

Glitching the Algorithm: Reclaiming Performative Agency through Real-Time Distilled AI and Experimental Fine-Tuning

Chenxi Bao
DynamiX
Beijing, China
cloudingcxb17@gmail.com

Yuting Wu
Individual
Beijing, China
yutingwu@outlook.com

Baisen Wang
DynamiX
Beijing, China
wbs2788@gmail.com

Abstract

Current deep learning audio models are frequently perceived as autonomous generators that threaten to replace musicians, largely due to their offline, high-latency nature and tendency to produce aesthetically "correct" but sterile music. This paper presents a novel performative framework that transforms the ACE Step 1.5 model into a real-time, playable digital instrument. By employing model distillation and a streaming generation architecture, we achieve sufficiently low latency for live performance. Furthermore, to counter the aesthetic homogenization of AI music, we trained personalized LoRA models on the authors' own experimental tracks. During live performance, musicians navigate this personalized latent space using three continuous parameters: prompt interpolation, temperature, and Classifier-Free Guidance (CFG) which are dynamically morphing between granular glitch textures and ambient soundscapes, pushing the model toward creative unpredictability rather than algorithmic perfection.

1 REVIEW CONTEXTUALIZING THE PRACTICE FIELD

Recent advancements in deep learning have catalyzed a revolution in text-to-audio generation. Latent diffusion models and auto-regressive transformers, such as MusicGen, Stable audio, and ACE-Step, have demonstrated remarkable capabilities in synthesizing high-fidelity music from textual prompts. However, within the context of live musical performance and Human-Computer Interaction (HCI), the current landscape reveals a critical bottleneck: these systems are fundamentally designed as offline, non-causal generators. The inference process requires significant computational time, resulting in high latency that renders them incompatible with real-time stage use. Consequently, musicians are forced into a "prompt-and-wait" workflow, stripped of the temporal agency required to actively perform or physically intervene during the sound synthesis process.

Beyond technical constraints, the field faces a significant aesthetic challenge. The massive datasets used to pre-train these foundational models predominantly consist of conventional, commercially structured music. As a result, their latent spaces are heavily biased toward generating harmonically "correct," polished, and highly homogenized outputs. While impressive for general consumer applications, this predictability is antithetical to experimental and avant-garde electronic music practices. For artists working with glitch, noise, and non-idiomatic sounds, standard AI generators often produce results that are technically flawless but artistically sterile, while experimental genres rely on continuous and microscopic manipulation of timbre rather than standard chord progressions. Therefore, our practice is situated at the intersection of these two unresolved challenges: overcoming the latency barrier to transform an AI model into a playable instrument, and resisting algorithmic homogenization to reclaim a personalized, experimental sonic aesthetic.

2 METHODS

To transform the generative AI from an offline algorithm into a playable stage instrument, our methodology integrates model optimization, personalized fine-tuning, and interactive interface design.

2.1 Real-time Engine: Distillation and Streaming

High inference latency is a major barrier to using AI in live performance. We chose the ACE Step 1.5 model as our foundation. Generating high-fidelity audio normally takes at least 8 denoising steps with this model, making it too slow for live interaction. We applied model distillation to compress this process and reduced the requirement to a single inference step. We then built a streaming generation pipeline. The model now processes and outputs continuous audio chunks rather than waiting for a full track to render. This creates an uninterrupted, real-time audio stream fast enough for stage use. This immediate auditory feedback is essential for maintaining rhythmic synchronization and allowing the musician to react naturally to the generated sound.

2.2 Aesthetic Personalization via LoRA

We used Low-Rank Adaptation (LoRA) to fine-tune the distilled model and avoid the predictable aesthetic of commercial AI generators. The training data consisted entirely of the two authors' original experimental tracks. Training the model on our own sonic signatures like non-idiomatic textures, noise, and glitch helped steer it away from standard musicality. This custom LoRA ensures the real-time output reflects our personal styles. It transforms the AI from a generic tool into an extension of our own artistic identity.

2.3 Performative Interface and High-Dimensional Control

We built a custom graphical user interface to make the system performable. It exposes the model's architecture

as a set of real-time controls. The performer interacts with three primary hyperparameters. First, a slider controls the prompt interpolation ratio. The system generates audio based on two distinct text prompts, such as "sharp rhythmic clicks" versus "cold ambient textures". The performer can crossfade between these prompts to morph the sound dynamically. Second, a dial controls temperature, which dictates the volatility and intensity of the music. Finally, we repurposed the Classifier-Free Guidance (CFG) scale as a tool to balance chaos and order. Altering the CFG intensity changes the binding strength between the text prompt and the audio. High CFG forces the model to obey the prompt, while lowering it pushes the system into unpredictable sonic hallucinations and experimental breakdown. These interactive dimensions allow the performer to actively play the latent space and reclaim creative agency. We mapped these software controls to physical MIDI hardware. Turning a physical knob directly alters the generation process, translating digital mathematics into tactile musical gestures.