

Algorithm Based GEN Video to Partial GEN Audio Synthesis

Xiaosha Li

xli5@berklee.edu

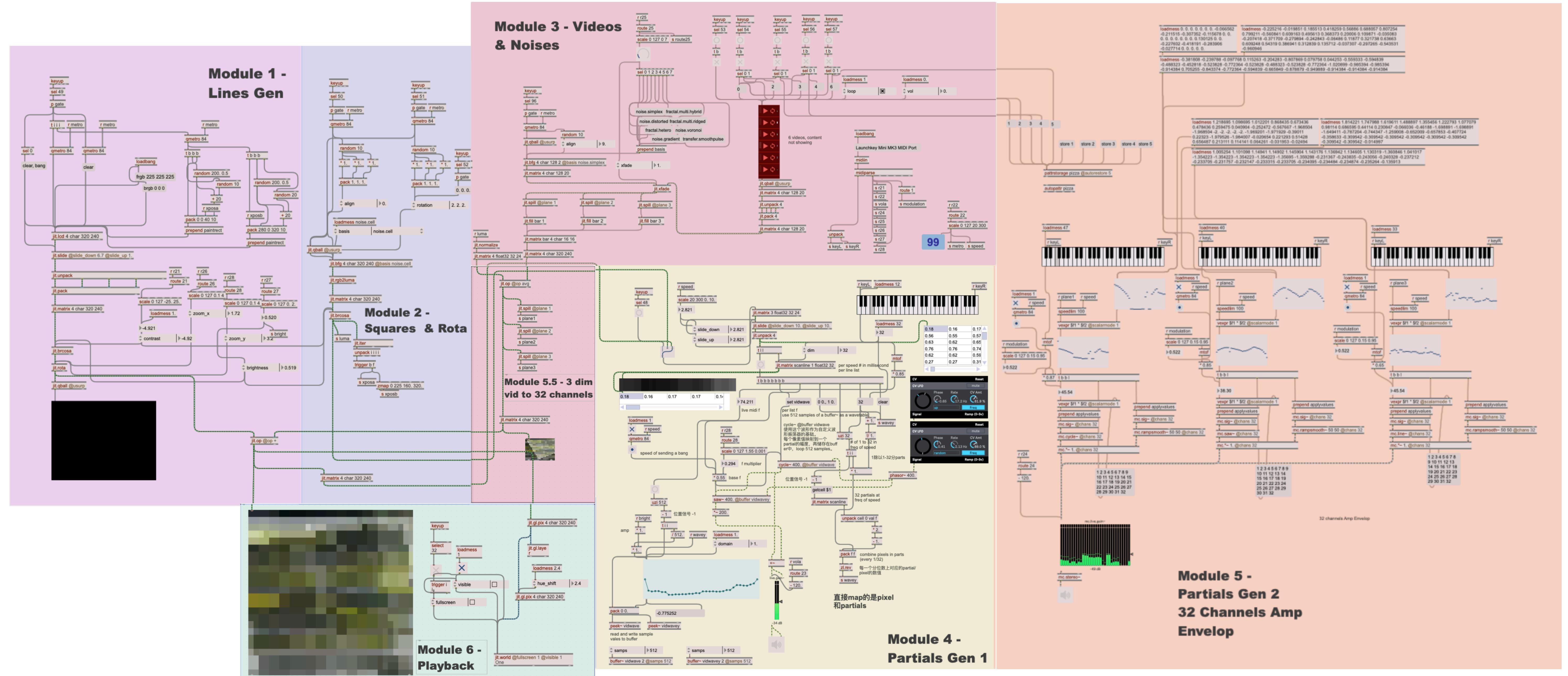
Abstract

This paper presents a novel Gen-to-Gen synthesis system that converts generative video signals into multichannel audio through six interconnected modules. The low-level architecture processes video pixel streams to generate partial audio synthesis, while higher-level modules map RGB channels from five video sources to control a 32-channel amplifier with user-defined sound shapes. The system features MIDI and

Core Synthesis Pipeline

$$A_{out}[n] = \sum_{k=1}^{32} \left(\frac{1}{512} \sum_{i=16(k-1)}^{16k} V_{frame}[i] \right) \cdot W_k[n]$$

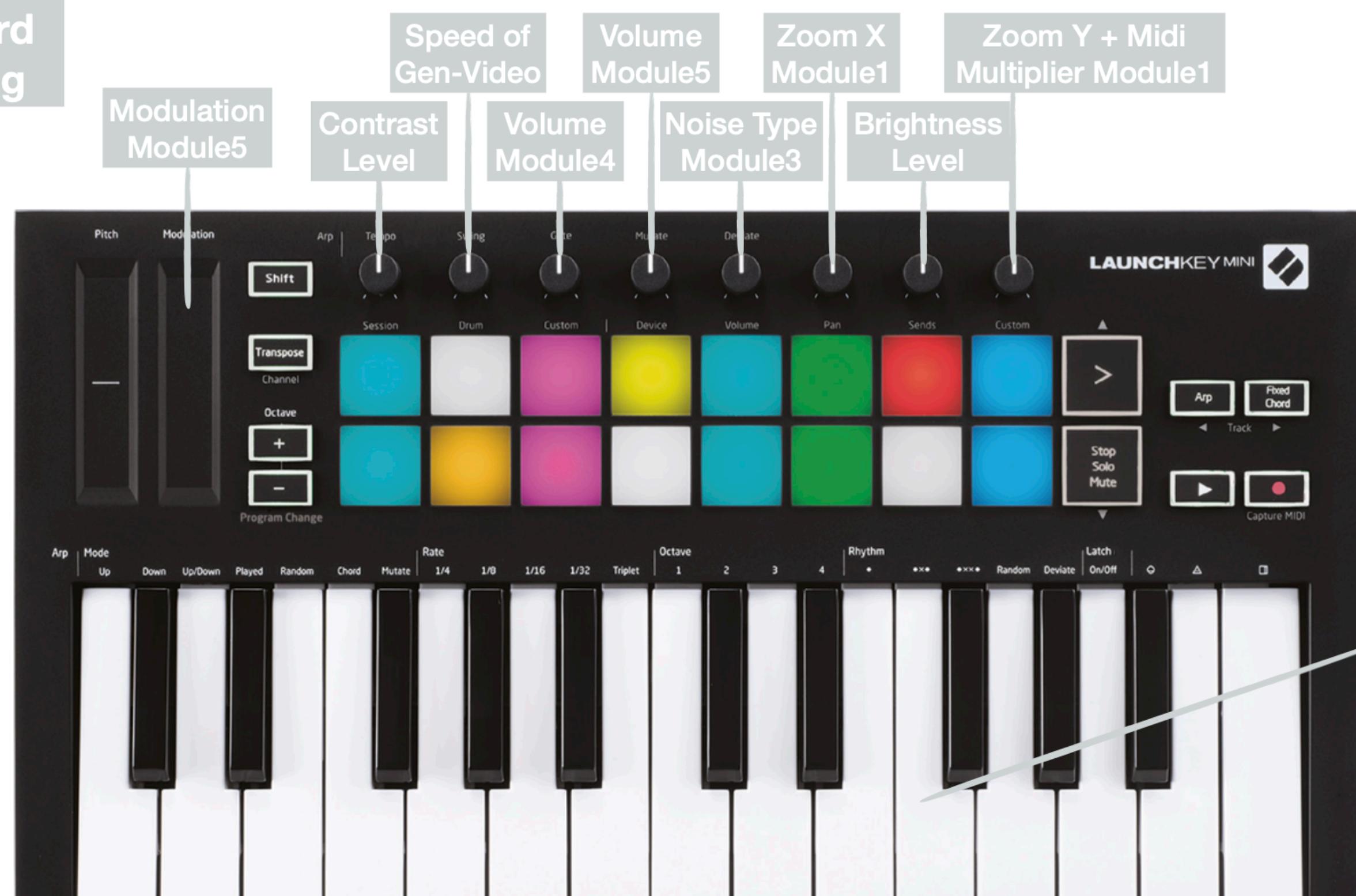
Gen Video to Gen Audio Pipeline



Demo Screenshots

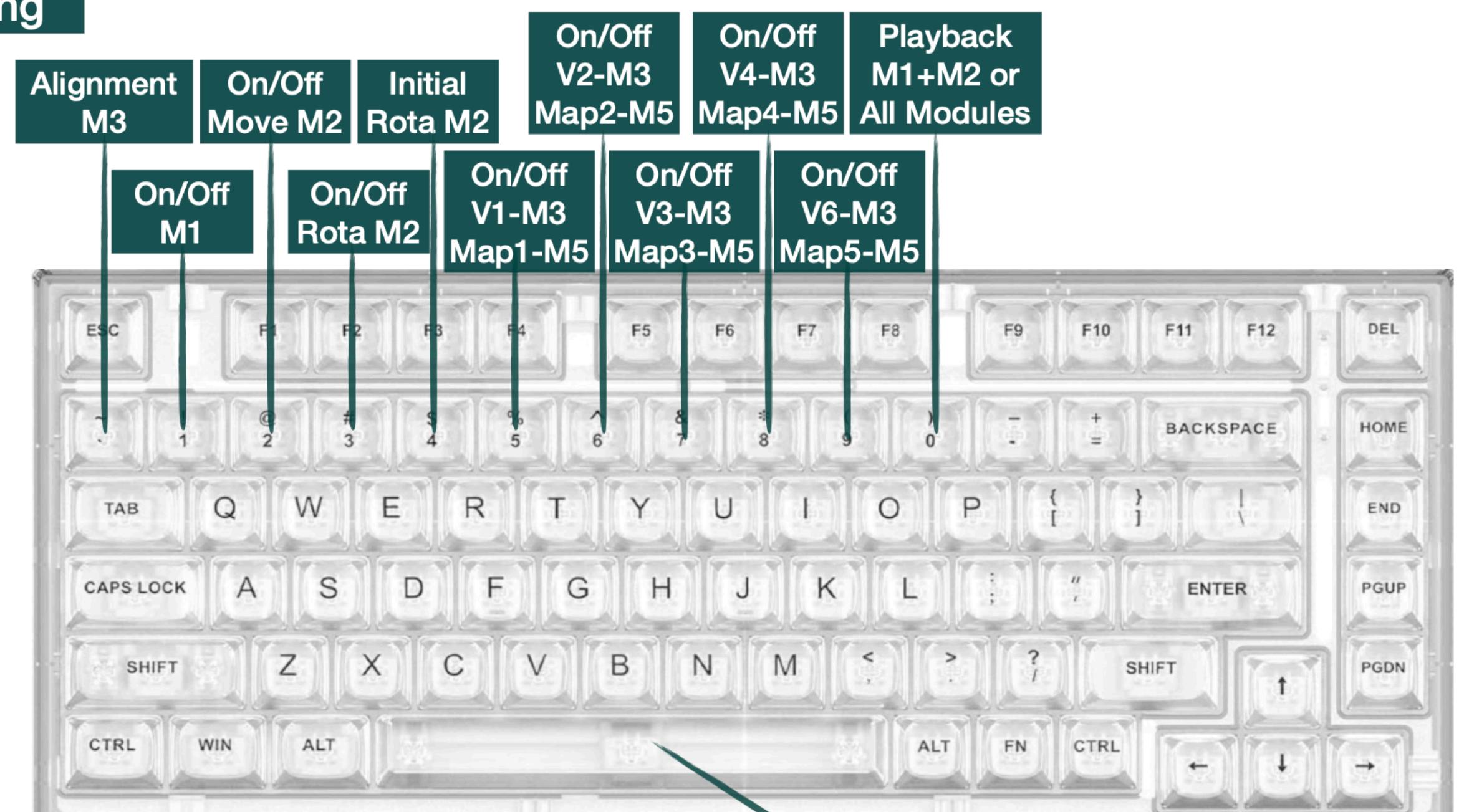


Keyboard Mapping



Initial Keyboards

Keyboard Mapping



Entering Full Screen

Control Implementation

The system integrates MIDI and computer keyboards for real-time interaction and parameter control, enabling dynamic adjustments during performances. The **MIDI keyboard** modulates key parameters, including contrast, brightness, speed, modulation depth, volume, and noise selection. It also supports zoom adjustments and MIDI-triggered events for audiovisual synchronization. The **computer keyboard** toggles module states, playback, and geometric transformations. It enables switching video-to-audio mappings and controlling rotation and movement in various modules. This dual-control scheme balances real-time modulation via MIDI with discrete toggling via the computer keyboard, offering an intuitive and flexible interface for live performance.

Conclusions

This work demonstrates effective video-to-audio synthesis through parametric mapping of visual features to spectral components. Future work will explore machine learning-based mapping strategies and expanded gesture control.