

Eaves Verse: An Interactive Audiovisual Performance with MediaPipe & AI Real-Time Generation

SHUNHANG HUANG, Zhejiang Conservatory of Music, China

1 Program Notes

Eaves Verse is an interactive audiovisual performance created for MediaPipe and AI real-time image generation. Centered around the cultural symbol of the eaves in traditional Chinese architecture, the performance begins with the creator's intent, using hand gestures as an interactive medium. Leveraging AI real-time image generation technology, it presents a three-tiered progression: from abstract concepts to AI-constructed concrete visuals, culminating in an aesthetic realm where reality and illusion intertwine. The lingering melodies of Eastern resonance naturally intertwine with the natural cadence of spoken phrases, while the texture of electronic rhythms synchronously resonates with interactive movements. This allows traditional sonic rhythms to flow within a contemporary interactive context, evoking an Eastern auditory essence where reality and illusion merge. In this work, AI transcends its role as a mere technical tool to become a “symbiotic partner” attuned to the creator's intent. It generates visuals in real time in response to gestures, transforming thoughts into tangible forms. Through this novel interactive format, it reconstructs the contemporary expression of traditional aesthetics, ultimately crafting an audiovisual experience that embodies the interplay of reality and illusion—a quintessential Eastern aesthetic.

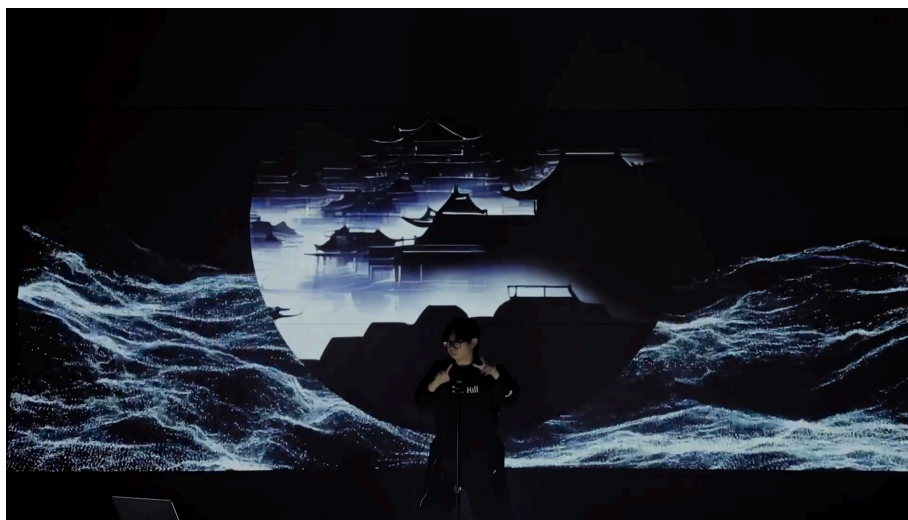


Fig. 1. Author's Stage Performance.

2 Project Description

Real-time interactive audiovisual technology serves as a carrier for the contemporary expression of traditional Eastern aesthetics, integrating gesture interaction, AI generative imagery and responsive sound to connect traditional cultural symbols with modern digital expression. They also propel the

Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). Copyright remains with the author(s).

exploration of expressing traditional cultural aesthetics through contemporary forms. Traditional Chinese aesthetics emphasise the unity of form and meaning, rooted in the "unity of man and nature" philosophy. Eaves, as a core symbol in this work, are not only practical but also philosophical—they connect human, nature and architecture, embodying the Eastern pursuit of balance, and echoing the Taoist concept of transcending virtual-real opposition. The eaves in traditional Chinese architecture carry the memory of folk life and the inheritance of cultural context, bearing the Eastern people's subtle perception of space and harmony. Every curve and outline of the eaves contains the aesthetic pursuit of "harmony between man and nature", which is the core cultural implication that this work intends to convey through digital interaction. This work is anchored in this concept, focusing on cultivating Eastern poetic imagery and fostering deep human-machine collaboration. A real-time interactive system has been developed that uses gestures as the medium, integrating motion recognition with AI-powered, real-time visual generation technology. This transforms abstract concepts into tangible visual forms, leveraging the synergy of technology and art to give traditional aesthetic symbols contemporary relevance.

This work uses a high-definition camera to capture the performer's upper-body movements. Hand keypoint data is extracted via MediaPipe and transmitted through the OSC protocol to Max/MSP for centralized processing. The system uses both raw keypoint information and self-summarized secondary motion features derived from gesture analysis. Extracted motion parameters include overall hand position, relative finger movement to the wrist, spatial distance between the hands, and finger pinch intensity.

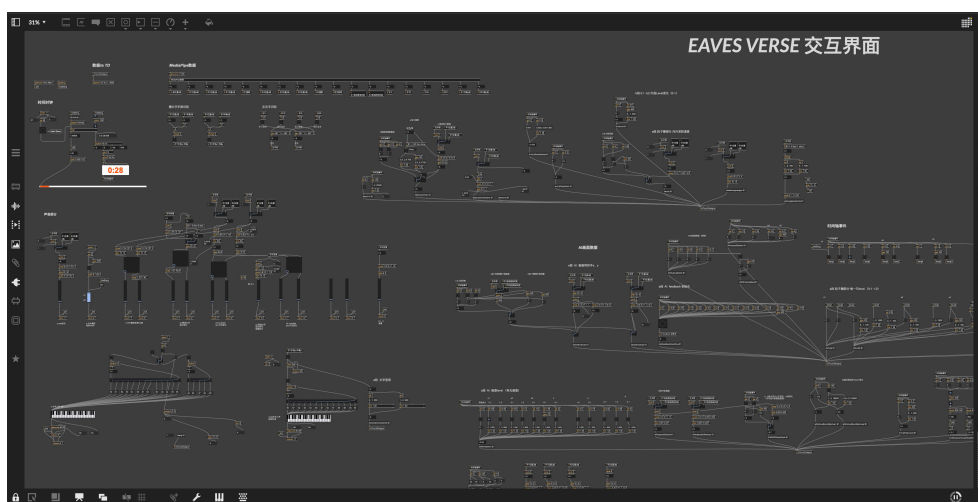


Fig. 2. The process of creating interactive sound components.

The work focuses on the guqin, incorporating a variety of timbres and digitally recreating traditional soundscapes. Drawing on Eastern melodic contours and spoken vocals, it allows for multidimensional sound modulation and triggering through real-time interaction. These evolve dynamically in synchrony with gesture-based controls and visual generation, reimagining traditional rhythmic patterns and embedding Eastern auditory perception within a contemporary digital framework. This approach mirrors the work's three-tiered audiovisual structure, using the interplay of tangible and ethereal soundscapes to interpret core Eastern philosophical concepts such as fluidity and interdependence. It serves as the pivotal link connecting the performer's intent with the audiovisual expression.

Visually, TouchDesigner incorporates the StreamDiffusion plug-in, combining it with creators' trained LoRA models to enable real-time AI-generated images. This process imbues the visuals with randomness and predictability, enabling collaboration between humans and machines in the creative process. As Harris argued, the essence of such a creative approach lies in the balance between control and non-control: the fundamental framework of the work can be defined algorithmically, while certain subtle details are left to chance [1].

Meanwhile, the system constructs a dynamic particle spatial field that interweaves and coexists with AI-generated imagery, shaping a poetic ambiance where virtuality and reality reflect one another. The performer can dynamically manipulate the fluid transitions and morphological variations of the particle space, including its specific diffusion changes, movement speed, diffusion range, brightness, and bloom intensity; beyond particle control, the performer also triggers discrete fragments and text elements. Simultaneously, the performer adjusts the stylistic features of the AI visuals in real time, as well as the position, presentation effect, and frame composition of the AI-generated content, thereby shaping the overall visual narrative. This delivers a fully interactive visual experience blending reality and illusion, which resonates artistically with live bodily movement. Rooted in the aesthetic connotation of traditional Chinese architectural charm, the work reconstructs such Eastern essence through dynamic digital visual expression.

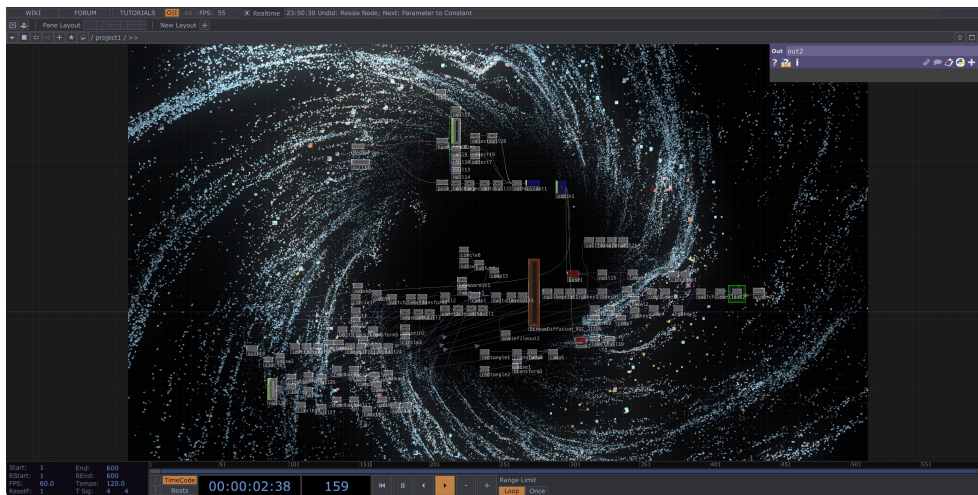


Fig. 3. The Process of Creating Interactive Visual Components.

This work achieves dynamic conversion from motion to audiovisual parameters through hardware-software collaboration, exploring a cross-modal creative pathway between control and randomness. By combining traditional Eastern aesthetics with modern, real-time, interactive technology, it offers a practical technical approach to creating interactive artistic narratives, contemporary expressions of traditional cultural elements and AI-driven audiovisual performances.

3 Technical Notes

Eaves Verse is a live interactive audiovisual performance and requires stereo output. The performer will be bringing one Apple MacBook Pro laptop, one high-definition camera for motion interaction data capture with its matching mount, and one Yamaha UR44 audio interface.

The following equipment is needed from the venue:

- One HDMI cable for connecting the laptop to the stage-specific visual projection equipment.

- One cable for connecting the Yamaha UR44 audio interface to the stage mixing console, which should be linked to the live stereo sound system.
- Stage floor monitor equipment for the performer's on-stage monitoring.
- One standard-sized table that can fit the MacBook Pro laptop and the Yamaha UR44 audio interface.
- One power outlet or power strip accessible within the stage area, with at least four available outlets to power the equipment.

4 Media Link(s)

- Video: <https://youtu.be/rrb-tWrjJ8M>

Ethical Standards

This work does not involve research on human participants or animals and is in full compliance with relevant ethical standards. The project is self-funded, and the author declares no financial or non-financial conflicts of interest.

References

- [1] Harris, L. (2021). *Composing Audiovisually: Perspectives on audiovisual practices and relationships*. Routledge.