Interactive Sound Wave: Visualizing Audio and Emotion Through Real-Time Sound–Gesture Interaction in Technology-Enhanced Learning

Naphatsanan Suwannawong, King Mongkut's University of Technology Thonburi, Thailand
Yanee Posuk, King Mongkut's University of Technology Thonburi, Thailand

The Asian Conference on the Social Sciences 2026
Official Conference Proceedings

Abstract

This study developed Interactive Sound Wave, a technology-enhanced learning artifact that visualizes audio and emotion through real-time sound–gesture interaction. The study had 3 objectives: (1) to develop an interactive learning artifact that transforms learners' sound input and hand gestures into real-time particle visualizations, (2) to evaluate the quality of the artifact, and (3) to assess user satisfaction with the artifact among 30 participants. A research-and-development approach was implemented in 3 phases. Phase 1 focused on designing and building the system to capture audio features (e.g., rhythm and spectrum) and hand-gesture signals from a camera, then mapping these inputs to particle motion and visual effects to provide immediate audiovisual feedback. Phase 2 evaluated media quality through expert review using a 5-point Likert-scale form covering course suitability, interaction consistency, clarity of control, visual quality, and classroom feasibility. Phase 3 conducted a try-out session with users followed by a 5-point satisfaction questionnaire. Results indicated that overall media quality was good. Experts rated interaction consistency as very good, while clarity of user experience and control was rated good; other criteria were rated good. Overall user satisfaction was very high, with engagement and interaction rated the highest, followed by the clarity of linking sound–gesture–visuals to support concept understanding. The study demonstrated that real-time, responsive audiovisual interaction can promote learner engagement and help make abstract audio concepts more observable and discussable, providing a practical classroom-ready artifact and an evaluation framework for technology-enhanced learning in multimedia and creative media education.

Keywords: Interactive Sound Wave, visualizing audio, real-time sound–gesture interaction, technology-enhanced learning

Introduction

In contemporary education, particularly within multimedia and creative media disciplines, there is an increasing demand for instructional approaches that bridge the gap between abstract concepts and learners' perceptual understanding. Sound, as an intangible and time-based phenomenon, presents significant challenges for learners, as it cannot be directly observed and is often experienced only through hearing. Traditional teaching methods often rely on visual materials and conceptual explanations, which may not fully capture the experiential nature of sound.

Recent developments in Technology-Enhanced Learning (TEL) have provided new opportunities to address these limitations. Interactive systems that integrate real-time data processing, visualization, and user input can transform abstract phenomena into more accessible and explorable experiences (Pauline-Graf et al., 2021). Furthermore, constructivist learning theory emphasizes that learners construct knowledge through active engagement and interaction (Erawati & Adnyana, 2024), suggesting that learning environments should support direct manipulation and exploration.

Combining sound and gesture in interactive experiences can support experiential learning by helping users understand how actions and sounds are connected. Rather than focusing on precise analytical representation, such systems emphasize perceptual and intuitive understanding. According to multimedia learning theory, integrating visual and auditory elements enhances comprehension, particularly when learners actively engage with both modalities (Meena & Arya, 2023).

This study proposes Interactive Sound Wave, a real-time interactive system that integrates gesture-based control and audio input to generate dynamic visual responses. The system is designed to support learners in exploring the relationships between gesture, sound, and visual behavior through interaction.

Objectives

1) To develop Interactive Sound Wave: Visualizing Audio and Emotion Through Real-Time Sound–Gesture Interaction in Technology-Enhanced Learning.
2) To evaluate the quality of Interactive Sound Wave: Visualizing Audio and Emotion Through Real-Time Sound–Gesture Interaction in Technology-Enhanced Learning.
3) To assess user satisfaction with Interactive Sound Wave: Visualizing Audio and Emotion Through Real-Time Sound–Gesture Interaction in Technology-Enhanced Learning among 30 participants.

Literature Review

Interactive Audiovisual Systems in Learning

Interactive audiovisual systems refer to digital environments in which users can manipulate input variables such as sound or motion and receive immediate visual and auditory feedback (Sharkey, 2008). These systems are grounded in interactive media design, where real-time responsiveness is a key factor influencing user engagement and learning effectiveness.

In educational contexts, such systems are often aligned with constructivist learning theory, which posits that knowledge is actively constructed through interaction and experience (Erawati & Adnyana, 2024). Through real-time interaction and feedback, interactive systems help users learn through experience and understanding the concepts. This is particularly important in domains involving abstract phenomena, such as sound, where learners benefit from multimodal representations.

Audio Visualization and Data Mapping

Audio visualization is the process of transforming sound signals into visual representations. For audio features such as amplitude, frequency spectrum, and rhythm can be represented visually through motion, scale, color, and distortion (Usher, n.d.). This mapping process allows users to perceive sound in a visual form, thereby enhancing comprehension. Studies have shown that visualizing sound structures can help improve pattern recognition and deepen conceptual understanding (Adiletta & Thomas, n.d.). However, many existing systems rely on passive visualization, where users cannot influence the output. This limits the potential for interactive learning. In contrast, real-time visualization systems use continuous data streams to dynamically update visual outputs. In node-based environments such as TouchDesigner, audio data is processed through CHOP (Channel Operators) and mapped to TOP (Texture Operators), enabling real-time audiovisual synthesis. Dynamic particle systems are often created using techniques such as noise generation, displacement mapping, and feedback loops (Graf et al., 2021).

Gesture Interaction and Motion Tracking

Gesture-based interaction enables users to control digital systems through physical movement. This form of interaction is a key area within human–computer interaction (HCI), particularly in the domain of embodied interaction, where bodily actions serve as input mechanisms. Modern gesture recognition systems, such as MediaPipe, utilize computer vision and machine learning to detect and track hand movements from standard RGB camera input (Zoppis et al., 2025). These systems generate spatial data (e.g., x, y, z coordinates of hand joints), which can be used to control system parameters. Gesture interaction enhances usability by enabling intuitive control and reducing the need for traditional input devices. According to Yusof and Ismail (2020), direct manipulation interfaces improve user engagement by providing immediate feedback and continuous interaction. This makes gesture-based systems particularly suitable for creative and exploratory learning environments (Jain et al., 2021).

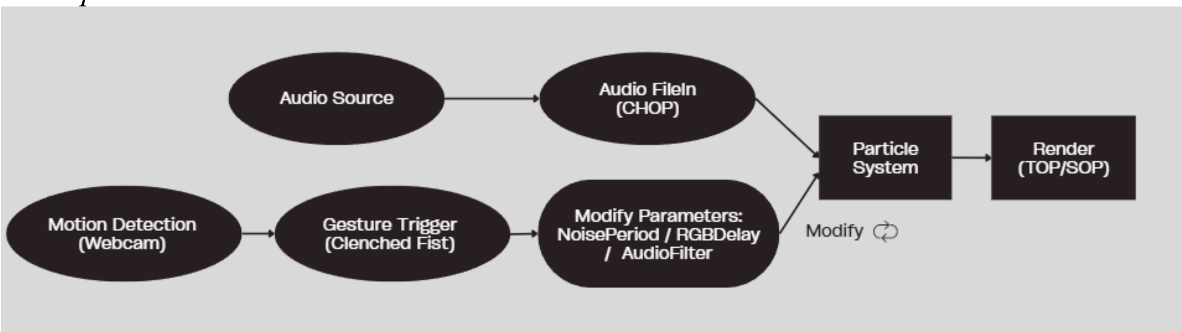
Technology-Enhanced Learning and Multimodal Interaction

Technology-Enhanced Learning (TEL) involves the use of digital tools to support and improve learning processes. Interactive systems that integrate multiple input modalities such as sound and gesture are particularly effective in TEL environments (Rahmah et al., 2022), as they provide rich, immersive experiences.

Multimedia learning theory suggests that combining visual and auditory information can significantly enhance learning outcomes (Daniela, 2021). Allowing learners to engage with sound and visual elements at the same time can help them better understand complex relationships between different variables. In creative media education, tools such as TouchDesigner provide opportunities to explore abstract concepts through real-time interaction and visualization. Through the integration of audio, gesture, and visual feedback, these interactive systems can enhance engagement, motivation, and overall understanding. In creative media education, tools such as TouchDesigner allow learners to explore abstract concepts through real-time interaction and visualization. By integrating audio, gesture, and visual feedback, such systems support engagement, motivation, and conceptual understanding (Potts et al., 2022).

Figure 1

Conceptual Framework



Note. This figure illustrates the overall system flow, including audio input, gesture detection, parameter modification, and visual output generation. The system integrates sound and motion data to produce real-time particle-based visualizations.

Research Method

Research Design

The study was conducted in three phases. Phase 1: System Development. The system was developed based on theoretical frameworks and implemented using TouchDesigner. The conceptual framework (Figure 1) illustrates the flow from audio input and gesture detection to visual output. The technical workflow integrates audio processing (AudioFileIn, AudioFilter), gesture tracking (MediaPipe), and visual generation (Noise, Displace, Feedback). Phase 2: Expert Evaluation. Five experts evaluated the system using a 5-point Likert scale. Data were analyzed using mean and standard deviation. Phase 3: User Testing. Thirty participants interacted with the system. Satisfaction was measured using a questionnaire and analyzed statistically.

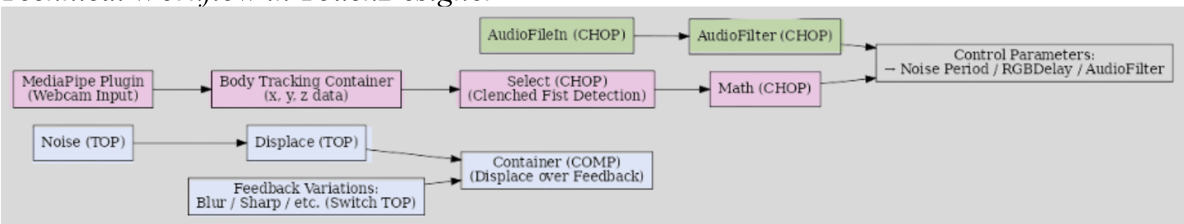
System Overview

The system is designed as a real-time interactive audiovisual platform that integrates three main subsystems: 1) Audio Input and Processing System 2) Gesture Detection and Control System and 3) Visual Generation System. As illustrated in the conceptual framework (Figure 1), audio and gesture inputs are processed simultaneously and mapped to visual outputs in real time.

Technical Implementation

Figure 2

Technical Workflow in TouchDesigner



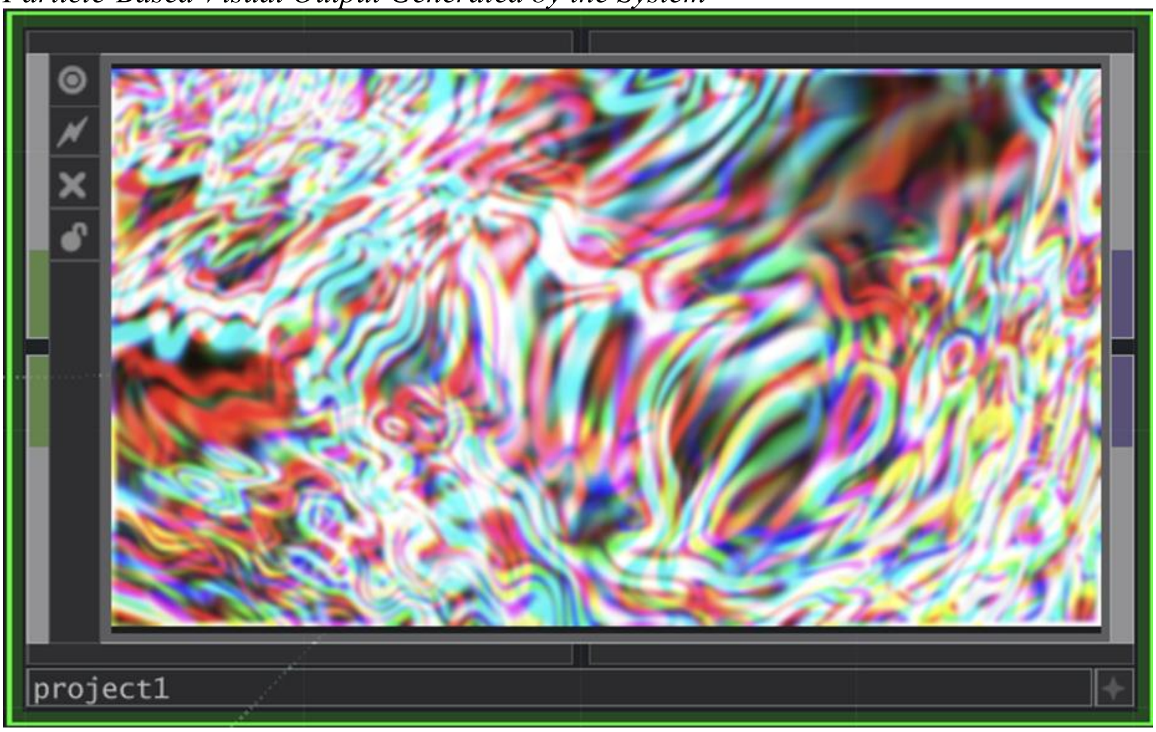
Note. This diagram shows the detailed node-based structure of the system, including audio processing, gesture detection, and visual generation components. The workflow demonstrates how data flows through CHOP and TOP operators to produce real-time visual outputs.

Visual Creation System (Particles and Feedback)

The visual system is built using a particle-based approach in TouchDesigner. The following components are used 1) Noise TOP generates base visual patterns, which act as the initial state of the particle system. 2) Displace TOP modifies the noise texture to create wave-like distortions, simulating motion influenced by audio signals. 3) Feedback TOP introduces temporal continuity by feeding previous frames back into the system, creating motion trails and echo effects. 4) Switch TOP enables dynamic switching between visual styles (e.g., blur, sharpness), controlled by gesture input. 5) Container COMP serves as the main structure that organizes and integrates all visual nodes. This system allows the generation of continuously evolving visual patterns that respond dynamically to input data.

Figure 3

Particle-Based Visual Output Generated by the System



Note. This figure presents an example of the dynamic visual output generated using Noise, Displace, and Feedback TOPs. The visuals evolve continuously based on audio and gesture inputs.

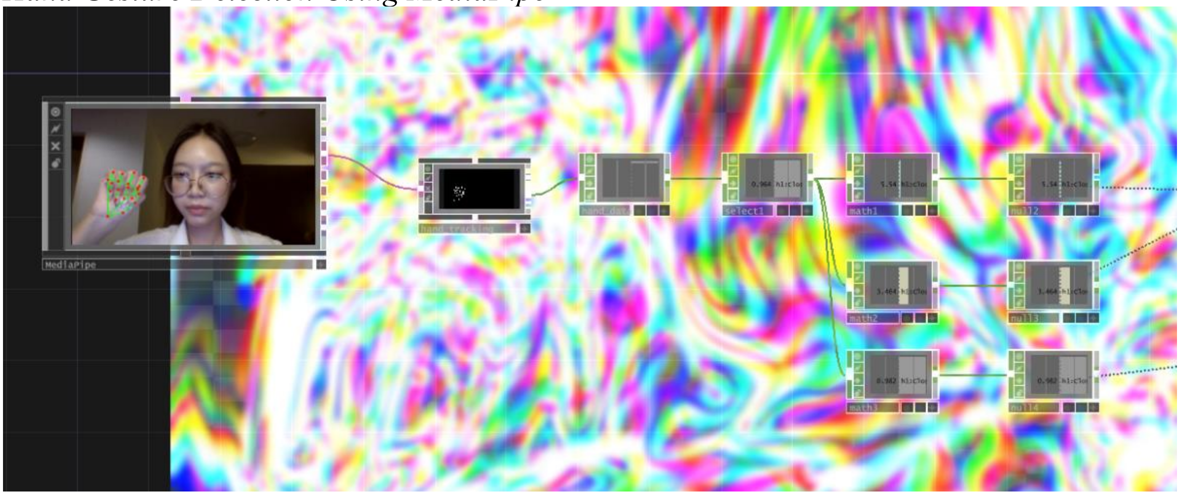
Audio Input and Control System

The audio subsystem processes sound input and extracts features for visual mapping. 1) AudioFileIn CHOP imports audio signals into the system. 2) AudioFilter CHOP modifies audio properties such as volume and frequency distribution. Audio features are then mapped to visual parameters, such as Motion intensity, Particle speed, Color variation. This mapping enables the system to visually represent changes in sound in real time.

Gesture Detection System (MediaPipe Integration)

Gesture detection is implemented using MediaPipe, which processes webcam input to track hand movements. 1) MediaPipe Plugin captures real-time video input. 2) Body Tracking Container extracts spatial coordinates (x, y, z) of hand joints. 3) Select CHOP filters specific gestures, such as clenched fist detection. 4) Math CHOP scales gesture data into usable control values. The processed gesture data is then used to control system parameters, including Noise Period, RGB Delay, AudioFilter values. This enables users to control and explore the system naturally through hand movements.

Figure 4
Hand Gesture Detection Using MediaPipe



Note. This figure illustrates the real-time hand tracking process using a webcam. The system detects hand landmarks and extracts spatial coordinates (x, y, z) for gesture-based control.

Data Mapping and Real-Time Interaction

The system integrates audio and gesture data through parameter mapping, which defines how input variables influence visual outputs in real time.

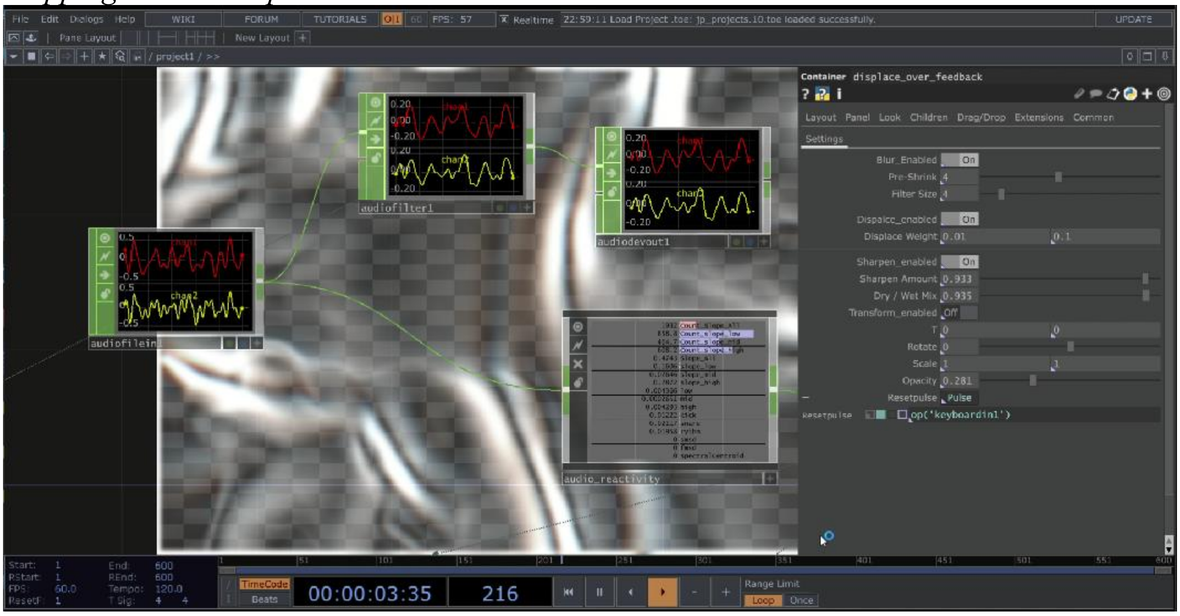
For example:

- Increased audio amplitude results in higher motion intensity and stronger visual distortion within the particle system.
- A clenched fist gesture activates increased visual complexity, including more chaotic movement patterns and enhanced color variation.

This mapping enables users to directly observe how changes in input affect visual behavior, reinforcing cause-and-effect relationships. The system operates in real time, ensuring immediate feedback, which is critical for user engagement and supports experiential learning.

Figure 5

Mapping Between Input Data and Visual Parameters



Note. This diagram shows how audio and gesture data are mapped to visual parameters such as motion intensity, color variation, and distortion effects in real time.

Results

The results of this study are presented in accordance with the research objectives. Regarding the first objective, the Interactive Sound Wave system was successfully developed as a real-time interactive audiovisual platform. The system enabled the integration of audio input and gesture-based interaction, resulting in dynamic particle-based visualizations. It demonstrated stable real-time responsiveness and allowed users to intuitively manipulate visual outputs through sound and hand gestures, enhancing the perceptibility of abstract audio concepts.

The evaluation results for the second objective showed that the developed media was considered to be of good quality and received very high user satisfaction ratings. As shown in Table 1, the overall media quality assessed by five experts was at a good level ($\bar{x}$ = 4.20, S.D. = 0.47). Interaction consistency received the highest rating, suggesting that the mapping between sound, gesture, and visual output was coherent and effectively designed. Other aspects were consistently rated at a good level, reflecting balanced performance in usability, visual design, and educational applicability.

In addition, the user satisfaction results in Table 2 revealed an overall very high level ($\bar{x}$ = 4.48, S.D. = 0.50). Engagement and interaction received the highest rating, indicating that the system effectively captured user interest and encouraged active participation. The integration of sound, gesture, and visual elements also helped users better understand the concepts presented in the system. In addition, users gave high ratings for ease of use, enjoyment, and suitability for learning activities. Overall, the findings indicate that the system is both engaging and effective for use in technology-enhanced learning environments.

Figure 6

User Interaction With the Interactive Sound Wave System

Note. This figure shows participants interacting with the system using sound and hand gestures in a real-time environment.

Results of Media Quality Assessment and User Satisfaction

The results of media quality assessment and user satisfaction are presented as follows. The media quality was evaluated by five experts in the fields of education and information technology, while user satisfaction was assessed with 30 participants.

Table 1

Media Quality Evaluation Results by Experts

Item	Quality Assessment Level		
	$\bar{x}$	S.D.	Interpretation
Interaction consistency (sound–gesture–visual mapping)	4.65	0.35	Very Good
Clarity of user experience and control	4.25	0.44	Good
Visual quality and aesthetics	4.10	0.50	Good
Suitability for learning context	4.05	0.52	Good
Classroom feasibility	3.95	0.55	Good
Total	4.20	0.47	Good

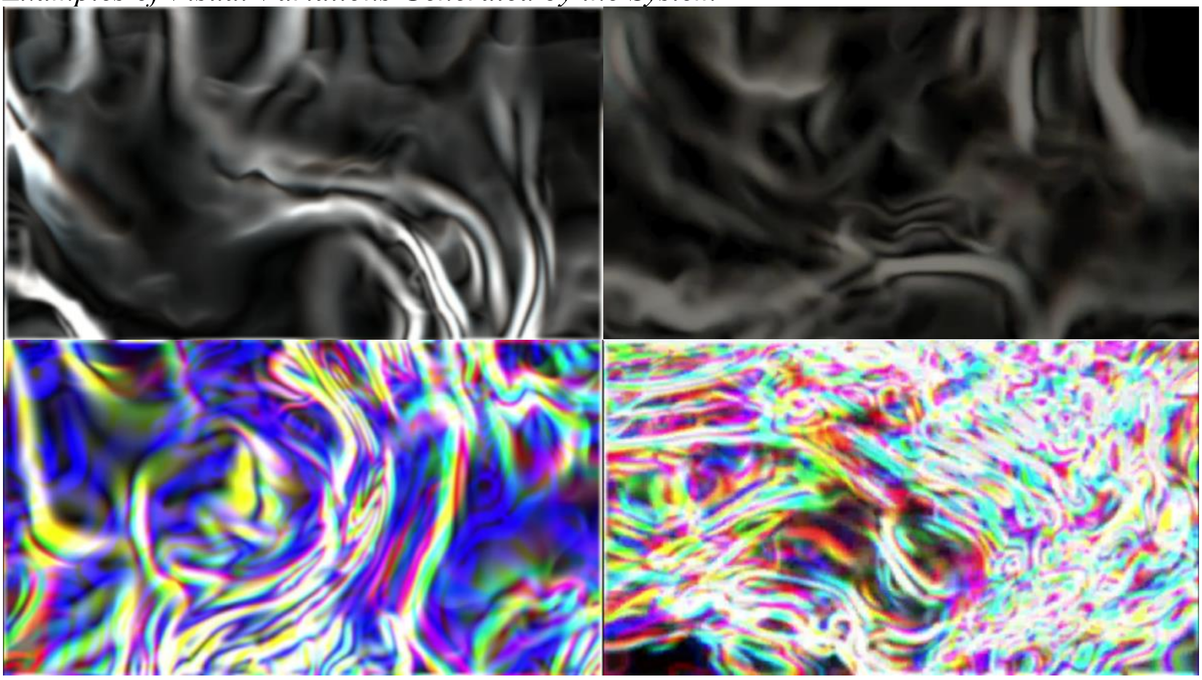
From Table 1, the result of overall media quality is at a good level. The mean was 4.20 and the standard deviation was 0.47. When considering each aspect, it was found that Interaction consistency (sound–gesture–visual mapping) had a very good level of quality. The mean was 4.65 and the standard deviation was 0.35. The topics Clarity of user experience and control, Visual quality and aesthetics, Suitability for learning context, and Classroom feasibility were at a good level. The mean values were 4.25, 4.10, 4.05, and 3.95. The corresponding standard deviations were 0.44, 0.50, 0.52, and 0.55.

Table 2
User Satisfaction Results

Item	Quality Assessment Level		
	$\bar{x}$	S.D.	Interpretation
Engagement and interaction	4.82	0.38	Very High
Clarity of linking sound–gesture–visuals for concept understanding	4.70	0.42	Very High
Ease of use and control	4.35	0.55	High
Enjoyment and learning experience	4.30	0.57	High
Applicability in learning activities	4.25	0.60	High
Total	4.48	0.50	High

From Table 2, the result of overall user satisfaction is at a very high level. The mean was 4.48 and the standard deviation was 0.50. When considering each aspect, it was found that Engagement and interaction had the highest level of satisfaction. The mean was 4.82 and the standard deviation was 0.38. The topic Clarity of linking sound–gesture–visuals for concept understanding also showed a very high level. The mean was 4.70 and the standard deviation was 0.42. The topics Ease of use and control, Enjoyment and learning experience, and Applicability in learning activities were all at a very high level. The mean values were 4.35, 4.30, and 4.25. The corresponding standard deviations were 0.55, 0.57, and 0.60, respectively.

Figure 7
Examples of Visual Variations Generated by the System



Note. This figure presents different visual outputs generated under varying audio and gesture inputs, demonstrating the system's responsiveness and diversity.

Discussion and Conclusion

The findings of this study support the effectiveness of integrating real-time sound–gesture interaction within a technology-enhanced learning environment. The developed Interactive Sound Wave system successfully transformed abstract audio concepts into observable and interactive visual representations. This is evident from the high level of user engagement and interaction ($\bar{x}$ = 4.82), indicating that users were actively involved in manipulating sound and gesture inputs. This reflects constructivist learning principles, where learners develop understanding through active participation and direct experience (Pauline-Graf et al., 2021). In this context, the system supports intuitive and perceptual learning by enabling users to explore the relationship between sound and visual behavior through real-time interaction.

The results of the expert evaluation further confirm the quality and applicability of the developed media. The overall media quality was rated at a good level ($\bar{x}$ = 4.20), with interaction consistency receiving the highest score ($\bar{x}$ = 4.65). This suggests that the mapping between sound, gesture, and visual output was coherent and effectively designed. These findings are consistent with multimedia learning theory, which highlights the importance of integrating visual and auditory elements to enhance understanding (Meena & Arya, 2023). The high ratings in visual quality and interactivity also indicate that the system provides an engaging and intuitive user experience, which is a key factor in promoting active learning.

From an educational perspective, the system can be applied as an interactive learning tool in multimedia, sound design, and creative media education. The system allows learners to explore fundamental sound-related concepts such as intensity, rhythm, and dynamic variation through hands-on interaction and real-time feedback. By observing how changes in sound and gesture affect visual outputs, learners can better understand the relationships between actions and outcomes, an important aspect of experiential learning. For example, users can experiment with different sound inputs and immediately observe how the visual patterns respond, helping them develop a deeper understanding through direct experience. In addition, the system can be used in classroom demonstrations, workshops, or exploratory learning activities to encourage engagement, collaboration, and creative experimentation. Although the system does not aim to provide precise analytical representations of sound, it effectively supports conceptual understanding and perceptual awareness, making it particularly suitable for introductory and practice-based learning environments.

However, some aspects, such as suitability for learning context and classroom feasibility, were rated at a good level rather than very good. This may indicate that while the system is technically effective, further refinement is needed to better align it with specific instructional objectives and classroom conditions. The system does not aim to represent sound in a precise scientific or measurable form, but rather to support perceptual and experiential understanding through interaction. Future development could focus on incorporating structured learning activities, guided interaction scenarios, or curriculum integration to enhance its pedagogical effectiveness. Additionally, integrating analytical features such as visual indicators of amplitude or frequency could extend the system toward a hybrid learning approach that supports both conceptual and analytical understanding.

In conclusion, this study demonstrates that real-time, multimodal interaction combining sound visualization and gesture control can serve as an effective approach in technology-enhanced learning. This research highlights the potential of real-time sound–gesture interaction in supporting experiential learning. Through direct interaction, users are able to explore the relationships between sound, gesture, and visual behavior in a more engaging and intuitive way. The findings also suggest that interactive systems of this kind can help make abstract concepts more accessible in multimedia education. Future studies may explore broader user groups, varied learning environments, and expanded system features to further develop and validate this approach.

Acknowledgement

The authors sincerely appreciate the support provided by the Department of Computer and Information Technology, Faculty of Industrial Education and Technology, King Mongkut's University of Technology Thonburi, Bangkok, Thailand.

References

Adiletta, M. J., & Thomas, O. (n.d.). *An Artistic Visualization of Music: Modeling a Synesthetic Experience*.

Daniela, L. (2021). Smart Pedagogy as a Driving Wheel for Technology-Enhanced Learning. In *Technology, Knowledge and Learning* (Vol. 26, Number 4, pp. 711–718). Springer Science and Business Media B.V. https://doi.org/10.1007/s10758-021-09536-z

Erawati, N. K., & Adnyana, P. B. (2024). IMPLEMENTATION OF JEAN PEAGET'S THEORY OF CONTRUCTIVISM IN LEARNING: A LITERATURE REVIEW. *Indonesian Journal of Educational Development (IJED)*, *5*(3), 394–401. https://doi.org/10.59672/ijed.v5i3.4148

Graf, M., Opara, H. C., & Barthet, M. (2021). *An Audio-Driven System For Real-Time Music Visualisation*. http://arxiv.org/abs/2106.10134

Jain, R., Jain, M., Jain, R., & Madan, S. (2021). Human Computer Interaction – Hand Gesture Recognition. *Advanced Journal of Graduate Research*, *11*(1), 1–9. https://doi.org/10.21467/ajgr.11.1.1-9

Meena, Y. K., & Arya, K. V. (2023). Multimodal interaction and IoT applications. In *Multimedia Tools and Applications* (Vol. 82, Number 4, pp. 4781–4785). Springer. https://doi.org/10.1007/s11042-022-14278-w

Pauline-Graf, D., Mandel, S. E., Allen, H. W., & Devnew, L. E. (2021). Assumption validation process for the assessment of technology-enhanced learning. *Contemporary Educational Technology*, *13*(4). https://doi.org/10.30935/cedtech/11071

Potts, D., Dabravalskis, M., & Houben, S. (2022, February 13). TangibleTouch: A Toolkit for Designing Surface-based Gestures for Tangible Interfaces. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3490149.3502263

Rahmah, A., Santoso, H. B., & Hasibuan, Z. A. (2022). Critical Review of Technology-Enhanced Learning using Automatic Content Analysis Case Study of TEL Maturity Assessment Formulation. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 13, Number 1). www.ijacsa.thesai.org

Sharkey, P. M. (2008). *The 7th International Conference on Disability, Virtual Reality and Associated Technologies with ArtAbilitation : proceedings : 8 to 11 of September, 2008, Maia, Portugal*. ICDVRAT, University of Reading.

Usher, M. E. (n.d.). *VISUALIZATION OF SOUNDS: IMPROVING THE LISTENING EXPERIENCE THROUGH DIGITAL VISUALIZATION*.

Yusof, C. S., & Ismail, A. W. (2020). Virtual Block Augmented Reality Game Using Freehand Gesture Interaction. *International Journal of Innovative Computing*, *10*(2). https://doi.org/10.11113/ijic.v10n2.266

Zoppis, Q., Pujades, S., Nigay, L., & Bérard, F. (2025). Efficient Finger Model and Accurate Tracking for Hover-and-Touch, Mid-air and Microgesture Interaction Efficient Finger Model and Accurate Tracking for Hover-and-Touch, Mid-air and Microgesture Interaction Microgesture Hover-And-Touch Mid-Air (A) (B) (C) (B) (A, C). *The 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25), September 28-October 1, 2025, Busan, Republic of Korea*, *1*, 1–14. https://doi.org/10.1145/3746059.3747701ï