HoloBeat Studio: Futuristic Music Production with AI

Sunpreet Singh
Computer Science
University Of Victoria

Turner Gullickson
Computer Science & Geography
University Of Victoria

Abstract
This project introduces a music production creativity support tool (CST) that simplifies the song-making process using futuristic qualities such as a holographic simulation (via a webcam) and an AI assistant. Designed for beginners, children, and even professionals in the music industry, the tool eliminates the complexities of traditional music software by offering an intuitive and step-by-step workflow. The tool consists of two physical structures: one representing the stages of music production (Guitar → Drums → Bass → Recording) and another for note and rhythm visualization (G, C, D). The user interacts through an encoder for instrument selection, hand gestures for note activation, and voice commands for AI engagement. The AI assistant listens in real-time, generating loops and layering tracks based on user actions, eventually producing a complete song. This CST is designed to make music creation accessible and engaging for all user levels. Beginners benefit from its simplicity, children enjoy its block-based visualization and interactivity, and professionals can create rough drafts effortlessly. Collaboration is also a key feature, enabling multiple users to contribute simultaneously. By bridging physical interaction with digital creativity, this tool redefines music production as a collaborative and futuristic experience making it both innovative and practical.

CCS Concepts
- Hardware → Sensor devices and platforms → Microcontrollers (e.g., Arduino)
- Human-centered computing → Interaction paradigms → Gestural input
- Applied computing → Arts and humanities → Sound and music computing
- Computing methodologies → Artificial intelligence → Natural language interfaces

Author Keywords: Creativity Support Tool, Music Production, Artificial Intelligence, Accessibility

ACM Reference Format: Sunpreet Singh, Turner Gullickson, 2024. Enhancing Music Creativity: A Gesture-Based Creativity Support Tool. In *Proceedings of the ACM SIGCHI Conference on Music Accessibility* (C&C '24). ACM, Victoria, BC, Canada, 10 pages. https://doi.org/XXXXXXX.XXXXXXX

Introduction

Music production is the process of creating music using software, hardware, and other tools and it encompasses composing, arranging, recording, and editing. As a creative and technical art form, it requires patience, practice, and dedication (AllRoundClub, 2023). However, traditional music production tools have often been complex and intimidating, particularly for beginners and children (AllRoundClub, 2023). Professional software typically demands technical expertise and significant financial investment, creating barriers for those interested in exploring music creation.

Our Creativity Support Tool (CST) addresses these barriers by simplifying the music-making process, making it more accessible, interactive, and engaging. With an user-friendly interface, our CST empowers users to explore music production without the steep learning curve or financial constraints of traditional tools.

The primary goal of our CST is to provide an intuitive and engaging platform for creating music that eliminates the financial, spatial and technical barriers associated with traditional music production tools. Traditional setups for home studios often demand significant financial investment and physical space, making them inaccessible to many aspiring musicians. For instance, creating a basic home studio can cost between $1,000 and $3,000, while professional-grade setups can exceed $10,000 and require equipment such as microphones, acoustic treatment, and ample space (The Techy Life, 2024). As well, music creation often appears overwhelming to beginners due to the complexity of software and the steep learning curve required to master technical skills and find their unique sound (WhippedCreamSounds, 2024). Even professionals may find the multifaceted nature of music production challenging, highlighting the need for more accessible and supportive tools (Dsokolovskiy, 2016).

Our CST addresses these three barriers by leveraging compact, holographic interactions (simulated via webcam) and AI-driven features to eliminate the need for costly and space-intensive hardware. Users can create music through simple gestures, such as hand movements or head nods, while enjoying immediate visual and auditory feedback—all within a space-efficient and affordable framework. Using features like a block-based interface, real-time visualizations, and AI assistance, the CST simplifies music-making, making it fun and approachable for children, educators, beginners, and hobbyists. It also fosters collaboration by enabling multiple users to contribute in real time, bridging the gap between technology and artistic creativity.

Related Work

Projects that exist that have similar elements to our CST are Digital Audio Workstations, AI-driven music tools, tangible user interfaces for music, and gesture-based systems. Digital Audio Workstations (DAWs) like Ableton Live and FL Studio are widely used for professional music production but are complex and lack physical interaction unlike our CST which offers a hands-on encoder, and gesture controls (WhippedCreamSounds, 2024). Similar to AI-driven tools like Amper Music and AIVA, our CST uses AI for loop generation but integrates physical interaction and a step-by-step workflow for a more intuitive experience. Interactive tools like Reactable and Music Blocks focus on tactile and visual engagement, which our CST provides. However, they do not have AI-powered features and collaborative capabilities. Additionally, gesture-based systems like Leap Motion inspire our

use of gestures for control, but our CST combines this with AI and real-time feedback to create a structured and engaging music creation process.

Our CST builds upon these existing tools by prioritizing accessibility, integrating advanced features, and fostering collaboration. Unlike tools aimed at professionals, our CST is designed for beginners, children, and casual creators. It incorporates futuristic elements, such as holographic-like interactions using a webcam and real-time AI assistance for loop generation and layering. Unlike conventional tools, our CST guides users through a structured, step-by-step music creation process—organizing production into clear stages like Guitar, Drums, Bass, and Recording which makes the workflow intuitive. Additionally, while many tools focus on individual creation, our CST supports real-time collaboration, allowing multiple users to contribute simultaneously, whether playing instruments or recording vocals, or both.

High-quality music tools are often expensive, limiting access for students, hobbyists, and educators. Our CST addresses this by providing an affordable alternative using accessible components like Arduino and a webcam. Unlike existing tools, which can lack engaging interaction, our CST combines tactile elements, gesture recognition, and real-time AI responses to create an immersive and enjoyable music-making experience.

Design Fiction

In a future where creativity meets cutting-edge technology, music production is revolutionized by holographic interfaces and AI-powered assistance. Our Creativity Support Tool (CST) envisions a world where users can create music by simply moving their hands and speaking to an AI assistant. The tool eliminates physical and technical barriers, enabling anyone—regardless of skill level—to engage with music production effortlessly and collaboratively.

The World of the CST

1. **People and Society**:
   Music is no longer confined to professionals or enthusiasts with access to expensive equipment. Instead, it becomes an inclusive activity that fosters creativity, education, and collaboration. People interact with music tools as naturally as they would with everyday objects, making music creation part of daily life.
2. **Technology**:
   The CST leverages holographic technology (currently mimicked via a webcam) to let users interact with notes and instruments in the air. Gestures like hand movements and head nods trigger sounds, while an AI assistant listens in real-time, processing commands through Natural Language Processing (NLP). Users can say, "Hey AI, create a loop," and the AI generates harmonious sounds based on their actions.
3. **Economy**:
   Affordable and accessible, the CST democratizes music production by replacing costly software and equipment with an intuitive system. Its design encourages creativity in schools, homes, and communities, bridging the gap between affordability and high-tech innovation.
4. **Politics**:
   A world driven by accessibility and equity ensures tools like the CST are available to everyone, fostering global collaboration and breaking down barriers in creative industries. The CST supports people with all abilities, and is especially helpful in engaging students or children who are non-verbal or use alternate forms of

communication. By using the hand and face gestures, they can create music and have fun.

The Role of the CST

The CST is designed as a diegetic prototype, representing how futuristic music tools might function:

- Holography: Users interact with floating notes in the air, mimicked by a webcam for this prototype, creating a simulated holographic experience.
- AI Integration: The AI assistant not only listens to users but adapts, generating loops, layering sounds, and even adding effects upon verbal commands.
- Physical Blocks: The CST's tangible design, with its modular blocks and LED feedback, bridges the digital and physical worlds, making the process engaging and visually intuitive.

User Scenarios

1. Beginner Music Creator:
    - Mira, a 12-year-old, explores music production for the first time. Standing before a holographic interface, she uses gestures to play notes while saying, "Hey AI, create a loop." The AI generates a track from her gestures, inspiring her to try drums and bass next. The tool builds her confidence without overwhelming her.
2. Professional Prototyping:
    - Arun, a music producer, uses the CST to quickly sketch ideas. He layers loops by moving between blocks (Guitar → Drums → Bass) and asks the AI, "Can you layer the drum loop over this track?" Within minutes, he has a rough draft to refine later in a studio.
3. Collaborative Jam Session:
    - Emma and Jack collaborate on a song. Jack selects instruments using the encoder while Emma controls notes through gestures. They build a full track together, then Jack records vocals on the top block while the AI adds effects, creating a polished sample.

Framing the Purpose

This CST is built on design principles that support exploration, accessibility, and collaboration:

1. Support Exploration: Users can try different instruments, sounds, and techniques by interacting with holograms or gestures.
2. Low Threshold, High Ceiling: The CST is simple enough for beginners to use but sophisticated enough for professionals to prototype ideas.
3. Collaboration: Multiple users can interact simultaneously, sharing creativity in real-time.
4. Simplicity: A clear, step-by-step process—Guitar → Drums → Bass → Recording—removes the complexity of traditional tools.

Imagined Future

The CST creates a future where:

- Music production is a universal activity: Accessible for children, beginners, and professionals alike.
- Technology meets creativity: Holograms and AI empower users to express themselves without the constraints of physical tools.
- Collaboration thrives: Music creation becomes a shared experience, fostering creativity across borders and communities.

Prototype Design and Implementation

Figure 1: HoloBeat Studio Prototype

Our prototype utilizes webcam-based gesture recognition to enable an interactive music production experience. This interaction revolves around two modular towers, each composed of blocks that represent musical notes or instruments. The system is designed to be intuitive and collaborative, allowing users to create music through simple hand and head movements. Future iterations will enhance this by enabling multiple users to collaborate efficiently in real-time, adding a layer of social interaction (Vaucelle & Pachet, 2010). The modular blocks are essential for providing visual feedback and facilitating an intuitive understanding of the music creation process.

The modular blocks form the physical interface of the prototype, offering a tactile and visual connection to the music production process. There are two main types of blocks: Note Blocks and Instrument Blocks, each serving a distinct role in the user experience.

The note blocks are associated with three fundamental musical notes: D, C, and G. These notes were selected for their widespread use in musical composition, as they form the foundation of many common chord progressions, making them both easily recognizable and accessible to users of all skill levels (Sánchez & García, 2016). These notes are ideal for beginners while still being useful for more experienced musicians looking to experiment with basic harmonies.

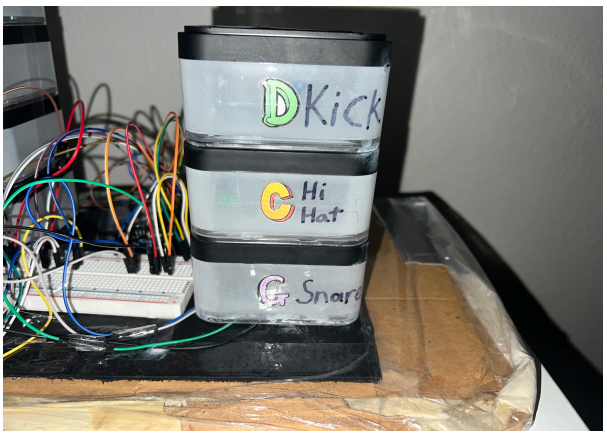

Figure 2: Close up of Note Blocks

Each note corresponds to a specific hand or head gesture. When the user performs a corresponding movement—whether with their hand or by nodding their head—the corresponding note block lights up, providing immediate visual feedback. This real-time feedback allows users to understand the connection between their movements and the musical output, reinforcing the interactive nature of the system (Vaucelle & Pachet, 2010). The simplicity of using only three notes ensures that users are not overwhelmed while still offering a broad range of musical possibilities.

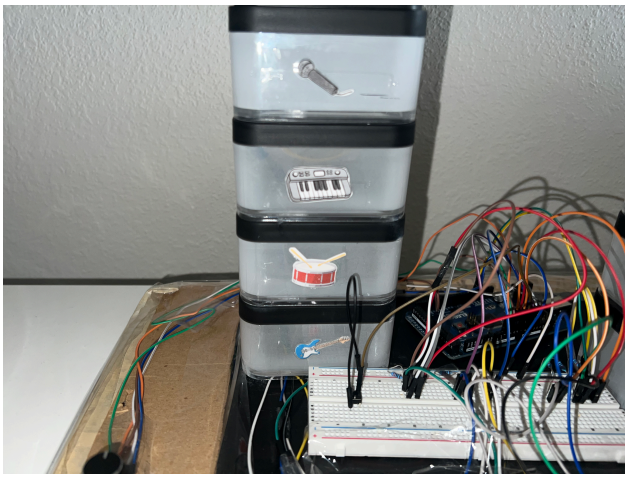

Figure 3: Close up of Instrument Blocks

The second set of blocks, known as Instrument Blocks, represents four core instruments: bass, guitar, drums, and vocals. These instruments were selected because they are the essential components of most popular music. The bass provides harmonic foundation, the guitar adds melody and rhythm, the drums bring rhythm and percussion, and vocals contribute the lyrical elements. By selecting one of these blocks, users can control which instrument is active and which sounds will be triggered by their hand and head movements (Hermann & Hunt, 2010).

The simplicity of having only four instruments ensures that users can focus on the creative aspects of song-making without feeling overwhelmed. As the user interacts with the system, they can layer these instruments together, building complexity and depth as they go. The modularity of the blocks allows for flexibility and experimentation, making it easy to create unique compositions (McCormack et al., 2004).

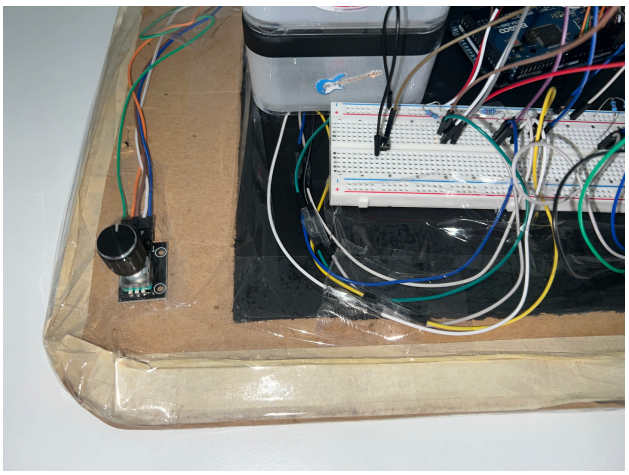

Figure 4: Close up of Rotary Encoder

The rotary knob, modelled after an equipment which is commonly used by many music professionals, provides an interaction point, making navigation through musical instruments feel more natural. Familiar controls like these are known to lower learning barriers and make experiences more seamless, encouraging creativity by letting users focus on their goals (GoodFirms, 2024).

The workflow of the system is designed to be straightforward. When a user selects an instrument by rotating the rotary encoder and pressing it down, they can "play" that instrument by performing hand and head gestures through a webcam. For example, selecting the guitar block and raising a right hand will trigger the sound of the guitar strumming the C note, while a nod of the head will produce the D note. As the user performs each gesture, the corresponding instrument and note block light up, providing visual feedback and helping the user to connect their physical movements with the music they are creating. This simplicity ensures that users are not overwhelmed, making it easier to experiment and build songs step by step (Sánchez & García, 2016).

As the user interacts with the system, an AI module works in the background, listening to the input and capturing information about the notes, rhythm, and style of the user's gestures. The AI system uses this data to generate music alongside the user. By layering sound in real-time, the AI collaborates with the user to create a song which enhances the creative experience and enables seamless, hands-free music production (Hermann & Hunt, 2010). The AI listens to each note played, adjusting the layers of sound to fit the style and rhythm of the user's input, gradually building a complete composition.

A key feature of the prototype is its AI-driven music generation. The AI listens to the user's inputs in real-time, capturing not only the notes played but also the rhythm, tempo, and overall style of the music being created. Based on this information, the AI synthesizes additional sound layers for each instrument, creating a rich, evolving composition that aligns with the user's actions (McCormack et al., 2004).

The advantage of integrating AI into our system is that it removes the need for users to manually program or arrange individual tracks, allowing them to focus on the creative aspects of music production. The AI listens, adapts, and generates sound layers that complement the user's choices, making the process of music creation feel natural (Hermann & Hunt, 2010). Whether the user is a beginner experimenting with basic rhythms or an experienced musician layering complex patterns, the AI's role is to enrich and accelerate the

creative process, ensuring that every interaction results in a meaningful addition to the composition.

The implementation of the CST involves assembling physical structures, connecting electronic components, and running Arduino and Python code. Here's the step-by-step process:

Hardware Setup

- **Block Towers:**
  - 4-Block Tower: Represents music production stages: Guitar (Pin 12), Drums (Pin 11), Bass (Pin 10), Recording (Pin 9).
  - 3-Block Tower: Represents notes and drum sounds: G (Bottom), C (Middle), D (Top).
- **Connections:**
  - AI Button: Pin 8 (Breadboard).
  - Encoder:
    - CLK → Pin 3, DT → Pin 2, SW → Pin 4, Gnd to the breadboard ground.
- **Assembly:**
  - Insert LEDs into holes drilled in glass blocks, secure with tape, and glue the blocks together.
  - Mount the blocks on a stable cardboard base.

Software Setup

- **Arduino Code:**
  - Upload the provided Arduino Code to handle LED control, encoder input, and AI button functionality: https://github.com/gillsunpreet/gillsunpreet/blob/main/DesignFiction_MusicProductionTool/ArduinoCode_DesignFiction/musiProduction/musiProduction.ino
- **Python Script:**
  https://github.com/gillsunpreet/gillsunpreet/blob/main/DesignFiction_MusicProductionTool/PythonCode_DesignFiction/handGesturePlay.py

  Install dependencies:
  pip install mediapipe opencv-python pygame pyserial(Run this command in visual studio terminal)

  - Run the script to enable:
    - Webcam-based gesture recognition (mimicking holography).
    - Communication with Arduino for LED synchronization and stage selection.
    - Preloaded loops (G, C, D notes and drum sounds) for AI-mimicked functionality.

Make sure you close the serial monitor window in Arduino IDE before running your Python script.

AI Integration
The AI assistant simulates real-time feedback by using a predesigned dataset of loops. Pressing the AI button:

- Generates a guitar loop in stage 1.
- Layers drums and bass in subsequent stages.
- Activates the final recording phase.

Final Steps

1. Complete all connections and upload the code.
2. Use the encoder to select stages, gestures to interact with notes, and the AI button to progress through the song-making process.

This implementation integrates hardware and software seamlessly, offering an interactive music creation experience.

Critique

CSI Scores:

- **Collaboration**: 7 (Score for enabling multiple users to interact and contribute, e.g., real-time collaboration in playing instruments and recording vocals.)
- **Enjoyment**: 8 (The interactive holographic features and AI assistance make the tool enjoyable and engaging.)
- **Exploration**: 6 (Users can experiment with different instruments and notes, though the range of options is limited.)
- **Expressiveness**: 5 (The CST allows users to express creativity, but the reliance on pre-defined loops and limited notes restricts full expressiveness.)

Our CST demonstrates significant strengths, particularly in its collaborative potential, allowing multiple users to work together. For instance, one user can manage instrument selection while others perform gestures or record vocals, enhancing teamwork in music creation. Its intuitive design incorporates a step-by-step music creation process and gesture-based controls that are both easy to use and engaging, especially for beginners and children. Additionally, the holographic-inspired interface and AI-driven loop generation provide a fun and immersive experience, making music production accessible and enjoyable.

However, the tool also has notable limitations. Its overdependence on AI to generate loops and layers reduces opportunities for users to develop skills in playing real instruments, which can limit creative growth. Furthermore, the current version supports only three notes (G, C, D) and basic drum sounds (snare, hi-hat, kick), which restricts its versatility and creative possibilities for advanced users. The expressiveness of the tool is also constrained, as the reliance on a preloaded dataset and lack of manual input options limit the ability to create unique loops, thereby impeding full creative freedom. Lastly, the block-based physical interface, while visually appealing, presents challenges for portability and scalability in larger collaborative use cases.

Future Work

1. Advanced AI Integration:
    - Develop a fully trained NLP-based music assistant to act as a virtual producer, guiding users and offering more freedom to customize loops, instruments, and effects.
2. Digital Interface:
    - Replace physical blocks with a digital system where users can preselect instruments and notes, increasing flexibility and scalability for diverse music styles. Or attach digital screens to the blocks for dynamic instrument and note selection.
3. Expanded Features:
    - Add more notes and instruments to support various genres.
    - Introduce rhythm and tempo controls for better customization.
4. Broader User Reach:
    - Include advanced features, such as layering options and customizable effects, to make the tool suitable for professionals beyond prototyping, without increasing complexity.
    - Create educational versions for schools to teach rhythm, melody, and collaboration.
5. Refinements from Feedback:
    - Improve gesture recognition accuracy and system portability based on user testing.

Conclusion

Our CST transforms music creation into an accessible, interactive and fun process. By combining intuitive controls, AI, and gesture-based interactions, it enables users to compose songs effortlessly without needing a background in music or being skilled in a musical instrument. Whether you are a musical beginner experimenting with simple notes or a professional looking for a way to quickly sketch new ideas, Holobeat Studio offers a versatile platform that adapts to the user's music, not the other way around.

This tool will continue to be refined. In future improvements we will enhance the adaptability and engagement while also making the tool more versatile and responsive. By integrating AI with physical interaction, we ensure that the tool remains a bridge between creativity and technology, allowing the users to easily express their musical passion. In the future, we envision HoloBeat will be an indispensable tool for all musical artists.

References

[1] "Music Production for Kids: 11 Tips to Get Your Kid Started." *Allround Club*, https://allroundclub.com/blog/music-production-for-kids-11-tips-to-get-your-kid-started. Accessed 30 Nov. 2024.

[2] "The Techy Life." *How Much Is a Home Studio?* The Techy Life, https://thetechylife.com/how-much-is-a-home-studio/. Accessed 2 Dec. 2024.

[3] "Whipped Cream Sounds." *Guide to Music Production.* Whipped Cream Sounds, https://www.whippedcreamsounds.com/guide-to-music-production/. Accessed 2 Dec. 2024.

[4]Sokolovskiy, Dmitry. *Producer's Fatigue.* DSokolovskiy, https://dsokolovskiy.com/blog/all/producers-fatigue/. Accessed 2 Dec. 2024.

[5] Hermann, T., & Hunt, A. (2010). *Music and Artificial Intelligence: A Survey*. In *The Cambridge Companion to Electronic Music* (pp. 232–250). Cambridge University Press.

[6] McCormack, J., Hutchings, P., & Hutchings, P. (2004). *Automated Music Composition and AI: Approaches and Applications*. *Proceedings of the International Joint Conference on Artificial Intelligence* (IJCAI).

[7] Sánchez, J. L., & García, D. (2016). *Accessibility of Music Production Tools for Novices: Towards a Universal Music Production Platform*. *International Journal of Human-Computer Interaction*, 32(4), 293-308.

[8] Vaucelle, G., & Pachet, F. (2010). *Gestural Interaction for Music Production*. Proceedings of the International Conference on New Interfaces for Musical Expression (NIME).

[9] "RMCAD." *Collaboration in the Digital Age: Online Tools for Music Producers.* Rocky Mountain College of Art + Design, https://www.rmcad.edu/blog/collaboration-in-the-digital-age-online-tools-for-music-producers/. Accessed 2 Dec. 2024.

[10] GoodFirms. (2023). Familiarity in product design and marketing: Why users prefer familiar designs. GoodFirms. Retrieved October 27, 2024, from https://www.goodfirms.co/blog/familiarity-in-product-design-and-marketing