DJESTHESIA: Tangible Multimedia for DJs

Eduardo Castelló Ferrer
eduardo.castello@ie.edu
CyPhy Life, Robotics & AI Lab,
School of Science & Technology, IE University
Madrid, Spain

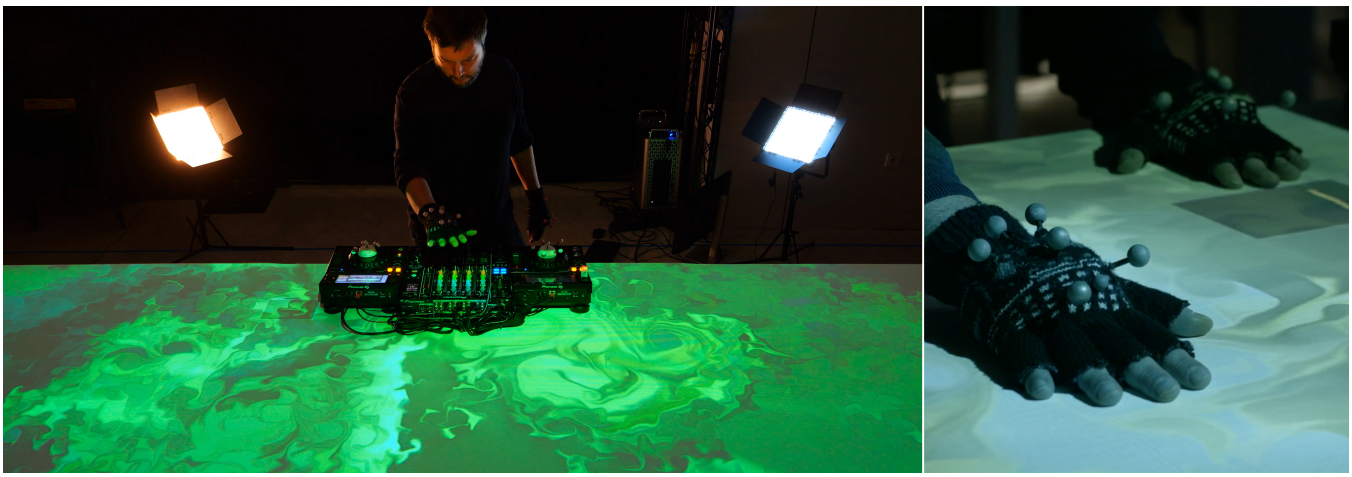

Figure 1: DJESTHESIA. Left: Top view of the setup where the DJ manipulates real-time visuals using motion capture gloves over a projection surface. Right: Close-up of gesture-based control of music and visuals.

Abstract

DJESTHESIA uses tangible interaction to craft real-time audiovisual multimedia, blending sound, visuals, and gestures into a unified live performance. The project supports four interaction modes: I) Knob changes music, where standard DJing is performed. II) Music changes visuals, where changes in the audio parameters done through the mixer have a direct impact in the visualizations representing the music (e.g., color palette). III) Gesture changes visuals, where gestures and body movements give the possibility to interact physically with the visual representation of the music (e.g., grab, release, throw). IV) Gesture changes music, where, gestures can convey information to an audio composition software to alter aspects of the music being played (e.g., EQs). The aim of DJESTHESIA is to transform the DJ into both a performer and a performance.

ACM Reference Format:
Eduardo Castelló Ferrer. 2025. DJESTHESIA: Tangible Multimedia for DJs. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Real-Time Live! (SIGGRAPH Real-Time Live! '25), August 10-14, 2025.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3721243.3735980

SIGGRAPH Real-Time Live! '25, Vancouver, BC, Canada

ACM ISBN 979-8-4007-1545-7/2025/08
https://doi.org/10.1145/3721243.3735980

1 Introduction

Tangible interaction research aims to bridge the gap between the physical and digital worlds, allowing users to manipulate digital information through physical actions such as touching, moving, and gesturing. Along these lines, Tangible User Interfaces (TUIs) are physical objects that users can interact with to control digital content and applications. In the musical performance field, TUI-based research has mainly focused on audio manipulation [Patten et al. 2002] and expressive control [Heap and MiMU 2015], rather than using visuals as an interface to generate and control the music. DJESTHESIA[1,2] aims to fill this gap by fully engaging audiences through synchronized visual elements that complement electronic music performance by using TUIs.

2 DJESTHESIA

DJESTHESIA supports four interaction modes (Fig. 2) which can all be active simultaneously: **Mode I**: Knob changes music. This first mode is standard in most DJ setups. When a DJ uses a mixer and decks, the DJ changes the music by physically turning knobs, for example to change parameters such as filters or equalizers. **Mode II**: Music changes visuals. A low-latency audio pipeline routes the sound from the mixer to an audio composition software on a computer (Fig. 3). Then, a Fast Fourier Transform (FFT) analyzer extracts real-time frequency data from the mix and sends it to a web-based fluid dynamics simulation, which is finally projected onto the DJ

[1]Demo video: https://youtu.be/Y5tWjAqEkEU
[2]Code repository: https://github.com/IERoboticsAILab/djesthesia

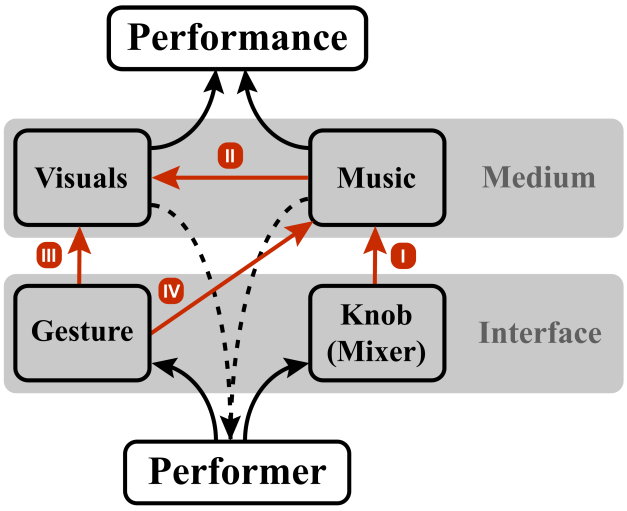

Figure 2: System Overview. Four interaction modes that connect performer and performance: I. Knob changes music, II. Music changes visuals, III. Gesture changes visuals, and IV. Gesture changes music.

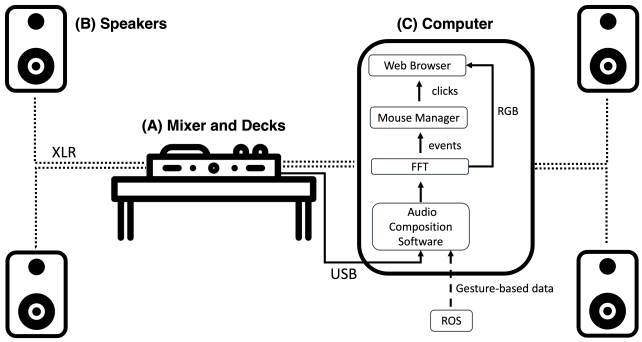

Figure 3: Audio Workflow

working space (Fig. 4). **Mode III**: Gesture changes visuals. A Motion Capture (MoCap) system is used for spatial gesture tracking. The DJ wears gloves (Fig.1 Left) with custom MoCap markers (a.k.a., TUIs) to interact with the generated visuals (e.g., grab, release, summon). Simultaneously, TUIs added to the decks allow the DJ system to act as an interface to change parameters of the simulation (e.g., splat radius, fluid density). **Mode IV**: Gesture changes music. The same gesture-based interaction is used to craft the music. By using two-handed gestures (Fig.1 Right), a DJ can draw EQ curves that will be applied to the music—even curves that would be impossible to achieve in a mixer—and modify other parameters such as volume levels. These modalities give DJs a tangible interface to sculpt music their spatial environment.

2.1 Audio

Fig. 3 depicts the audio workflow. Audio files (e.g., MP3) are loaded into the decks and channeled into the mixer (**A**). The output of the mixer is connected to four speakers in the room by corresponding XLR cables (**B**). The resulting mix is also channeled through a USB

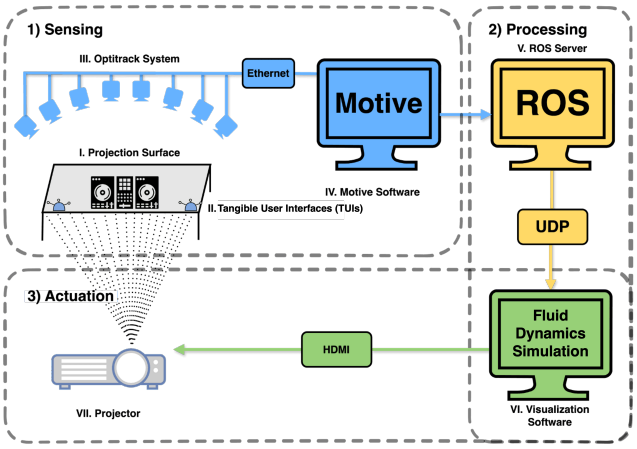

Figure 4: Visual Workflow

connection to a computer hosting all the audio processing elements of the project (**C**). First, the audio signal is channeled to an audio composition software (i.e., Ableton Live 11) where a parametric EQ plugin listens to gesture-based data coming from the Robot Operating System[3] (ROS) server (Mode IV). Then, the signal is routed to the FFT, which filters the 0-250 Hz band and triggers an event every time the resulting signal goes above the -2 db mark (i.e., a "kick"). This event is processed by a script (i.e., Mouse Manager) that converts it into mouse "clicks" in the web browser window. These "clicks" produce the pulses in the fluid dynamics simulation. In parallel, the FFT also analyses the low (0-250 Hz), mid (250-1250 Hz), and high (>1250 Hz) frequency bands and maps their signal values onto RGB values. Finally, this information is sent to the web browser to update the color of the visuals (Mode III).

2.2 Visuals

Fig. 4 shows the visual information workflow. **1. Sensing:** The system uses a MoCap camera system (OptiTrack) to track the position and orientation of the TUIs. For the predominant interaction modality, TUIs are placed inside gloves to track the hands and gestures of the DJ. As a supplemental modality, they are also placed on top of the mixer or decks to capture the interactions between the DJ and the equipment. This information is sent to a tracking software (Motive). **2. Processing:** A server running ROS acts as an intermediary, processing the MoCap data and sending it to the visualization engine software via UDP. **3. Actuation:** The visualization engine is a WEBGL fluid dynamics simulation[4] (running on a web browser) programmed in Javascript (JS). The computer processing this information is connected to a short-throw projector via HDMI and located at the ceiling of the performance area. Finally, the generated visuals are projected onto the table where the DJ is performing.

References

Imogen Heap and Team MiMU. 2015. MiMU Gloves: Wearable Technology for Musical Expression. In *Proceedings of the NIME Conference*. MiMU Limited.

James Patten, Ben Recht, and Hiroshi Ishii. 2002. Audiopad: A Tag-based Interface for Musical Performance. In *Proceedings of the SIGGRAPH Conference*. MIT Media Lab.

[3]https://ros.org

[4]https://paveldogreat.github.io/WebGL-Fluid-Simulation