Real-Time Neuro-Augmented Cinema via Generative AI

Antoine Bellemare-Pepin[1,*]
Philipp Thölke[1*]
Yann Harel[1]
Karim Jerbi[1,2,3]

1 CoCo Lab, Department of Psychology, Université de Montréal

2 Mila (Quebec AI research Institute), Montreal, QC, Canada

3 UNIQUE Center (Quebec Neuro-AI research Center), QC, Canada

Abstract

In this paper, we present a novel system that integrates real-time neurofeedback into the creative process of generative AI, enabling seamless interactions between users and AI systems. By leveraging the user's cognitive variability, the system allows for continuous and fluid co-creation, moving beyond the traditional prompt-based interactions common in generative AI workflows. We achieve this using electroencephalography (EEG) to continuously monitor the user's brain activity, which then acts as a control signal for a visual generative AI model. We focus specifically on Lempel-Ziv complexity, a measure of signal diversity that have previously been associated with mental states, task engagement and phenomenological richness. The proposed architecture includes an EEG feature extractor and a generative AI pipeline, working in tandem to dynamically alter the visual content of a pre-existing movie based on the user's brain activity. This approach offers a new dimension of complexity and complicity in the interaction between humans and AI. Future work will explore the integration of more sophisticated bio-signals and multi-modal feedback, aiming to further enhance the depth and richness of the embodied creative experience. This work serves as a proof of principle for integrating biotechnology and generative AI in the emerging field of adaptive cinema. A playlist with video illustrations of the system in action can be found at youtube.com/playlist?list=PLMu36WzSQKiVeBnrUdwUAoUhqLqGX3_bw.

[*]Equal Contribution

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

1 Introduction

The rapid development and proliferation of generative AI models have led to the creation of content with impressive quality, often indistinguishable from human-made outputs (Samo and Highhouse, 2023). Despite these progresses, the means through which human users guide and act on generative AI systems remains largely constrained. Typically, users interact with these models through a process of iterative tweaking of text prompts and generation parameters in order to achieve the desired output, alternating between commanding (prompting) and receiving (e.g. watching) generated material. An alternative approach is to continuously provide the generative model with data streams controlled by the user, allowing for fluid and intuitive interactions that doesn't rely on deliberate manipulation of textual prompts (Menon, 2024). Recent studies have demonstrated how generative AI can constitute a tool for the exploration of altered perceptual states (Suzuki et al., 2017), promote embodied learning (Memarian and Doleck, 2024), as well as enhance user engagement through content personalization (Yu et al., 2024; Bag et al., 2022; Thölke et al., 2024). In this paper, we present a system architecture aimed at advancing the state of embodied and co-creative interfaces by steering the output of a visual generative AI system using biosignals, specifically electroencephalography (EEG). This system leads to a more integrated form of interaction, bypassing the continual reliance on textual prompting by enabling users to influence AI-generated content through their cognitive and emotional states in real-time.

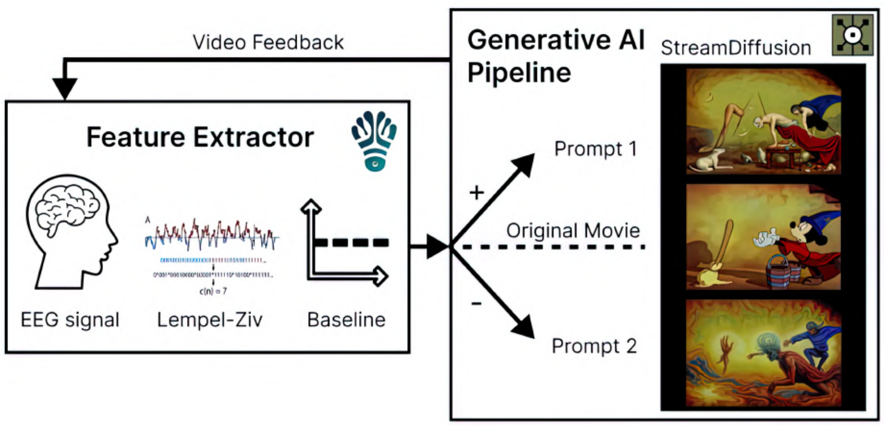

Figure 1: Architecture of the Neuro-Augmented Cinema pipeline showing the EEG feature extractor and generative AI modules, highlighting the continuous user feedback loop.

2 System Design

The co-creative system we propose consists of two main components: an EEG feature extractor and a generative AI pipeline. The EEG feature extractor processes the user's brain signals to extract meaningful features, which are then used to control the generative AI pipeline. The AI system is designed for a single user, who enters into a feedback loop with the AI to augment or modify the appearance of an existing movie. The system operates in real-time with low latency, allowing for a smooth and continuous interaction between the user and the AI.

2.1 EEG Feature Extractor

The EEG feature extractor is implemented using *goofi-pipe*, an open-source framework that we designed for real-time processing of biosignals (github.com/PhilippThoelke/goofi-pipe) and which is able to handle a wide variety of EEG devices and other input sources. *goofi-pipe* receives data that can be streamed from any EEG device through LSL. The signal first undergoes preprocessing (filtering and artifact removal) to ensure data integrity and improve the signal-to-noise ratio. Following preprocessing, Lempel-Ziv complexity (Lempel and Ziv, 1976) is calculated, which quantifies the signal's diversity based on its algorithmic compressibility. Lempel-Ziv complexity correlated well with subjective experience during testing but can easily be replaced or augmented by other common measures, such as spectral power or aperiodic slope. The extracted features are normalized (z-transform) based on the individual's running baseline levels, allowing for a personalized and adaptive interaction with the AI system. Specifically, the normalization process results in values where negative indicates levels lower than baseline, zero corresponds to the baseline, and positive values represent levels higher than baseline. EEG features are updated at a user-adjustable sampling frequency, defaulting to 30 Hz. The normalized features are then transmitted to the generative AI pipeline via Open Sound Control (OSC) protocol, facilitating real-time interaction.

2.2 Generative AI Pipeline

The generative AI pipeline is implemented in *TouchDesigner*, a visual programming environment designed for real-time graphics and multimedia. In this setup, a predetermined movie is played and movie frames are sent to *StreamDiffusion* (Kodaira et al., 2023). In conjunction with *Stable Diffusion (SD) Turbo*, a distilled SD model enabling high-quality images from just 1-4 diffusion steps (Sauer et al., 2023), we augment the movie frames in real-time with a diffusion-based image-to-image paradigm. The AI system is pre-configured by the user with two prompts that are attuned to the movie's content and capture complementary emotional states. The normalized EEG Lempel-Ziv complexity determines which prompt is applied and the intensity of the effect. Negative complexity values apply the first prompt to modify the visuals, zero leaves the movie unchanged, and positive values trigger the second prompt. The effect's strength scales continuously with EEG complexity.

3 Discussion

The proposed system represents an early and simple demonstration towards embodied human-machine co-creation, which combines real-time neurofeedback and generative AI. By injecting the user's brain activity into the creative process in real-time, the system enables a more fluid interaction that does not require to alternate between commanding and receiving modes. Instead, the user can engage in a continuous, flow-like experience where their mental state directly influences the AI-generated content. While the current implementation is limited in the depth of the embodied experience, particularly due to the simplicity of the EEG features used, the system is designed in a modular fashion that facilitates the use of more advanced and potentially AI-driven bio-features. Currently, the system utilizes only a single dimension of the user's mental state, which is far from capturing the richness and complexity of human cognition and emotion. Despite these limitations, this system demonstrates the potential for more natural and embodied interfaces with generative AI, where the individual feels like an integral part of the creative process rather than an external commander. A playlist with video recordings of the system in action can be found at youtube.com/playlist?list=PLMu36WzSQKiVeBnrUdwUAoUhqLqGX3_bw.

3.1 Broader Impact

The broader aim of this research is not to replace current means of interacting with such systems in a creative way, but to augment this process with a sense of embodiment and synergy. The system, akin to a real-time remixing tool, reinterprets existing media based on fluctuations of one's mental state. This work serves as a proof of principle for integrating biotechnology and generative AI in the emerging field of adaptive cinema. This integration has the potential to drive the evolution of adaptive media experiences, enabling more personalized, immersive, and responsive content that dynamically adjusts to the viewer's mental and emotional state in real-time.

3.2 Future Work

Future work will focus on enhancing the system's capabilities by incorporating more advanced brain features or other bio-signals. It is likely that the quickly advancing field of AI applied to brain data will lead to highly capable brain decoders, reaching increasingly high levels of embodied experience. Additionally, the image prompts could be guided dynamically by the user's state, not just to reflect their current emotions but to advance or broaden the experienced narrative. Expanding the system to other modalities, such as audio, is another promising direction for further research. These advancements will contribute to developing more sophisticated and natural co-creative systems that can more fully integrate the user's subjective experience into the generative process.

References

Bag, S., Srivastava, G., Bashir, M. M. A., Kumari, S., Giannakis, M., and Chowdhury, A. H. (2022). Journey of customers in this digital era: Understanding the role of artificial intelligence technologies in user engagement and conversion. *Benchmarking: An International Journal*, 29(7):2074–2098. Publisher: Emerald Publishing Limited.

Kodaira, A., Xu, C., Hazama, T., Yoshimoto, T., Ohno, K., Mitsuhori, S., Sugano, S., Cho, H., Liu, Z., and Keutzer, K. (2023). StreamDiffusion: A Pipeline-level Solution for Real-time Interactive Generation. arXiv:2312.12491 [cs].

Lempel, A. and Ziv, J. (1976). On the complexity of finite sequences. *IEEE Transactions on information theory*, 22(1):75–81.

Memarian, B. and Doleck, T. (2024). Embodied AI in education: A review on the body, environment, and mind. *Education and Information Technologies*, 29(1):895–916.

Menon, A. (2024). From Mind to Machine: An Embodied Approach to Image Creation with Generative AI. Master's thesis, OCAD University.

Samo, A. and Highhouse, S. (2023). Artificial intelligence and art: Identifying the aesthetic judgment factors that distinguish human- and machine-generated artwork. *Psychology of Aesthetics, Creativity, and the Arts*, pages No Pagination Specified–No Pagination Specified. Place: US Publisher: Educational Publishing Foundation.

Sauer, A., Lorenz, D., Blattmann, A., and Rombach, R. (2023). Adversarial Diffusion Distillation. arXiv:2311.17042 [cs].

Suzuki, K., Roseboom, W., Schwartzman, D. J., and Seth, A. K. (2017). A deep-dream virtual reality platform for studying altered perceptual phenomenology. *Scientific reports*, 7(1):1–11.

Thölke, P., Bellemare-Pepin, A., Harel, Y., Lespinasse, F., and Jerbi, K. (2024). Bio-Mechanical Poet: An Immersive Audiovisual Playground for Brain Signals and Generative AI. In *Proceedings of the 15th International Conference on Computational Creativity*.

Yu, J., Dickinger, A., So, K. K. F., and Egger, R. (2024). Artificial intelligence-generated virtual influencer: Examining the effects of emotional display on user engagement. *Journal of Retailing and Consumer Services*, 76:103560.

A Supplementary Material

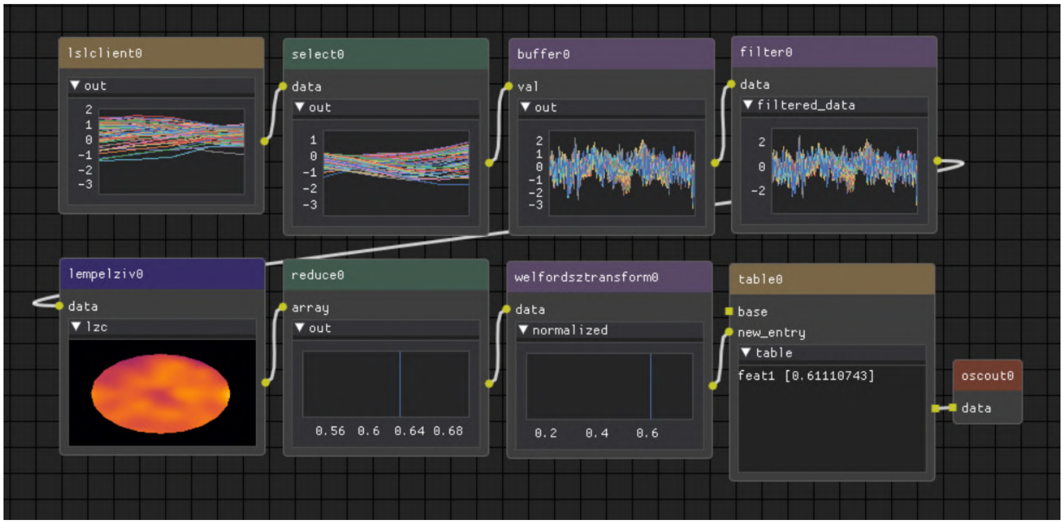

Figure A.1: Screenshot of the *EEG Feature Extractor* implemented inside *goofi-pipe*. Lempel-Ziv complexity is continuously computed on a 3s buffer of filtered EEG data, and then sent to the *Generative AI Pipeline* via OSC.

Figure A.2: Screenshot of the *Generative AI Pipeline* implemented inside *TouchDesigner*. The EEG-derived control value is used to steer the image-to-image diffusion model, which augments the frames of a movie in real-time.