

The Allegory of the Cave

Thomas Boulousis

Department of Performing Arts
Kingston University London
thomasboulousis@mail.com

Abstract

Contemporary music generation tools connect music and text through written prompts, opening up new avenues for interrogating literary-inspired musical narrative. *The Allegory of the Cave*, a programmatic composition for solo violin, explores this dynamic by adapting Plato’s famous allegory into music. The composition is divided into five movements, alternating between being written solely by me and being created with input from Meta Research’s MusicGen. The system is prompted with recordings of two principal themes representing the cave and the outside world, alongside descriptive text derived from the allegory by an LLM, to generate short audio fragments. These are curated, transcribed, and transformed into the second and fourth movements, which are in turn anticipated musically by the surrounding movements. Using an older model with less stylistic consistency resulted in interesting and unexpected yet less refined output, compelling me to move from a curatorial approach to a more actively interpretative role.

1 CONTEXT

The Allegory of the Cave is a programmatic composition for solo violin which combines music and literary storytelling with AI-generated material by adapting Plato’s famous allegory. Music and storytelling have a long, shared history that has prompted varied discourse about music’s narrative potential (Almén, 2008; Antović, 2009; Kregor, 2015; Zbikowski, 2008). Programmatic music engages directly with musical storytelling by portraying non-musical material—typically literature—without words (Kregor, 2015; Scruton, 2001). Contemporary music generation tools can also produce music influenced by literature through prompts. With this composition, I set out to explore what this possibility holds for programmatic music.

2 COMPOSITION PROCESS

I divided Plato’s allegorical narrative into five parts, each corresponding to a movement of the composition: the introduction to the cave, the cave’s strange shadows, the prisoner’s escape, the outside world, and the return to the cave. Movements two and four were developed with direct input from Meta Research’s MusicGen (Copet et al., 2024), and the others were composed solely by me in response. I created and recorded two principal melodic themes representing the cave and the outside world (figures 1 and 2). To create musical coherence, I made the two themes appear similar at first and separated them through contrasting musical treatment. I formed isolated vignettes with the cave motif and extended the outside world motif into longer melodic lines.



Figure 1: Cave motif



Figure 2: Outside-world motif

In parallel, I used a locally run Minstral 8B language model (Liu et al., 2026) to generate descriptive text detailing shadows seen by prisoners within the cave and the objects encountered outside (table 1). These descriptions, alongside the recorded motifs, served as prompts for MusicGen to generate sixty 8-second audio samples (available [here](#)). Despite ostensible connections to concrete descriptions, these samples lacked typical meaning-making devices such as the imitation of real-world sounds (Castelões, 2009) or culturally mediated tropes (Antović, 2009). I thus treated them as interchangeable variations on generic shadows or objects.

Table 1: Examples of LLM-generated descriptive text

Shadows	Earthly Objects
A slithering serpent-like form.	A stately oak tree with twisted roots.
A shadowy ballerina in mid-leap.	A delicate crystal obelisk.

I then curated and transcribed selected samples to build the primary musical substance of the second and fourth movements, which depict the prisoner’s perceptions inside and outside the cave. Afterwards, I expanded the material to create a coherent musical dramaturgy. I employed the same musical material in the non-AI movements and used those movements to foreshadow the

musical ideas explored in the AI-mediated second and fourth movements.

3 DISCUSSION

The composition process of the second and fourth movements was conceived as predominantly curatorial, where the selection and transcription of audio samples would involve as few changes as possible. I would compose a movement, MusicGen would respond with its own, and I would then react with another movement. However, as List (1974) observed, adapting audio into symbolic notation always involves ambiguity. In my case, even though MusicGen produced mostly convincing violin-like gestures, the results included uncertain rhythms and pitches and were often impossible to play by a single violin. Thus, rather than strictly copying the samples, I imitated their perceived intent by, for example, replacing impossible double stops with similar alternatives and mapping microtonal pitches onto a mostly 12-tone equal temperament framework.

Furthermore, the specific version of MusicGen I used (musicgen-melody) is a relatively limited model that frequently generates artefacts such as noise, inconsistent structure, and stylistic indeterminacy (Zhao et al., 2025). Rather than lamenting them, I viewed these deficiencies as a licence to freely interpret the generated sound and experiment with structure. I used this freedom to organise the movements so they more closely align with the structure of the allegory's narrative, for example by following Plato's move from reflections, to physical objects, and finally to the sun, through the three contrasting sections of the fourth movement.

In the end, I intervened more than I would have if I had adhered to my original plan. Transcribing generated audio for solo violin was challenging in itself, and adding more instruments would have been even more demanding. The samples only vaguely linked to the prompts, requiring significant intervention. Larger models might have presented a different challenge by generating formulaic tropes drawn from dominant Western musical datasets (Mehta et al., 2025), though exploring their possibilities could also be worthwhile.

In negotiating knowledge, Plato's *Allegory of the Cave* has much to say about its acquisition; crucially, that genuine learning is an active process of transformation and practice (Plato, 1997, *Republic*, 514a–541b). By enabling effortless creation, generative AI offers a tempting alternative to traditional composition that, when passively over-relied upon, could cause certain skills to atrophy (e.g., Zhai et al., 2024). Nevertheless, it seems that achieving the best results involves critically evaluating AI output, leveraging its unpredictability, and ultimately exerting just as much effort to make the process productive.

4 PROGRAM NOTES

Deep inside a dark cave, strange prisoners are chained so that they can only see shadows cast on a wall by a fire behind them. One prisoner is freed and, though initially blinded by the firelight, slowly adapts. Emerging outside, he learns to understand reflections, earthly objects, the night sky, and finally the sun itself. Wanting to share this truth, he returns to the cave, but the others reject the revelation and ultimately kill him.

Thus unfolds the narrative of Plato's famous allegory, which invites us to ponder the nature of knowledge and education. This story serves as the source material for this programmatic composition for solo violin. I used a music AI tool prompted with recordings of two melodies representing the cave and the outside world, alongside AI-generated text describing shadows on the cave walls or entities encountered outside. The AI system produced short audio fragments that I curated, transcribed, and transformed, to form the musical substance of the work. In this way, I explored how generative AI can mediate between text and music by enabling new ways of engaging with literary narrative.

ACKNOWLEDGMENTS

This work is supported by the UK Technē Arts and Humanities Research Council Doctoral Training Partnership [grant number AH/R01275X/1].

AUTHOR DECLARATIONS

For the paper, I used Microsoft Copilot to review grammar and syntax. For the composition, I used Meta Research's MusicGen and Mistral AI's Ministral 3 8B as outlined. MusicGen was trained on licensed music recordings.

REFERENCES

- Almén, B. (2008). *A theory of musical narrative*. Indiana University Press. <https://doi.org/10.2307/j.ctt2005szf>
- Antović, M. (2009). Towards the semantics of music: The twentieth century. *Language & History*, 52(1), 119–129. <https://doi.org/10.1179/175975309X452003>
- Bribitzer-Stull, M. (2015). *Understanding the leitmotif: From Wagner to Hollywood film music*. Cambridge University Press. <https://doi.org/10.1017/CBO9781316161678>
- Castelões, L. E. (2009). A Catalogue of Music Onomatopoeia. *International Review of the Aesthetics and Sociology of Music*, 40(2), 299–347. <http://www.jstor.org/stable/20696544>
- Copet, J., Kreuk, F., Gat, I., Remez, T., Kant, D., Synnaeve, G., Adi, Y., & Défossez, A. (2024). *Simple and controllable music generation* (Version 3). arXiv. <https://doi.org/10.48550/arXiv.2306.05284>
- Hugging Face. (2023). *MusicGen*. https://huggingface.co/docs/transformers/en/model_d oc/musicgen
- Hugging Face. (2024). *MusicGen Melody*. https://huggingface.co/docs/transformers/en/model_d oc/musicgen_melody
- Kregor, J. (2015). *Program music*. Cambridge University Press.
- Liu, A. H., Khandelwal, K., Subramanian, S., Jouault, V., Rastogi, A., Sadé, A., Jeffares, A., Jiang, A., Cahill, A., Gavaudan, A., Sablayrolles, A., Héliou, A., You, A., Ehrenberg, A., Lo, A., Eliseev, A., Calvi, A., Sooriyarachchi, A., Bout, B., Rozière, B., ... Robert, T. (2026). *Ministral 3* (Version 1). arXiv. <https://doi.org/10.48550/arXiv.2601.08584>
- List, G. (1974). The Reliability of Transcription. *Ethnomusicology*, 18(3), 353–377. <https://doi.org/10.2307/850519>

- Mehta, A., Chauhan, S., & Choudhury, M. (2025). *Missing melodies: AI music generation and its “nearly” complete omission of the Global South* (Version 3). arXiv.
<https://doi.org/10.48550/arXiv.2412.04100>
- Mistral AI. (2026). *Ministral-3-8B-Instruct-2512*. Hugging Face.
<https://huggingface.co/mistralai/Ministral-3-8B-Instruct-2512>
- Plato (1997). *Plato: Complete Works*. Edited by J.M. Cooper and D.S. Hutchinson. Indianapolis: Hackett Publishing Company
- Scruton, R. (2001). Programme music. In *Grove Music Online*. Oxford University Press.
<https://doi.org/10.1093/gmo/9781561592630.article.22394>
- Zbikowski, L. M. (2008). Metaphor and music. In R. W. Gibbs Jr. (Ed.), *The Cambridge handbook of metaphor and thought*, 502–524. Cambridge University Press.
<https://doi.org/10.1017/CBO9780511816802.030>
- Zhai, C., Wibowo, S. and Li, L.D. (2024) ‘The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review’, *Smart Learning Environments*, 11, 28.
<https://doi.org/10.1186/s40561-024-00316-7>
- Zhao, Y., Yang, M., Lin, Y., Zhang, X., Shi, F., Wang, Z., Ding, J., & Ning, H. (2025). AI-Enabled Text-to-Music Generation: A Comprehensive Review of Methods, Frameworks, and Future Directions. *Electronics*, 14(6), 1197.
<https://doi.org/10.3390/electronics14061197>