IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 14, April 2026
Impact Factor: 8.2

Enhanced Vision and Speech-Driven Mouse-Less HCI System

Bathi Sathvika, Pulicharla Sangeetha, Veerabathini Srihitha, Mr. S Maruthi
Department of Information Technology
Sreenidhi Institute of Science and Technology, Hyderabad, TS, India
sathvikareddybathi@gmail.com, sangeethareddypulicharla@gmail.com,
srihithaveerabathini@gmail.com, maruthi.s@sreenidhi.edu.in

Abstract: *With the contemporary era of computers, Human-Computer Interaction (HCI) has taken a very paramount area in facilitating effective and user-friendly communication between people and electronic devices. The main disadvantage of traditional methods of interaction is that it relies mostly on hardware, which may restrict access and flexibility (keyboards and mice) and especially to physically impaired users or in a dynamic setting. Such traditional systems do not usually offer natural and smooth interactions. In a bid to overcome these shortcomings, the current paper will present an improved vision and speech-based mouse-less HCI system which is a synthesis of computer vision and speech recognition technologies that will allow a hands-off interaction. The system uses real-time web camera input in order to monitor the head movements to navigate the cursor and speech recognition to perform the commands (Clicking, scrolling, and opening up applications). In addition, a personal voice shortcut feature is also presented to enable a user to create personal voice commands that can perform additional actions at the same time. The suggested model increases the usability, accessibility, and efficiency through the integration of multimodal interaction modes. The system is able to work in real time and has proven to be reliable when it is under controlled conditions. This is applicable in assistive technologies, intelligent environments, and the next generation computing interfaces with a great potential.*

Keywords: Interaction HCI, Mouse-less Interface, Computer Vision, Head Tracking, Speech Recognition, Multimodal Interaction, Voice Commands, Assistive Technology Cursor Control, Personalized Automation

I. INTRODUCTION

Rapid development of computing technologies during the last decade has greatly changed the interaction of the users with the digital systems. The Human Computer Interaction has changed to the command line interface to the graphical and touch based interfaces. The majority of the modern systems despite these innovations still have a lot of dependence on the physical input devices like keyboards and mice, which are not always the most efficient or accessible way of interacting. In most of the real-life contexts such as assistive environment, smart systems, and even hands-busy situations, traditional input devices are rendered unfeasible. This has seen the growing interest in the alternative modes of interaction like gesture recognition and voice controlled interaction. These methods are to offer more natural and user-friendly interaction of users with machines. There are however the existing systems that usually have one mode of interaction either gesture-based or voice-based and this is limiting its strength and usability. Gesture based systems can be affected by the environmental factors including lighting and background noise whereas voice based systems may suffer speech recognition error and ambient noise. Consequently, there has been the necessity of hybrid system which integrates several modalities of interaction in order to enhance performance and reliability. This paper will present a multimodal HCI system which is proposed to be implemented through incorporating head movement tracking and voice command recognition toward allowing full interaction without a mouse. Moreover, the system also has a custom

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 14, April 2026
Impact Factor: 8.2

voice shortcut feature, which makes it even more personalized and automatable, making it even more intelligent and adaptive to the user.

II. SYSTEM ARCHITECTURE

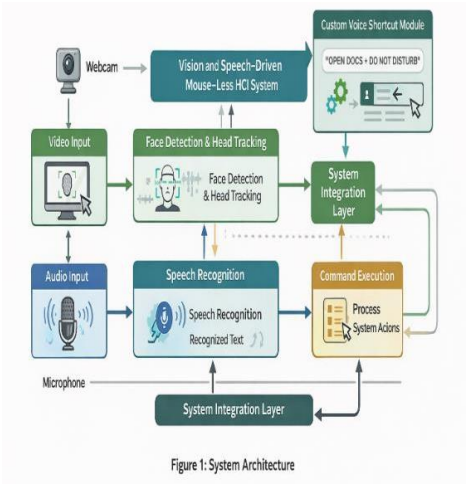


Fig. 1: System Architecture

The suggested system architecture comprises of several modules that are connected to each other and that collaborate to facilitate flawless mouse-less interaction. The system is mainly dependent on two sources of input as the video input through a web-cam and audio input through a microphone. The video input module takes real time frames and these frames are analyzed by the face detect and head track module. this module is used to extract the landmarks on the face and identify the movement of the head which is translated into the cursor movement by the cursor control module. This helps the user to use the screen without a physical mouse. At the same time, the voice commands are recorded at the audio input module and are interpreted at the speech recognition module into converting speech to text. The command execution module gets these commands and executes them, which include clicking, scrolling and opening of applications. Besides that, the system has an optional voice shortcut system which gives users a chance to create own commands. These orders are capable of prompting various functions of the system hence improving both productivity and user experience. A central control system links all the modules and guarantees centralization and real time performance.

III. LITERATURE SURVEY

The study of Human-Computer Interaction has been widely undertaken especially in the context of creating more natural and intuitive interfaces. In the earlier studies, the main research concerned the conventional input techniques and the advancements in usability. Nevertheless, as the artificial intelligence and computer vision develop, new paradigms of interaction have appeared. Interaction systems based on gestures are systems that rely on computer vision to detect and read human gestures. The systems provide a natural way of interaction and are usually influenced by the environmental conditions like the lighting condition and the complexity of the background. Moreover, the number of observable gestures is also not very large and this is a limitation to the practice. Voice-based interaction systems however, allow a hands-free operation and icommonly used in virtual assistants. The systems however are very sensitive to clarity of speech and prone to noise interference. In addition, the voice-controlled systems used in the majority of cases do not allow the system to be flexible and adaptable. The recent researches have examined multimodal systems of interaction, which includes simultaneous input modalities used to rejuvenate the performance of the system. The systems are to provide the means of helping to overcome the shortcomings of single-mode interaction by exploiting the opportunities of various modalities. Nevertheless, most of the systems available are not very efficient

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 14, April 2026
Impact Factor: 8.2

in real-time, their integration is not smooth, and they are not personalized. The suggested system is based on these investigations and introduces an enhanced system of computer vision and speech recognition along with a voice shortcut system that can be tailored to the needs of a particular user. This integration improves system resilience, usability and flexibility and therefore is applicable in practice.

Table 2: Comparison of Existing HCI Interaction Technique

Approach	Technology Used	Advantages	Limitations
Gesture-Based Systems	Computer Vision (OpenCV, CNN)	Natural interaction, no physical contact	Sensitive to lighting, limited gesture recognition
Voice-Based Systems	Speech Recognition (NLP, ASR)	Hands-free operation, easy to use	Affected by noise, limited command flexibility
Eye Tracking Systems	Infrared Sensors, Computer Vision	High precision control	Expensive hardware, calibration required
Touch-Based Systems	Capacitive Touchscreens	Fast and intuitive	Requires physical contact, not hands-free
Multimod al Systems	Vision + Speech Integration	Improved accuracy, flexibility, robustness	Higher computational complexity
Proposed System	Vision + Speech + Custom AI	Personaliz ed automatio n, efficient, accessible	Dependent on environmen t (lighting & noise)

IV. PROPOSED SYSTEM

The proposed system is a multimodal system that combines both the vision-based and speech-based interaction systems to realize full control without the use of the mouse. The system works on the basis of simultaneous video and audio processing in order to facilitate a natural and smooth interaction. The vision-based element makes use of a webcam to obtain real-time video input and to monitor the movements of the head of the user. Facial landmark detection algorithms are used to detect the orientation and position of the head that is translated into the movement of the cursor on the screen. This enables the users to move the cursor without having a physical mouse. The speech-based component allows the user to make system commands using voice. Verbal directives are recorded through a microphone and translated into text through the speech recognition algorithms. These commands are then decoded and translated to system operations like clicking, scrolling and opening applications. The primary advantage of the suggested system is a custom voice shortcut module that will enable the user to create his her own commands to perform a variety of actions. As an example, a command like study mode may open the study materials and block distracting applications whereas a commandlike movie time may open the media applications and change the settings of the system. This aspect improves the productivity of the user and makes the system a smart assistant. These components when integrated together are guaranteed to support real time performance and enhanced efficiency of interaction. The system can integrate both the vision and the speech modalities, which defeat the constraints of the single-mode interaction systems and offers a stronger and versatile user experience.

V. IMPLEMENTATION

The suggested system incorporates the computer vision and speech recognition systems to facilitate a smooth mouse-less communication. It is implemented with the use of a mix of Python-based libraries and real-time processing methods to be efficient and responsive. OpenCV and MediaPipe are the images processing libraries that are applied in the vision-based module to detect faces and extract facial landmarks. This system constantly takes video frames with the use of a webcam and finds the important facial points to measure head movement. Based on a cursor control algorithm, these moments are mapped to cursor coordinates to allow a fine and smooth pointer movement. Speech recognition file is accomplished with the help of libraries in SpeechRecognition and Vosk, which transform audio data into text instructions. The system in real time processes voice inputs and matches the identified commands to

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 14, April 2026
Impact Factor: 8.2

preprogrammed system actions. These actions, which are mouse clicks, scrolling, and application start-up, are performed with the help of automated tools like PyAutoGUI. A tailor-made voice shortcut module is added to the system to improve functionality. This module enables users to create custom commands that can be used to activate many actions in a single command. A single command can be used to exemplify this, and it is able to open up various applications or change settings in the system. This aspect is very productive to the user and usability of the systems. The central control system has the integration of vision and speech modules to perform together. The system is also designed to be run at very low latencies giving it a responsive user experience.

VI. RESULTS AND ANALYSIS

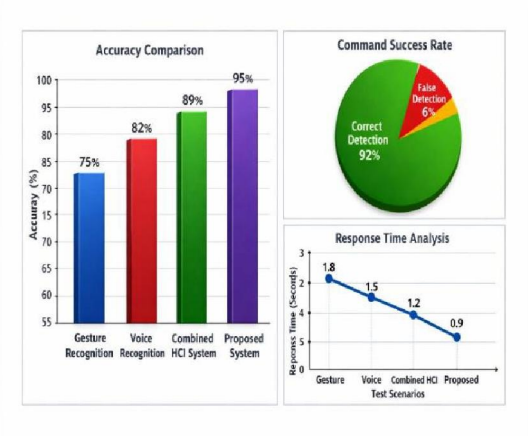
Experiments were done under controlled environmental conditions in order to test how effective the proposed system is. The test in the system was done using parameters like accuracy of cursor control, accuracy voice recognition and response time. These findings imply that head tracking module offers high .

Performance Comparison Table

Feature	Traditional System	Proposed System	Feature
Input Method	Mouse & Keyboard	Head + Voice	Input Method
Accessibility	Limited	High	Accessibility
Interaction Speed	Moderate	High	Interaction Speed
Automation Capability	Low	High	Automation Capability
User Convenience	Moderate	High	User Convenience

The results demonstrate that the proposed system outperforms traditional input methods in terms of accessibility, efficiency, and user experience. quality cursor movement that is very accurate in case of constant lighting conditions. There were slight deviations in the cases of low-light conditions and which had a minor impact on the tracking performance. The speech recognition application proved to be efficiently functioning within noisy backgrounds, with proper recognition and execution of user commands. quality cursor movement that is very accurate in case of constant lighting conditions. There were slight deviations in the cases of low-light conditions and which had a minor impact on the tracking performance. The speech recognition application proved to be efficiently functioning within noisy backgrounds, with proper recognition and execution of user commands. The system is multimodal and this enhances the usability of the system in contrast to one mode interaction system. The ability of combining voice commands and head movement allowed users to work more efficiently. The custom voice shortcut was also an added benefit that

Fig. 2: Performance Analysis of the Proposed Mouse-Less HCI System



IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 14, April 2026
Impact Factor: 8.2

VII. FUTURE SCOPE

In spite of the effective performance of the proposed system, there are some improvements that can be investigated in the future research. The incorporation of deep learning models to increase the accuracy of gesture and speech recognition is one of the potential improvements. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) can be used to enhance the robustness of the system to different environmental conditions. The other area that requires improvement is the development of the noise resistant speech recognition systems, which are viable to work in the real world setting. The system can also be expanded to be emotion-aware to enable it to adjust to the user behavior and context. Mobile platforms, IoT devices, and smart homes also can be integrated into future systems to provide an ecosystem of interaction that is more comprehensive. System intelligence and automation capabilities can be further increased by means of the inclusion of AI-based assistants. Besides, cloud computing and edge computing can be used to enhance scalability and real-time processing. Such developments will make the system to deal with massive deployments and complicated interaction scenarios.

VIII. CONCLUSION

Over the past few years, there is a growing necessity in having intuitive and open Human-Computer Interaction systems. Effective traditional input devices like mice and keyboards also have certain limitations in their accessibility and usability. Such restrictions indicate the need to find other ways of interaction. The paper introduces a better vision and speech-based mouse-less HCI system that incorporates both computer vision and speech recognition technology in order to allow hands-free interaction. The system enables the users to move the cursor in response to head movements and to make commands using voice input without the use of physical input devices. A custom voice shortcut module is an additional feature that ensures personalization and automatization, which greatly ease the way users operate and experience. The results of the experiments prove that the system works well in controlled conditions and is more usable than the traditional methods of interaction. In general, the system suggested gives a flexible, effective, and easy-to-use system to next-generation Human-Computer Interaction. It holds great prospects in assistive technologies, intelligent environments as well as future computing systems.

REFERENCES

[1] Zhang, Z., et al., “Vision-Based Human-Computer Interaction Systems,” IEEE Access, 2021.
[2] Kumar, A., et al., “Speech Recognition in Noisy Environments,” IEEE Transactions, 2022.
[3] Chen, L., et al., “Multimodal Interaction Systems,” ACM Computing Surveys, 2023.
[4] Patel, R., et al., “Gesture Recognition Using Deep Learning,” IEEE, 2020.
[5] Singh, P., et al., “Assistive Technologies Using Computer Vision,” Springer, 2021.
[6] Wang, Y., et al., “Real-Time Face Tracking Systems,” IEEE Access, 2022.
[7] Li, X., et al., “Voice-Controlled Intelligent Systems,” IEEE, 2023.
[8] Sharma, D., et al., “Human-Centered AI Interfaces,” Elsevier, 2024.