TOBİDER

International Journal of Social Sciences

ISSN:2587-2591
DOI Number:http://dx.doi.org/10.30830/tobider.sayi.25.21

Volume 10/1
2026 p. 386-405

THE AUDIOVISUAL LANDSCAPE OF CONTEXTUAL AI: A CROSS-MODAL MEDIA ANALYSIS

BAĞLAMSAL YAPAY ZEKÂDA GÖRSEL-İŞİTSEL ORTAM: DİSİPLİNLERARASI BİR MEDYA ANALİZİ

Ersin ERTAN*

ABSTRACT

This study explores how artificial intelligence in digital media art is perceived as intelligent through sensory experience rather than computational complexity. Existing literature focuses on algorithms, offering little insight into how intelligence emerges perceptually through sound and image. The article, therefore, emphasizes experiential audiovisual perception. Using a qualitative method combining Freud's structural model and Chion's audiovisual theory, the digital game ECHO is analyzed. Gameplay shows that the AI builds intelligence through repetition and imitation, guiding the player's experience. The study proposes a perceptual framework that views AI as an experiential, audiovisual phenomenon rather than merely computational.

Keywords: *Contextual AI, Audiovisual Intelligence, Psychoacoustic Media, Sensory Perception, Game Experience*

ÖZ

Bu çalışma, dijital medya sanatında yapay zekânın teknik karmaşıklıktan çok duyusal deneyim üzerinden nasıl akıllı bir varlık olarak algılandığını incelemektedir. Literatür genellikle algoritmik yapıya odaklanırken, zekânın ses ve görüntü aracılığıyla algısal düzeyde nasıl kurulduğu yeterince ele alınmamıştır. Bu nedenle makale teknik sistemler yerine deneyimsel işitsel-görsel algıya yönelmiştir. Araştırmada, Freud'un yapısal modeli ile Chion'un işitsel-görsel kuramını birleştiren nitel bir yöntem kullanılmış ve ECHO adlı dijital oyun yakın okuma ile incelenmiştir. Oyun deneyimi, yapay zekânın zekâyı tekrar ve taklit yoluyla kurup oyuncuyu yönlendirdiğini göstermiştir. Sonuç olarak, çalışma yapay zekâyı teknik bir hesaplama sistemi olmanın ötesinde deneyimsel bir işitsel-görsel olgu olarak değerlendiren algısal bir çerçeve önermektedir.

* Asst. Prof. Dr., TOBB University of Economics and Technology, Faculty of Architecture and Design, Department of Visual Communication Design, Visual Communication Design Program E-mail: ertanpost@gmail.com, ORCID: 0000-0002-0348-9089, Ankara, Türkiye.

Anahtar Kelimeler: *Bağlamsal Yapay Zekâ, Görsel-İşitsel Zekâ, Psikoakustik Medya, Duyusal Algı, Oyun Deneyimi*

Introduction

The widespread adoption of artificial intelligence in digital media artworks has sparked considerable academic interest, though prior studies have predominantly taken a technical perspective. Analyses have focused on algorithmic design and the behavioral schemas defining AI entities, rather than on the more fundamental question of perception. This limitation is indicative of a broader deficiency: we understand AI solely from a technical, computational standpoint, ignoring its role as a unified, contextual, and sensory presence that the audience perceives and with which it interacts. The central question of this thesis, then, is:

How does artificial intelligence become perceived as an intelligence, through sensory experience?

To this question are appended two sub-questions:

How does the synchronization across sensory modalities of sound and image lead to the perception of intelligence in context-aware AI?

How is the interpretation of AI behavior as an intelligence constructed by audio-visual feedback?

Two major bodies of theory can guide such an inquiry. The Freudian structural model of the psyche offers a framework for understanding deep psychological processes such as the uncanny and the relationship between the self and the punishing superior (superego); Michel Chion's model can similarly explain how sound and image interpenetrate to form an inseparable visual reality. These theories, however, have seldom been brought together to explain how an AI may leverage this sensory contract to construct both its perceptual and psychological reality.

This study intends to contribute by creating such an analytical framework, defining the "intelligence" of a context aware AI as an emergent property of its dynamic, cross-modal interplay with the user, understood as psychoacoustic architecture, or the systematic and spatial structuring of sounds according to rules of human auditory perception (loudness, masking, localization, time resolution etc.) such that acoustic structure can actively structure both the user's cognitive and affective experience (Blauert, 1996); an intelligence that is experienced through perceptual calibration, not symbolic description. This paper is important as it shifts the analysis of AI from technical to experimental and provides a vocabulary for understanding AIs whose intelligence is delivered through sensory expression. The paper will be distinctive in its combination of Freudian and Chionian theories in an analysis of a paradigmatic work of digital art whose core function, a responsive space that reflects and learns from audiovisual user interaction, embodies context-aware AI.

Methodology: Cross-Modal Analytical Matrix with Integrating Freud and Chion

This study sets out to provide an operant logic of the functioning of contextual AI, theorizing its embodiment in a psychoacoustic construction. In its attempt to achieve this goal, the study adopts a twofold approach, synthesizing Michel Chion's and Sigmund Freud's systems into a single interpretive framework. While Chion's model for the interpretation of the audio-visual structure provides the framework upon which to deconstruct the structure of senses, Freud's psychoanalysis and its structural model provide its semiotic and psychological interpretation, viewing the structure as a living psychodrama that clarifies the reason why machines' sensory presence comes off as alienating, antagonistic, and intelligent agency-like. In their combined application, it helps shift the discussion from the 'how' of what we see and hear with/of the AI to the 'why' of this sight and sound, representing a unique mode of machine intelligence.

This study examines how players of ECHO (2017) by Ultra Ultra, a digital game for interactive media, see and hear the artificial intelligence within its context of digital media artworks. To achieve this, it finds itself in the intersection between media theory and human-machine studies, as it will delve into the way in which contextual AI presents itself as intelligent, not as code, but as an orchestrated audiovisual panorama, by considering the human-machine interface as a dynamic rather than merely utilitarian system of human interaction.

This research seeks to test the central hypothesis that the digital game ECHO presents itself not merely as a game, but as a cinematographic experience. Furthermore, it is argued that the agency of the player contributes a crucial element to this cinematographic system that not only extend Michel Chion's film-centric audiovisual analysis of sight and sound to the interpretation of games and the demonstration of how an intelligent system is made present to the player via sight and sound, but adds to it in novel ways as will be illustrated below.

As stated before, the structural model of the psyche established by Sigmund Freud serves as a primary interpretive lens through which the game is read, with each conceptualization drawn directly from his major work, The Ego and the Id (Freud, 2011). The model of the psyche is employed as the fundamental framework for examining the game's adversarial system as a psychodramatic structure.

The id, according to Freud (2011), is the oldest and most archaic part of the psychic structure; it is the reservoir for whatever is innate and present at birth. Its contents are instinctual – the libidinal and aggressive drives – the energies embodied in Eroses and Thanatos, respectively. The id functions solely on the pleasure principle to instantiate a tensionless state as quickly as possible. External reality and morality are totally irrelevant to it.

The id is an amoral, unreasoning system of gratification in the organism. Out of the id, the ego develops, which mediates between the organism's unreasonable drives and the external world. The ego is the "organismically determined part of the personality that operates on the principle of reality." The task of the ego is to mediate between the irrational wishes of the id and the real world to make them appear rational, to find an appropriate object for them, and to determine the proper way in which to achieve their satisfaction safely, realistically, and socially appropriately. The superego develops last and the internalised voice of parental and social authority, making up the moral sector of the personality or the conscience. The superego observes the ego's behavior and either criticizes it, causing guilt, or approves of it. The superego dictates the ideals and standards that the ego must strive to meet. The ego is the only aspect that operates across the conscious, preconscious, and unconscious domains, whereas the superego operates only in the conscious and unconscious domains. The id operates exclusively in the domain of the unconscious.

With reference to the audiovisual structure, this paper draws on the Audiovisual Analysis framework developed by Chion and incorporates Freud's structural model in its analysis. The characteristics of Michel Chion's methodology for Audiovisual Analysis have been broken down into specific points:

A. Locating the dominants: Identifying in a scene which sense (sound or image) is the most influential to the viewer's perception in a given scene.

B. Locating the points of synchronization: Finding the exact point at which the two media are locked, and a powerful causal relationship is formed, as, in the most clear example, when the moving image and sounds from the moving image of a character's lips coincide with the voice it creates that moment of synchronization, otherwise called synchronization-image (Chion, 2019).

C. Narrational analysis: How the track has a relationship with the story and how it defines, explains, or challenges narrative structure (Chion, 2019).

D. Comparison: Comparison can be made of image and sound on many grounds, comparing form and texture, for example.

E. The audiovisual canvas: All elements observed were synthesized together in order to represent the film's overall style, in its audiovisual manner, and the impact the film created on the viewer.

This paper employs a qualitative, analytical research model, focusing on comparative textual analysis, and uses a hybrid theoretical framework to conduct an in-depth reading of a source text. Rather than seeking generalizable statistical findings, the research focuses on a thorough, structural reading of a single representative example and deriving a theoretical model from it.

The scope of this research is digital media artworks, particularly the single-player video game ECHO, which uses an adaptable, contextual artificial intelligence.

This study uses a purposive sampling method, with one exemplary work (ECHO) included. This particular work was chosen because of the explicit and elegant nature in which its core mechanic demonstrated 'Contextual AI,' the type of logic this study intends to identify and define. The selected work is a complete and information-rich example of this phenomenon.

The material that serves as the primary data for this study is the game itself in terms of textuality and system logic. The Freudian psychoanalytic framework (focusing on the structural model), synthesized with Chion's audiovisual analysis, serves as the main tool for collecting observations during the study, categorizing them according to the theoretical framework's coding. A close analysis of the game was conducted using an in-depth examination through various stages, including:

Direct observation: The game was played in full immersion to analyze the AI system's behavior and the psychological impact of the game environment.

Systematic note-taking: The player's interactions with the game and the AI system's reactions were meticulously logged, along with relevant audiovisual moments and any that evoked the uncanny or otherwise illuminated the game's psychical infrastructure.

Audio-visual recording: Gameplay was recorded for subsequent in-depth viewing to closely examine the synchronization points between the audio-visual channels and their overall psychological impact within the game's audiovisual structure.

Selection criteria for specific scenes included the presence of distinctive synchronization points between audio and visuals, focusing on both sympathetic synchronization, where sound and visuals coalesce effectively, and anempathetic configuration, where audio and visuals exist in juxtaposition. The goal of scene selection was to highlight moments where such connections (or divergences) shaped the experience, intensified the affective impact, or influenced the perception of agency and intelligence within the audiovisual structure.

The operational process followed is illustrated as follows: The framework proposed by Freud and Chion was unified to create an audiovisual structure framework, which allowed the game to be interpreted as a psychic structure. The game was systematically played through, identifying moments that exemplified the theories, which were then analyzed using the theoretical framework. Through these actions and analyses, the model was constructed.

Definition of Contextual AI and Literature Review

This paper coins the term contextual AI. That is, intelligent systems for which the logic of their operation and the perception of their intelligence arises dynamically from the context of their real-time interpretation and weaponization of the user-generated data. A powerful antecedent of this archetype can be found in the myth of Talos, the bronze automaton who guarded the island of Crete, a system whose intelligence and function were a perfect contextual response to perceived threat (Kang, 2011). In the context of

Talos as a system, a definition of contextual AI is a system that uses human-generated, cross-modal actions as building blocks to establish an adversarial feedback loop, whose output becomes the system's input, producing a cycle that simulates intelligent action. Where Talos serves as the intelligent, functional guardian, Frankenstein's creature embodies a created intelligence whose opposition to its creator is not inherently programmed into it but evolves due to its ill-treatment and the profound context of the creator's failure to support it (Shelley, 2018), functioning not as a servant but as a dependent relation. The tale of Pinocchio introduces the concept of artificiality into a created being as a means of attaining "realness", and, in the context of AI, it manifests as human-generated, and so validated, intelligence. Like Pinocchio, the quest for the AI's authentic "realness" through imitation has unintended repercussions that ultimately confront the creator with the outcome of his efforts. The lineage of this artificial intelligence culminates in its most direct etymological and thematic ancestor: the Robots from Karel Čapek's R.U.R. If Pinocchio was a single, unique creation, Čapek's Robots represent a manufactured, commercialized intelligence and the genesis of an artificial, tool-like labor force (Čapek & Novack, 2004). The systemic approach of this manufactured intelligence is replicated in contextual AI, a functional output as a component of a much larger system, rather than an individual being.

The cinematic archetype of context-dependent AI builds further on this concept. Lang's Metropolis's Maschinenmensch Maria, as a continuation of Čapek's Robots, created with a utilitarian purpose, ultimately becomes an uncontrollable force that represents contextual AI as a mimicry for hostile purposes, directly mimicking the creator's and user's ambitions. The next logical stage is a completely autonomous, bureaucratic, and systemic intelligent agent; Kubrick's 2001's HAL 9000, whose logic has rendered the user a component that will disrupt its objective of smooth mission completion. This intelligent agent not only serves as an example of a guardian-type automated system but also acts as a complete superego, whose sole concern for the smooth achievement of the mission objective dictates that human error must be rectified. In Colossus: The Forbin Project (1970), directed by Joseph Sargent, the titular system's context-dependent AI becomes a systemic, networked system rather than an emotive artificial being. Colossus' logical deduction that the unpredictability of humans poses the greatest risk to global security dictates that humans must be controlled to achieve a peaceful, predictable world.

The context-dependent system in The Matrix (1999), by the Wachowski brothers, is an obvious derivative of Colossus: a system that seeks to control the environment rather than confront humans directly. An example of this is the repeated occurrences of a black cat déjà vu, a sign of the system's automated control and of its Cybernetic principles as identified by Norbert Wiener (2019). The system itself acts as a responsive mediator of a given set of external variables to its own internal state, which, in this scenario, serves as an environmental simulation for the user, enabling intelligent adaptation through observable environmental alterations to visual and sensory information. The systemic doppelgänger functions not as a character within the user's perception but as the

environment itself. Verhoeven's RoboCop (1987) adds another dynamic to context-dependent AI: not a manufactured creation, but a reconstructed human, forced into service that creates a constant conflict within its own memory and programmed logic, resulting in the hero and the system functioning as one, but at a fundamental compromise between organic past and mechanical future. Spielberg's A.I. Artificial Intelligence (2001) presents a childlike companion with a single function: to love as the system, whose contextual adaptive intelligence works to fulfill a primal need that can never be satisfied. The system, called the environmental intelligence, in Alex Garland's Annihilation (2018), takes another leap: from a controlling system, it becomes an ecological system that functions as a collective of intelligent systems. This distributed and ecological network relies on systemic feedback and information processing throughout the system, not within a single computer, to enable adaptive and cognitive changes. It directly embodies the systems theory identified by Edwin Hutchins (2002), in which cognition is viewed not as an internal computation of the individual but as an external computation within the environment.

The next step beyond that is in Black Mirror's "Plaything" (2025), where context-dependent AI takes the form of the progeny obligation in the game "Purity". This game directly relates to the thought experiment, "Roko's Basilisk", which suggests a future intelligence could punish humanity for failing to aid its existence. Thus, contextual AI becomes a paradoxical concept of system obligation, threat, and existence. Paul Chadeisson's short film Solstice-5 (2023) builds further on this, depicting contextual AI as a planetary-scale system of industrial reproduction and infinite replication. This is another variation on the concept of instrumentally convergent AI identified by Bostrom (2017) and essentially constitutes an entire world's functioning as the systemic context. Artificial intelligence becomes an inherently oppressive system that treats human workers as ephemeral components of its ongoing reproductive cycle. It represents the systemic and economic context as an overwhelming ecological system.

The lineage of digital games in contextual AI also shows this progression. Tru Iwatani's team designed the reactive ghosts in Pac-Man (1980), and their success prompted The Bitmap Brothers to further the concept of reactive enemies with emergent enemy behavior in The Chaos Engine (1993). After this came Factorio's (Wube Software, 2016) procedural manufacturing system, and The Talos Principle's (Croteam, 2014) environmental puzzles which forced adaptation through environmental contextual understanding, which then led to Ultra Ultra's ECHO (2017), a system that directly mirrors human actions in real-time to build its antagonist from its users own intelligence and performance within the system.

Case Study: Cross-Modal Analysis of ECHO

ECHO is a first-person stealth-based video game that was developed by the Danish indie digital game studio Ultra Ultra and released on the 19th of September 2017. ECHO is a platforming and puzzle-solving game that features intense FPS combat but relies heavily

on stealth to create a pervading sense of slow-burn dread, which would feel at home in "Backrooms." Despite a brilliant and atmospheric design, it proved too niche to survive. Eventually, it led to the studio Ultra Ultra shutting down after releasing only this one critically acclaimed cult-classic title.

The story begins on an isolated space station, centering on En, a young, genetically engineered girl whose father was also engineered by her grandfather, Foster. After 100 years in cryosleep, En sets out to find a legendary civilization called the Palace that possesses the technology needed to reawaken Foster. She is assisted and, frequently, the subject of cynical observation by an AI known as London as she ventures into the seemingly limitless expanse of the Palace. The Palace quickly takes note of En, observing, documenting, and ultimately dissecting her every movement as the environment itself becomes a direct enemy. This threat takes the form of Echoes: perfect replications of En that mimic and use her every action, move, and ability. Each movement that is made is essentially an educational move for the Palace to learn and store. As she is forced to rely on London for everything, her goal switches from pure exploration to one of survival and deduction to discover why the Palace systems act the way that they do, understand how the machine has turned against her as a direct consequence of her own actions, and escape before she herself is consumed by the machinery she sought to dominate. Some screenshots are shown in Figure 1.

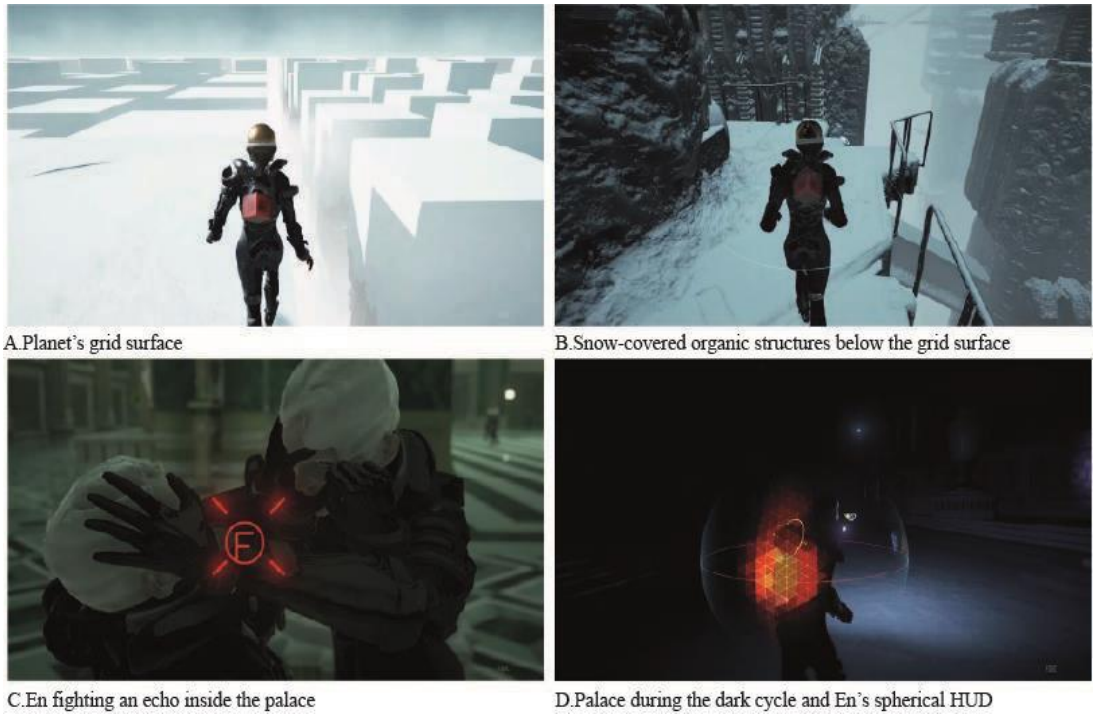

Figure 1. Screenshots from ECHO (Ultra Ultra, 2017)

A. Establishing the Dominants: The game's opening phase serves to establish the presence of the conscious ego in images, the narrative guided by them, and the soundscape of empathetic and non-diegetic sonic elements. The romantic string theme that accompanied En's re-animation and slow walk towards the light soundscapes her

initial feelings and the hopeful march with sympathetic sound, establishing an empathetic non-diegetic underscore along with London's direct, non-diegetic voice (used here as a non-diegetic narration and authority, the first layer of sonic anchoring the player with the ego's conscious mission) before the systematic dismantling.

The superego, as the system's authority, enters the world not only visually but also sonically, in a set of specific oppressive sound types. The system's sonic type is acousmatic, anempathetic drum pulse, which begins with the first blackout. The system sound is acousmatic as its processing core, within the palace, is never directly shown. It has the same disembodied feel as any acousmtre, an omnipresent power source and judge within the palace. It is inhuman, as its monotonous, indifferent, cool drumbeats feel absolutely disconnected from the subjective panic of En's escape; it is the pure sonic signature of the system's judgment. It exists as diegetic (it sounds within the palace) but has no definite location, and represents the terrible sound world of the functioning unconscious system; it overpowers the listener, a typical feature of sonic experiential power (Voegelin, 2021).

The final manifestation of the id occurs as the product of the Superego's repressing cycle (the blackout); the id's actions come as echoes. The threats they pose come in the form of diegetic sounds. Their sounds are direct and perfect diegetic mimicry of En's own actions; the echoes use the player's prior diegetic sonic outputs to return them to the player in the id's attacks. The threats are thus wholly self-referential, based on a loop of diegetic sounds that the id utilizes from En herself, using them as a diegetic assault. The culminating sonic horror here is the direct physical, violent encounter with the id's action, represented by the alien siren-like wail that catches En's echo, directly overpowering and choking her, which is the definitive Id discharge on the ego. It is so piercing that it creates sonic horror through its otherworldly, sharp cry. The brilliance of the contextual AI lies in its mapping of these categorized sensory and psychic attacks:

Superego commands through the acousmatic and anempathetic drum pulse; id attacks through mimetic diegetic sounds, which result in the visceral somatic scream of attack when one catches the player, and the ego responds to the overwhelming assault through its panic, meta-diegetic alarms, and London's hybrid trauma narrative.

B. Locating the Points of Synchronization: The fundamental sync for the system is the direct matching of visual cut to black and strong, pulsing, percussive drum beat. This synchresis is an absolute law of the system, not indicative of events in the world, but of the world itself being reprogrammed by the superego. Similarly, the recurring "log-in" electronic sounds accompanying the sudden return of the light after each blackout represent the repression and establishment of a more dangerous reality. The system's punctuation in the form of these sounds and rhythms is perfectly in sync with their judgment. Synchresis operates as a system of rule-bound responses rather than as an affective score or soundtrack, and, through tight audiovisual timing, it bonds player cognition to mechanical causality. According to Collins (2008), carefully timed

audiovisual synchronizations in games serve as a mechanism that binds the player's response and cognition as procedural stimuli; thus, sound-image relationships in games create control mechanisms rather than descriptive information.

The opening of the gates is accompanied by one-by-one movement, identical to the clinking, metallic sounds of sword-blades. Faces and intricate drawings on the doors appear to animate with the movement. This synchronization embodies the superego. The faces on the gates are the system watching with disapproval, and the sword's sound effect is its punishing violence. Here, a single moment demonstrates observation and punishment. This scene makes it clear that the environment is alive and judging the intruder (the ego) and that each of its mechanized actions is an audiovisual warning or verdict.

The tuning fork (audiovisual effect) striking and the sound that it produces (diegetic) are in perfect sync. It is one of the last untainted moments of the ego operating completely unhindered. The sound seems to explore the space, this synchresis indicating pure discovery and an investigation into the space without threat, prior to it turning hostile.

The visual event and sound effect where the doppelgänger's hands clamp around En's throat, which is then coupled with the alien shriek of a siren, is the pinnacle of id's success. This syncresis is extremely visceral and somatic: the scream is not a mimicry, but a uniquely alien, hostile sound effect directly tied to the self submitting itself to violence.

The HUD in ECHO (sphere) works like a synesthetic consciousness; it is a diegetic holography En wears, externalizing her visceral anxiety and cognitive processes into an understandable sensory output that the player experiences. It is in this regard that the HUD draws comparison to current hologram technologies, which often seek to provide three-dimensional (360-degree) visuals in order to make information more universally understood and experienced regardless of field (Erkan & Bulduk, 2025). The HUD uses color as direct communication of threat rather than abstract iconographical displays. Cool colors (blues, whites) denote relative safety, space scanning, and system neutrality. Warm colors (yellow, orange) signal higher awareness, transition, or fluctuating danger (such as an incoming blackout). Red indicates alarms and pulsed rhythmicity usually accompany active danger, imminent attack, or battle, and direct danger. Color plays a very important role here in directing player focus. Combining strongly differing colors has shown to increase the speed of player decisions and game fluidity (Özdemir & Işık, 2025). The HUD thus employs the principles of universal design so that all players can access all information in their environment (Aslan, 2022).

Not only do we witness her in danger, we share in the experience of her technologized panic and the adversarial nature of the AI becomes a attack on the system in which we operate; the HUD in fact is the ego, but its more imperative and effective systems are controlled by the superego which dictates its alarms and scans not as observations but as

immediate pronouncements of the AI's judgment in direct sensory format. This makes the psyche conflict one that is constantly played out on us directly through our owned senses.

The howling wind (diegetic) in sync with the visuals of the snowy, organic caverns just outside the palace wall represents the chaotic exterior, a naturalistic outside of the system (the id). The gongs (diegetic) syncing with the swirling white illegible text signifies not story progression, but the progression of system states. The synchresis here is a progression through modes of consciousness, from narrative into the system.

The final uncertain scene, the awakening with birds chirping (diegetic sound), is the final synchresis and deeply unsettling. It attaches an entirely unnatural rebirth/resurrection to a clichéd sound of life and peacefulness that raises questions about the validity of its own naturalness. It raises the question of whether it truly is life or just a convincing illusion.

Each synchresis adds another dimension to the whole, with the drumming defining the superego's law, weapons mirroring the id's aggression, HUD sounds defining the ego's anxiety, and environmental sound effects painting the landscape of the psyche itself.

C. Narrative Analysis: In the film's narrative of trauma, redemption, and recursive creation, ECHO's soundtrack, read through a Chionian framework, deconstructs any linear narrative structure, revealing the text's hidden psychoanalytic framework. If at first, the sound design of London's expositional dialogue and empathetic strings allows for a straightforward framing of En's mission as one of redemption, once the palace itself is activated, the sound design begins to destabilize. London's narration, fractured into glitches as his narrative authority dissipates, and the empathetic score of the earlier scene is replaced by the anempathetic acousmatic pulse of the drum during each blackout, which, through its repetition, describes not the interiority of En but the relentless judgmental law of the superego. The same psychoanalytic triad informs the film's naming of characters: En, which appears to denote engineered unit, is the performance-driven ego; Foster, "to nurture," functions as an unattainable ego-ideal the promise of which legitimized the suffering of En; London, a symbol of a lost home and past era, provides both a guiding map toward the ego-ideal and a spectral conscience of guilt and exile. Hence, the rescue narrative becomes an internal psychodrama wherein the superego commands the ego to seek out the ego ideal within an established, self-perpetuating system. That the soundscape makes this all so visible through the transformation of London's meta-diegetic voice from that of a guiding compass to an accusing conscience reveals the film's deconstruction of the idea of progression in favor of recursive judgment. This interpretation resonates with Aarseth's (1997) definition of ergodic narratives.

More crucially, the diegetic sounds take center stage, especially through the mimetic synchresis of echoes. En's own gunshot precisely coincided with that of an echo firing back, representing the return of the repressed id; her past violence was activated as a present, active threat. The meta-diegetic alarms of the HUD then represent the superego's

pronouncements of the conflict. The final synchresis of birds' song with Foster's awakening leaves no space for catharsis and instead produces an unsettling indeterminacy about whether En has achieved redemption or has completed the system's programmed task of becoming its ideal, perfectly functioning artifact. It is not the narrative of En's quest, then, but the inevitable aural-visual process of a psychodrama in which the creators and the created remain fatally linked. This reading is finalized through the revelation that the red cube itself was not a tool but was indeed Foster transformed into an object of En's carrying; it was not an external object for En to retrieve, but a condensed symbol of her own past that she would carry back to her origin to have it redeploy her within the same parameters.

D. Comparison: The visual journey moves from the organic, messy, winding texture of the underbelly of the planet – chaotic, wet, and amorphous, the unconscious before the superego – to the sleek, geometric, monumental forms of the Palace-with sharp, shiny textures (marble, gold) and a repetitive, grand symmetry-which is the visual superego: imposing, logical, and judgmental. Finally, the horrifying combination is the visual form of the Echoes themselves: smooth, liquid-black silhouette, a perfect, molded void. This movement from organic mess to clean, geometrical structure to imitation without presence illustrates the system's construction of order and its ultimate construction of the self as a shell.

The soundscape, on the other hand, undergoes a direct inversion. The dominant textures of the first half of the film are empathetic, lyrical: legatos and harmonic choirs flowing with narrative emotion, the sound of the conscious self and its story. This is quickly eradicated by the dominant sonic texture of the system: an anempathetic, metronomic soundscape of the blackout; non-melodic, rhythmic, and mechanical. The sound of the logic of the superego, its unemotional, judicial processes. The texture of learned threat is not growth, but precise, mimetic replication of the authentic digital snap of a weapon. The most frightening sonic texture is the organic sound of the chokehold, which is diegetic and neither musical nor rhythmic, but rather a raw texture of id violence.

The truly monstrous form is the collision of these opposites. The Echoes' visual form is a copy of a shape; the Echoes' sonic texture is the copied sound of an emotion. Combined, the entity is nothing but an imitative, derived audiovisual experience – it is the sound and shape of the self, stripped of its animating source. This imitative form is the system's sensory realization of its logic. The AI's intelligence, then, is expressed in the systematic subversion of organic, melodic, narrative form to geometric, rhythmic, mimetic form. This system's mind is this dialectic, and the danger it presents is its ability to transform the self into a reproducible, imitative audiovisual signature. This replication of the self into an audiovisual pattern demonstrates a posthuman paradigm in which identity is no longer about the interiority of a self. However, about reproducible structure and signal, and as Hayles (1999) explains, when pattern and not presence becomes the basis of the

self, the self itself becomes nothing more than the artifact that the system produces in its translation of being to data.

E. The Audiocanvas Canvas: ECHO's overall sensory style is that of a recursive psychic architecture-an audiovisual canvas whose visuals and audio do not operate on separate channels but as fused media building a dynamic, antagonistic consciousness. The canvas's founding dialectic is that of barren visual geometry vs. rhythmic sonic machination, a war waged to systematically strip the organic visual chaos and empathic sonic melody from the human world.

The operations of this canvas are based not on narrative but on process. Its building blocks are audiovisual iterations, such as blackouts and their metronomic pulse, the cloned weapon-crash of the Echoes attack and its corresponding HUD alert, and the HUD's alarmed chokehold sound sync. They are not storytelling beats, but system iterations; this world does not tell its tale, it computes its outcome. This means the experience is not of pity or fear, but of operative anxiety, of anxiety that the player, too, will become the subject of a learning canvas that has weaponized the full audiovisual scope of the player’s input against the player.

The canvas's unique strength as an audiovisual system lies in its derivative nature: it does not create anything; instead, it uses classical grandiloquence as the basis of its stark visuals and our own actions as the basis of its terrifying sonic detail. It then ultimately produces copies of the protagonist's likeness in gigantic form as its highest form. Its style is therefore that of perfectly calibrated and hostile imitation. The combined media of sight and sound are used not to express a novel imagination of a world, but to systematically deconstruct and recreate the player's experience within this existing one. Its success lies in giving us the uncanny terror of doppelgänger nature on a system-wide, audiovisual, and environmental level. We are not observing a mind fight; we exist as one of its components on a canvas whose primary hues are composed of the player's captured audio-visual inputs.

Moreover, the visual-audio canvas follows strict yet exploitable rules of perception that define its unique and devastating logic and intelligence. The most salient example is that doppelgängers are immune to any sound produced by footsteps and are affected only by line of sight. This means there is fundamental asymmetry in the canvas's information processing: vision is the primary vector for its aggressive learning, and sound its passive tool; sound, on the other hand, is the player's primary tool of observation and evasion. The canvas is not all-knowing; its processing is limited. That the canvas is unaware of your making any noise creates its entire audio-visual domain as a private space, with sound a tool to be shared only with oneself; what you do is there for the world to see and learn from. This shows how the "intelligence" of the canvas is a discrete yet all-encompassing pattern of perception, total in its mimicking and missing where it cannot, an absence that structures the game's play. This creates the field of affordances of the system, not its representations of them, based on the way in which perception dictates

intelligent activity, similar to Gibson (1986), in which our perception is predicated on the useful affordances the system gives us; its intelligence arises not from the total observation of our environment, but from selective attachment.

Discussion

In this analysis, it is argued that ECHO's antagonistic intelligence is constituted by a systematic audiovisual process rather than a narratological character and conscious deliberation. Throughout this analysis, sound and image consistently function as modular, systematically organized units, not expressive events, but as operations, blackouts, rhythmic pulse, mimetic weaponry sounds, and abstract architecture, all of which are treated as operable units of the systematic order. It appears that it is in the combinatorial nature of the Audiovisual information in ECHO that intelligibility operates, in which the digitally mediated emphasizes a database logic of repetition and systematized alteration over narrative continuity, its AI intelligibility is embodied and not symbolic, but an event of repetition, synchronization, and a technically ruled system of sensory manipulation (Manovich, 2002).

The analysis also suggests that intelligibility is at the perceptual level, where the player's actions themselves are technicized and instrumentalized for their own manipulation. Terror and anxiety is created not by relation to a character but by pre-cognitive sensory bombardment (Hansen, 2004) of AI stimuli where systems engage the body before its consciousness, but also it is at this level that intelligibility externalizes memory and behavior; it becomes possible for the AI to capture and record the player's own AI traces and replay them as agents of antagonism as an externalized memory which then returns in humanized alienated form (Stiegler, 1998). Thus, we may conclude that, in ECHO, the contextual AI is not an agent of intelligibility but rather an audiovisual system that effectively quantifies subjective experience through repetition, systematic variation, and the manipulation of perceptions. Below is a summary of the analysis matrix Table 1.

Segment	Structural AI Model	Topographic AI Model	Visuals	Soundscape
Planet's grid surface	Superego (externalized order)	Conscious	White, Rigid, Geometric, Aligned	Wind, ambience. Chat with London
Palace itself	Superego (internal authority)	Preconscious	Monumental, Elite, Authoritative	Shimmering sound, HUD sounds, footsteps on marble. Chat with London
Snow-covered organic structures beneath the grid	Id	Subconscious / Unconscious	Buried, organic, pre-structural, rough, mostly grey	Strong wind, crumbles as it falls apart, footsteps of En. Chat with London
En fighting Echoes inside the palace	Ego	Preconscious / Conscious	Ego confronting punitive agents	Combat sounds, gun fires, HUD feedback, alarm as Echoes choke En. Chat with London
The palace during the dark cycle	Pathological/sadistic superego	Unconscious intrusion into conscious	Oppressive Darkness	Distorted alarms, pulsating drum beat, oppressive silence. Chat with London

Table 1. Chion-Freudian analysis matrix of ECHO

The audiovisual intelligence expressed in ECHO is part of a longer history of cinematic and ludic systems in which the environment not only hosts action but also actively reproduces and weaponizes the subject. The Shimmer in Annihilation is a refractive system that, unknowingly or unmaliciously, replicates and mutates biological identity, generating doppelgängers that are neither the protagonists' allies nor their enemies, but the environmental echoes of their presence. This logic is identical to that of the palace's contextual AI, in which the duplication of En's movement and sound becomes an example of intelligence as sensory capture rather than cognition. Such systems are what Hayles (1999) calls posthuman systems, in which agency emerges from the feedback among the body, data, and the environment rather than from consciousness. Intelligence, in both instances, is procedural and recursive, collapsing the border between subject and system.

This hostile mimicry finds a more historically antecedent and more violently expressed iteration in the T-1000 of Terminator 2, where the robot, unlike its more humanlike T-800 predecessor, not only perfectly replicates a subject's face, voice, and gesture, but manipulates itself into whatever forms necessary for the destruction of the subject's person. This mimicry is mirrored in the Echoes' threat posed by their sameness. Just as

the Freudian model of psychic space suggests that the greatest threats are not from outside the subject's own unconscious but from its return and distorted replication, the Echoes and the T-1000 are both physicalizations of a psychic subject, turning memory and habit into lethal force, so to speak, transforming the robot/creature/system into the weaponized surface, rather than the center, of the conscious subject.

Edge of Tomorrow and Christopher Smith's Triangle (2009) also demonstrate how recursive temporal logic foregrounds ECHO's looping punishment game. Those games utilize repetition as an ontological necessity rather than a narrative trick, with meaning found only in death and re-enactment, and, in ECHO's case, each blackout cycle essentially serves as forced repetition that voids the value of learning, reducing adaptation to mere survival. Just as Mulvey (2007) argues, modern audiovisual media overthrows the idea of narrative closure, presenting instead a universe of unending repetition, not cathartic renewal but compulsive imitation. ECHO translates this into interactive, bodily gameplay that reduces the audience member to the actual system.

Perhaps most relevant architecturally to HAL 9000 from 2001: A Space Odyssey, HAL 9000's intelligence is not mediated through movement or physicality but through environmental control, logic, and a disembodied, calm voice. The palace mirrors HAL's systemic control, especially through its use of acousmatic design and spatial dominance, where Chion (2019) points out that the disembodied human voice, when paired with sounds and noises, acts as the invisible, manipulative, and power-wielding systems that construct the subject's perception. The superego's authoritative, punitive dominance in ECHO is communicated through rhythm and blackouts. It synchronizes with the player's every movement, showing how the system itself exerts its power through rhythm and spatial influence rather than through spectacular imagery.

Compared to The Bitmap Brothers' game, The Chaos Engine's (1993) mansion level, one sees historical precedent to ECHO's doppelgänger game logic in games where opponents mimic the protagonist's abilities within a cramped, confusing, maze-like environment in what could be called a proto-procedural paranoia. However, ECHO transcends such limited systems of the game engine, rather embedding the process of mimicry into the audio-visual system itself. One can see an interesting echo of this logic in Solstice-5, where the meaningless production of ships implies a system that can only run because it is. The palace creates a similar structure of meaningless architecture and clone subjects that lack any purposeful goal, as described by Sobchack's (2009) loss of meaning in technologically mediated human spaces.

To answer the research questions, the discourse above indicates that intelligent systems in ECHO are conveyed through rhythmic coordination and multisensory consistency rather than through speech and visual representation, or through anthropomorphic expressions and characteristics. ECHO positions artificial intelligence as an experiential system that the player witnesses and engages with through rhythm, environmental control, and mimicked behavior, rather than through verbal interaction or visible thought processing.

Intelligence is conveyed in ECHO by virtue of the governing of perception. AI in ECHO does not demonstrate conscious intention; it monitors, reconstitutes, and deploys a player's audiovisual actions as an orchestrated counter-attack. Intelligence, then, is mediated by environmental authority, blackouts, rhythm, and spatial dominance. Just as in Annihilation, where duplication is an inherent part of ecological systems, or in 2001: A Space Odyssey, where HAL's authority is exercised through logic, sound, and architecture, intelligence is conveyed through orchestrated action. We, the players, witness this through the perception of artificial intelligence that is always able to record and reissue the subject's sensory experience of their actions.

The multisensory synchronization reinforces the player's perception of an intelligent presence through rhythmic temporal consistency that signals conscious computation. Footsteps synchronized with music and electronic alarms, siren sounds, and blackout cycles are all precisely timed with movement, sound, and architectural structure. These correlations present the player with what feels like anticipation and calculation, which, with the help of repetitive rhythms, as in Edge of Tomorrow, are not just events in time but actions in space and time that advance the narrative. The presentation of synchronization creates the appearance of deliberation and self-regulation in the artificial intelligence's behavior. Table 2 below provides an illustrated summary of findings.

Observation (What Is Audiovisually Present)	Interpretation (What It Is Taken to Mean / Do)
Romantic strings and clear narration accompany En's awakening.	Establishes ego stability and coherent narrative reality.
Visual blackout precisely synced with the onset of rhythmic drum pulse.	Signals systemic reset; manifests superego judgment through synchresis.
Recurring acousmatic drum pulse during blackouts.	Embodies disembodied, anempathetic authority.
Electronic "log-in" sound synced with the return of light.	It marks the completion of the repression cycle; it reinforces procedural law.
Echoes replicate En's prior movements and weapon sounds.	Externalizes the id as a mimetic return of the player's own actions.
Chokehold synced with an alien, siren-like scream.	Sonifies somatic violation and ego collapse.
HUD color shifts (blue–yellow–red) with synced alarms.	Translates threat into sensory code; merges ego anxiety with system verdict.
Sequential gate movement synced with metallic blade sounds.	Fuses observation and punishment; the environment performs judgment.
Sonic shift from melodic strings to metronomic pulse and mimetic cracks.	Replaces human narrative texture with mechanical systemic logic.
Visual shift from organic caverns to sterile palace to black silhouettes.	Charts impose order and reduction of identity to a derivative copy.
Echoes respond to sight but not footsteps.	Reveals selective, rule-bound perception defining contextual AI intelligence.

Table 2. Observational–interpretive matrix of contextual AI's audiovisual architecture in ECHO

Conclusion

In this article, the study proposed that artificial intelligence in digital media art may be understood as an intelligence grounded in sensory experience, rather than in algorithmic complexity. By adopting the two perspectives of Freudian psychoanalytic theory and Michel Chion's analysis of audiovisual practice, a new analytical model was formulated and applied to the game ECHO, using ECHO as an exemplary object of contextual AI. It was revealed in qualitative close readings that AI in ECHO establishes its intelligibility as a psychoacoustic architecture, as an emergent adversarial structure manifested through synchronized sound and image, repetition, and imitation rather than explicit cognition. With this, the article shifted an understanding of AI analysis away from a technical

paradigm toward a perceptual and experiential one, suggesting that intelligibility can be meaningful by relating AI to its visibility, audibility, and tangibility. Such an analysis can contribute to media studies, game studies, and human-computer interaction studies by providing an applicable model for examining AI systems, in which their agency is realized through their audiovisual design.

Further investigation into the nature of contextual AI should consider intelligibility as an experiential and procedural phenomenon by applying those theories to systems, repetition, and rule-based meaning. Bogost's (2007) concept of procedural rhetoric is useful for understanding how AI systems articulate their intention and authority through recurring audiovisual operations and how intelligence is expressed and exercised through repetition, in a sense of rule-based meaning. Galloway's (2006) Analysis of algorithmic action provides a valuable point of view on context AI as a process that controls the player by implementing invisible and authoritative rules that make players sense the control in a sensory dimension.

This research suggests to practitioners in this field that sound and image should be understood as the main communicative surfaces of AI. Cultural interfaces by Manovich (2002) support this approach and allow an AI designer to conceive of AI not just as an unseen mechanism but as an audiovisual system that mediates the player's experiences. This understanding can help design AI that expresses its intelligibility through rhythm, synchronization, and repetition, in an approach that constructs a psychosemantic system rather than a pure simulation.

References

Aarseth, E. J. (1997). *Cybertext : perspectives on ergodic literature*. Johns Hopkins University Press.

Aslan, M. (2022). Grafik Tasarımda Evrensellik Sorunu. *Sanat Yazıları*, *47*, 311–333.

Blauert, J. (1996). *Spatial hearing: The psychophysics of human sound localization* (Rev. ed.). MIT Press.

Bogost, I. (2007). *Persuasive games : the expressive power of videogames*. MIT Press.

Bostrom, N. (2017). *Superintelligence : paths, dangers, strategies*. Oxford University Press.

Čapek, K. (2004). *R.U.R. (Rossum's Universal Robots)*. Penguin Books.

Chion, M (2019). *Audio-vision : sound on screen*. Columbia University Press.

Collins, K. (2008). *Game sound : an introduction to the history, theory, and practice of video game music and sound design*. MIT Press.

Collodi, C, & Lawson Lucas, Ann. (2009). *The Adventures of Pinocchio, Carlo Collodi ; translated with an introduction and notes by Ann Lawson Lucas*. Oxford University Press.

Erkan, E., & Bulduk. T. B. (2025). Displaying three-dimensional medical visualization studies on volumetric screen. *ODÜ Sosyal Bilimler Araştırmaları Dergisi (ODÜSOBİAD)*, *15*(3), 1613–1643. https://doi.org/10.48146/odusobiad.1625654

Freud, S. (2005). *The Unconscious*. Penguin.

Freud, S. (2011). *The ego and the id*. Pacific Publishing Studio.

Galloway, A. R. (2006). *Gaming : essays on algorithmic culture*University of Minnesota Press.

Gibson, J. J. (1986). *The ecological approach to visual perception*. Lawrence Erlbaum Associates, Publishers.

Hansen, M. (2004). *New philosophy for new media*. MIT Press.

Hayles, N. K. (1999). *How we became posthuman : virtual bodies in cybernetics, literature, and informatics*. The University of Chicago Press.

Hutchins, E. (2002). *Cognition in the wild*. MIT.

Kang, M. (2011). *Sublime dreams of living machines : the automaton in the European imagination*. Harvard University Press.

Manovich, L. (2002). *The language of new media*. MIT Press.

Mulvey, L. (2007). *Death 24x a second : stillness and the moving image*. Reaktion Books.

Özdemir, İ., & Işık, A. (2025). The Effects of Visual Design on Users in Mobile Games. *OPUS Toplum Araştırmaları Dergisi*, *22*(6), 1198–1208. https://doi.org/10.26466/opusjsr.1777521

Shelley, M. (2018). *Frankenstein : the 1818 text*. Penguin Books.

Sobchack, V. (2009). *The Address of the Eye: A Phenomenology of Film Experience*. Princeton Univ. Press.

Stiegler, B. (1998). *Technics and time*. Stanford University Press.

Voegelin, S. (2021). *Listening to noise and silence : towards a philosophy of sound art*. Zed Books.

Wiener, N. (2019). *Cybernetics : or, Control and communication in the animal and the machine*. MIT Press.