International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 807-811
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0113

AI Presentation Tool Using Gesture Recognition

Pradnya K. Bachhav[1], Ayesha Shaikh[2], Pooja Gupta[3], Shiba Shaikh[4], Anushka Jadhav[5]
[1]Assistant Professor, Department of Computer Engineering, Guru Gobind Singh College of Engineering and Research Centre, Nashik, Maharashtra, India.
[2,3,4,5]Student, Department of Computer Engineering, Guru Gobind Singh College of Engineering and Research Centre, Nashik, Maharashtra, India.
Emails: *shaikhaayesha26@gmail.com[2]*

Abstract
The proposed idea AI Presentation Tool Using Gesture Recognition introduces an innovative AI-driven presentation tool that utilizes hand gestures for controlling PowerPoint slides, aiming to transform the presentation experience. By enabling presenters to interact with their slides through natural movements, the tool eliminates the reliance on traditional physical devices, such as remotes or keyboards, fostering a more fluid and engaging atmosphere. The system employs advanced computer vision and deep learning techniques to accurately recognize hand gestures in real-time, providing a hands-free solution that is particularly beneficial for users with physical disabilities. The proposed tool not only simplifies the act of presenting but also enhances accessibility, making it an ideal fit for diverse environments, including classrooms, corporate meetings, conferences, and virtual settings. By leveraging recent advancements in AI technology, the proposed idea aspires to redefine the standards of interactivity and inclusivity in presentation tools, making presentations more dynamic and accessible to a wider audience.
Keywords: *AI, Presentation Tool, Computer Vision, Classroom Teaching, Speech Recognition, Machine Learning.*

1. Introduction
Presentations are a vital part of professional and educational communication. How- ever, traditional tools like clickers and keyboards often interrupt the natural flow, diverting focus from the audience to controlling slides. These devices can limit interactivity and present accessibility challenges for individuals with disabilities, who may find such tools cumbersome or unusable. To overcome these challenges, we propose an AI-powered presentation tool that uses gesture recognition for hands-free slide control. By leveraging advanced computer vision and deep learning, this system interprets simple hand movements in real time, enabling presenters to navigate slides smoothly without physical controllers. This not only enhances the presenter's connection with the audience but also makes presentations more inclusive and engaging. The tool's intuitive interface ensures a dynamic presentation environment, allowing users to focus entirely on their message. Its inclusivity extends to professionals, educators, and individuals with disabilities, providing an equal platform for delivering impactful presentations. Beyond slide control, the system could support complex interactions, such as triggering multimedia elements or adjusting settings through specific gestures. As AI and machine learning continue to revolutionize human-computer interaction, this project represents a step towards more intuitive, hands-free technology, transforming not only how we present but also how we engage with digital content across various domains, from education to entertainment and assistive technology.

2. Literature Survey
[1] Explores gesture-based presentation control using machine learning, improving engagement and accessibility. Challenges include misinterpreted gestures and low-light performance. Future work suggests refining recognition algorithms for broader adaptability. [2] Focuses on gesture recognition in education and presentations, enhancing hands-free interactivity. Limitations include environmental

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 807-811
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0113

dependency, requiring further adaptation for diverse settings. [3] Uses CNNs and RNNs for real-time gesture recognition, achieving high accuracy but struggling with fast or subtle movements. Future work aims to refine sensitivity for better gesture distinction. [4] Integrates gesture, voice commands, and multimedia control in presentations, reducing cognitive load. Challenges include complex gestures and environmental variations. Further refinements needed for edge cases. [5] Combines gesture and touch technologies for collaborative presentations, enhancing engagement. However, multi-user differentiation remains a challenge, requiring optimization for smoother interaction. [6] Uses Kinect for depth-sensing gesture recognition, excelling in low-light environments but limited by hardware dependency. Future improvements should focus on more flexible, device-independent solutions. [7] Automates presentation control with computer vision-based hand gestures, enhancing convenience. Lighting and camera quality affect accuracy, suggesting hardware optimizations and algorithm refinement. [8] Introduces edge detection for subtitle recognition in videos, enabling efficient extraction even with complex backgrounds. Struggles with fast-moving text and low contrast, requiring additional image processing techniques. [9] Evaluates open-source speech recognition for social robots, highlighting real-time capabilities and noise sensitivity. Suggests combining multiple models for robustness in interactive applications. [10] Develops a machine learning-based gesture recognition system, using visual interest points for accurate tracking. Enhances real-time hand gesture detection in video sequences.

3. System Architecture

The Figure 1 architecture of the AI Presentation Tool is meticulously designed to integrate advanced technologies, ensuring a seamless and efficient user experience. This architecture is organized into four primary layers: The User Interface Layer, the Application Logic Layer, the Data Management Layer, and the Hardware Layer. Each layer contributes distinct functionalities while collectively enabling the tool's advanced capabilities.

1. **User Interface Layer:** The User Interface Layer serves as the primary interaction point for users, offering an intuitive and user-friendly graphical interface. Built using the Tkinter library, this layer provides essential controls that allow users to:

- Configure preferences for gestures, voice commands, and subtitle settings, View and navigate through slides during presentations.
- Manage and customize gestures and voice commands for specific presentation actions.
- This layer emphasizes simplicity and accessibility, ensuring that even users with minimal technical expertise can leverage the tool effectively.

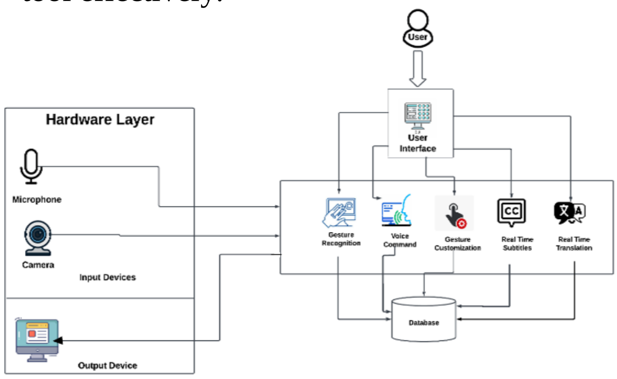

Figure 1 System Architecture

2. **Application Logic Layer:** The Application Logic Layer forms the functional core of the AI Presentation Tool, integrating multiple modules to deliver its innovative features. Key components include:

- **Gesture Recognition Module:** Powered by MediaPipe, this module interprets user gestures in real-time and maps them to specific actions, such as advancing or reversing slides.
- **Voice Command Module:** Using the Speech Recognition library, this component processes spoken commands to perform various tasks, such as starting or pausing the presentation.
- **Gesture Customization:** This feature empowers users to define and save personalized gestures for specific actions, enhancing adaptability and user control.
- **Real-Time Subtitles:** By transcribing spoken

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 807-811
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0113

content, this module improves accessibility, particularly for audiences with hearing impairments or those in noisy environments.

- **Real-Time Translation:** This feature enables multilingual presentations by translating spoken content into various languages on the fly, broadening the tool's usability across diverse linguistic contexts.

3. **Data Management Layer:** The Data Management Layer underpins the application by managing and organizing user data. It utilizes a robust database to store:

- User profiles, including preferences for gestures and voice commands. Gesture definitions and associated mappings to actions. Voice command mappings and language settings. This centralized storage ensures a personalized user experience while enabling data analytics to track and enhance user engagement.

4. **Hardware Layer:** The Hardware Layer integrates the physical components required to run the AI Presentation Tool, ensuring real-time performance. It includes:

- **Input Devices:** A camera for detecting gestures and a microphone for processing voice commands. These devices serve as the primary input sources, enabling hands-free and voice-driven interactions.
- **Output Devices:** A display screen for displaying presentation slides and a computing system to execute the application's functionalities. These components ensure a smooth and responsive user experience, even during complex interactions.

The interaction within the AI Presentation Tool begins with user inputs through gestures or voice commands, captured by the camera and microphone, respectively. The Application Logic Layer processes these inputs in real time, where the Gesture Recognition Module utilizes MediaPipe to identify specific gestures and map them to predefined actions, such as advancing slides or pausing the presentation. Simultaneously, the Voice Command Module employs the SpeechRecognition library to interpret spoken instructions and trigger appropriate responses. Any additional configurations or actions initiated through the Tkinter GUI, such as setting preferences or selecting slides, are seamlessly integrated into the system. Processed inputs are further validated and personalized by fetching relevant data from the Data Management Layer, including user profiles, gesture mappings, and voice command settings. Finally, the interpreted actions are executed and reflected on the output devices displaying slides, updating subtitles, or translating content ensuring a responsive and interactive user experience.

4. Results and Discussion

4.1. Performance Evaluation

To assess the effectiveness of our AI-powered Presentation Tool, we evaluated key parameters related to system responsiveness, accuracy, and user interaction. The evaluation criteria were inspired by prior research and adapted to measure gesture recognition, voice command responsiveness, and real-time translation accuracy, shown in Table 1.

Table 1 Parameter Description & Measured Time

Parameter	Description	Measured Time
GRT (Gesture Recognition Time)	Time taken to recognize and process a gesture input.	0.2s
VRT (Voice Recognition Time)	Response time for processing voice commands and executing corresponding actions.	2.0s
RST (Real-time Subtitles Processing)	Delay in generating subtitles from spoken input.	1.0s
RTT (Real-time Translation Latency)	Time taken to translate speech into another language.	2.0s
UILT (User Interface Load Time)	Time required the application dashboard to fully load.	0.8s

These results indicate that the tool provides fast and efficient recognition and translation, maintaining

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 807-811
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0113

real-time interaction with minimal delay, Shown in Figure2.

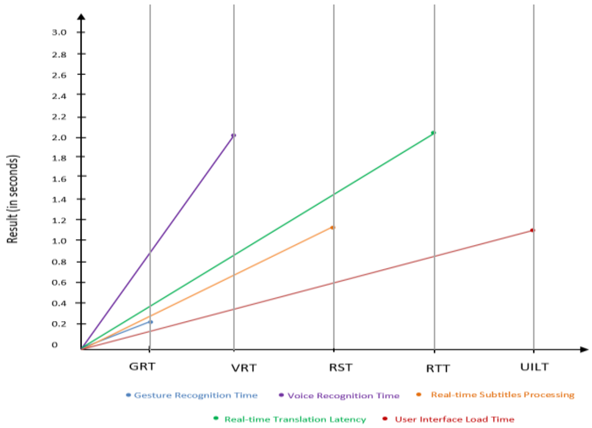

Figure 2 Results

4.2. Accuracy Analysis

We conducted a real-time user testing phase with three participants. Each participant performed multiple trials for gesture recognition and voice commands. The accuracy results are summarized below in Table 2:

Table 2 Accuracy Analysis

Feature	User 1	User2	User3	Accuracy
Gesture Recognition	9/10	10/10	10/10	96.67%
Voice Command	7/10	10/10	8/10	83.33%
Real-time Subtitles	10/10	7/10	10/10	90.00%
Real-time Translation	8/10	8/10	10/10	86.67%
Overall Accuracy	85.00 %	87.50 %	95.00% /	89.17%

4.3.Discussion

Gesture recognition achieved a remarkable accuracy of 96.67%, making it a highly reliable feature for seamless presentation control. Voice command accuracy stood at 83.33%, demonstrating strong performance, though some variations were observed due to differences in speech patterns. Real-time subtitles maintained an impressive 90.00% accuracy, ensuring clear and precise text generation, while real-time translation reached 86.67%, reflecting slight challenges in handling speech variations and language processing complexities. Overall, the system performed efficiently, with an average accuracy of 89.17%, and optimized processing speeds minimized delays in UI loading and recognition, ensuring a smooth and interactive user experience.

Future Scope

The future scope of the AI-powered presentation tool lies in enhancing gesture recognition accuracy, real-time adaptive learning, and multilingual support to create a seamless and intuitive user experience. Advancements in machine learning and deep learning will enable personalized gesture training, adaptive voice commands, and improved real-time translation for diverse linguistic and cultural needs. Integrating the tool with AR/VR environments and IoT devices can revolutionize interactive presentations, making them more immersive and engaging. Additionally, AI-driven analytics can provide insights on presenter performance and audience engagement, offering real-time feedback for improvement. Future developments will also focus on accessibility features, such as sign language recognition and assistive voice control, ensuring inclusivity for individuals with disabilities. As AI continues to evolve, this tool has the potential to redefine human-computer interaction in professional, academic, and virtual presentation spaces, making communication more efficient, engaging, and accessible.

Conclusion

The AI-driven presentation tool utilizing gesture recognition and voice commands represents a major leap in interactive presentation technology by offering a hands-free, inclusive solution that enhances engagement and accessibility. By integrating advanced AI, computer vision, and natural language processing, the tool redefines how users interact with presentations, making it especially beneficial for individuals with disabilities. Its modular design, real-time gesture control, voice interactions, and support for dynamic multimedia elements cater to diverse settings, including corporate and educational environments. With features like real-time subtitle generation and multilingual translation, the tool is adaptable for global use, making a significant contribution to human-computer interaction and assistive technologies.

References

[1]. R. M. Shakya, S. P. Sharma, and N. Shrestha, "The Use of Hand Gestures as a Tool for Presentation," International Journal of Advanced Computer Science and Applications, vol. 14, no. 11, 2023.

[2]. S. R. Patel, "Gesture Recognition: Applications in Education and Presentations," IEEE Transactions on Human-Machine Systems, vol. 53, no. 3, pp. 250-260, 2023.

[3]. M. K. Jha and A. D. Gupta, "A Study on Gesture Recognition in Interactive Presentations," ITM Conference Proceedings, ICACC 2022.

[4]. A. Lee and B. Chan, "Enhancing Presentations with Gesture Recognition Technologies," IEEE Access, vol. 10, pp. 500-510, 2022.

[5]. J. P. Brown, "The Future of Interactive Presentations: Gesture and Touch," IEEE Transactions on Education, vol. 65, no. 2, pp. 120-130, 2022.

[6]. L. Zhang, "Utilizing Kinect for Enhanced Presentation Control," IEEE Transactions on Multimedia, vol. 24, pp. 1-10, 2022.

[7]. Salonee Powar, Shweta Kadam, Sonali Malage, and Priyanka Shingane, "Automated Digital Presentation Control using Hand Gesture Technique," 2022.

[8]. College of Computer and Cyber Security, Communication University of China, "Video Subtitle Location and Recognition Based on Edge Features," 2022.

[9]. O. Mubin, J. Henderson, and C. Bartneck, "A Comparative Analysis of Real-Time Open-Source Speech Recognition Tools for Social Robots," 2020.

[10]. Toshiyuki Kirishima, Kosuke Sato, and Kunihiro Chihara, "Real-Time Gesture Recognition by Learning and Selective Control of Visual Interest Points," March 2005