

# Title: (un)Stable (dis)Connection

MISAGH AZIMI, Victoria University of Wellington

## 1. PROGRAM NOTES

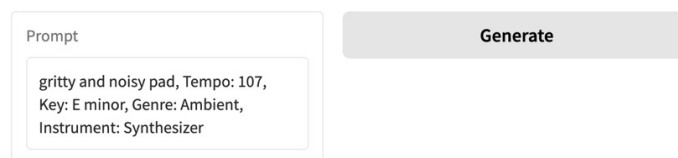
This improvisation piece explores the integration of a fine-tuned text-to-audio latent diffusion model into a real-time electronic music performance environment. By framing Generative AI (GenAI) as a co-creative agent, I combine the system's unpredictable sonic output with traditional improvisation methods in Ableton Live. The result is a live performance that highlights the unique sonic possibilities of AI-generated audio fragments shaped through spontaneous interaction.

The AI model autonomously generates sonic ideas (fragments) upon receiving prompts in real-time, while the performer integrates, loops, and manipulates them on the spot. The interplay yields a sonic world rich with unpredictable textures—ideal for ambient, drone, or other experimental styles of contemporary electronic music.

## 2. PROJECT DESCRIPTION

In this live performance, the musician inputs short text prompts into a fine-tuned AI model that instantly generates new audio generates. These fragments are then imported into Ableton Live, where the performer manipulates them in real time—producing an evolving, co-created soundscape. Fine-tuning an open-source text-to-audio latent diffusion model [1] on the performer's own recordings aims for stylistic coherency while preserving creative freedom. Rather than treating AI as a mere style imitator, this approach frames it as a genuine creative partner that fuels spontaneous collaboration.

The system architecture follows a three-step process. First, the performer enters a short descriptive prompt into a simple UI (Fig. 1).



Prompt

gritty and noisy pad, Tempo: 107,  
Key: E minor, Genre: Ambient,  
Instrument: Synthesizer

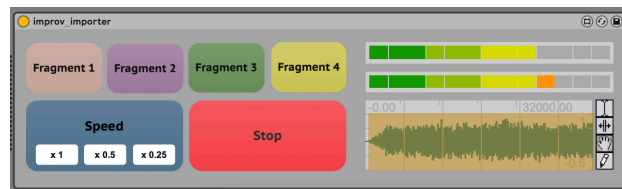
Generate

Fig. 1. A simple Gradio UI to enter prompts

This text is then fed to the fine-tuned text-to-audio diffusion model, which generates a stereo fragment based on the provided description. Immediately after

generation, a lightweight Python script trims any silence, ensuring the audio is ready to be used. Once processed, the fragment is automatically imported into Ableton Live via a custom Max for Live device (Fig. 2).

Fig. 2. Custom Max patch to buffer and play fragments



A MIDI controller (Fig. 3) is used to trigger playback, apply effects, and manipulate loops in real time, allowing the artist to layer, stretch, and remix the newly created material with minimal latency.

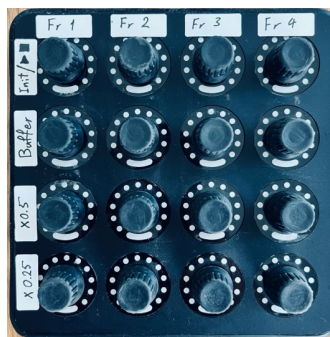


Fig. 3. MIDI mapped parameters for haptic interaction

### 3. PERFORMANCE NOTES

This piece runs between five and ten minutes and can be performed in either stereo or quadraphonic format, depending on the available venue setup—both options are fully supported. The performer will bring a laptop running Ableton Live, an audio interface, and DI-boxes; hence, no additional audio equipment is required from the venue beyond the main speakers and XLR cables. A stable internet connection is needed to facilitate real-time text prompts and audio generation in Google Colab. If the conference would like the audience to observe the AI prompting and inference process during the performance, an HDMI connection to a projector or screen should be provided.

### 4. MEDIA LINK(S)

- Video and Audio: <https://misaghazimi.com/research/sao-improv>

## ETHICAL STANDARDS

The model used in this case study was trained exclusively on Creative Commons–licensed audio (CC-0, CC-BY, or CC-Sampling+). The dataset employed for fine-tuning the model consists solely of the author's intellectual property, ensuring that all materials are ethically sourced.

## REFERENCES

- [1] Zach Evans, Julian D. Parker, C. J. Carr, Zack Zukowski, Josiah Taylor, and Jordi Pons. 2024. Stable Audio Open. <https://doi.org/10.48550/arXiv.2407.14358>