From "Gaze" to "Interaction": The Paradigm Shift in Aesthetic Appreciation of Immersive Digital Art

Noah Bennett [1]

[1] College of Software, Zhejiang University; Ningbo, China

Abstract

Since the twentieth century, the "gaze" has functioned as a core category of aesthetic appreciation, predicated on the unidirectionality of vision, the stillness of the spectator, and the fixity of time and space — a framework central to analytical traditions from Laura Mulvey's gaze theory to Guy Debord's critique of the spectacle. The emergence of digital immersive art — encompassing virtual reality (VR), augmented reality (AR), embodied interactive installations, and holographic immersive spaces — is now systematically dismantling this contemplative gaze paradigm, propelling aesthetic appreciation toward an "interaction" paradigm characterized by embodied interactivity, dynamic co-construction, and perceptual multimodality. Drawing on Oliver Grau's history of immersive art, Maurice Merleau-Ponty's embodied phenomenology, Claire Bishop's criticism of participatory art, and Helmut Leder's model of aesthetic information processing, this paper systematically compares the fundamental differences between the "gaze paradigm" and the "interaction paradigm" across four dimensions — perceptual structure, bodily role, spatiotemporal experience, and meaning production — and proposes a Three-Layer Interaction Model of Immersive Appreciation (TLIMIA) that distinguishes three strata of immersive digital art appreciation: a perceptual immersion layer, an embodied interaction layer, and a social co-construction layer. The paper further argues that this paradigm shift is neither a simple negation of the gaze tradition nor an inevitable elevation of aesthetic depth, but rather simultaneously opens multiple entry points for appreciation while generating new tensions around critical reflection, deep reading, and slow contemplation. The findings carry direct reference value for the design of immersive arts education, museum curatorial practice, and the methodological construction of digital art criticism.

Keywords: *gaze , interaction , immersive digital art , embodied appreciation , paradigm shift , TLIMIA model*

1 Introduction: The Shattered Glass Window

In On Painting, written in 1435, Alberti compares the picture plane to a transparent glass window, through which the viewer gazes upon a rationalized visible world. This metaphor precisely describes the fundamental posture of Western classical art appreciation: the viewer, as a visual subject occupying a fixed position, receives artistic information through a one-way visual channel from a safe distance and in a passive manner. From the Renaissance studio to the modernist white-cube gallery, this mode of appreciation centered on contemplative gazing continued for more than five hundred years, and later received systematic ideological analysis in the critical theories of Mulvey (1975), Debord (1967), and others. Meanwhile, the development of digital technology has provided a new media foundation for narrative immersion and interaction [1].

However, when a visitor enters teamLab Borderless, her body immediately becomes the trigger for the work's visual responses: her movement path determines the direction in which digital flowers bloom, her encounters with others generate new visual events, and she is unable to obtain a panoramic gaze of the work from any fixed position. The Alberti-like glass window has been completely shattered. The appreciator no longer stands outside the window, but is instead situated within a fluid, responsive, and embodied visual universe.

This paper aims to systematically analyze the paradigm shift in art appreciation from "gaze" to "interaction," including its historical roots, theoretical dimensions, structural characteristics, and two-way value assessment. The central argument of this paper is that immersive digital art appreciation is not a simple negation of the tradition of the gaze. Rather, through technological mediation, it reactivates the embodied dimension that has long been suppressed by visual centrism in Western theories of art appreciation, while also generating a new structure of appreciation centered on perceptual immersion, embodied interaction, and social co-construction. This new structure requires corresponding critical language and educational frameworks in response.

2 The Theoretical Genealogy and Internal Limitations of the Gaze Paradigm

2.1 The Threefold Origins of Gaze Theory

As a concept in the theory of art appreciation, the "gaze" has three intertwined historical origins. The first is the tradition of visual science: since the Renaissance, the system of linear perspective has constructed the human eye as a rationalized, point-like, and disembodied visual receiver, reducing the act of seeing to a one-way transmission of optical information. The second is psychoanalysis and the politics of the gaze: in Visual Pleasure and Narrative Cinema, Mulvey (1975) reveals that the narrative structure of mainstream cinema encodes gendered power relations within visual pleasure by positioning women as objects "to be looked at" and men as the subjects of the gaze [2]. The third is the critique of spectacle: Debord (1967) understands modern capitalist culture as a "society of the spectacle," arguing that passive gazing is a perceptual form of capitalist alienation, in which genuine social participation is replaced by a spectacularized one-way visual regime [3].

2.2 The Structural Presuppositions and Limitations of the Gaze Paradigm

Although these three theoretical strands differ in their critical orientations, the gaze model they share contains several structural presuppositions. When confronted with immersive digital art, these presuppositions are fundamentally challenged at the technological level.

First, there is visual one-directionality. The gaze is one-way: the viewer looks, the artwork is looked at, and information flows from the work to the viewer. The presence of the viewer does not alter the artwork itself. Immersive art, by contrast, breaks this one-directionality through real-time interactive technologies: the viewer's actions become input variables for the real-time generation of the work, and the perceptual relationship is constituted bidirectionally.

Second, there is the absence of the body. The gaze presupposes a "pure visual subject" that exists only from the shoulders upward, while the body—as a moving, weight-bearing entity that resonates with its environment—is suspended. The exhibition system of the museum, including prohibitions on touching, prescribed routes, and enforced silence, constitutes the material institution of this bodiless gaze. Immersive art, however, recalls the entire body, making it the basic unit of perception and action.

Third, there is the fixity of time and space. Classical gazing takes place within fixed time and space: the artwork is stable, the exhibition has a definite schedule, and the act of viewing follows a linear route. The spatiotemporal structure of immersive art is dynamic, contingent, and even infinite—each entry produces a different experience, and the work forms an open-ended collaboration with time itself.

Fourth, there is the one-way transmission of meaning. In the tradition of the gaze, meaning primarily flows from the artwork to the viewer. The task of the critic is to decode the meaning embedded in the work by the artist, while the task of the viewer is to receive this already encoded information. The rise of relational aesthetics, as proposed by Bourriaud (2002), has already called this model into question [4]. Immersive art further returns the agency of meaning production to participants, enabling them to become co-authors in the true sense.

3 The Theoretical Foundations of the Interaction Paradigm: Embodiment, Participation, and Immersion

3.1 Grau's History of Immersive Art: The Technology Is New, but the Intention Is Ancient

Grau's foundational study provides one of the most important historical perspectives on immersive art [5]. He traces a genealogy of immersive art from the mural environments of ancient Roman villas in Pompeii, to the ceiling paintings of Baroque Jesuit churches, and further to the nineteenth-century panorama, revealing that the human desire to envelop the viewer's body within a visual environment is not a new invention of the digital age, but a persistent impulse running throughout Western art history. The critical implication of Grau's historical study is that digital immersive technology is merely the latest technological carrier of this ancient desire, rather than a revolution in aesthetic essence. The perspective of media archaeology also helps to reveal this historical continuity [6].

3.2 Merleau-Ponty's Embodied Phenomenology: The Body as the Primordial Site of Consciousness

In Phenomenology of Perception (1945), Merleau-Ponty offers a fundamental critique of the Cartesian mind–body dualism in Western philosophy [7]. He argues that perception is not a cognitive operation performed by a conscious subject outside the body, but a pre-reflective contact dynamically constituted between the living body, or le corps propre, and the world. The body is not a container that carries consciousness; rather, it is the primordial site where perception itself takes place. This argument has direct methodological implications for understanding the appreciation of immersive art: when the act of appreciation mobilizes the entire living body as it moves through space, what is activated is precisely the embodied structure of perception described by Merleau-Ponty, rather than the mere reception of retinal information.

3.3 Bishop's Critique of Participatory Art: The Tension Between Democratization and Aesthetic Quality

In Artificial Hells, Bishop (2012) critically evaluates the political and aesthetic claims of participatory art [8]. She points out that advocates of participatory art often replace aesthetic judgment with ethical justification, namely the democratizing potential of participation, thereby avoiding rigorous criticism of the artistic quality of the work. Her central critique is that simple participation does not equal genuine political emancipation, nor does it automatically amount to a profound aesthetic experience. This critique is equally applicable to immersive digital art: interaction itself does not constitute aesthetic value. What matters is whether the interaction generates an aesthetic event with conceptual depth and emotional resonance.

4 A Systematic Comparison Between the Gaze Paradigm and the Interaction Paradigm

Based on the theoretical resources discussed above, this paper systematically compares the structural differences between the gaze paradigm and the interaction paradigm across four core dimensions, as shown in Table 1.

Table 1. Four-dimensional systematic comparison of the gaze paradigm and the interaction paradigm

Comparative Dimension	Gaze Paradigm	Interaction Paradigm
Perceptual Structure	Single visual channel; one-way reception; separation between perception and action	Multisensory integration; bidirectional constitution; perception as action
Role of the Body	The body is suspended; the viewer is a fixed visual point of support; touching is prohibited	The body is the basic unit of meaning production; embodied movement triggers the artwork
Spatiotemporal Experience	Linear viewing route; fixed exhibition duration; a single optimal viewing position	Nonlinear pathways; each experience is unrepeatable; multi-centered space
Meaning Production	From artwork to viewer; critics decode predetermined meanings; viewers receive	Viewers co-produce meaning; social interaction constitutes the final form of the artwork
Subject Position	Fixed subject; binary opposition between gazer and object of gaze	Fluid subject; multiple identities as participant, trigger, and co-author
Critical Language	Formal analysis; intentionalism; historical tracing; semiotics	Process description; relational criticism; ethical evaluation; experience economics
Typical Site	White-cube gallery; traditional museum; theater seating	Immersive exhibition venue; digital art festival; VR art space

Note: The two paradigms are analytical ideal types. In actual artistic practice, there are numerous hybrid forms. Immersive art also contains elements of gazing, while traditional museums are increasingly introducing interactive elements.

5 The TLIMIA Model: A Three-Layered Interactive Structure of Immersive Appreciation

5.1 Proposal and Structure of the Model

This paper proposes the Three-Layer Interaction Model of Immersive Appreciation, abbreviated as TLIMIA, in order to systematically describe the internal structure of immersive digital art appreciation and the relationships among its different layers, as shown in Figure 1.

The first layer is the Perceptual Immersion Layer, abbreviated as PIL. This layer corresponds to the initial perceptual restructuring that occurs when the appreciator enters an immersive space. The coordinated integration of multisensory stimuli breaks the boundaries of everyday perception and produces a sense of de-thresholding, that is, the temporary dissolution of the boundary between everyday reality and the artistic space. Leder et al.'s aesthetic information-processing model (2004) places perceptual analysis at the initial stage of aesthetic appreciation [9]. The PIL layer corresponds precisely to the high-intensity activation of this stage: visual enveloping, auditory surround, and, in some works, tactile participation jointly construct an entry experience into a new sensory world. The PIL layer is the entrance to immersive experience, and its quality directly determines the depth of participation in the subsequent layers.

The second layer is the Embodied Interaction Layer, abbreviated as EIL. This layer corresponds to the dynamic interaction between the appreciator and the immersive environment. The viewer's bodily movements, gestures, and choices trigger the real-time responses of the work, transforming appreciation from passive reception into active exploration. Merleau-Ponty's theory of embodied cognition [7] finds its most direct artistic realization here: the appreciator understands the work through the spatial practice of the body, rather than excluding the body through conceptual analysis. The EIL layer is the core level that distinguishes immersive appreciation from traditional gaze-based appreciation, and it is also the main field in which aesthetic meaning is generated through action.

The third layer is the Social Co-construction Layer, abbreviated as SCL. This layer corresponds to the collective construction of artistic meaning through interactions among multiple participants within the immersive space. Encounters among strangers trigger new visual events, and social interaction becomes a generative variable of the artwork. Bishop's critical analysis of social relations in participatory art (2012) [8] suggests that the SCL layer is both the highest developmental form of immersive appreciation, when social relations truly become critical aesthetic events, and its most fragile layer, when social participation degenerates into mere entertainment-oriented interaction.

Three-Layer Interaction Model of Immersive Appreciation

Model Layer	Core Experiential Content	Theoretical Support	Aesthetic Risk
Third Layer: Social Co-construction Layer (SCL)	Participants jointly construct the meaning of the artwork through mutual interaction; social relations become generative variables	Bishop (2012); Bourriaud (2002)	Entertainment replaces aesthetic depth
↑ Nested within			
Second Layer: Embodied Interaction Layer (EIL)	Bodily movement triggers real-time responses; appreciation becomes action; perception is generated through bodily practice	Merleau-Ponty (1945/1962)	Technological spectacle obscures conceptual depth
↑ Nested within			
First Layer: Perceptual Immersion Layer (PIL)	Multisensory integration; temporary dissolution of the boundary of reality; sense of de-thresholding	Grau (2003); Leder et al. (2004)	Sensory overload blocks reflection
↕ In dialogue with			
Gaze Paradigm (Comparative Baseline)	Static visual reception; one-way information transmission; suspension of the body	Mulvey (1975); Debord (1967)	Passivity; political gaze

Core Proposition of TLIMIA: The three-layered nested structure of immersive appreciation (PIL → EIL → SCL) is not a simple replacement for the gaze paradigm. Rather, while retaining certain gaze-based aesthetic elements, it systematically activates the dimensions of embodied perception and social co-construction.

Figure 1. The TLIMIA model: three-layer structure of immersive appreciation and its relationship to the gaze paradigm

Note: The three-layered model does not represent a linear sequence of stages, but rather a set of mutually nested analytical dimensions. Different artworks and different appreciators may vary in their degree of activation across the three layers. The depth of color reflects the level of embodied participation, with darker blue indicating a higher degree of embodiment.

5.2 Theoretical Claims and Critical Implications of the Model

The core theoretical claim of the TLIMIA model is that a high-quality experience of immersive art appreciation requires the coordinated activation of all three layers, rather than the extreme intensification of a single layer. If only the PIL layer, or sensory spectacle, is strengthened while the substantive bodily participation of the EIL layer is absent, the result is "spectacularized immersion": visitors remain internal spectators amid intense visual stimulation, which is essentially no different from the consumption of spectacle criticized by Debord. If only the EIL layer, or bodily interaction, is strengthened while lacking the social dimension of the SCL layer and the aesthetic-perceptual depth of the PIL layer,

the result is "gamified interaction": participation is reduced to mere operational behavior and lacks the conditions necessary for the emergence of an aesthetic event.

Pine and Gilmore's (1998) argument on the experience economy is also applicable to the commercialization risks of immersive art [10]. When the SCL and EIL layers in the three-layered model are reduced to sensory services for commercial entertainment, immersive appreciation degenerates into experience consumption, and the non-instrumental value of art faces systematic erosion by commercial logic.

6 Two-Way Evaluation: Emancipatory Potential and Critical Tension

6.1 The Aesthetic Emancipatory Potential of the Interaction Paradigm

The interaction paradigm of immersive digital art possesses genuine emancipatory potential on three levels. First, embodied access: by mobilizing the whole body rather than merely training the eye, it provides an affective entry point for audiences who do not possess a traditional art-historical background, helping to break down the knowledge-based class barriers of art appreciation. This resonates with Rancière's concept of the emancipated spectator [11]. Second, spatiotemporal democratization: the reproducibility of digital immersive art, in which the same work can operate simultaneously in multiple cities, breaks the geographical monopoly of traditional large-scale installation art. This reproducibility is rooted in the digital logic of new media [12]. Third, creative democratization: when the participant's body becomes a generative variable of the artwork, the boundary of artistic creation opens toward the public, forming a genuinely decentralized structure of creation.

6.2 Critical Tensions of the Interaction Paradigm

However, Bishop's critique (2012) reminds us that interaction does not equal emancipation, and participation does not equal a critical aesthetic experience [8]. When the interactive design logic of immersive art moves closer to the user-engagement mechanisms of the game industry, driven by instant feedback, a sense of achievement, and entertainment, the distinctive artistic functions of defamiliarization, estrangement, and critical cognition risk being diluted by the logic of consumption.

A more fundamental tension lies in the structural conflict between the high-intensity perceptual activation of immersive experience and the critical distance required for deep aesthetic reflection. Leder et al.'s aesthetic information-processing model shows that genuinely profound aesthetic appreciation requires multiple stages, including perception, the integration of implicit memory, and cognitive mastery [9]. Sensory overload may activate the perceptual stage while systematically blocking the subsequent stages. This means that well-designed immersive art must consciously find a balance between the intensity of immersion and the space for reflection. This is precisely the core methodological challenge facing contemporary digital art curatorial practice.

7 Conclusion

Through a systematic comparison between the gaze paradigm and the interaction paradigm, the construction of the TLIMIA model, and a two-way evaluation of the emancipatory potential and critical tensions of immersive digital art, this paper establishes the following core conclusions:

First, the paradigm shift from gaze to interaction has genuine structural significance, rather than being merely a technological update at the level of medium. The systematic reorganization of perceptual structure, bodily role, spatiotemporal experience, and meaning production constitutes a truly paradigmatic transformation in the practice of art appreciation.

Second, the TLIMIA model reveals that high-quality immersive appreciation requires the coordinated activation of three layers: perceptual immersion, embodied interaction, and social co-construction. Any design that excessively strengthens one single layer while neglecting the others risks degenerating into spectacle consumption or gamified entertainment.

Third, the interaction paradigm is not the terminator of the tradition of the gaze, but its interlocutor. Grau's historical research suggests that immersion and gaze are two complementary forms of human aesthetic desire. Digital technology merely provides new conditions for realizing this complementarity, while the intersection of artificial intelligence and art further extends this proposition into new and more complex dimensions [13]. The core task of contemporary art appreciation education is to help audiences move consciously between the two paradigms: to be able both to immerse themselves within the work and to step back from it, and to reconstruct the Archimedean point of criticism after embodied participation.

Peng Feng (2015) reminds us that the beauty of nature in environmental aesthetics lies in "non-separation": the viewer and the natural environment are not positioned in an oppositional relationship of gaze, but rather in a symbiotic

relationship of perception [14]. What immersive digital art pursues may be precisely the re-realization of this ancient state of non-separation through technological media. The task of criticism is to distinguish whether this re-realization is merely a sensory temptation produced by technological spectacle, or a genuine form of aesthetic symbiosis.

References

[1] Ryan, M.-L. (2001). Narrative as virtual reality: Immersion and interactivity in literature and electronic media. Johns Hopkins University Press.

[2] Mulvey, L. (1975). Visual pleasure and narrative cinema. Screen, 16(3), 6–18. https://doi.org/10.1093/screen/16.3.6

[3] Debord, G. (1994). The society of the spectacle (D. Nicholson-Smith, Trans.). Zone Books. (Original work published 1967)

[4] Bourriaud, N. (2002). Relational aesthetics (S. Pleasance & F. Woods, Trans.). Les Presses du réel. (Original work published 1998)

[5] Grau, O. (2003). Virtual art: From illusion to immersion (G. Custance, Trans.). MIT Press.

[6] Huhtamo, E., & Parikka, J. (Eds.). (2011). Media archaeology: Approaches, applications, and implications. University of California Press.

[7] Merleau-Ponty, M. (1962). Phenomenology of perception (C. Smith, Trans.). Routledge & Kegan Paul. (Original work published 1945)

[8] Bishop, C. (2012). Artificial hells: Participatory art and the politics of spectatorship. Verso Books.

[9] Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. British Journal of Psychology, 95(4), 489–508. https://doi.org/10.1348/0007126042369811

[10] Pine, B. J., & Gilmore, J. H. (1998). Welcome to the experience economy. Harvard Business Review, 76(4), 97–105.

[11] Rancière, J. (2009). The emancipated spectator (G. Elliott, Trans.). Verso Books. (Original work published 2008)

[12] Manovich, L. (2001). The language of new media. MIT Press.

[13] Weibel, P. (Ed.). (2019). Being human: Facing AI. ZKM | Center for Art and Media.

[14] Peng Feng. (2015). Perfect Nature: The Philosophical Foundation of Contemporary Environmental Aesthetics. Peking University Press.