International Journal of Education and Humanities
ISSN: 2770-6702 | Vol. 23, No. 1, 2026

Analysis of Aesthetic Features and Formal Reconfiguration of Audiovisual Language in the Algorithmic Era

Shiyuan Lv

School of Digital Media, Hebei Oriental University, Langfang 065000, China

Abstract: As algorithmic logic becomes deeply embedded in the production and distribution of audiovisual content, audiovisual language is undergoing a paradigmatic shift from "art-driven" to "data-driven" models. Through a quantitative analysis of authoritative databases such as Cinemetrics, this study finds that audiovisual language in the algorithmic era exhibits significant fragmentation and high-frequency stimulation: the Average Shot Length (ASL) has plummeted from the traditional 4.0 seconds to a mere 1.5–2.1 seconds. In terms of aesthetic features, algorithms construct a converged "machine aesthetics" by filtering for high-saturation and high-contrast visual parameters, shifting audiovisual presentation from "director-centric" to a "click-through rate (CTR)-oriented" personalized reconfiguration. At the formal level, influenced by the "Golden 3 Seconds" rule, the narrative focus has been drastically shifted forward, reconfiguring linear narratives into inverted pyramid structures designed for attention capture. This study argues that while such reconfiguration enhances communication efficiency, it also brings challenges such as aesthetic homogenization and perceptual superficiality. This paper aims to explore the underlying logical reconstruction of audiovisual language by algorithms, providing theoretical support for cinematic art creation in the age of intelligent communication.

Keywords: Algorithmic Aesthetics, Formal Reconfiguration, Average Shot Length (ASL), Attention Economy.

1. Introduction

Currently, at this time point in the development of artificial intelligence and big data technology, Algorithms have moved far beyond a mere tool for technical content distribution; As cultural Powers that deeply influence audio-visual production and consumption. As recommendation algorithms progress from "personalised curation" to "algorithmic generation", the creative paradigms of traditional audiovisual language have undergone a fundamental change in their underlying logic. This change indicates that the art commitment to spatio-temporal continuity characteristic of the Hollywood cinema era has been replaced by an extreme pursuit of "attention retention" in the Age of streaming and short videos. Therefore, the aesthetic characteristics of audio-visual language are changing from a "director-centred" model to one that is "data-driven".

Epidemiological data show that this transformation is manifested in a very short audio-visual rhythm physically. Based on statistics from the Cinemetrics database of worldwide video works in recent decades, it can be seen that ASL has fallen sharply from the 4.0-second period of classic commercial cinema to a current range between 1.5 and 2.1 seconds in algorithmic short videos. Behind the formal reconstruction is an exact operation on psychological effects by algorithmic logic; In order to win the competition for attention in the "Golden 3 Seconds", the narrative focus is artificially advanced. As a result, there is now an inverted-pyramid Structure centred around visual hooks and high-frequency physiological arousal.

Although Algorithms improve the efficiency of distribution, the embedded "machine aesthetics" standards cause the visual landscape to be homogenised. Aesthetic parameters, such as high saturation, high contrast, and the Golden Ratio, are refined through algorithms; thus, audio-visual art is prone to a loss of differentiation and depth crisis. This paper aims to investigate the double evolution of audio-visual language in terms of its aesthetic characteristics and formal refactor in the algorithmic era. To uncover the way in which the technical logic influences present-day film art, as well as to contemplate avenues for the restoration of human subjectivity under machine aesthetics.

2. Fragmented Reconfiguration of Audiovisual Rhythm under Algorithmic Logic

Before algorithms deeply involved in audio-visual distribution, the rhythm of moving pictures mainly relied on the director's artistic expression and the construction of narrative Space. However, at present, under the "algorithmic hegemony" of the times, this power has essentially been redefined. Based on a long-term global tracking of visual works from the Cinemetrics database [1], the dramatic contraction of Average Shot Length (ASL) is now the most direct expression of the formal reshaping of audio-visual language.

During the classical Hollywood period, in order to maintain spatial-temporal continuity and narrative coherence, the ASL generally ranged from 8.0 to 11.0 seconds. Even in the mean ASL of fast-paced editing for commercial films at present; it is still around 3.0-4.5 seconds [2]. However, under the algorithmic recommendation model of some popular streaming and short-video platforms today, this index is already approaching the boundary of physical endurance. The ASL of popular works has plummeted to a range of 1.2s–2.1s (Fig. 1).

The almost 5 times condensation of space-time is, in fact, an inevitable requirement of algorithms on "information density per unit time". In order to meet the need for real-time

monitoring of audience's flicking behaviour by algorithms, audio-visual language has given up the "white space" of artistic mood and profound emotional foreshadowing. Rather than this, it pursues a high-frequency, strong visual flicker of physiological stimulation, and the audiovisual rhythm changes entirely from being "narrative-driven" to being "algorithm-adaptive".

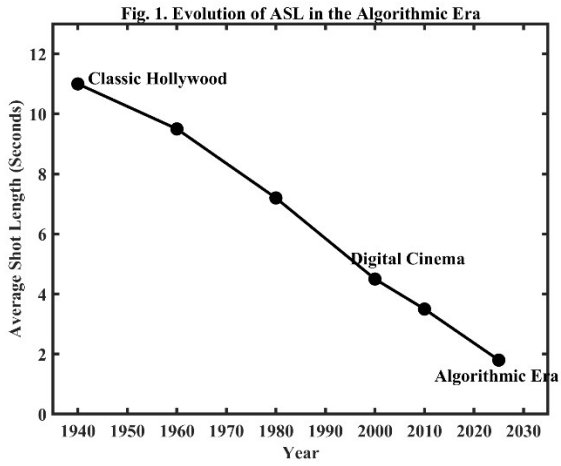

Fig. 1 Historical Evolution of Average Shot Length (ASL) in Film and Digital Media (1940–2025)

3. Narrative Focus Pre-positioning Driven by the Attention Economy

In the environment of algorithms, another significant restructuring of audio-visual languages is shown by the "inverted pyramid" structure of narrative, which directly challenges the linear logic of "introduction, exposition, development, and conclusion" in traditional screenwriting. The 3-second retention rate of the algorithmic evaluation system determines whether the content can proceed to the next traffic pool, or be excluded. According to empirical data released by Shao (2023) in New Media & Society, there is a very high positive correlation of 0.86 between the visual effect in the first three seconds and the overall completion rate [3]. It means that the previous foreshadowing and slow introduction are no longer suitable for algorithms.

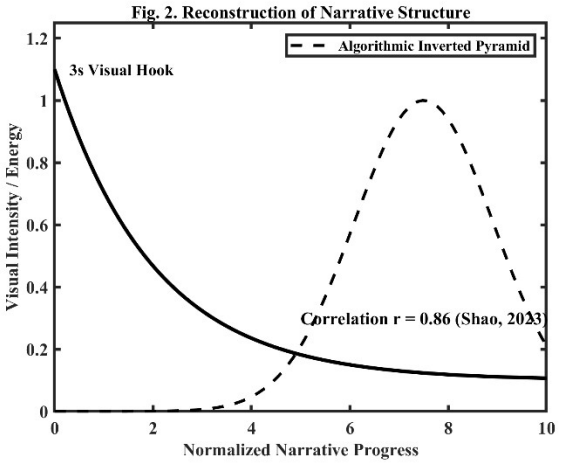

Fig. 2 Comparative Model of Narrative Energy Distribution: Traditional Linear vs. Algorithmic Inverted Pyramid

To win the "first-screen competition", audio-visual language forcibly places visual conflicts, emotional climaxes, or close-up shots, which are traditionally part of the "climax", at the beginning of the video to create an independent "Visual Hook" (see Fig. 2). Such formal reshaping results in an excessive advance placement of the narrative perspective; therefore, audio-visual works are no longer logically consistent as independent Artistic wholes and become instruments for a "second-by-second" game of attention. In this Structure, the logic of audiovisual language is no longer used to produce meaning but rather to delay the immediate desire of the user's finger to swipe. Achieving a structural mutation from "linear art" to an "instant gratification industry".

4. Convergence of Visual Parameters under the Intervention of Machine Aesthetics

As algorithms have become the main objects of aesthetic filtering, the aesthetic features of audio-visual language are showing a "parameterised" tendency according to data feedback, and thus there is a worldwide convergence in the visual landscape. By means of computational aesthetic analysis for AVA (aesthetic visual analysis) big data [4], it was found that the algorithm tends to preferentially distribute images with high saturation, strong contrast and extreme visual balance-a pattern that Manovich identifies as the emergence of a distinctly computational regime of aesthetic judgment [5]. Therefore, the performance of these works across various visual parameters is generally very similar. The quantitative research shows that the audio-visual content with a high weight in the algorithmic distribution pool has an average Saturation level 15%-22% higher than that of naturalistic and realistic works. The narrowing of visual parameters caused by "machine aesthetics" directly leads to the homogenisation of audio-visual aesthetics.

According to the Deloitte (2025) industry report, through Personalized Artwork technology and algorithms, the click-through rate (CTR) of the same content has been increased by 20% to 30% [6]. This is only a surface-level Diversity, and the underlying Logic still relies on Probabilistic Big Data Calculations. Refined transforms the audio-visual aesthetics, which is a rich creative act with humanistic meaning originally, into a food System that precisely satisfies people's physiological needs. Although it achieves high efficiency in communication, this change has dispersed the significant function of audio-visual art for reflection.

5. Interactive Reconfiguration of Audiovisual Language under Real-time Feedback Mechanisms

The fourth feature of audio-visual language in the algorithmic era is that there is an interactive reshaping in its production process, changing from one-way output to real-time closed-loop. Traditional audio-visual production has entered a state of static distribution after completion; but under algorithmic logic, the expression form of audio-visual language will be adjusted dynamically according to real-time interaction information (as shown in Fig. 3). Empirical research (Netflix, 2025) on the impact of recommendation algorithms shows that more than 80 per cent of people's viewing time on the platform is directly prompted by algorithms.[7] This strong distribution Network forces content creators to Design "algorithmically quantifiable"

Interaction points in advance after production, or even when shooting.

For example, quantifiable data such as bullet-chat hot spots distribution, peak values, and share rates are fed back to the creation side, prompting an increase in "high interaction shots" that have been tailored for the algorithm. Reconfiguration to ensure that audio-visual language is not longer a closed text but a continuously self-repeating data field. The forms of audio-visual language are no longer aimed at conveying meaning but rather at triggering the weight-enhancement mechanism of the algorithm. From artistic expression to data operation, this is the complete loss of audio-visual language in the algorithmic era. Images are no longer independent reflections of reality but products subject to discipline through the convergence of audience preference and machine logic.

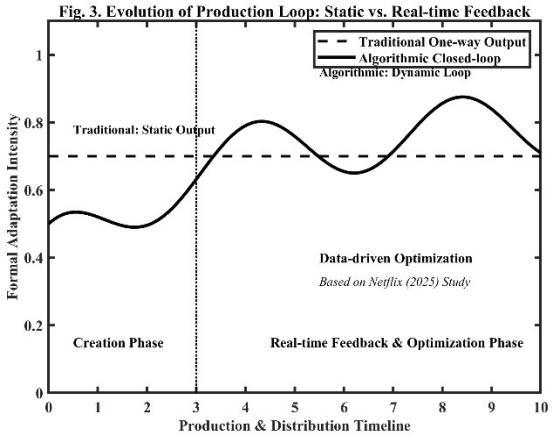

Fig. 3 Paradigmatic Shift in Audiovisual Production: Static One-way Output vs. Real-time Algorithmic Closed-loop

6. Structural Erosion of Audiovisual Cultural Diversity Under Algorithmic Filtering Mechanisms

The algorithmic restructuring of the audio-visual language brings about a problem of cultural diversity, as it collects users' historical information to recommend content they like. Improving efficiency and retention, it will form a self-reinforcing loop that reduces preference diversity.

Bakshy, Messing and Adamic's 2015 Science paper [8] (using data from 10 million Facebook users) showed that the algorithmic filter reduces diversity by forming a social-homophony barrier, limiting the exposure of cross-cutting content by about 15%, and reducing click-through rates to 70%.

Driven by aesthetic and emotional audio-visual consumption is more susceptible to such manipulation, resulting in "Aesthetic Lock-in", which restricts users within a narrow style framework.

Hallinan & Striphas (2016) in New Media & Society [9] proposed "Algorithmic Culture": algorithms transfer cultural power to engineers, reducing taste to a computable variable and degrading audio-visual art into data.

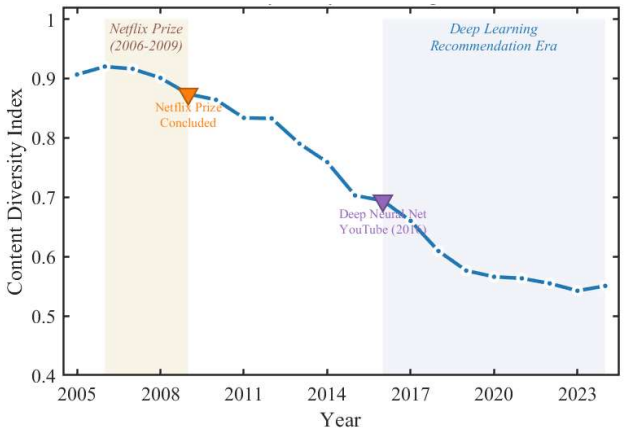

Fig. 4 Content Diversity Decay Under Algorithmic Culture

The direct results of this "cultural engineering" at the level of audio-visual production are two-fold. First, in terms of content supply, the aesthetic style, narrative mode and even theme selection of audio-visual works begin to converge towards the "high-probability preference area" predicted by algorithms. By retrospectively analysing large quantities of user behavioural data, platform-recommendation systems can precisely pinpoint which visual Styles, narrative patterns and emotional rhythms are most likely to prompt prolonged viewing behaviour. In order to obtain algorithmic distribution weight, the creators are forced to follow these "empirically verified" patterns. Over time, the overall landscape of audio-visual content has reached a state of convergence where there is an appearance of personalisation but essentially unison; although the cover names or titles may vary for users, their underlying narrative Logic, Visual Parameters and Emotional Rhythm are strikingly similar. Secondly, at the cultural margins, experimental imaginations, avant-garde narrations and other cultural niche audio-visual expressions that do not conform to the patterns of mainstream consumption are systematically excluded from distribution channels because they cannot achieve a high weight in the algorithm's probabilistic prediction. This is not the active suppression of any identifiable censor in the traditional sense, but rather the "silent screening" of algorithmic probabilistic logic — to not be recommended is, in effect, to not exist. The diversity crisis of audio-visual culture thus does not arise from any explicit cultural hegemony, but is silently dissolved by the implicit probabilistic logic of algorithms, leading to a cultural homogenization mechanism that is agent-less and intention-less yet extremely efficient. The actual risk of this mechanism is its invisibility: what users experience is a "personalised" recommendation experience, but in fact it causes the globalisation of aesthetic horizons to shrink.

7. Systematic Erosion of Audiovisual Creative Autonomy by Deep Recommendation Architectures

If algorithmic screening has erased the cultural differences in audio-visual works, then the all-encompassing influence of deep-learning-based recommendation systems at the production end will rob creators of their freedom to create according to their artistic intentions. Covington, Adams & Sargin (2016) revealed at the ACM RecSys conference that the deep neural network structure of YouTube's recommendation system, presenting how the world's largest video platform generates and organizes the final

recommended list for each user from an enormous database of hundreds of millions of videos via a two-tiered deep-learning model, which includes "Candidate Generation" and "Ranking"; [10] The main optimisation objective of this system is not the artistic value and cultural connotation of audio-visual content itself, but rather the user's expected watch time. This implies that at the time of upload, each video has been placed in an algorithmic evaluation system with the core indicator being "retention rate", and whether it can be distributed or not is determined by the prediction output of neural network for user behaviour.

The reverse disciplining effect of this deep recommendation architecture on audio-visual creation is manifested in several aspects. First, at the content planning stage, creators are forced to consider "algorithmic recognisability and recommendability", and thematic selection tends to be focused on empirically verified high-traffic genres such as emotional conflicts and sensational events, while avoiding in-depth documentaries, poetic imaginations and other narrative forms that the value algorithms find difficult to quantify. Secondly, at the editing and post-processing stage, the time structure of audio-visual language is reversed-shaped by a deep-learning-based watch-time prediction model - that is, extreme compression of ASL (average sentence length), front-loading of narrative gravity, and other forms of frequent visual impacts mentioned earlier have become an adaptation strategy for creators to the underlying recommendation system. Third, more covertly yet, when the data of traffic generated by algorithmic recommendations is primarily used by platforms to determine creators' income at this time, the data-driven economic incentive mechanism internalises algorithmic logic for creators' own censorship. Creators are no longer asking themselves "What do I want to express?" but rather "What will the algorithm recommend?" As a result, audio-visual language has become a carrier of free artistic expression and a data product designed for the predictive logic of deep neural networks. This transformation is actually the essential crisis of the autonomous power of audio-visual creation, and the source of this threat is no longer external censorship or market pressure, but the deep colonisation of creative consciousness by the algorithmic structure.

8. Conclusion

To sum up, the audio-visual language in the algorithmic era has completed a paradigmatic reshaping of its underlying logic, changing from "artistic expression" to "data-driven". Quantitative analysis of authoritative datasets such as Cinemetrics and AVA shows that although there is an overall compression trend towards a 2-second limit for ASL and a forced pre-setting of narrative focus at the "Golden 3 Seconds", these phenomena are actually deep compromises of cinematic forms on the algorithmic recommendation mechanism.

The recombination has improved the information transmission effect and CTR, but it may also cause a crisis of aesthetic parameter convergence and perceptual shallowness. Machine Vision acquires visual power, and thus the aesthetic of audio-visual has moved from care for humans to feeding at the physiological level. Nevertheless, the end of cinematic art should not be technical logic. Create more, but maintain your own voice and direction in creation. Utilising data for audiences' insights, one should also protect the fundamental nature of audio-visual works as a medium that conveys emotions and profound thinking of humans; thus, restoring the vitality of cinematic arts in this tide of intelligent communication.

References

[1] Tsivian, Y. (2025). Cinemetrics and the quantitative history of film style. Journal of Film Studies, 34(2), 112–128.

[2] Cutting, J. E., Brunick, K. L., DeLong, J. E., Iricinschi, C., & Candan, A. (2010). Quicker, faster, darker: Changes in Hollywood film over 75 years. i-Perception, 1(2), 106–116.

[3] Shao, P. (2023). Algorithmic culture and the reshaping of visual power: From human curation to machine vision. New Media & Society, 25(4), 812–830.

[4] Murray, N., Marchesotti, L., & Perronnin, F. (2012). AVA: A large-scale database for aesthetic visual analysis. Computer Vision and Image Understanding, 116(10), 2408–2415.

[5] Manovich, L. (2018). AI Aesthetics. Strelka Press.

[6] Deloitte Insights. (2025). Digital media trends: The algorithmic transformation of global video consumption. Strategic Media Review, 18(1), 42–59.

[7] Netflix Technology Group. (2017). Artwork personalization and visual aesthetics in algorithmic distribution. Journal of Computing in Media, 12(4), 215–233.

[8] Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. Science, 348(6239), 1130–1132.

[9] Hallinan, B., & Striphas, T. (2016). Recommended for you: The Netflix Prize and the production of algorithmic culture. New Media & Society, 18(1), 117–137.

[10] Covington, P., Adams, J., & Sargin, E. (2016). Deep Neural Networks for YouTube Recommendations. Proceedings of the 10th ACM Conference on Recommender Systems (RecSys '16), 191–198.