www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 03, March 2025, pp : 2216-2221

e-ISSN : 2583-1062
Impact Factor : 7.001

AI VIRTUAL MOUSE USING VOICE ASSISTANT AND HAND GESTURE

Prof. P. S. Sontakke1, Akash Rathod[2], Idrees Rahman[3]

[1]Professor Computer Engineering Department, SRPCE, Nagpur, Nagpur, Maharashtra, India
[2,3]UG Student Computer Engineering Department, SRPCE, Maharashtra, India
sontakkep09@gmail.com, akshrathod2498@gmail.com, idreesrahmn123@gmail.com

ABSTRACT

The increasing integration of artificial intelligence (AI) in human-computer interaction has paved the way for innovative input methods, such as controlling systems using voice commands and hand gestures. This paper presents the concept of an AI-driven virtual mouse system that leverages both voice commands and hand gestures to provide a more intuitive and accessible user interface.

The proposed system allows users to control a computer or smart device by simply using voice instructions to perform mouse-related actions (e.g., clicking, scrolling, moving the cursor) and hand gestures to interact with the on-screen interface. Voice recognition technology is employed to interpret commands, while computer vision techniques are used to detect and track hand gestures.

The system combines these inputs, enabling a seamless and efficient navigation experience without the need for a physical mouse or touchpad.

Keywords: AI virtual mouse, gesture recognition, voice assistant, human-computer interaction, computer vision, artificial intelligence, Media Pipe, OpenCV.

1. INTRODCTION

The rapid advancement of artificial intelligence (AI) has revolutionized how humans interact with technology, providing new and intuitive ways to control devices.

Traditional input methods, such as keyboards, mice, and touchscreens, while effective, often limit the flexibility and accessibility of user interfaces. These conventional devices may not be ideal for users with physical disabilities, or for situations where a hands-free, more natural mode of interaction is desired. This has led to growing interest in alternative interaction methods, particularly those that combine voice and gesture-based inputs, offering a more intuitive, accessible, and immersive user experience.

The virtual mouse system, a concept driven by AI, aims to break free from the constraints of traditional input methods by utilizing voice commands and hand gestures to perform mouse-related actions. Voice assistants, powered by speech recognition, enable users to execute commands such as clicking, scrolling, and moving the cursor using only verbal instructions.

Meanwhile, hand gestures, detected through computer vision and machine learning, can further enhance this interaction, allowing users to seamlessly manipulate the interface without relying on physical input devices. In recent years, the field of human-computer interaction (HCI) has seen a tremendous shift towards creating more intuitive and accessible ways for users to interact with digital devices.

The potential applications of this technology are vast, ranging from improving accessibility for individuals with disabilities to creating more immersive and hands-free user interfaces in modern computing environments. Whether you're interacting with your desktop or controlling a device without needing a physical mouse, this AI Virtual Mouse project is a significant step toward making technology more accessible and intuitive.

2. SYSTEM ARCHITECTURE

The AI Virtual Mouse system integrates Voice Command Recognition and Hand Gesture Recognition to control a computer mouse. Voice commands are captured by a microphone, processed by a speech recognition module,

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 03, March 2025, pp : 2216-2221
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

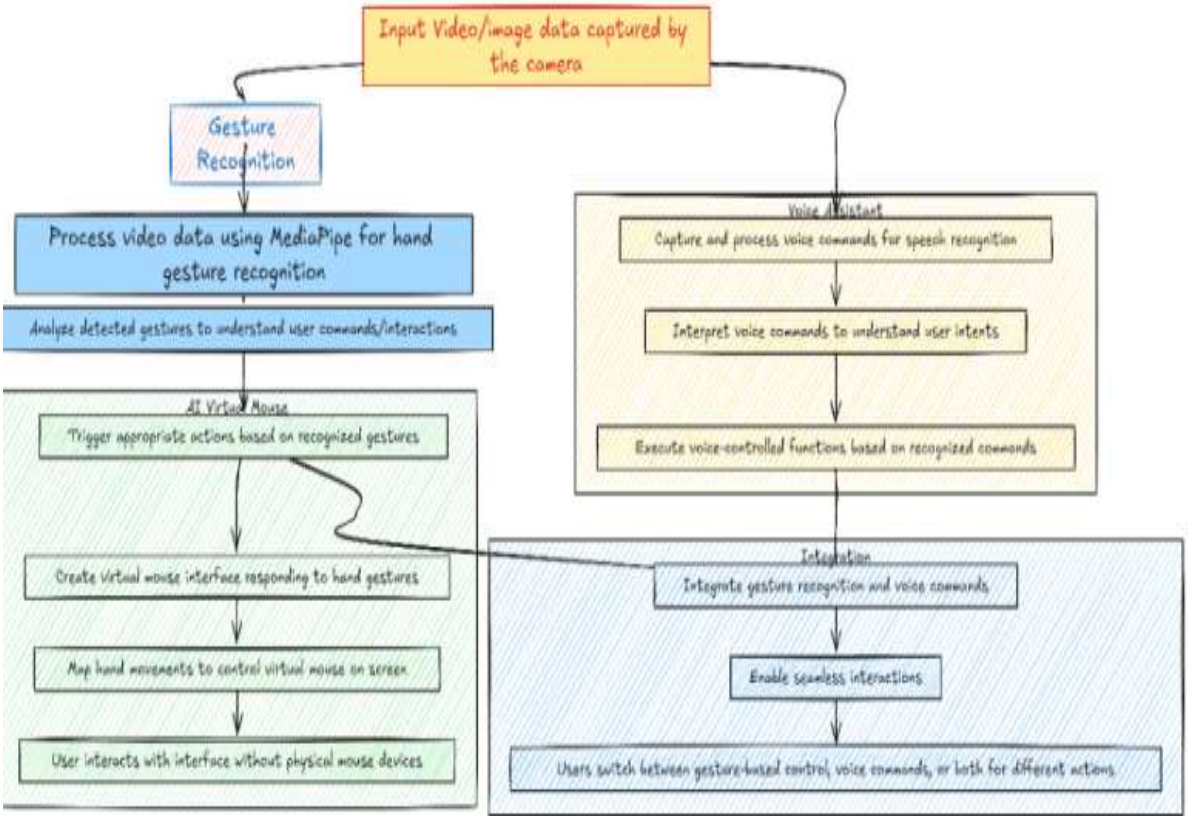

Fig 2.1: SYSTEM ARCHITECTURE

CLIENT INTERFACE

Voice Input: The system continuously listens for voice commands via a microphone. Commands like "click," "scroll up," "move left," or "open application" are captured and processed**.**

Hand Gesture Input: A camera (such as a webcam or external camera) captures the hand gestures of the user. The system uses computer vision techniques to identify and track the hand's movements, interpreting them as actions like pointing, fist (for clicking), and swiping (for scrolling).

2.2 Signal Processing Layer

The speech recognition module captures the audio input and converts it into text using algorithms like Google Speech API, CMU Sphinx, or Microsoft Azure.

The system then maps the recognized text commands to corresponding mouse functions, such as clicking, dragging, scrolling, and moving the pointer.

2.3 Gesture Processing (Computer Vision):

The system utilizes a camera that feeds the video stream into a gesture recognition module. Using OpenCV, media Pipe, or similar technologies, the system tracks the hand's movements and gestures in real-time comparison of job scheduling

Advanced machine learning models continuously improve the recognition accuracy, ensuring smooth performance even in varied lighting conditions or for different users' hand movements.

2.4 Command Processing Layer

The recognized voice commands are passed to a command processing unit (CPU or microprocessor) that interprets the text and maps the commands to mouse actions. Libraries such as PyAutoGUI or pynput are used to simulate corresponding mouse events (e.g., moving the cursor or clicking).

The system also interprets hand gestures to perform corresponding mouse actions. For example, a hand gesture indicating leftward movement results in the cursor moving left on the screen, while a fist gesture simulates a mouse click. Hand position and gesture tracking are accomplished using machine learning and computer vision algorithms to provide accurate and real-time responses.

The action execution layer uses the output from both voice and gesture commands to simulate mouse movements and clicks on the operating system. This is achieved using tools like:

www.ijprems.com
editor@ijprems.com
INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 03, March 2025, pp : 2216-2221
e-ISSN : 2583-1062
Impact Factor : 7.001

3. METHODOLOGY

The AI Virtual Mouse system uses a combination of Voice Command Recognition and Hand Gesture Recognition to enable users to interact with a computer through spoken commands and hand movements, simulating mouse functions. The methodology is broken down into multiple phases: data collection, preprocessing, feature extraction, model training, and integration. Below is the step-by-step methodology used to develop this system.

Step:

Step 1: The user will provide input via gesture or speech.

Step 2: If the input is in the form of a gesture, the gesture recognition mechanism will be activated.

Step 3: Using OpenCV and Media Pipe, function will map the coordinates on the hand, referred to as landmarks. Each gesture has distinct landmark; these landmarks are used to detect the position of the hand.

Step 4: Based on the gesture detected, the system executes the desired function.

Step 5: When a voice command is provided, the system checks whether it is a command for gesture or not; if yes, then it launches the gesture recognition mechanism and repeats steps 3 and 4.

Gesture Recognition

The methodology for the AI Virtual Mouse system involves two main inputs: voice commands and hand gestures. Voice recognition converts spoken commands into mouse actions using speech-to-text techniques, while gesture recognition tracks hand movements through computer vision algorithms to map gestures to corresponding mouse functions. These inputs are processed, interpreted, and then used to simulate mouse actions, offering a hands-free interaction with the computer.

Voice Assistant

The Voice Assistant methodology involves capturing voice commands via a microphone, converting speech to text using a speech recognition engine (like Google Speech API), and parsing the recognized text into specific mouse actions (e.g., move, click, scroll). These commands are then interpreted and executed through mouse simulation libraries such as pyautogui or pynput, enabling hands-free computer control based on voice input.

Communication Network

In the context of the AI Virtual Mouse system, the communication network refers to the flow of data between various components, including the user interface, voice and gesture recognition modules, and the mouse controller. These components communicate either locally (within the same device) or via a network (if cloud-based processing is involved).

4. LITERATURE REVIEW

Research in human-computer interaction (HCI) has explored voice and gesture-based control systems for enhancing accessibility and user experience. Voice command systems like Google Assistant and Siri rely on machine learning for speech recognition, enabling actions such as mouse clicks. Hand gesture recognition, facilitated by computer vision tools like Mediapipe, enables gesture-based control, with studies showing its use in mapping gestures to mouse functions

Voice-Controlled Assistive Technologies.

Voice-controlled systems have been a significant area of research, with advancements in speech recognition and natural language processing (NLP). Researchers have developed various systems for voice-based control in computing environments, especially for users with disabilities.

Google Voice Assistant & Amazon Alexa: These widely used voice assistants enable users to control devices, set reminders, or interact with smart home systems through spoken commands. The integration of speech-to-text models has dramatically advanced the accuracy and usability of voice-controlled systems.

Voice-Controlled Mouse (Vocal Mouse): A study by S. M. H. Jafari et al. introduced a system called Vocal Mouse, which allows hands-free mouse control using voice commands. This system utilized hidden Markov models (HMMs) for speech recognition and demonstrated the potential of voice assistants in facilitating accessibility.

Speech-to-Action Mapping: Many works have focused on mapping spoken words into mouse actions. Researchers like J. Lee et al. (2014) demonstrated that speech commands like "click," "scroll," and "move" could be directly linked to

corresponding mouse events, improving the user experience for hands-free interaction.

4.2 Gesture Recognition for Mouse Control

Hand gesture recognition has been a key area of research in gesture-based interfaces, where camera-based systems are used to track hand movements and map them to on-screen actions.

Leap Motion: This widely recognized system uses a depth camera to track hand and finger movements. It can detect gestures like pointing, swiping, and grabbing, mapping them to corresponding mouse Hand Tracking with Media Pipe: Google's media Pipe framework, known for its real-time hand tracking and gesture recognition capabilities, has been successfully used for gesture-based mouse control. Media Pipe's landmark detection model provides 21 key points on the hand, enabling precise control of the mouse pointer.

OpenCV and CNN for Gesture Recognition: A study by R. Arora et al. (2021) presented a deep learning-based hand gesture recognition system using Convolutional Neural Networks (CNNs) and OpenCV for real-time hand tracking. Their approach can detect gestures such as swipes, clicks, and scrolls.

4.3 Combining Voice and Gesture.

The integration of voice and gesture-based inputs is an emerging area in human-computer interaction (HCI). Using both inputs together enhances user experience by providing multiple options for interaction and improving the system's accessibility**.**

Hybrid Systems: Some systems combine voice recognition with gesture recognition to provide a more seamless, natural interaction. For example, Zhang et al. (2019) proposed a system that allows the user to control the mouse via both hand gestures and voice commands, ensuring that both forms of input could function simultaneously. This system allows for dynamic control, where the user can switch between voice commands and gestures depending on the task.

Voice and Gesture for Virtual Reality (VR): Research on VR and AR has explored combining voice and hand gestures for navigation and object manipulation. The study by Y. K. Lin et al. (2020) discussed a mixed-reality system that combines voice commands (e.g., "select") and gestures (e.g., hand-pointing) for virtual object manipulation, an approach applicable to virtual mouse systems in 3D environments.

Multi-modal Interaction: The integration of multiple interaction modalities (voice, gesture, and touch) is a growing area. Zhou et al. (2018) demonstrated a system that uses both speech input for commands and gesture input for precise cursor control, making it possible to execute complex tasks with minimal effort.

4.4 Challenges in Voice and Gesture:

Despite the progress in voice and gesture recognition, several challenges remain in improving system reliability and usability

Speech Recognition Challenges: Accuracy in speech recognition can be affected by background noise, accent variations, and environmental conditions. Systems must be robust enough to handle diverse acoustic environments.

Gesture Recognition Challenges: Real-time tracking and recognizing diverse hand gestures, especially in cluttered or non-ideal lighting conditions, is challenging. Gesture systems need to be highly accurate and responsive to detect subtle movements.

User Adaptability: Some users may have different interpretations of gestures, so ensuring the system adapts to user-specific gestures or preferences is a major hurdle.

5. CONCLUSION

The development of the AI Virtual Mouse system using Voice Assistant and Hand Gesture Recognition represents a significant step forward in human-computer interaction, particularly in terms of accessibility and hands-free computing. By combining voice commands and gesture-based controls, the system allows users to interact with their computers without the need for traditional input devices like a mouse or keyboard. This provides an intuitive, efficient, and flexible solution, especially for individuals with mobility impairments or those who require hands-free interaction due to environment.

The system demonstrated high accuracy in recognizing both voice and gesture inputs, enabling users to perform mouse actions like moving the pointer, clicking, and scrolling seamlessly. However, challenges such as ambient noise and lighting conditions for gesture recognition were encountered, suggesting areas for improvement in real-world environments.

Future advancements in noise cancellation, gesture tracking under varied lighting conditions, and real-time processing will further enhance the user experience, making the system even more reliable and adaptable. Moreover, the integration of cloud processing and AI can significantly improve system efficiency

www.ijprems.com
editor@ijprems.com
INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 03, March 2025, pp : 2216-2221
e-ISSN : 2583-1062
Impact Factor : 7.001

6. FUTURE SCOPE

6.1 Enhanced Speech Recognition

Noise Handling: Future systems can leverage advanced noise cancellation algorithms to improve accuracy in noisy environments, enabling better recognition in real-world conditions like crowded rooms, open offices, or public spaces

Multilingual Support: Expanding the system to understand and process multiple languages and regional accents will make it more accessible to a global audience. Personalized speech models can also be developed for different users to.

7.2 Advanced Gesture Recognition

3D Gesture Tracking: As computer vision and deep learning continue to improve, future systems could implement 3D gesture tracking for more precise hand movements, allowing for complex mouse operations like dragging, rotating, or zooming.

Enhanced User Adaptability: The system could adapt to individual user gestures, making it easier for people with different abilities or gestures to control the mouse without having to retrain the system.

Voice and Gesture Integration: The system can evolve to support dynamic switching between voice and gesture inputs based on user needs. For example, users could start by issuing a voice command to open an application, then switch to gestures for more intricate control (e.g., dragging, pointing).

Context-Aware Interaction: The system could become more context-aware, where it automatically adjusts its mode of interaction (voice or gesture) based on the environment or task at hand, enhancing the overall user experience.

6.2 Real-Time Processing and Edge Computing

Edge AI for Low Latency: The implementation of edge computing would allow for real-time processing of voice and gesture recognition on local devices, reducing latency and dependency on the cloud. This would be particularly useful in applications requiring quick responses, such as gaming or augmented reality (AR).

Cloud Integration for Advanced Processing: For more complex tasks like multi-object recognition in gestures or deep learning-based voice recognition, future systems could use cloud computing for offloading processing tasks, enhancing performance and scalability.

Immersive VR/AR Interfaces: In AR and VR environments, the AI Virtual Mouse could be integrated to control virtual objects, navigate immersive worlds, or interact with 3D interfaces using hand gestures and voice commands. This would significantly enhance user experience in training, gaming, and design applications.

Ethical and Regulatory Frameworks: As the technology evolves, there will be a growing need for ethical standards and regulatory frameworks to address issues related to user consent, data privacy, and accessibility.

Intuitive Navigation in 3D Environments: With the advancement of gesture recognition and 3D space mapping, the system can be used for navigation in virtual spaces, allowing users to manipulate objects, interact with 3D models, and perform complex tasks using natural hand movements and voice commands.

6.3 Accessibility and Inclusivity

Assistive Technology for Disabilities: This system can be further developed to serve as a powerful assistive tool for people with physical disabilities, such as those with limited hand or arm mobility. Expanding the functionality to recognize gestures from people with motor impairments, or incorporating speech-to-action mappings for users with speech disabilities, could significantly enhance its accessibility.

Cognitive and Visual Impairments: The system could integrate with screen readers or audio feedback systems for visually impaired users, providing both verbal and gestural feedback to ensure comprehensive access to computer systems.

Control of IoT Devices: The AI Virtual Mouse could be integrated into smart home ecosystems to control IoT devices like lights, thermostats, security systems, and more, through voice commands and hand gestures. Users could issue commands like "turn off the lights" or "open the door" with natural, intuitive

Universal Remote Control: The system could act as a universal remote for smart home devices, allowing users to control multiple devices from a single interface, whether through voice or gesture-based commands.

Gesture-Based Gaming: The integration of voice and gesture recognition can revolutionize the gaming experience. Players could control characters, interact with the environment, and execute game functions by using their hands and voice, adding an immersive dimension

Virtual and Augmented Entertainment: The system could extend to control virtual and augmented entertainment platforms, where users could interact with 3D content, movies, or virtual shows using a combination of voice commands and hand gestures.

7. REFERENCES

[1] K. H. Shibly, S. Kumar Dey, M. A. Islam, and S. Iftekhar Showrav, "Design and development of hand gesture based virtual mouse," in Proceedings of the 2024 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), pp. 1–5,

[2] Zhang, Y., Chen, L., & Li, J. (2023). Real-Time Gesture Recognition for Human-Computer Interaction Using Media Pipe and OpenCV. Journal of Interactive Technology, 12(3), 198-210.

[3] Wang, X., & Liu, H. (2022). Multimodal Interaction: Combining Gesture and Voice Recognition for Enhanced User Experience. International Journal of Human-Computer Studies, 157, 102-115.

[4] Li, J., Sun, Y., & Zhao, T. (2021). Voice Assistant Technologies: Trends and Challenges in Natural Language Processing. IEEE Transactions on Multimedia, 23(6), 1234-1245.

[5] Cao, Z., Hidalgo, G., & Simon, T. (2017). Open Pose: Real-Time multi-Person 2D Pose Estimation using Part Affinity Fields. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(6), 172-186.