

“The Terms Remain” (A Fixed Audio-Visual Composition)

Bob L. T. Sturm

Division of Speech, Music and Hearing
School of Electrical Engineering and Computer Science
KTH Royal Institute of Technology, Stockholm, Sweden
bobs@kth.se

Abstract

Generative AI technologies, such as large language models and image and music synthesis models, are being made broadly accessible as online platforms sporting simple interfaces and cheap subscription plans. Each, however, comes with a *Terms of Service* (ToS) that is not so broadly interpretable. My audio-visual composition provides an accessible interpretation of the ToS of a specific commercial prompt-based AI music generation platform.

1 Method of Composition

1.1 Text/Libretto

I first render the 8,846-word Terms of Service (ToS) of the prompt-based AI music platform Udio¹ as a PDF. I upload the PDF as context to OpenAI’s ChatGPT platform (model 4o, paid license),² and then prompt: “Explain each section of these terms of service in a way that only the modern American composer Robert Ashley can understand.” The chatbot processes the PDF and prompt, and outputs text that I copy to construct a libretto. An example output is shown below.

ChatGPT 4o ▾

2. Who May Use the Services

(a soft telephone filter is applied to the voice now)

You must be old enough to lie. Not just any lie, but the kind that sticks.
Thirteen. That’s the number. If you’re not, go back to sleep.
If you’re between, say, thirteen and whatever makes you legal, you need someone older
to nod on your behalf.
Someone with a driver’s license. Or a memory of failure.
If that’s not you, you don’t belong here.

I repeat this process about ten times, curating the lines I like for each of the ten sections of the ToS, and create a libretto from the materials (which ends up being one tenth the size of the ToS).

1.2 Voice/Narration

¹ <https://www.udio.com>, last accessed June 16, 2025. A prompt-based AI music platform is an online application that serves music content generated from free-form textual descriptions entered by users, e.g., “Uptempo stadium rock song about love and hotdogs with funky bass guitar and boogie-woogie piano”. At this time there is no guarantee that the generated music will match the given description.

² The OpenAI ChatGPT platform (<https://chatgpt.com>, last accessed June 16, 2025) is an online application that interacts with a user through a free-form text interface. Using next-token prediction, it simulates chatting with a person. It has been trained on massive amounts of text, e.g., billions of pages on the world wide web.

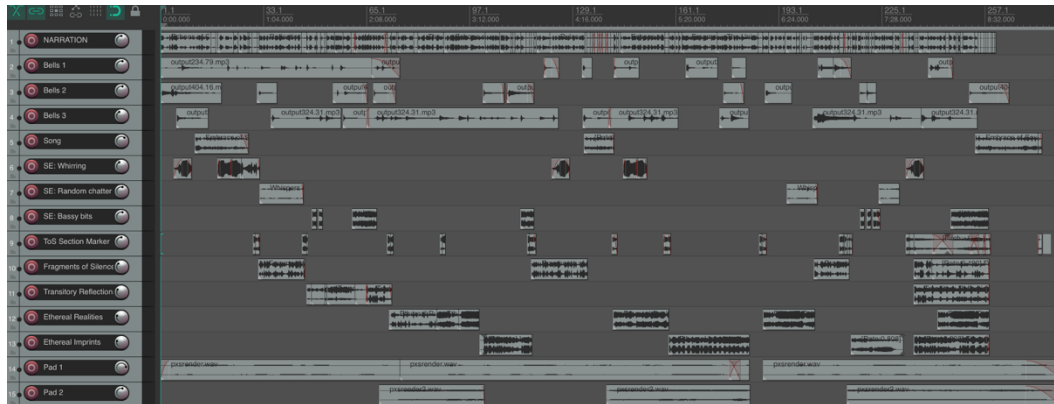
I extract a 38-second portion of an online interview of Robert Ashley,³ and upload it as context to be extended by Udio (paid license) for another 32 seconds (this is a setting of Udio’s “v1.5 Allegro” model). I use the text-prompt “Interview, Thomas Buckner, Perfect Lives, Robert Ashley”, and copy-paste portions of my libretto into the “lyrics” field. Udio generates two sound files for each request. I either download one that I find successful or ask Udio to generate again. I repeat this process for the entire libretto to assemble the recorded narration.

1.3 Audio/Music

I prompt Udio to generate music audio material for the work. A song appearing at various times is generated by Udio from the prompt: “humming machine, mechanic, fans, soundscape, field recording”. Other sound materials are generated by Udio from the prompt: “Robert Ashley Perfect Lives Blue Gene Tyranny Thomas Buckner Improvement El/Aficionado Atalanta Acts of God experimental electroacoustic quiet sparse music of changes john cage”. Each section of the ToS is demarcated by sound material generated by Udio from the prompt: “noise music, robust tupperwares, commercial”.

1.4 Composition

I assemble the audio materials (41 sound files) using a digital audio workstation (Reaper, paid license), applying effects and other transformations as I want, e.g., multiband compression, pitch shift, reverberation, and ducking. The image below shows the fifteen tracks of this project.

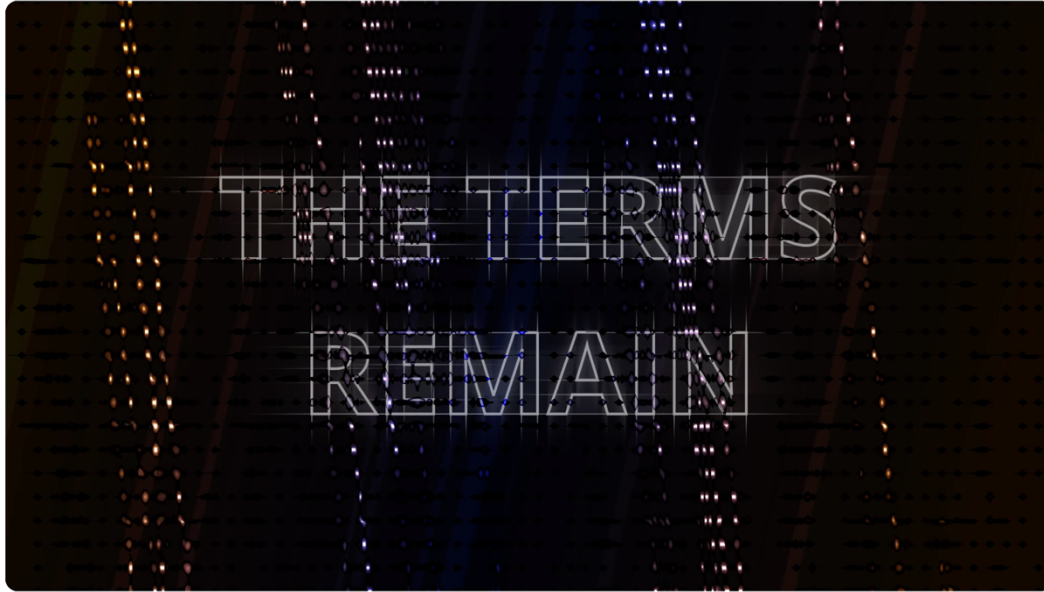


I build the piece around the narration in Track 1. Tracks 2-4 are synthetic bell-like sounds I generate from the recorded narration using a custom program. Track 5 contributes portions of the song. Tracks 6-8 contribute some sound effects. Track 9 contributes sonic demarcations for the beginning of each of the 10 sections of the ToS. Tracks 10-13 contain the sonic materials for each of the ToS sections. Finally, Tracks 14 and 15 contribute background drones I create by excessively time stretching some of the music generated by Udio.

1.5 Video

I use Davinci Resolve (paid license) to create a video accompanying the audio. The video demarcates each section of the ToS and highlights specific phrases of the narrator. A still of the video is shown below.

³ “Robert Ashley: You Can't Call It Anything Else But Opera: Robert Ashley at home in conversation with Frank J. Oteri, March 13, 2001”, <https://youtu.be/Ib-vMUbddEM>, last accessed June 16 2025.



2 Reflections

Robert Ashley (1930-2014) has had a significant impact on my musical aesthetics since I became familiar in 1995 with his opera *Improvement: Don Leaves Linda*. He speaks like people I know (Midwest USA), assembles layers of text and subtext wonderfully, and has a unique style that I enjoy. So, I wondered, how might he interpret the ToS of Udio? Then I wondered: How well can ChatGPT and Udio mimic Ashley’s inimitable voice? This audio-visual work answers that question. I find the output of ChatGPT to provide a surprisingly good imitation of Ashley. I can recognize in the generated text things I would expect him to say. The narration generated by Udio sounds like him speaking in an interview, but is missing a few subtle aspects of his characteristic “drawl”. The music generated by Udio is by and large quite different from his own, which is essentially slowly unfolding serial-composed minimalism accompanied by amplified voices. Perhaps some portions of the generated music are reminiscent of *In Sara, Mencken, Christ and Beethoven There Were Men and Women*—but that music was actually created by Paul DeMarinis. Nonetheless, my goal of this work was not to create music that sounds like it was made by Ashley. I consciously avoided replicating aspects of his style, such as a serial approach to composition, a chorus, call and response, and his idiosyncratic delivery of text. (I do use drones, however.)

(None of the text or ideas in this document were created by or with an LLM.)

Acknowledgments

This work is an outcome of a project that has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (MUSAiC, Grant Agreement No. 864189).