LapiDude: Biometric-Authenticated Multilingual Framework for Secure Desktop Automation

RANGANI HIMABINDU, 1
SHAIK ROOHI REHANA BEGUM,2
SINGAMSETTY TARUNI,3
VALLA BHARGAV,4
PONUGUPATI PRATAP,5
PULIPATI SYAMBABU6

1 Assistant Professor, Department of CSE(AIML), Bapatla Engineering College, Bapatla 522101, AP, India
2,3,4,5,6 Student, Department of CSE(AIML), Bapatla Engineering College, Bapatla 522101, AP, India.

***Abstract*—** Over the years, voice-based interaction has become an important interface paradigm in today's computing systems. Voice interactions has become widespread with mobile devices and structured on smart assistants, desktop environments still depends on manual input via keyboards and pointing devices.

Additionally, since most desktop automation tools do not even check whether the user sending commands is who he/she claims to be, there are security implications when performing operations on a system level are executed through automated processes. Also, another gap of current systems which makes them not much efficient is that they do not provide multilingual corpus since a majority of available voice interaction framework are designed to be helpful only for English based users.

This work proposes a new biometric-authenticated multilingual framework, LapiDude for secure voice-driven desktop automation. It utilizes speaker verification models on top of multi-lingual command recognition to allow users to control desktop applications with natural spoken phrases, but only allows authorized speakers to issue commands. We use deep speaker embeddings from a pretrained voice representation to perform speaker verification. It uses cosine similarity to match incoming voice samples against stored administrator voiceprints in order to authenticate identity.

The spoken command gets transformed to text post successful authentication and subsequently passed through a multilingual keyword-based intent mapping mechanism that can understand a command in English, Hindi and Telugu. The recognized commands are then mapped to the relevant desktop automation functions (e.g., launching applications, controlling processes, window handling).

The framework works fully on local hardware, which improves privacy and decreases the response latency. Experimental assessment reveals that we perform secure recognition (average similarity < 0.65) and average command execution latency < 200 ms, the results demonstrate that the proposed framework is an efficient and secure mechanism for enabling multidimensional verbal-based desktop pages.

Key Words - Voice Biometrics, Desktop Automation, Multilingual Voice Commands, Human–Computer Interaction, Speaker Authentication, Pattern Recognition

1. INTRODUCTION

During the last several decades, there has been a tremendous progress in human–computer interaction. Although early computing systems did not leave much room for wayward keystrokes, users were vastly restricted by the manual-oriented language which controlled relational databases via command-line interfaces that required precise input text. User-friendly Graphical User Interfaces (GUI) made interaction much easier through visual elements like icons, windows, and menus. These advances not withstanding, interaction in desktop environments is still done largely via keyboard and pointing device.

Voice interaction helps the users to execute functions through voice commands instead of manual input devices. This approach is useful when user needs the hands-free interaction or when traditional input devices are inconvenient to use.

In the last few years, it has become common to interact with smartphones, virtual assistants and smart home devices using voice. However, It is still limited in the desktop systems. The most serious part is that many smart voice assistants do not implement the secure authentication methods.

Another challenge is the language diversity. Current voice systems are available only in English commands, and are thus not usable by multi-national people. Users were still able to work in their native languages with the systems; therefore, multilingual support is something desirable for increasing accessibility.

To overcome the above problems, we proposed a new secure desktop automation framework — LapiDude which uses voice-based biometric authentication and natural language processing in multiple languages (mainly Telugu, Hindi and English) integrated together to develop an automated model for a human-computer interaction. It authenticates the speaker before executing commands and then processes the voice command in English, Hindi and Telugu. The framework ensures the privacy protection and fast system response.

2. RELATED WORK

Traditional desktop automation solutions have been mostly built around user keyboard shortcuts, scripts and GUI-based operations **[1]**. These techniques automate tasks that would otherwise need to be carried out by users, but still require users to interact directly with the system through traditional input devices.

This has led to the introduction of some voice-based systems that facilitate speech-driven interaction with computers **[2]**. But many of these systems rely on cloud-based speech processing services **[3]**. User voice data in such systems are sent to servers that may be remote, where they are recognized and processed for commands. This is a compelling approach with powerful computation; however, it comes at the cost of privacy, security, and reliance on high-speed networks **[4]**.

The other aspect we hope to figure out is user authentication, something which many voice automation tools leave off. Numerous voice systems react to all detected commands without identifying the speaker **[5]**. This may allow unauthorized users to execute system commands.

Moreover, most of the available tools work only in English **[6]**. This lack of multilingual support renders them less approachable in global contexts where users might find it easier to engage with systems using their local tongue.

These limitations have made it evident the necessity for a desktop automation system which will combine secure authentication, multilingual interaction and local processing in one framework.

Table1. Comparison of Existing Systems

Study / Tool	Methodology	Limitation	Ref
Traditional Desktop Automation	Automation using keyboard shortcuts, scripts and GUI-based operations	Requires manual interaction through traditional input devices	[1]
Cloud-based Voice Assistants	Speech recognition and command processing performed using cloud servers	Raises privacy concerns and depends on high-speed internet connectivity	[2]
Voice Command Systems	Enables users to interact with computers using speech-driven commands	Many systems process commands without verifying speaker identity	[3]
Multilingual Voice Interfaces	Supports interaction with systems using different languages	Many existing tools support only English, limiting accessibility for global users	[4]

3.PROPOSED SYSTEM AND METHODOLOGY

LapiDude, a secure voice-controlled desktop automation framework. This framework combines the aspect of speaker biometric authentication and multi-language command interpretation to prevent an unauthorized command from being executed on the system.

Table 2: Modules of the Proposed Voice-Controlled Desktop Automation System

Module	Description
Audio Capture	Records voice input from the microphone.
Speaker Authentication	Verifies the identity of the user through voice biometrics
Command Interpretation	Converts voice commands to text and identifies the intended action.
Automation Engine	Executes system commands such as opening applications or managing processes.

Next, the system runs entirely on user's devices and therefore minimizes privacy leakage and eliminates latency due to network communication. It enables the voice commands in English, Hindi, and Telugu so users can easily interact with the system using languages they already know.

3.1. System Architecture

The overall architecture of the proposed framework is shown in **Fig. 1**.

The process begins when the user provides a voice command through the microphone. The audio signal is captured and processed to extract voice features. These features are used to generate a speaker embedding that represents the identity of the speaker.

The generated embedding is then compared with the stored administrator voiceprint to verify whether the speaker is authorized. If authentication is successful, the system proceeds to interpret the spoken command and execute the corresponding automation task.

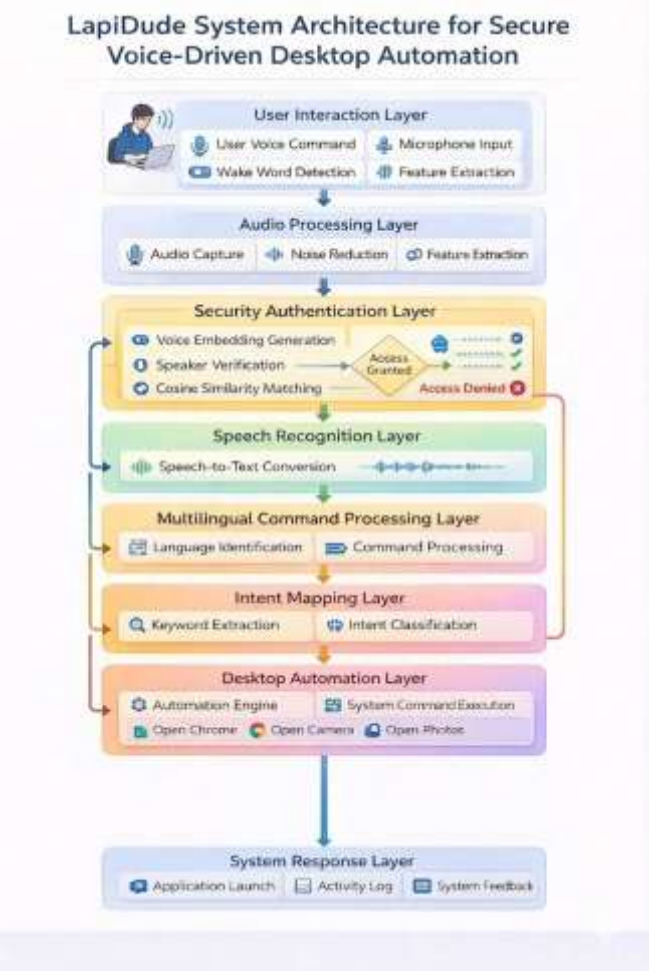

Fig. 1. LapiDude System Architecture

3.2. Mathematical Formulation

Speaker verification is performed using **cosine similarity** between the incoming voice embedding and the stored administrator embedding.

$$S = \frac{E_{input} \cdot E_{admin}}{\| E_{input} \| \times \| E_{admin} \|}$$

Where:

- E_{input} represents the embedding of the incoming voice signal
- E_{admin} represents the stored administrator embedding
- S represents the similarity score between the two vectors

If the similarity score satisfies:

$$S \geq 0.65$$

the speaker is considered authenticated.

3.3. Algorithms

Algorithm 1: Speaker Authentication

Input: Voice signal V

Output: Authentication status

1. Capture voice signal from the microphone.
2. Preprocess the audio signal to remove background noise.
3. Generate speaker embedding E_{input}.
4. Retrieve stored administrator embedding E_{admin}.
5. Compute cosine similarity between the embeddings.
6. If similarity score ≥ 0.65

   Return Authorized
7. Else

   Return Access Denied.

Algorithm 2: Multilingual Intent Mapping

Input: Transcribed command text C

Output: Automation action

1. Convert the command text to lowercase.
2. Remove punctuation and unnecessary symbols.
3. Tokenize the command into individual words.
4. Identify action keywords such as open, close, or run.
5. Identify the target application or system operation.
6. Match the keywords with predefined command templates.
7. Execute the corresponding automation function.

4. RESULTS AND DISCUSSION

LapiDude framework proposed was tried to be deployed as a voice recognition desktop assistant that could be able to do a biometric voice recognition and desktop multilingual automation. Several modules are included in the system to facilitate safe human-computer interaction: voice capture, speaker verification, command interpretation and automation execution. The findings prove that the assistant is capable of identifying users and processing the system commands via voice recognition.

4.1 System Interface

The main interface of proposed LapiDude assistant is located in Fig. 2. Status information on the interface has indicators of system conditions such as status of police and crime solutions, language preference, microphone listening, and connectedness of the server. Using the microphone module, the user is able to talk to the assistant or type commands in the command input field. The activity log panel also contains a list of the commands issued and response to the system.

Fig. 2. LapiDude multilingual desktop assistant main interface.

4.2 Voice Profile Registration

Figure 3 represents voice profile registration system that will enable a new user to be enrolled in the system. During the registration, a voice sample of the user is captured by the system and a voice embedding is formed which represent the unique vocal characteristics of the speaker. The voice profile that was generated is stored locally and is utilized to carry out authentication in future interactions.

Fig. 3. Voice profile of an authenticated user is registered.

4.3 Biometrics Voice Authentication

When a user enters voice, the system compares the voice embedding of the voice he has entered with a voice profile that is stored in the system. At the point when the degree of similarity exceeds the established limit, the system will provide access and enable the assistant. This would ensure that only those who are supposed to manipulate the automation system can do so.

Fig. 4. The effective voice verification of a registered customer.

4.4 Desktop Automation Performance

Figure 5 indicates that the proposed system can be automated. On a successful authentication the assistant accepts the commands and takes system actions as necessary. One can use voice commands to activate such applications as Chrome, Camera, Photos, and Microsoft Excel. This confirms the fact that the system can be able to automatically automate the desktop in real time and interact with voice.

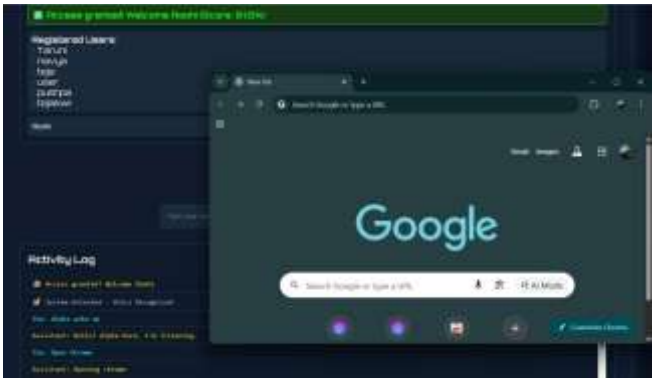

Fig. 5. Desktop automation performed using voice.

4.5 Activity Log Monitoring

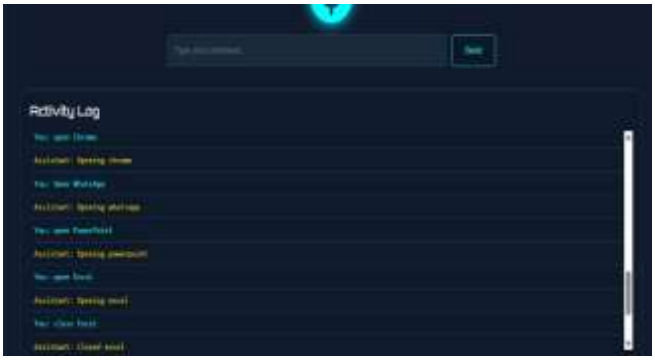

Fig. 6. Activity log containing the voice commands and system responses that have been adopted.

Fig. 6 is a list of the performed command activity log. The log shows all the interactions between the assistant and the user that included the given orders and system responses. The feature makes the users monitor the activity of the system and make sure that command is followed.

5.FUTURE SCOPE:

The system is an initial model, and can be extended with further research to include more regional languages in it so as to provide better accessibility for the users. Also the recognition rate can be improved by using several voice samples in the speaker authentication module.

But one area where there is room for improvement is to use more advanced natural language processing techniques so that users could include more complex, natural-sounding requests. Moreover, automation capabilities could also be added, to automate a wider range of desktop applications and OS functions.

CONCLUSION:

In this paper, LapiDude was described as a multilingual biometric-authenticated framework to be used in automation of desktops securely. The suggested system will involve speaker verification and multilingual voice command processing of the voice command in order to allow safe and efficient hands-free communication with desktop environments.

The structure of the system ensures that the speaker is an authorized user and therefore cannot execute system commands due to the verification of the identity of the speaker before any automation task is performed. The system is entirely run on local hardware, ensuring that users remain private and the system is also fast in response.

The results of the experiments prove that the system yields good authentication, low latency, and effective resource utilization. The features render the framework applicable to the practice of the implementation of secure voice-controlled desktop applications.

REFERENCES

[1] S. Furui, Digital Speech Processing, Synthesis, and Recognition, 2nd ed., CRC Press, 2018.

[2] . Snyder et al., *"X-vectors: Robust DNN embeddings for speaker recognition,"* in Proc. IEEE ICASSP, 2018, pp. 5329–5333.

[3] L. Wan, Q. Wang, A. Papir, and I. Lopez-Moreno, *"Generalized End-to-End Loss for Speaker Verification,"* in Proc. IEEE ICASSP, 2018.

[4] T. Kinnunen and H. Li, *"An Overview of Text-Independent Speaker Recognition: From Features to Supervectors,"* Speech Communication, vol. 52, no. 1, pp. 12–40, 2010.

[5] B. Logan, *"Mel Frequency Cepstral Coefficients for Speech Recognition,"* in IEEE Workshop, 2000.

[6] A. Graves, A. Mohamed, and G. Hinton, *"Speech Recognition with Deep Recurrent Neural Networks,"* IEEE Signal Processing Magazine, 2013.

[7] J. H. L. Hansen and T. Hasan, *"Speaker Recognition by Machines and Humans: A Tutorial Review,"* IEEE Signal Processing Magazine, vol. 32, no. 6, pp. 74–99, 2015.

[8] P. Viola and M. Jones, *"Rapid Object Detection Using a Boosted Cascade of Simple Features,"* IEEE CVPR, 2001. (optional general ML reference)

[9] M. Abadi et al., *"TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems,"* 2016.

[10] Al Sweigart, Automate the Boring Stuff with Python, No Starch Press, 2015.

[11] J. R. Pierce, An Introduction to Information Theory: Symbols, Signals, and Noise, Dover Publications, 2013.

[12] T. Dean and M. Wellman, Planning and Control, Morgan Kaufmann, 1991.

[13] S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 3rd ed., Pearson, 2010.