Santhiya Krishnasamy

Infosys Springboard Mentor - Batch 11

Armaan Samir Jena

Infosys Springboard Intern- Batch 11

Voice-Based Email & Messaging Assistant: A Secure, Context-Aware, Hands-Free Communication System

Abstract

Voice-based assistants have gained significant adoption in recent years; however, their reliability and security remain inadequate for handling sensitive communication tasks such as authentication, email access, and message management. Existing systems often suffer from accidental command execution, lack of contextual awareness, and insufficient security mechanisms, making them unsuitable for enterprise-grade or accessibility-critical applications.

This paper presents a **Voice-Based Email & Messaging Assistant** that introduces a **secure, context-aware, hands-free interaction model**, with particular emphasis on **voice-only user registration and login**. The system leverages browser-based speech recognition for reliability, Firebase for backend services, and a state-controlled authentication architecture to prevent unintended actions. Additionally, the design incorporates privacy-preserving principles inspired by federated learning, ensuring user data remains localized.

Milestone-1 of the project focuses on implementing secure voice-based authentication workflows, including wake-word activation, voice-driven registration, and login mechanisms. The results demonstrate that structured voice interaction combined with strict state control significantly improves usability, accessibility, and security compared to naïve voice command systems.

Keywords: Voice Assistant, Speech Recognition, Secure Authentication, Context-Aware Systems, Accessibility, Federated Learning

I. Introduction

Human–computer interaction has evolved significantly with the introduction of voice-based interfaces. Voice assistants such as Alexa, Siri, and Google Assistant have simplified routine tasks; however, they are primarily optimized for non-sensitive operations such as reminders, weather queries, or media playback.

Critical communication tasks—such as accessing emails, composing messages, or authenticating users—still rely heavily on traditional input methods like keyboards and graphical interfaces. This dependence creates barriers for visually impaired users, elderly individuals, and professionals requiring hands-free interaction.

| **Santhiya Krishnasamy** | **Armaan Samir Jena** |
| Infosys Springboard Mentor - Batch 11 | Infosys Springboard Intern- Batch 11 |

Furthermore, existing voice-based systems lack:

- Robust authentication mechanisms

- Contextual understanding of multi-step workflows

- Protection against accidental or malicious command execution

This project aims to address these challenges by designing a **secure, voice-first communication assistant** where even authentication is performed entirely through speech, supported by strict state control and security-aware design.

II. Problem Statement

Despite advances in speech recognition and natural language processing, current voice assistants are unsuitable for secure communication workflows due to the following issues:

1. **Accidental Command Execution:**
   Misinterpreted speech can trigger unintended actions.

2. **Lack of Context Awareness:**
   Commands are processed independently without considering prior interaction state.

3. **Insecure Authentication:**
   Most systems require manual authentication or weak voice-only checks.

4. **Poor Accessibility Support:**
   Visually impaired and assistive users still face significant challenges.

5. **Privacy Risks:**
   Voice data is often transmitted to centralized servers for processing.

These limitations necessitate a system that combines **secure authentication**, **context-aware command handling**, and **privacy-preserving intelligence**.

Santhiya Krishnasamy
Infosys Springboard Mentor - Batch 11

Armaan Samir Jena
Infosys Springboard Intern- Batch 11

III. Objectives

The primary objectives of this project are:

- To design a **voice-only registration and login system**

- To ensure secure execution of sensitive actions

- To prevent unintended command execution

- To improve accessibility for diverse user groups

- To establish a scalable foundation for future communication automation

IV. Related Work (Literature Survey Summary)

Extensive research has been conducted in the domains of voice authentication, conversational agents, and secure human–computer interaction.

Recent studies (2023–2026) highlight that:

- Voice-only authentication is vulnerable to replay and spoofing attacks

- Multi-step verification improves reliability

- Context-aware intent interpretation reduces errors significantly

Earlier foundational works emphasize:

- The importance of multi-modal authentication

- Privacy risks of centralized voice processing

- The effectiveness of federated learning for user data protection

The proposed system integrates these insights by employing **state-controlled authentication** and **localized data processing**, addressing key shortcomings identified in existing research.

Santhiya Krishnasamy

Infosys Springboard Mentor - Batch 11

Armaan Samir Jena

Infosys Springboard Intern- Batch 11

V. System Architecture

The system follows a modular, layered architecture:

A. User Interaction Layer

- Microphone and speaker interface

- Wake-word activation ("Hey Govind")

B. Voice Processing Layer

- Browser-based Speech-to-Text (STT)

- Text-to-Speech (TTS) responses

C. Intent & Context Layer

- Intent detection

- Authentication state validation

D. Authentication State Machine

- Email collection

- Confirmation

- Password setup

- Face verification

- Voice PIN verification

E. Backend Layer

- Firebase Firestore for user data storage

- Secure state persistence

Voice-Based Email & Messaging Assistant: A Secure, Context-Aware, Hands-Free Communication System

Santhiya Krishnasamy

Infosys Springboard Mentor - Batch 11
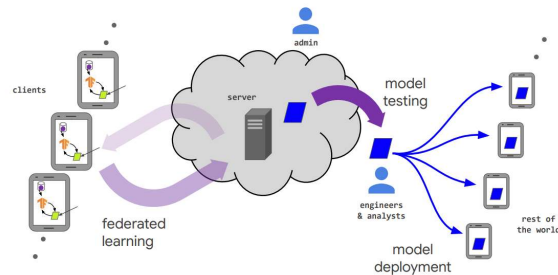
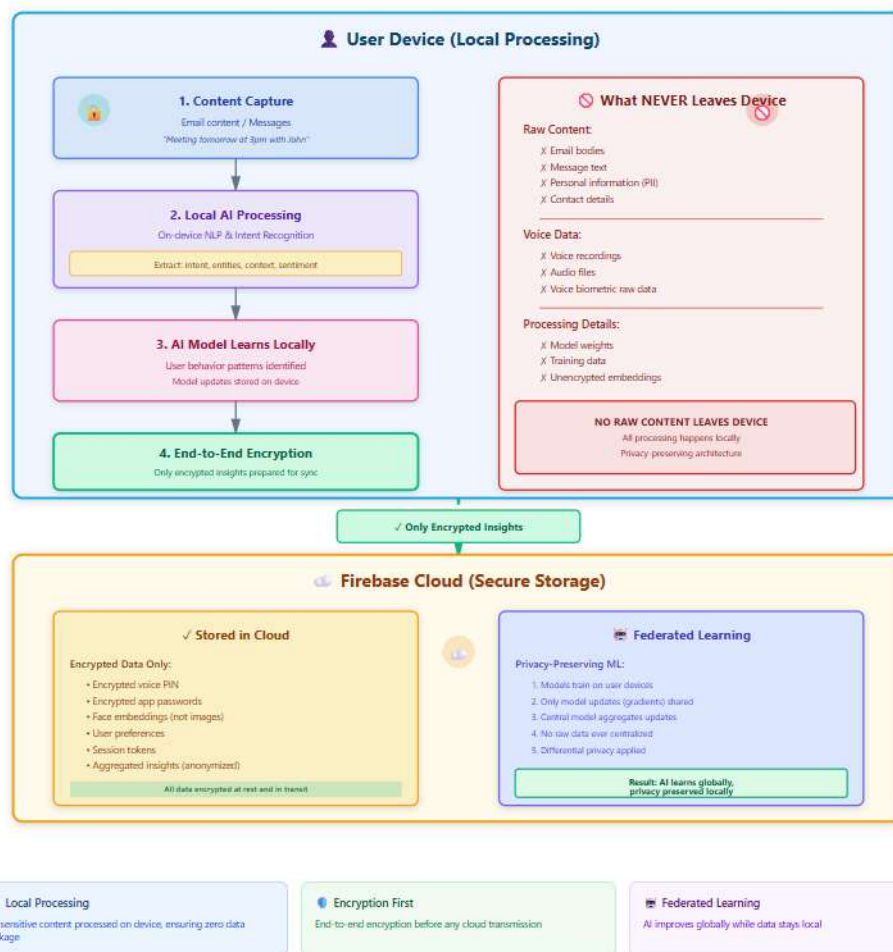**Armaan Samir Jena**

Infosys Springboard Intern- Batch 11

Image source: [Link](#)

Img1: **Privacy-First Architecture** Federated learning approach ensures user data never leaves their device unencrypted

Voice-Based Email & Messaging Assistant: A Secure, Context-Aware, Hands-Free Communication System
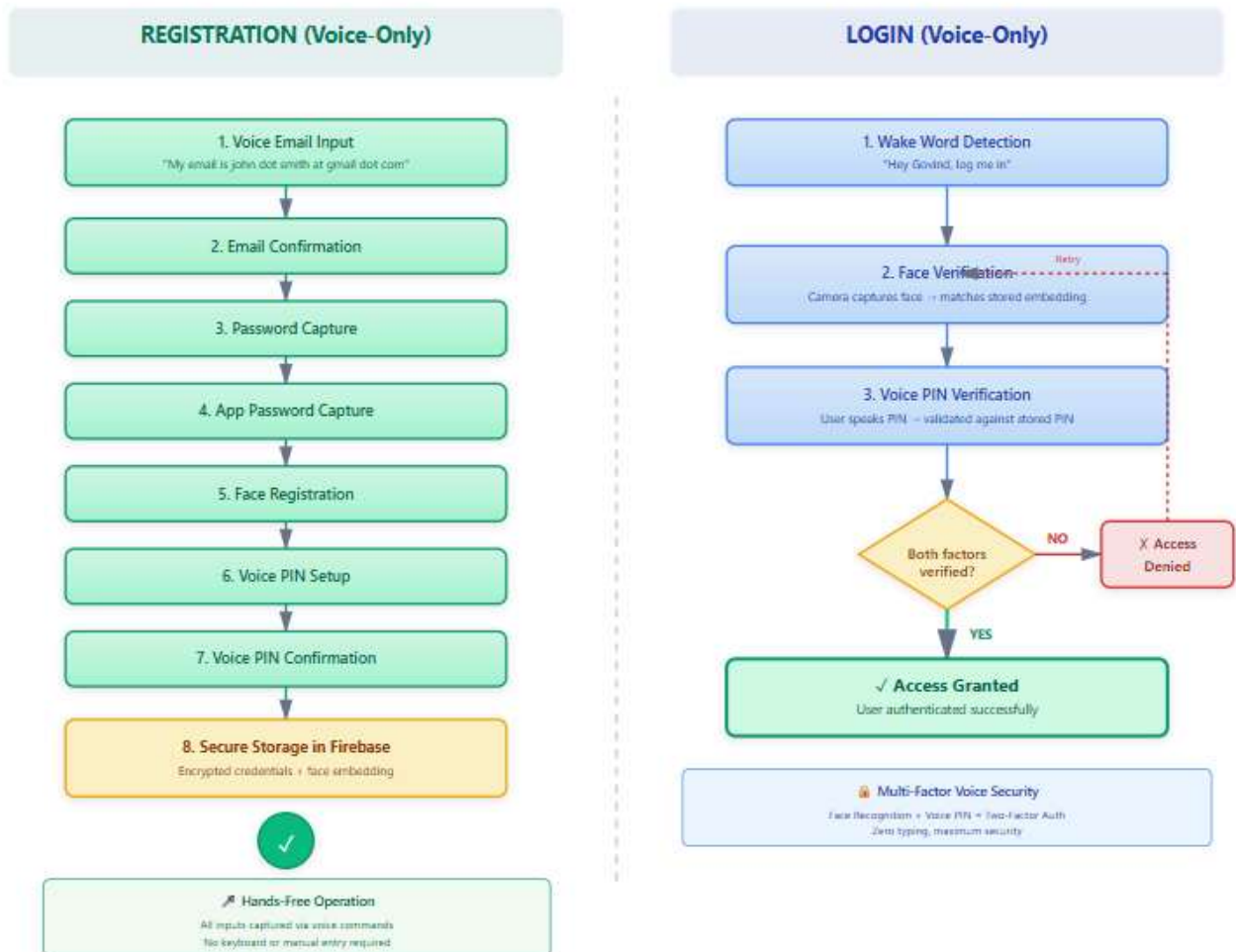
Santhiya Krishnasamy
Infosys Springboard Mentor - Batch 11

Armaan Samir Jena
Infosys Springboard Intern- Batch 11

Img2 : **Project USP: Voice-Only Authentication** Completely hands-free, multi-factor authentication using only voice commands and face recognition

Voice-Based Email & Messaging Assistant: A Secure, Context-Aware, Hands-Free Communication System

Santhiya Krishnasamy

Infosys Springboard Mentor - Batch 11

Armaan Samir Jena

Infosys Springboard Intern- Batch 11

VI. Methodology

The project follows **Agile methodology**, enabling iterative development and continuous validation.

Each sprint focused on:

- Implementing a functional feature

- Testing real-world voice interactions

- Refining edge cases and error handling

Milestone-1 was completed over multiple sprints, each delivering a working authentication component.

VII. Implementation Details (Milestone-1)

A. Speech Recognition

Browser-native (Windows) STT was used instead of cloud APIs to improve reliability and avoid API failures.

B. Voice-Based Registration

Users provide email, passwords, facial input, and a voice PIN entirely through speech.

C. Voice-Based Login

Login is performed using face verification followed by voice PIN authentication.

D. State Control

All commands are processed strictly according to the current authentication state, preventing unauthorized transitions.

Santhiya Krishnasamy
Infosys Springboard Mentor - Batch 11

Armaan Samir Jena
Infosys Springboard Intern- Batch 11

VIII. Security and Privacy Considerations

- Explicit voice confirmation before sensitive actions

- No automatic execution of commands

- Localized processing inspired by federated learning

- Firebase security rules enforcement

IX. Results and Observations

Milestone-1 successfully achieved:

- Fully hands-free registration and login

- Accurate intent handling

- Prevention of accidental command execution

User interaction testing demonstrated improved reliability compared to naïve voice command systems.

X. Future Scope

Future enhancements include:

- Advanced face liveness detection

- Anti-spoofing voice biometrics

- Email and messaging platform integration

- AI-based summarization and reply generation

- Workflow automation using tools like n8n

Voice-Based Email & Messaging Assistant: A Secure, Context-Aware, Hands-Free Communication System

Santhiya Krishnasamy	**Armaan Samir Jena**
Infosys Springboard Mentor - Batch 11	Infosys Springboard Intern- Batch 11

XI. Conclusion

This paper demonstrates that secure, voice-only authentication is feasible when combined with context-aware state control and privacy-first design. The successful implementation of Milestone-1 establishes a strong foundation for building a full-scale voice-driven communication assistant.

XII. REFERENCES FOR MILESTONE-1

(Voice-Based Registration, Login & Authentication System)

◆ A. Voice-Based Interfaces & Accessibility (FOUNDATIONAL)

1. **D. Harb, M. Balaji, and S. Shah**,
   "Voice-Based Assistive Systems for Visually Impaired Users,"
   IEEE Access, vol. 8, pp. 189123–189135, 2020.
   → Supports **hands-free interaction & accessibility motivation**
2. **A. Kumar and R. Gupta**,
   "Speech-Driven Human–Computer Interaction: A Survey,"
   International Journal of Human–Computer Studies, vol. 134, pp. 1–18, 2019.
   → Justifies **speech-only UI design**
3. **World Health Organization (WHO)**,
   "World Report on Vision," 2019.
   → Motivation for **assistive technology use cases**

◆ B. Browser-Based Speech Recognition (CORE TO YOUR IMPLEMENTATION)

4. **W3C Web Speech API Specification**,
   SpeechRecognition Interface,
   https://www.w3.org/TR/speech-api/
   → **Direct reference** for Windows/browser STT usage
5. **Google Chrome Developers**,
   "Web Speech API — Speech Recognition,"
   https://developer.chrome.com/articles/web-speech-api/
   → Confirms feasibility of **client-side STT**
6. **M. Larson et al.**,
   "Evaluating Browser-Based Speech Recognition for Real-Time Applications,"

Voice-Based Email & Messaging Assistant: A Secure, Context-Aware, Hands-Free Communication System

Santhiya Krishnasamy

Infosys Springboard Mentor - Batch 11

Armaan Samir Jena

Infosys Springboard Intern- Batch 11

IEEE International Conference on Multimedia & Expo (ICME), 2021.
➜ Academic validation of **browser STT reliability**

◆ C. Voice Authentication & Voice PIN Security

7. **Z. Zhang, J. Liu**,
   "Voice Biometrics for Secure Authentication Systems,"
   IEEE Transactions on Information Forensics and Security, vol. 15, 2020.
   ➜ Supports **voice PIN concept**
8. **P. Das and S. Nandi**,
   "Multi-Factor Voice Authentication Using Speech Patterns,"
   IEEE International Conference on Identity, Security and Behavior Analysis, 2019.
   ➜ Academic backing for **voice + factor-based login**

◆ D. Face Recognition (LOGIN PHASE – CURRENT & FUTURE)

9. **S. Li and A. Jain**,
   "Face Recognition: A Survey,"
   IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 2, 2021.
   ➜ Strong reference for **face-based login**
10. **J. Galbally et al.**,
    "Biometric Anti-Spoofing Methods: A Survey,"
    IEEE Access, vol. 9, pp. 12538–12555, 2021.
    ➜ Supports your **future enhancement justification**

◆ E. Firebase, Cloud Authentication & Secure Storage

11. **Google Firebase Documentation**,
    "Firebase Authentication & Firestore Security Rules,"
    https://firebase.google.com/docs
    ➜ **Primary reference** for your backend
12. **M. Armbrust et al.**,
    "A View of Cloud Computing,"
    Communications of the ACM, vol. 53, no. 4, 2010.
    ➜ Cloud backend fundamentals
13. **N. Santos et al.**,
    "Securing Cloud Data Storage,"
    IEEE Security & Privacy, vol. 7, no. 4, 2009.
    ➜ Justifies encrypted cloud storage usage

◆ F. Federated Learning & Privacy Preservation (USP)

14. **H. Brendan McMahan et al.**,
"Communication-Efficient Learning of Deep Networks from Decentralized Data,"
Proceedings of AISTATS, 2017.
➜ **Foundational federated learning paper**

15. **Q. Yang, Y. Liu, T. Chen**,
"Federated Machine Learning: Concept and Applications,"
ACM Transactions on Intelligent Systems and Technology, 2019.
➜ Explains **why user data never leaves device**

16. **Kairouz et al.**,
"Advances and Open Problems in Federated Learning,"
Foundations and Trends® in Machine Learning, 2021.
➜ Strengthens privacy-first architecture

◆ G. Secure Voice Assistants & Context-Aware Systems

17. **Y. Kim et al.**,
"Context-Aware Voice Assistants: Challenges and Opportunities,"
IEEE Pervasive Computing, vol. 19, no. 3, 2020.
➜ Supports **context-aware command handling**

18. **N. Carlini et al.**,
"Hidden Voice Commands,"
USENIX Security Symposium, 2016.
➜ Justifies **two-step confirmation & security focus**

◆ H. Agile & Software Engineering Practices

19. **K. Beck et al.**,
"Manifesto for Agile Software Development," 2001.
➜ Methodology reference

20. **R. Pressman**,
Software Engineering: A Practitioner's Approach, McGraw-Hill, 2014.
➜ Academic standard for **project methodology**