



# AVATAR AGENCY: ALGORITHMIC SUBJECTIVITY IN AESPA

Sulgi Lie<sup>1</sup>

<sup>1</sup>*Binghamton Please provide Affiliation Details*

## Abstract

"What is an avatar?" The popular imagery of the 2009 film of the same name by James Cameron and its two sequels has significantly influenced the public's conception of the term "avatar." Although the term is derived from Sanskrit and refers to the incarnation of a divine being in human or animal form, popular culture instinctively associates it with the blue mythical creatures that inhabit the planet Pandora in Cameron's films and are called Na'vi. In the film's mythological narrative, the Na'vi—the indigenous inhabitants of the planet—are depicted as both natural beings and subjects of biotechnological experimentation. Their DNA can be engineered in a laboratory setting to be controlled by a human via a neural interface. As suggested by the name, the Na'vi are "natives" as an original biological species. However, as biotechnological clones, they assume the role of navigators ("Navi" in the abbreviation), who can also be controlled and moved by humans who, like the protagonist Jake Sully in Avatar, are paraplegic. In order to visually distinguish Sully as the film's hero from the mass of other Na'vi, his face is designed to be somewhat more anthropomorphic. However, there are no visual markers that would indicate an ontological difference between the original and the avatar.

## Keywords

Avatar, Fractality, K-Pop, Music Video, Multiverse, Recursion, Transmedia

"What is an avatar?" The popular imagery of the 2009 film of the same name by James Cameron and its two sequels has significantly influenced the public's conception of the term "avatar." Although the term is derived from Sanskrit and refers to the incarnation of a divine being in human or animal form, popular culture instinctively associates it with the blue mythical creatures that inhabit the planet Pandora in Cameron's films and are called Na'vi. In the film's mythological narrative, the Na'vi—the indigenous inhabitants of the planet—are depicted as both natural beings and subjects of biotechnological experimentation. Their DNA can be engineered in a laboratory setting to be controlled by a human via a neural interface. As suggested by the name, the Na'vi are "natives" as an original biological species. However, as biotechnological clones, they assume the role of navigators ("Navi" in the abbreviation), who can also be controlled and moved by humans who, like the protagonist Jake Sully in "Avatar", are paraplegic. In order to visually distinguish Sully as the film's hero from the mass of other Na'vi, his face is designed to be somewhat more anthropomorphic. However, there are no visual markers that would indicate an ontological difference between the original and the avatar.

Despite the film's efforts to preserve the disparities between the native Na'vi and the avatar Na'vi, as well as those between the avatar and its human navigator, Sully, through both narrative and ideological means, the very logic of the signifier "Na'vi" unveils the inherent indistinguishability of nativity and navigation, of nature and technology, within the context of biotechnological media culture in the present era. The fundamental question that emerges from this analysis is whether Sully exerts genuine control over the avatar, or whether the reverse is true. The avatar's operational ambivalence, characterized by its oscillation between external control and self-control, necessitates a reevaluation of the conventional definition of the avatar as a mere virtual representation of a pre-existing subject in digital space. This avatar, under the purview of a gamer or player, lacks an independent existence, existing only within the parameters of the control exerted by the subject. Instead, the focus should be on the avatar's specific agency, which can reach a point where the avatar, as a digital representative, liberates itself from its operator and adopts the form of a figure that functions as its own performer: the secondary takes the place of the primary, the copy the place of the original, and the virtual the place of the real.

The present article aims to trace the process of avatar autonomy by conducting a comprehensive analysis of the music videos produced by the K-pop girl group *aespa*. This analysis will serve to formulate a novel media

theoretical hypothesis, which is outlined as follows: The argument posits that the advent of autonomous avatars in *aespa*'s music videos is not exclusively attributable to content-related justifications stemming from the group's avatar storytelling. Rather, it is firmly embedded in the medial form of the music video itself. The music video, which emerged in the early 1980s on cable television and has since migrated from MTV to YouTube, has long been based on a montage that proceeds synchronously and simultaneously rather than diachronically and linearly. This approach predates the advent of digital special effects and algorithmic AI, which have since become ubiquitous in the field. The playback technique plays a pivotal role in this context, as it liberates the voice from its traditional reliance on visual media, akin to cinema. Instead, it serves to subjugate the image to the voice and the song, fostering an immersive auditory experience. It is yet to be demonstrated that, in contrast to a live television performance, which sustains the spatiotemporal integrity of the performance despite the dissociation of the recorded playback voice from the physical presence of the performer, the music video, through the use of non-linear editing techniques, creates the illusion that a single individual is present in disparate spatiotemporal domains concurrently and manifests in multiple forms, yet vocalizes with a singular voice.

However, it is imperative to first examine *aespa*'s interpretation of the concept of the avatar, which has been incorporated into the group's name as an abbreviation: The South Korean girl group *aespa* was founded in 2020 by SM Entertainment. It is considered the first multiverse group in K-pop history because, depending on the method of calculation, it comprises 4, 8, or 9 members. The four "authentic" members—Giselle, Karina, Ningning, and Winter—are replicated as their respective virtual alter egos, designated "ae Giselle, ae Karina, ae Ningning, and ae Winter." In the *aespa* concept, the "ae" in "ae Giselle, ae Karina, ae Ningning, and ae Winter" functions as an abbreviation for "Avatar Experience." However, it also phonetically approximates the term "AI," suggesting a correlation between the "ae" and "AI." This phonetic similarity prompts the hypothesis that *aespa*'s avatars might be classified as AI avatars. The second syllable, "spa," is derived from the term "aspect." While the four alter ego avatars appear to fulfill a conventional representative function in relation to their presumed originals, this is not the case for the ninth band member. "Naevis" does not function as an avatar, but rather as a virtual female creature devoid of a human original. A comprehensive analysis of *aespa*'s so far 20 music videos reveals that all nine members appear in various combinations. The linguistic congruence between Naevis and the Na'vi from "Avatar" is unmistakable, yet it is noteworthy that Naevis is the sole recipient of a music video for the song "Done," devoid of the involvement of the actual members. In this music video, Naevis occupies a central role, embodying the role of the protagonist.

The *aespa* multiverse incorporates a narrative element reminiscent of fairy tales. Within this narrative, Naevis assumes the role of an artificial intelligence that facilitates the entry of real-life members into the virtual world known as "Kwangya." In this realm, members are able to interact with their avatars, a concept that combines the physical and digital realms. Naevis functions as a medium in two distinct senses. Primarily, it serves as a technical conduit for data streams, referred to as "Synk." Additionally, it operates as an ethereal facilitator, a "navigator," between the domains of reality and virtuality. Naevis is confronted by an antagonistic entity known as the Black Mamba, a colossal serpent that aims to disrupt the medium and systematically interrupts the connection to "Kwangya" with the objective of "desync" *aespa* from their avatars. In *aespa*'s oeuvre, encompassing songs, music videos, and YouTube clips, the confrontation between Naevis and Black Mamba is persistently elaborated upon, thereby achieving a serial franchise effect. This phenomenon is emblematic of the cultural-industrial proliferation of sequels, prequels, and reboots in cinema, television, and streaming. However, the fact that this transmedial "worldbuilding" originates from music is what makes *aespa* unique. "Segyegwan" is the Korean term for this worldbuilding, which musicologist Wonseok Lee identifies in a recent text on *aespa* as a general trend in contemporary K-pop:

"While some dismiss this trend as a marketing gimmick, I argue that it is pivotal to predicting how K-pop will transform in the future. By incorporating the concept of segyegwan, K-pop has become a cultural space in which fans and artists interact beyond the scope of music. As such, K-pop is a prominent example of transmedia storytelling, and the prevalence of segyegwan demonstrates how K-pop challenges the conventional notion of musical genres. (Lee 2026, p. 197)

However, the "segyegwan" can only transcend the generic boundaries of music because it relies primarily on the transmedial nature of the music video before being disseminated through newer digital social media channels. Therefore, the avatar characters of the *aespa* multiverse are programmatically introduced in their inaugural music video, entitled "Black Mamba," which was released in 2020.

The purpose of this segment is to reiterate the previously stated hypothesis that the creation of avatars, which first appear in animated form in "Black Mamba" and subsequently stand on stage alongside the real members in the music video for "Savage" with ontological equality, is not solely due to a transhuman K-pop futurism but rather corresponds to the medial logic of the music video. This logic can be termed "fractality" or "self-similarity," a term coined by Jean Baudrillard. (Baudrillard 1994) The music video employs a fractal mode of image-sound connection, a phenomenon that is particularly evident in *aespa*'s clips. In stark contrast to the classical method of

cinematic montage, the cut in the music video does not mark a spatiotemporal difference between shot A and shot B. In cinema, shots are related to each other in a temporal relationship of before and after and a spatial relationship of here and there. Classical editing techniques, for instance, serve to establish a connection between two shots by means of the suturing of movement and line of sights. This phenomenon has been grammatically codified as "continuity editing" in the context of Hollywood cinema.

In contrast to this linear continuity editing, music videos still feature movement- and line-of-sight-continuity in a purely formal sense. However, shot B does not show the character in a temporal sequence but rather in a state that could be called parallel simultaneity. At the commencement of the "Black Mamba" music video, the sequence rapidly transitions between three distinct virtual locations, with Karina performing impeccable lip-syncing in rapid succession. This creates the illusion that Karina is not merely changing locations and costumes, but rather, is simultaneously present in all locations, exhibiting different versions or aspects of herself. The editing thus engenders the illusion of a multitude of Karinas who could potentially also be in other time periods yet are still synchronized with one another in the here and now.

This synchronization effect is essentially due to playback technology. However, unlike a live performance with playback on stage—as is familiar from traditional television—there is not just one singer lip-syncing to herself, but several singers who are identical to one another. This phenomenon is reminiscent of a song title by Björk that encapsulates the paradox of this fractal multiplication of the same self: "Army of Me" In this sense, the montage does not produce a distinction between a Karina here and a Karina there; rather, it exhibits shifting views of a Karina who has multiplied in parallel simultaneity: An Army of Karina. From a comparative media studies perspective, this simultaneity has less to do with cinematic montage than with the "switch-image" (Engell 2021) of television, which enables the disjunctive switching between different channels that continue to exist in parallel simultaneity even when they are not currently tuned in.

In contrast to the classical parallel montage of cinema, which alternates between two different spatial events and produces the illusion of narrative simultaneity only within the framework of this difference, the montage of the music video creates the impression of a parallel switching of a single event that splits into multiple variants or "aspects." In the context of television, it can be posited that Karina, through fractal replication, is performing the same song simultaneously on disparate television channels. This suggests that each switch is the "aspectualization" of an identical similarity rather than a cinematographic "perspectivization."

In consideration of the previously mentioned simultaneous multiplication of the same entity, one is reminded of a scene in David Lynch's "Lost Highway" in which the Mystery Man confronts the protagonist, played by Bill Pullman, in a face-to-face encounter and requests that he place a call to his own home. As Bill Pullman performs this task, the Mystery Man himself is the one to answer the phone. The concept of the multiverse posits that an individual can exist simultaneously in multiple parallel dimensions. In the case of *aespa*, this parallelism is not merely a marketing strategy but rather is attributed to the media specificity the music video. Historically, music videos have consistently produced avatars without hosts and serial copies without originals. In this context, it is not necessary to inquire about the originality of the various Karinas. These facsimiles do not serve to illuminate a particular perspective of the original; rather, they are originless copies that have supplanted the original, akin to Naevis.

The fractal montage in *aespa*'s music videos functions according to a logic of identical addition, multiplication, and proliferation of simulated simultaneities. For the film and media scholar Steven Shaviro, these are characteristic of the digital composite technologies of the music video in general. The fundamental question that emerges is whether the concept of montage retains its relevance in a context where its capacity to delineate spatiotemporal distinctions is progressively diminished. In his seminal work on digital music videos, Shaviro expounds on the preeminence of synchronous layering over diachronic sequencing, articulating his perspective as follows:

"This use of multiple layers highlights two general tendencies of digital music videos. The first is that digital media tend to present simultaneously, through compositing, the figurations that traditional films used to present sequentially, through associative editing. The horizontal and temporal relationship of one image to the next (montage) gives way to the vertical relationship among disparate elements presented at the same time (collage)." (Shaviro 2017, p. 32)

This transition from horizontal montage to a vertical collage of images, sounds, and text is of paramount importance for the videographic design of the television image and for the data interface of the digital image. The coherence and transparency of the classic cinematic reality effect plays only a secondary role in the music video, partly because the primary image-generation by a reality-recording camera is undermined by a multitude of composite elements: The music video under consideration features a combination of cartoon animation, digital special effects, graphic inserts of various types, live-action segments, and virtual blue and green screen backgrounds. Notably, these elements coexist without any of them asserting a distinct hierarchical status. The music video for "Black Mamba" appears as a condensed compendium of this vertical collage, so that one could

almost speak of a digital ornamentalism. According to Shaviro, the music video is also post-cinematic because the dominance of digital post-production over the reality captured by the camera causes the film theoretical distinction between the inside (diegesis) and the outside (enunciation) to collapse in favor of a total immanence of the image. Shaviro's argument is as follows:

"The second tendency is that digital video tends to collapse the distinction, generally upheld in the movies, between diegetic and nondiegetic elements, between story and narration, between naturalistic representation and self-conscious formal devices, and ultimately between what Christian Metz calls statement (énoncé) and enunciation (énonciation). In all these cases, the two terms are not on an equal footing, because the second is on a higher level (a metalevel) compared to the first. But digital video generally works to break down the hierarchy; different sounds and images are all presented on the same level, regardless of whether they have been actually captured by the audiovisual apparatus or simulated in software." (Shaviro 2017, p. 32)

Thus, according to Shaviro, music videos exemplify two significant phenomena. Firstly, they demonstrate the fractal neutralization of temporal markers. Secondly, they illustrate the leveling of the difference between the diegetic enunciate and the nondiegetic enunciation. A salient illustration of this phenomenon is evident in the persistent reframing and deframing exhibited in *aespa's* videos. This dynamic process entails the transformation of what is initially perceived as primary reality into a secondary image-within-an-image. Conversely, it involves the de-diegetization of secondary images-within-an-image through the process of deframing.

In the first scenario, a purported depiction of *aespa* is metamorphosed into a cellular telephone image, which is perceived by *aespa* themselves rather than by an external observer. In the second scenario, a cellular telephone image of *aespa*, as perceived by *aespa* themselves, is transmuted into an authentic representation of *aespa*. The fractal structure of the music video is formed by recursive images, which contain themselves as images (a concept known in art history as "mise-en-abyme"). It is imperative to distinguish between this recursivity and reflexivity, as these phenomena involve self-referential practices rather than self-reflexivity, a point that Shaviro underscores in his subsequent study on the music video, titled "The Rhythm-Image":

"More generally, the *mise-en-abyme* illustrates, and indeed embodies, the paradoxes of recursivity and self-reference that tend to arise in many realms of modern thought and practice (set theory, cybernetics, aesthetics, and so on). Today, the networks within which we find ourselves contained are commonly said to exhibit fractal self-similarity across scales." (Shaviro 2022, p. 102)

In the music video, the constant recordings—unlike the formal interventions of modernist cinema—do not function as self-reflexive figurations that call into question the image's relation to reality. Conversely, these devices, imbued with a sense of playfulness, serve to obfuscate the epistemological distinction between image and reality. The image, as depicted in the music video, is elevated to a state of equivalence with reality, thereby suggesting that reality itself can be regarded as nothing more than a manifestation of image.

The fractalization of cell phone images in both cases is indicative of a broader trend in contemporary society, particularly in Korea, where the digital age has progressed at an accelerated pace. In this rapidly evolving environment, the lines between appearance and essence are increasingly short-circuited, as images proliferate rapidly through social media and other digital platforms. Finally, it is important to return to the avatar in the end: The short-circuiting of image and reality signifies the avatar's transition from a mere stand-in to an algorithmic actor, endowing it with a novel ontological consistency. In specific moments within *aespa's* music videos, the avatar literally transforms the real members into its derivatives, thereby exemplifying the dynamic interplay between human and artificial identities in contemporary pop culture.

At the conclusion of the video for "Forever," the avatars, as it were, allegorically advance to become the primary enunciators of the diegetic world. A sudden reverse zoom reveals the seemingly real stage as the interior of a digital snow globe. Within this digital environment, the four members of *aespa* find themselves encapsulated, observed by their animated avatars from the outside. Through this fractal recursion, the avatar is elevated to an algorithmic subject that has long since emancipated itself from its anthropomorphic hosts.

### References

- Baudrillard, J. (1994). *Simulacra and simulation* (S. F. Glaser, Trans.). University of Michigan Press.
- Engell, L. (2021). *The Switch Image. Television Philosophy*. (Anthony Enns, Trans.). Bloomsbury.
- Lee, Wonseok (2025): "Next Level" aespa's Storyworld and Cultural Significance. In: Miller, Nicholas E. (Ed.): *Critical Approaches to Transmedia Storytelling in K-Pop*. Bloomsbury, pp. 197-214.
- Shaviro, S. (2017). *Digital music videos*. Rutgers.
- Shaviro, S. (2022). *The Rhythm-Image: Music Videos and New Audiovisual Forms*. Bloomsbury.