ISSN: 3107-6513

An AI-Enabled Gesture and Voice Based System for Natural Computer Interaction

[1]Mohamed Hussain A, [2]Sowmiyapriya V

[1]Student, B.Sc Artificial Intelligence and Machine Learning, Final year, Nehru Arts and Science College, Coimbatore, Tamil Nadu, India

[2]Assistant Professor, Iot and AIML,, Nehru Arts and Science College, Coimbatore, Tamil Nadu, India

I. ABSTRACT

Human–computer interaction has significantly evolved with the development of artificial intelligence, computer vision, and speech recognition technologies. Traditional computer systems mainly rely on input devices such as keyboards and mice, which require continuous physical interaction. Although these devices are effective, they may not always provide a natural or convenient way for users to communicate with computers. In many situations, such as when users have limited mobility or when hands-free interaction is required, traditional input methods become less efficient. Therefore, there is a growing need for more intuitive and touchless interaction systems that allow users to control computers in a simpler and more natural way.

This project presents a Gesture and Voice Based Interaction System that allows users to control a computer using hand gestures and spoken commands. The system uses a webcam to capture hand movements and a microphone to process voice instructions. Computer vision techniques are used to detect hand landmarks and recognize gestures, while speech recognition technology converts spoken commands into executable actions. The recognized inputs are mapped to system operations such as cursor movement, clicking, scrolling, and opening applications or websites. By integrating gesture recognition and voice control into a single platform, the proposed system provides a more flexible, accessible, and efficient approach to human–computer interaction.

Keywords: *Gesture Recognition, Voice Command System, Human–Computer Interaction, Computer Vision, Speech Recognition, Touchless Interaction*

II. INTRODUCTION

Human–Computer Interaction (HCI) has become an important area of research as computers continue to play a vital role in everyday activities. From education and communication to business and entertainment, people interact with computer systems on a regular basis. Traditionally, interaction with computers has been achieved through hardware input devices such as keyboards, mice, and touchpads. These devices allow users to provide commands and control system operations efficiently. However, they require constant physical contact and may not always offer the most natural or convenient form of interaction.

In many situations, traditional input methods can create limitations. For example, users may need to operate a computer while performing other tasks, or they may not have immediate access to input devices. In addition, individuals with physical disabilities or mobility challenges may find it difficult to use conventional devices effectively. As technology continues to evolve, there is a growing demand for alternative interaction methods that allow users to communicate with computers in a more natural, intuitive, and accessible way.

Recent advancements in artificial intelligence, computer vision, and speech processing technologies have opened new possibilities for developing touchless interaction systems. Gesture recognition is one such technology that enables computers to understand human hand movements captured through cameras. By identifying key points on the hand and analyzing their positions, a computer system can recognize different gestures and translate them into commands. Similarly, voice recognition technology allows computers to interpret spoken language and convert it into digital instructions that can trigger specific system actions.

Both gesture recognition and voice recognition have been widely used in various modern applications. For instance, voice assistants such as smart speakers and mobile assistants allow users to perform tasks through simple voice commands. Gesture recognition is commonly used in gaming systems, virtual reality environments, and smart devices to provide more immersive and interactive experiences. Despite the availability of these technologies, many existing systems implement gesture and voice control separately rather than combining them into a single integrated platform.

Integrating gesture recognition and voice command processing into one system can significantly enhance the user experience. Gesture-based control enables users to perform tasks such as moving the cursor, clicking, dragging, and scrolling without touching any input device. At the same time, voice commands allow users to open applications, perform searches, and control system operations through spoken instructions. The combination of these two technologies creates a more flexible

ISSN: 3107-6513

interaction environment where users can choose the most convenient method of control depending on the situation.

The Gesture and Voice Based Interaction System proposed in this project aims to develop a unified platform that enables users to control computer functions using both hand gestures and voice commands. The system uses a webcam to capture real-time video input and detect hand gestures using computer vision techniques. At the same time, a microphone captures spoken commands that are processed through speech recognition technology. The detected gestures and voice commands are interpreted and mapped to specific system actions such as cursor movement, clicking, scrolling, opening websites, and executing other tasks.

III. Problem Statement and Objective

Traditional computer systems mainly depend on physical input devices such as keyboards, mice, and touchpads for user interaction. Although these devices are widely used and reliable, they require continuous physical contact and manual operation. In many situations, this form of interaction may not be convenient or efficient. For example, users may need to operate a computer while performing other tasks, or they may require hands-free control in environments where physical interaction with devices is difficult.

Another important challenge is accessibility. Individuals with physical disabilities or mobility limitations may find it difficult to use traditional input devices effectively. This limitation can reduce their ability to interact with computers and access digital services easily. As technology becomes increasingly integrated into everyday life, it is important to develop systems that provide more inclusive and flexible interaction methods for all users.

With the advancement of technologies such as computer vision and speech recognition, it has become possible to design systems that allow computers to understand human gestures and spoken commands. However, many existing solutions focus on either gesture recognition or voice recognition independently, rather than integrating both interaction methods into a single platform. This lack of integration limits the flexibility and usability of such systems.

Therefore, there is a need to develop an intelligent and efficient system that combines both gesture recognition and voice command processing to enable touchless and hands-free interaction with computers. The proposed Gesture and Voice Based Interaction System aims to address this problem by allowing users to control computer operations through simple hand gestures and voice commands, thereby improving accessibility, convenience, and overall user experience.

IV. Related Work

Research in Human–Computer Interaction (HCI) has expanded significantly in recent years, particularly with the development of technologies that allow users to interact with computers without relying solely on traditional input devices. Gesture recognition and voice command systems have become important areas of study because they enable more natural and intuitive interaction between humans and machines. Several researchers have explored the use of computer vision and speech recognition technologies to develop systems that support touchless and hands-free computer control.

In 2023, Zhou et al. conducted a comprehensive study on gesture recognition technologies for human–computer interaction. Their research examined various computer vision methods used to detect and interpret human hand movements. The study highlighted the growing role of camera-based gesture detection in modern applications such as virtual reality, gaming, robotics, and smart devices. The authors emphasized that gesture recognition systems allow users to interact with digital systems in a more intuitive way by replacing traditional hardware-based input methods with natural hand movements.

In 2024, Kumar, Bharadwaj, and Mahanthesha proposed a system that integrates voice assistants with gesture recognition to control computer mouse operations. The system used computer vision techniques to detect hand gestures and speech recognition to interpret voice commands. The research demonstrated that combining gesture and voice interaction allows users to perform tasks such as cursor movement, clicking, and scrolling more efficiently. The study also highlighted that such systems can improve accessibility for individuals who have difficulty using conventional input devices.

Another research work by Shaikh, Walimbe, Shinde, and Sase (2024) introduced a gesture-based virtual human–computer interaction system. In their approach, a webcam was used to capture real-time hand movements, and computer vision algorithms were applied to recognize specific gestures. These gestures were then translated into computer commands such as mouse movement, clicking, and application control. The results showed that gesture-based systems can effectively simulate mouse functions and provide a convenient alternative to traditional input devices.

In 2025, Bhonde, Mongse, Naikwar, Dwivedi, and Mahulkar developed a gesture and voice-based personal computer control system designed to improve user interaction with desktop systems. Their system used hand landmark detection for gesture recognition and speech-to-text technology for interpreting voice commands. The study demonstrated that integrating both interaction methods allows users to control applications, navigate websites, and perform system operations without direct physical contact with input devices.

These studies clearly indicate that gesture recognition and voice-based technologies have strong potential to transform the way users interact with computer systems. However, many existing solutions focus primarily on either gesture-based interaction or voice-based control separately. Systems that effectively integrate both technologies into a unified platform are still relatively limited. Therefore, the proposed Gesture and Voice Based Interaction System aims to build upon these previous studies by combining gesture recognition and voice

ISSN: 3107-6513

command processing into a single system that provides flexible, efficient, and accessible computer control.

V. System Design

The Gesture and Voice Based Interaction System is designed to provide a natural and touchless way for users to interact with a computer. The system combines gesture recognition and voice command processing to control various computer functions. The overall design follows a modular architecture where each component performs a specific task, ensuring better efficiency, maintainability, and scalability. The system mainly consists of input modules, processing modules, command interpretation, and execution modules.

5.1 Input Module

The input module is responsible for collecting user data in the form of hand gestures and voice commands. A webcam is used to capture real-time video frames of the user's hand movements. These video frames serve as the input for the gesture recognition system. At the same time, a microphone captures spoken commands from the user. The audio signals are converted into digital data so that they can be processed by the voice recognition module. This dual-input approach allows the system to support both gesture-based and voice-based interactions simultaneously.

5.2 Gesture Processing Module

The gesture processing module is responsible for detecting and interpreting hand movements captured by the webcam. The system uses computer vision techniques to analyze the video frames and identify hand landmarks. Libraries such as OpenCV and MediaPipe are used to detect key points on the hand and track their positions. Based on the position and movement of these landmarks, the system determines specific gestures such as finger movements, pinch gestures, or open palm gestures. These gestures are then translated into commands for mouse operations such as cursor movement, clicking, dragging, and scrolling.

5.3 Voice Recognition Module

The voice recognition module processes spoken commands captured through the microphone. The system uses speech recognition techniques to convert audio signals into text. Once the speech is converted into text, the system analyzes the command and identifies keywords that correspond to specific system actions. For example, commands like “open Google,” “scroll down,” or “stop system” are recognized and mapped to predefined functions. This module allows users to control applications and system operations through simple voice instructions.

5.4 Command Interpretation Module

After the system detects gestures or voice commands, the information is sent to the command interpretation module. This module acts as the decision-making unit of the system. It evaluates the detected input and determines the appropriate action that should be performed. Logical conditions and predefined rules are used to map user inputs to system commands. This ensures that the system responds accurately and prevents unintended operations.

5.5 Execution and Automation Module

The execution module performs the actual system-level actions based on the interpreted commands. Automation libraries such as PyAutoGUI are used to simulate mouse and keyboard operations. These actions include moving the cursor, performing clicks, dragging objects, scrolling pages, and opening web applications. By automating these operations, the system enables users to control the computer without physically interacting with traditional input devices.

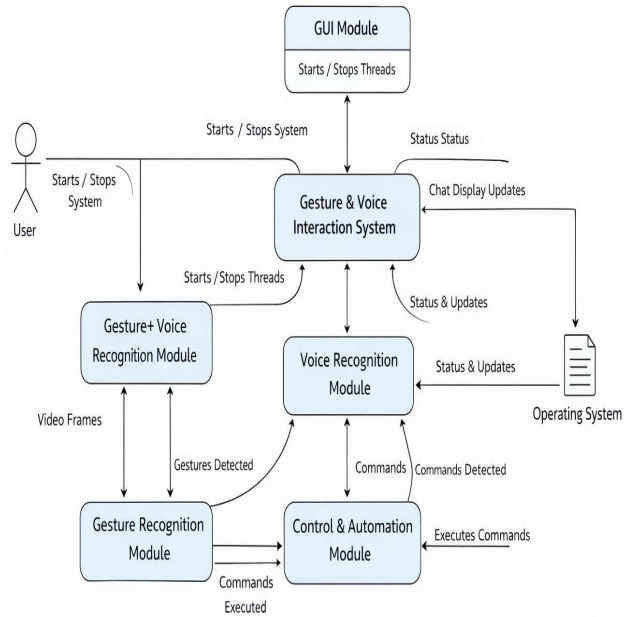
5.6 Output and Feedback Module

To improve user interaction and system transparency, the system provides feedback through both visual and audio outputs. A graphical user interface (GUI) developed using Tkinter allows users to monitor system activity, view recognized commands, and start or stop the interaction system. Additionally, a text-to-speech module generates audio responses to confirm executed commands. This feedback mechanism ensures that users are aware of the system’s actions and improves overall usability.

Fig. !. System Design of the Proposed Model

VI. Advantages of Proposed System

The proposed Gesture and Voice Based Interaction System offers several advantages compared to traditional computer interaction methods. By integrating gesture recognition and voice command technologies, the system provides a more



Data Flow Diagram

natural, flexible, and efficient way for users to interact with computers.

ISSN: **3107-6513**

One of the major advantages of this system is touchless interaction. Users can control various computer functions without physically touching input devices such as keyboards or mice. This makes the system convenient in situations where hands-free control is required or when traditional input devices are not easily accessible.

Another important benefit is improved accessibility. The system provides an alternative method of interaction for individuals with physical disabilities or mobility limitations who may find it difficult to use conventional input devices. By allowing users to control computers using simple hand gestures and voice commands, the system helps make technology more inclusive and user-friendly.

The system also supports dual-mode interaction, meaning users can choose between gesture control and voice commands depending on the situation. This flexibility enhances the overall user experience and allows users to perform tasks more efficiently. For example, gestures can be used for cursor movement and mouse operations, while voice commands can be used to open applications or perform searches.

Another advantage is real-time processing and responsiveness. The system is designed to detect gestures and voice commands quickly and execute the corresponding actions with minimal delay. This ensures smooth interaction and improves the usability of the system during continuous operation.

VII. Conclusion

The Gesture and Voice Based Interaction System presents a modern approach to improving human–computer interaction by enabling users to control computer functions using hand gestures and voice commands. Traditional interaction methods mainly depend on devices such as keyboards and mice, which require continuous physical interaction. The proposed system reduces this dependency by introducing a touchless and hands-free method of controlling computer operations. By combining gesture recognition and voice processing technologies, the system provides a more natural and intuitive way for users to interact with computers.

The system successfully integrates computer vision and speech recognition techniques to detect hand movements and interpret spoken commands in real time. Using tools such as OpenCV, MediaPipe, and speech recognition libraries, the system can perform actions like cursor movement, clicking, scrolling, opening websites, and executing other system functions. The addition of a graphical user interface and audio feedback further improves the user experience by providing clear interaction and system responses.

Another important contribution of the system is its ability to improve accessibility. Users who may have difficulty using traditional input devices can benefit from gesture-based and voice-based control methods. The system also demonstrates how modern technologies can be used to create flexible and efficient computing environments that support different user needs.

VIII. References

1. Bhonde, D., Mongse, K., Naikwar, L., Dwivedi, N., & Mahulkar, O. (2025). Gesture and voice-based personal computer control system. *International Journal on Advanced Electrical and Computer Engineering*, 14(1), 190–195.

2. Kumar, V., Bharadwaj, A. H., & Mahanthesha, U. (2024). Enhancing human–computer interaction: Integrating voice assistants and gesture recognition for innovative mouse control. *Journal of Propulsion Technology*, 45(2).

3. Shaikh, S., Walimbe, B., Shinde, A., & Sase, S. (2024). Gesture-based virtual human–computer interaction system. *Journal of Microelectronics and Solid State Devices*.

4. Zhou, H., Wang, D., Yu, Y., & Zhang, Z. (2023). Research progress of human–computer interaction technology based on gesture recognition. *Electronics*, 12(13), 2805.

5. Patil, N., Ansari, M. W., Jadhav, S. R., et al. (2024). Gesture voice: Revolutionizing human–computer interaction with an AI-driven virtual mouse system. *Turkish Online Journal of Qualitative Inquiry*.

6. Wachs, J., Kölsch, M., Stern, H., & Edan, Y. (2011). Vision-based hand-gesture applications. *Communications of the ACM*, 54(2), 60–71.

7. Mitra, S., & Acharya, T. (2007). Gesture recognition: A survey. *IEEE Transactions on Systems, Man, and Cybernetics*, 37(3), 311–324.

8. Oviatt, S. (2003). Multimodal interfaces. In *The Human-Computer Interaction Handbook*. Lawrence Erlbaum Associates.

9. Zhang, Z. (2012). Microsoft Kinect sensor and its effect. *IEEE Multimedia*, 19(2), 4–10.

10. Turk, M. (2014). Multimodal interaction: A review. *Pattern Recognition Letters*, 36, 189–195.