DAWZY: A New Addition to AI powered "Human in the Loop" Music Co-creation

Aaron C Elkins
San Diego State University

Sanchit Singh
San Diego State University

Adrian Kieback
San Diego State University

Sawyer Blankenship
San Diego State University

Uyiosa Philip
San Diego State University

Aman Chadha
San Diego State University

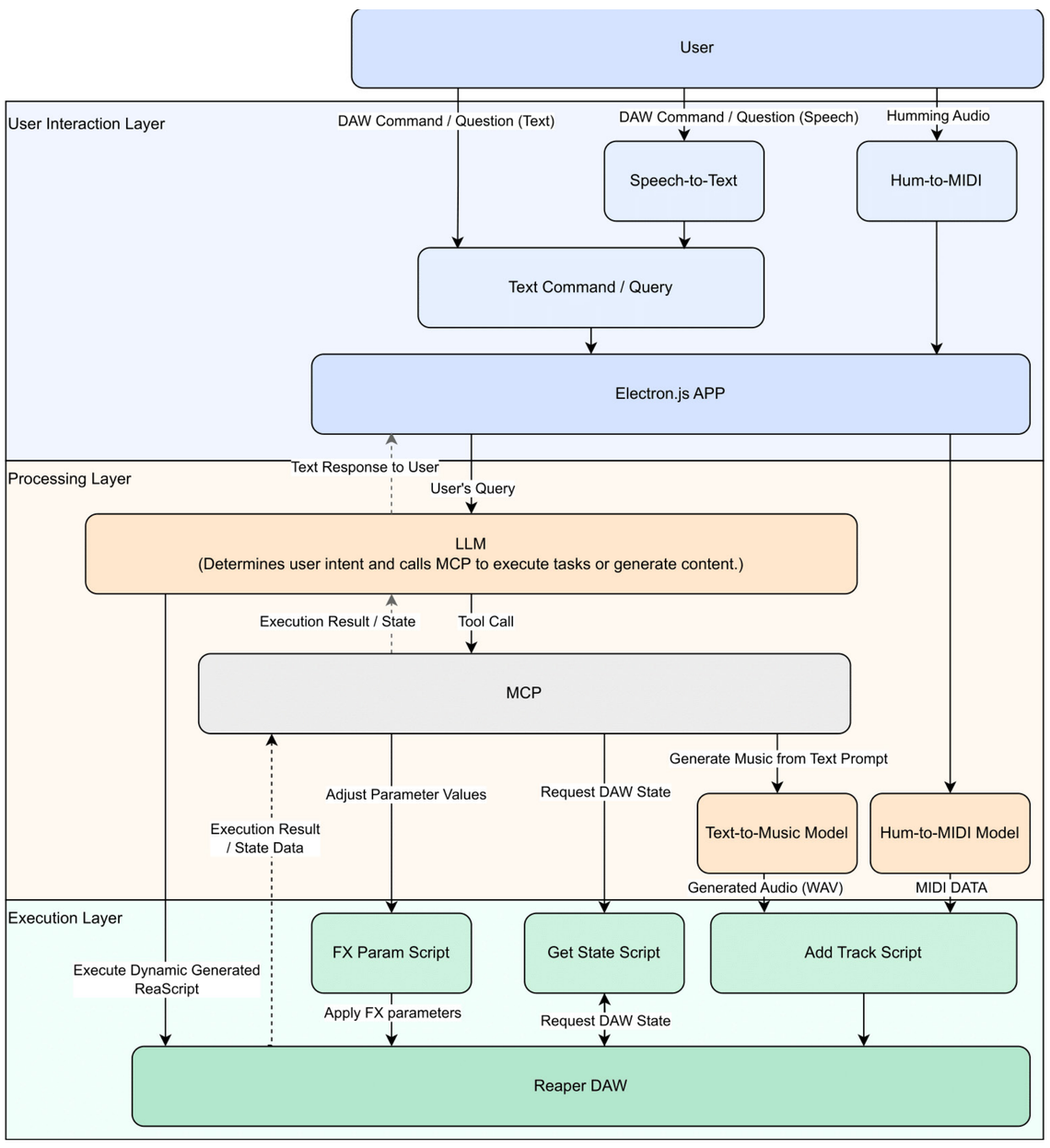
1. Motivation

Modern music production relies on **Digital Audio Workstations (DAWs)** that offer fine-grained control but **overwhelm users** with complex interfaces and technical steps. Translating creative intent—such as "*make the vocals warmer*" —into low-level edits interrupts artistic flow and requires deep domain knowledge.

DAWZY bridges this gap by turning **natural-language input (text, voice, or humming) into precise, reversible DAW actions**. Instead of navigating hundreds of buttons and sliders, users express intent directly and receive interpretable results with full control and transparency.

DAWZY preserves the DAW as the creative hub, supports **human-in-the-loop interaction**, and reduces the technical barrier for music producers at all experience levels.

2. System Overview



DAWZY consists of **three layers**:

- **User Interaction Layer**: Electron.js interface that captures **text, speech** (transcribed via Whisper), and **humming** (converted to MIDI via BasicPitch).
- **Processing Layer**: A **GPT-5-based LLM** interprets intent, **calls** Model Context Protocol **(MCP) tools** to **query session state, adjust FX parameters,** or **generate audio** via MusicGen.
- **Execution Layer**: Executes **safe, reversible ReaScripts** in REAPER through ReaPy, ensuring **grounded edits and undo functionality**.

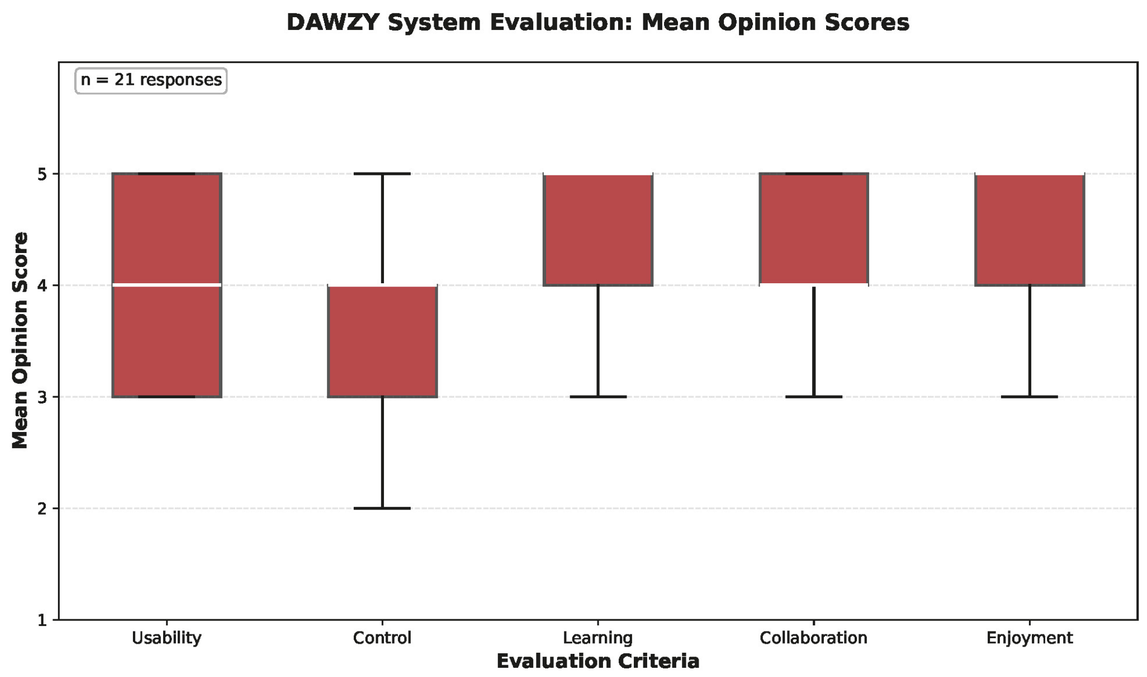
This architecture enables an **adaptive workflow** where users can **create, modify, and learn**—without losing creative control.

3. Evaluation

	DAWZY				Ableton-MCP
Task	GPT-5	Qwen 480B	GPT-OSS 120B	Claude 4.5 Sonnet	Claude
Wavy	3	2	2	3	0
Bouncy	2	1	1	2	2
Fade	2	1	2	1	0
Success Rate (out of 9)	78%	44%	56%	67%	22%

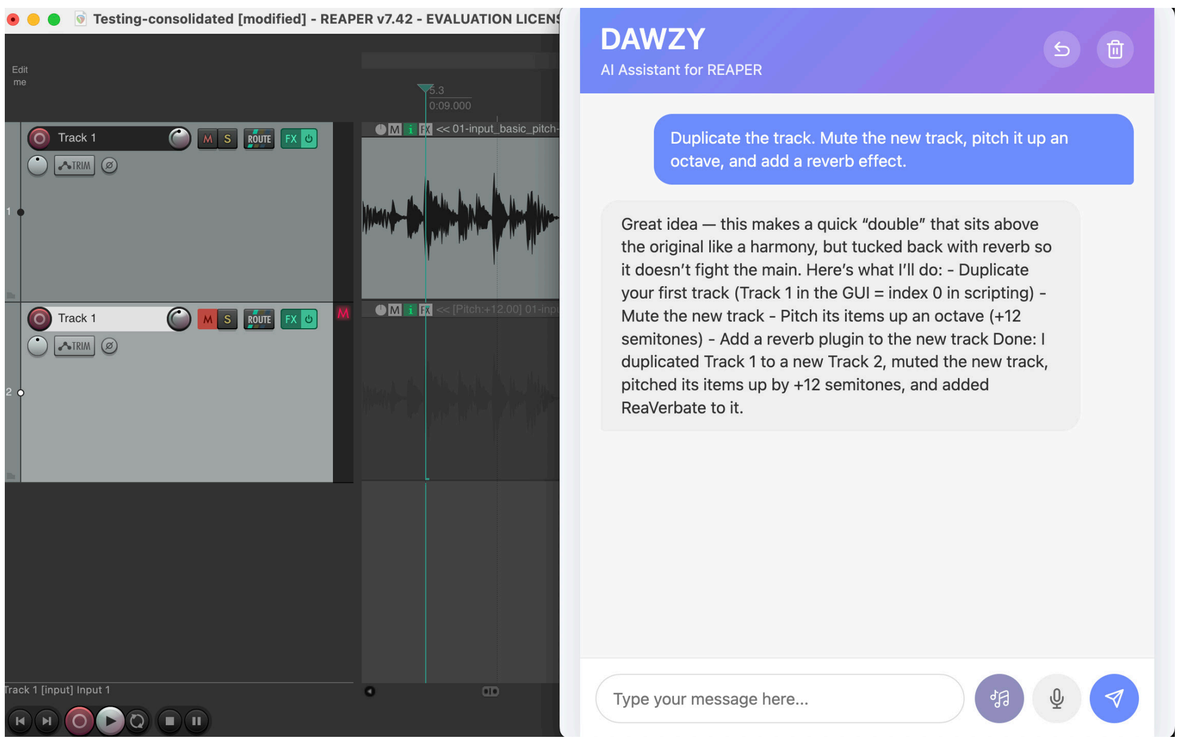
Quantitative Performance:
We tested DAWZY and comparable systems across three creative tasks: **Wavy, Bouncy, and Fade**.

GPT-5 achieved a 78 % success rate (7/9), outperforming open-source and commercial baselines by maintaining consistent, context-aware editing.



A **Mean Opinion Score (MOS)** evaluation with **N = 21** participants rated DAWZY **across five dimensions** on a **5-point scale**.
All categories scored **above neutral**, with the **highest ratings for Enjoyment** (M = 4.48, SD = 0.60*) **and Learning** (M = 4.38, SD = 0.74*). These results indicate that users found **DAWZY intuitive, engaging**, and **effective for creative collaboration**.

4. Demo



Code

Demo

Website