International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 02 | Feb 2026 www.irjet.net p-ISSN: 2395-0072

Lyra: A Multimodal Intelligent Desktop Interaction System

Siya Vaity[1], Prachi Das[2], Shrot Maurya[3],Manthan Bid[4]

[1] *Artificial Intelligence and Machine Learning, Viva Institute Of Technology*
[2] *Artificial Intelligence and Machine Learning, Viva Institute Of Technology*
[3] *Artificial Intelligence and Machine Learning, Viva Institute Of Technology*
[4]*Artificial Intelligence and Machine Learning, Viva Institute Of Technology*

Abstract - *Human-computer interface is taking a new direction towards more natural and touchless interface. In this paper, the author describes the Lyra, the intelligent desktop assistant that combines the voice recognition and real-time hand gesture control to allow operating the computer without hands. The system enables users to open applications, browse, handle files, type, manage windows and carry out mouse actions by voice command and gesture detected by the webcams. Lyra is a blend of speech to speech processing, command interpretation, computer vision and system automation as a means to develop an effective and convenient interaction model. The suggested system will enhance the productivity, user ease of access due to their physical disabilities, and facilitate hygienic touchless computing conditions. Real world experimental use has shown to have smooth cursor control, command recognition and low system response.*

Key Words*: *Voice Assistant, Gesture Recognition, Human–Computer Interaction, Computer Vision, Automation, AI Assistant

1. INTRODUCTION

Traditional computer interface relies on keyboards and pointers. These tools are effective but constrain the natural communication process and they might not fit in the accessibility-oriented or touchless environments. The recent developments in the field of Artificial Intelligence (AI), speech recognition, and computer vision have allowed more intuitive interaction mechanisms. This paper presents the concept of Lyra, a multi-modal AI assistant integrating the use of voice commands with the use of gestures to manage a desktop system. This is aimed at offering an uninterrupted, smart and natural interface that minimizes hardware input device physical addiction.

The key contributions of this work are:

- Desktop automation through voice.
- A hand gesture control mouse and window control system in real-time.
- Combination of speech and vision modules to one assistant.
- Safety and stability systems to avoid unintentional actions of the system.

1.1 Voice Assistant Module

This module works with speech recognition and automation of the system.
Components:

- Speech recognition engine
- Matching and interpretation unit of command.
- Web launcher and application.
- Search module of files and folders.
- Automated typing controller.
- Filter to avoid interference of critical systems.

1.2 Gesture Control Module

Hand tracking and gesture recognition This module is a camera-based application that tracks hands and recognizes gestures.
Components:

- Webcam capture system
- Hand landmark identification algorithm.
- Logic of gesture classification.
- Layers of web socket communication.
- System action executor using Python.

The two modules communicate with the operating system to run command on-the-fly

2. Voice Assistant Functionalities

2.1 Wake and Sleep Mechanism

Lyra starts with preprogrammed wake words and sleeps after not being used, which maximizes consumption.

2.2 Website and Application Control.

Users can also open the installed applications and websites through the natural voice commands.

2.3 File and Folder Management

The assistant locates popular directories and opens requested files or folders.

2.4 Information Services

Lyra has time and date information that is in real time.

2.5 Web Search and Media Playback.

The voice commands start web searches and playback of music provided through YouTube.

2.6 Automated Typing

The assistant typed what was dictated into active text fields.

3. Gesture Control System

The gesture system can allow the full control of the mouse and windows without physical device.

3.1 Cursor Movement

The palmcentermapping and motion smoothing enables the cursor to be controlled by an open palm gesture.

3.2 Scrolling

Finger patterns enable operating finger scrolling smoothly in an adaptive manner both uphill and downhill.

3.3 Window Management

The movements of the window and safe window closing are made possible by specified gestures.

3.4 Stability Enhancements

Exponential Moving Average (EMA) smoothing.
Dead-zone filtering
Cursor freeze during click
Adaptive speed control
The methods minimize jitter and enhance precision.

4. Communication Framework.

The gesture module is working with:

- HTTP Server to serve gestures interface.
- WS Server to transmit real-time data.
- Python Backend to do action at the OS level.

This guarantees low-latency and on-going communication between parts.

5. Performance Optimizations

It improves performance in the system by:

- Motion smoothing algorithms.
- Dynamic speed scaling
- Auto calibration
- Camera window persistence
- Port conflict prevention

These are done to provide stable tracking and hassle free control of the system.

6. Future Scope

- The future development can be:
- Multi-language support
- Offline speech recognition
- Adaptive gesture learning that is based on AI.
- IoT integration
- Command customization on a personal level.

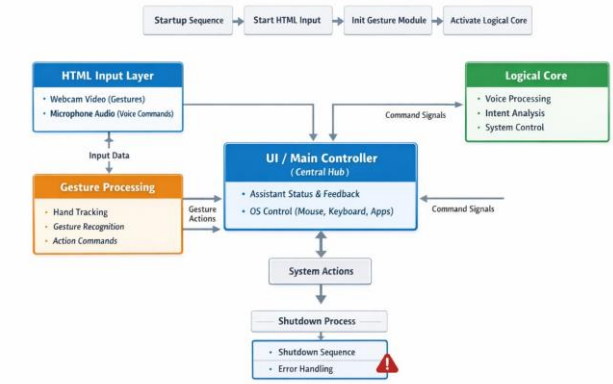

Fig -1: Flowchart

The diagram represents the overall workflow and architecture of a smart laptop/desktop assistant that operates using hand gestures, voice commands, and system controls. The system is divided into multiple functional layers that work together through a central controller.

1. Startup and Initialization Phase

The system begins with a Startup Sequence, which initializes the core components. This is followed by:

- Start HTML Input– activates browser-based input interfaces
- Initialize Gesture Module – prepares gesture detection mechanisms
- Activate Logical Core – enables voice processing and intent analysis

This phase ensures all modules are ready before user interaction begins.

2. HTML Input Layer

The HTML Input Layer acts as the primary data acquisition layer. It collects:

- Webcam video input for detecting hand gestures
- Microphone audio input for voice commands

These inputs are continuously fed into the system as raw input data.

3. Gesture Processing Module

The Gesture Processing unit interprets visual data received from the webcam. It performs:

- Hand tracking to detect hand position and movement
- Gesture recognition to identify predefined gestures
- Action command generation based on recognized gestures

The resulting gesture actions are forwarded to the central controller.

4. Logical Core

The Logical Core is responsible for intelligent decision-making. Its key functions include:

- Voice processing (speech recognition)
- Intent analysis to understand user commands
- System control logic to determine appropriate actions
- It sends command signals to the main controller after processing user intent.

5. UI / Main Controller (Central Hub)

The UI / Main Controller acts as the heart of the system. It

- Receives command signals from both the Gesture Processing Module and the Logical Core
- Provides assistant status and feedback to the user
- Controls operating system functions such as mouse movement, keyboard input, and application management

This module coordinates all interactions and executes system-level actions.

6. System Actions and Shutdown Process

Once commands are executed:

System Actions are performed (opening apps, controlling OS, etc.)

If required, the system enters a Shutdown Process, which includes:

- Shutdown sequence
- Error handling mechanisms to manage unexpected failures

This ensures a safe and controlled termination of the system.

7. CONCLUSIONS

Lyra shows how voice AI and gesture recognition can change the interaction at the desktop to a touchless experience. The system is a system that is based on the combination of speech automation, computer vision and real-time processing to form an intelligent human-computer interaction framework. Lyra has good prospects of being accessible, productive as well as computing next-generation environments.

ACKNOWLEDGEMENT

The author would also like to mention the help of academic materials and open-source technologies which helped in the creation of the Lyra assistant system.

REFERENCES

[1] S. Mitra and T. Acharya, “Gesture Recognition: A Survey,” IEEE Transactions on Systems, Man, and Cybernetics, vol. 37, no. 3, pp. 311–324, May 2007.

[2] Z. Zhang, “Hand Gesture Recognition Based on Computer Vision,” in Proc. International Conference on Artificial Intelligence and Pattern Recognition, 2019, pp. 45–50.

[3] R. Szeliski, Computer Vision: Algorithms and Applications, London, U.K.: Springer, 2011.

[4] D. Jurafsky and J. H. Martin, Speech and Language Processing, 3rd ed., vol. 1, Pearson Series in Artificial Intelligence, 2020.

[5] T. Starner, “Wearable Computer Vision System for Gesture Recognition,” U.S. Patent 6 577 742, Jun. 10, 2003.