
- Yamaguchi, S., & Ninomiya, K. (1998). What is umami? *Food Reviews International*, 14(2–3), 123–138.