

break me, genAI

MAYANK SANGANERIA, Independent, USA

KURT JAMES WERNER, SoundToys, USA

Additional Key Words and Phrases: generative AI, Music perception, synesthesia, midi, diffusion

ACM Reference Format:

Mayank Sanganeria and Kurt James Werner. 2024. break me, genAI. 1, 1 (December 2024), 3 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 PROGRAM NOTES

The act of listening to music is a profoundly complex experience, shaped by myriad factors including cultural background, context, environment, and personal emotional state [1]. As we journey through a piece, our perception of its elements evolves; a bass line that initially captures our attention may, after multiple repetitions, recede into the background of our consciousness [2, 9].

In the realm of audiovisual art, acknowledging this dynamic nature of musical perception offers an opportunity to forge deeper connections between sound and sight [3]. While recent years have seen a proliferation of works utilizing Music Information Retrieval (MIR) to create visuals responsive to audio [4, 11], these approaches often fall short in capturing the nuanced, context-dependent nature of our listening experience. The result can be a visual uniformity that fails to distinguish between vastly different musical styles.

The emergence of generative AI has opened new possibilities for creative audio visualization, ranging from fully automated systems [10] to interactive approaches that maintain significant human involvement [12]. "break me, generative ai" embraces the latter approach, implementing a multi-layered system for audiovisual creation that preserves artistic intent throughout the generation process. As shown in Figure 1, the piece begins with hand-drawn MIDI parameter curves mapping abstract perceptual qualities like "grittiness" and "rattle". These curves drive procedurally generated visuals inspired by the synesthetic works of Kandinsky and Klee [5], establishing a foundation deeply rooted in the artist's subjective experience of the music.

This initial layer undergoes further transformation through a series of AI-driven processes, as illustrated in Figure 2. Select segments are processed through StyleGAN [6], introducing machine-learned aesthetics (Figure 2b). The entire video is then run through a video-to-video model using Stable Diffusion [7] (Figure 2d) and AnimateDiff [8], guided by prompts echoing the original visual inspirations.

The resulting artwork emerges as a tapestry of ambiguity, challenging viewers to consider the relative influence of each layer on the final output. This layered approach raises important questions about artistic authorship and preservation: How much of the artist's original perception persists through the AI processing? To what extent do the influences of Kandinsky and Klee manifest in the procedural generation versus the diffusion models? Could any frames be mistaken for works by these master artists?

Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). Copyright remains with the author(s).

DOI: <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

Music Proceedings of the International Conference on New Interfaces for Musical Expression

Audiovisual Symposium 2024, 6 December 2024, Dalarna, Sweden

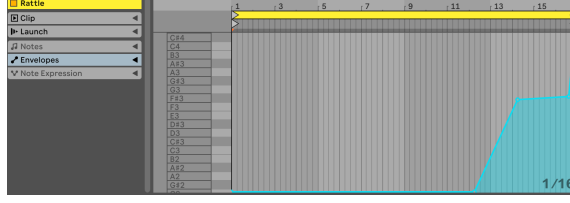


Fig. 1. An example of the manually drawn MIDI curve mapping perceptual qualities to parameters

"break me, generative ai" invites us to explore the frontiers of human-AI collaboration in art. Building on recent advances in both music visualization [11] and generative AI systems [12], the piece demonstrates how combining human perception with machine learning capabilities can create audiovisual experiences that more faithfully reflect the complex, ever-changing nature of musical perception. As you experience this piece, we encourage you to embrace the ambiguity, recognizing patterns while acknowledging the elusive nature of definitive answers in this blend of human creativity and machine-generated art.

2 PROJECT DESCRIPTION

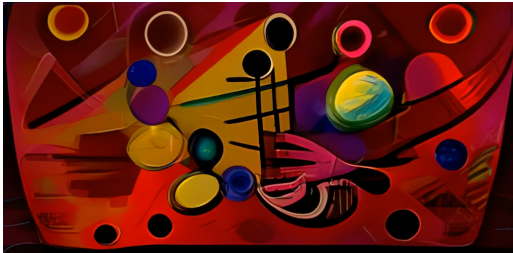
"break me, generative ai" is an audiovisual artwork that explores the complex nature of musical perception and its representation through visual means. Building on recent work in generative music visualization [10, 12], the piece combines hand-drawn MIDI parameter curves, procedurally generated visuals inspired by Kandinsky and Klee, and AI-driven processes to create a multi-layered visual experience.



(a) Procedurally generated frame



(b) StyleGAN transformation



(c) DiscoDiffusion transformation



(d) StableDiffusion transformation

Fig. 2. Progressive transformation of a single frame through different AI models

Key components of the project include:

- Hand-drawn MIDI curves mapping abstract perceptual qualities like "grittiness" and "rattle", as shown in Figure 1
- Procedurally generated visuals based on these curves (Figure 2a). See Video 1 in Media Links.
- AI processing using StyleGAN [6] on select segments (Figure 2b). See Video 2 in Media Links.
- Further transformation using DiscoDiffusion [13] (Figure 2c), Stable Diffusion [7] (Figure 2d) and AnimateDiff [8], guided by prompts echoing the original visual inspirations. See Video 3 in Media Links.
- Composing the final video through various shader effects applied through TouchDesigner [14]. See Final Video in Media Links.

The project aims to address limitations in current Music Information Retrieval (MIR) approaches to audiovisual art [11] by incorporating subjective human perception and machine learning capabilities. The result is a tapestry of ambiguity that challenges viewers to consider the relative influences of human creativity and AI processing in the final artwork.

3 PERFORMANCE NOTES

This is a fixed media audiovisual piece.

4 MEDIA LINKS

- Video 1 (Procedural Generation): <https://vimeo.com/326360052>
- Video 2 (StyleGAN Processing): <https://vimeo.com/395905172>
- Video 3 (DiscoDiffusion): <https://vimeo.com/922942605>
- Final Video: <https://vimeo.com/994689901>

5 ACKNOWLEDGMENTS

The authors would like to thank Stochastic Labs for their support of the creative community and enabling the knowledge transfer that enabled the progress on this piece for the past 2 years. We would also like to thank the contributors to Processing, openFrameworks, Disco Diffusion, Stable Diffusion and AnimateDiff.

REFERENCES

- [1] Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. MIT press.
- [2] Margulis, E. H. (2014). *On repeat: How music plays the mind*. Oxford University Press.
- [3] Cytowic, R. E., & Eagleman, D. M. (2009). *Wednesday is indigo blue: Discovering the brain of synesthesia*. MIT press.
- [4] Müller, M. (2015). *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications*. Springer.
- [5] van Campen, C. (2010). *The Hidden Sense: Synesthesia in Art and Science*. MIT Press.
- [6] Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4401-4410).
- [7] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 10684-10695).
- [8] Guo, Z., et al. (2023). AnimateDiff: Animate Your Personalized Text-to-Image Diffusion Models without Specific Training. *arXiv preprint arXiv:2307.04725*.
- [9] Braganca, G. F. F., Fonseca, J. G. M., & Caramelli, P. (2015). Synesthesia and music perception. *Dementia & neuropsychologia*, 9:16-23.
- [10] Ng, B. M., et al. (2024). Visualize Music Using Generative Arts. In *2024 IEEE Conference on Artificial Intelligence (CAI)*.
- [11] Lima, H., Santos, C., & Meiguins, B. (2021). A Survey of Music Visualization Techniques. *ACM Computing Surveys*, 54(7), Article 143.
- [12] Liu, V., Long, T., Raw, N., & Chilton, L. (2023). Generative Disco: Text-to-Video Generation for Music Visualization. *arXiv preprint arXiv:2304.08551*.
- [13] Crowson, K. (2022). Disco Diffusion. <https://github.com/alembics/disco-diffusion>
- [14] Derivative. (2024). TouchDesigner 2024.11180. <https://derivative.ca/>