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

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.

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**.

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.

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

V. System Architecture

The proposed system follows a **modular, layered, and security-first architecture** designed to ensure scalability, reliability, and safe execution of voice-driven commands. Each layer is decoupled to allow independent evolution, testing, and future extensibility.

A. User Interaction Layer

The User Interaction Layer acts as the **primary interface between the user and the system**, focusing on accessibility and hands-free operation.

- **Microphone Interface**
  Captures continuous or event-triggered voice input from the user through the browser or device microphone. Noise handling and input buffering ensure clarity and stability during voice capture.

- **Speaker Interface**
  Delivers system responses, confirmations, and alerts via synthesized speech, enabling full interaction without visual dependency.

- **Wake-Word Activation ("Hey Govind")**
  A wake-word detection mechanism ensures that the system listens actively only after intentional invocation. This reduces accidental triggers, preserves system resources, and improves privacy by preventing unintended command execution.

B. Voice Processing Layer

This layer converts raw audio signals into actionable data and delivers audible responses back to the user.

- **Browser-Based Speech-to-Text (STT)**
  Voice input is converted into text using browser-native or cloud-supported STT engines, enabling low-latency transcription and compatibility across devices without additional hardware requirements.

- **Text-to-Speech (TTS) Responses**
  System-generated responses, confirmations, and prompts are converted into natural-sounding speech, ensuring clarity and accessibility for visually impaired and hands-free users.

C. Intent & Context Layer

The Intent & Context Layer acts as the **intelligence core** of the system.

- **Intent Detection**
  Transcribed text is analyzed to determine user intent (e.g., login, read email, compose message). This abstraction allows flexible command phrasing rather than rigid keyword-based control.

- **Context Tracking**
  Maintains conversational context across multiple interactions, enabling multi-step workflows such as email composition or authentication sequences.

- **Authentication State Validation**
  Ensures that commands are evaluated against the user's current authentication and session state before execution, preventing unauthorized actions.

D. Authentication State Machine

This component enforces **secure, step-by-step user verification** using a deterministic state-based model.

- **Email Collection**
  Captures and validates the user's email address as the primary identifier.

- **Confirmation Phase**
  Verifies user intent and correctness of provided information before proceeding.

- **Password Setup**
  Allows secure password creation or verification when required.

- **Face Verification**
  Integrates biometric validation for enhanced security, particularly for high-risk operations.

- **Voice PIN Verification**
  Adds an additional voice-based authentication factor, strengthening identity assurance in hands-free scenarios.

The state machine ensures that transitions occur only in valid sequences, eliminating authentication bypass risks.

E. Backend Layer

The Backend Layer provides **secure data handling, persistence, and state management**.

- **Firebase Firestore for User Data Storage**
  Stores structured user data, authentication states, and integration metadata in a secure, scalable NoSQL environment.

- **Secure State Persistence**
  Authentication progress and session-related data are preserved across interactions, ensuring seamless recovery from interruptions or network delays.

In addition to the authentication-centric architecture introduced in Milestone-1, the backend now supports **service integration logic and session-aware command execution**, enabling real-world communication workflows.

F. Application Integration Layer (Milestone-2)

Introduced in Milestone-2, this layer enables **controlled and secure interaction with external communication services**.

- **Gmail Integration Using OAuth 2.0**
  Allows secure, permission-based access to user email accounts without exposing credentials.

- **Firebase-Based Session Validation**
  Ensures that only authenticated and active sessions can trigger third-party service actions.

- **Command Gating Based on Authentication State**
  All commands are evaluated against the Authentication State Machine before execution.

All messaging operations—such as reading or sending emails—are executed **only after successful verification**, preventing accidental or malicious actions.

G. Data Persistence & Session Layer

This layer ensures **long-term reliability and continuity** of user interactions.

- **Firestore Stores**:
    - User profile metadata
    - Connected email account details
    - Session tokens and authentication states
- **Firebase Authentication Maintains**:
    - Secure login sessions
    - Token-based access control
    - Automatic session expiration and renewal

This separation of authentication and application data improves security while maintaining high system performance.

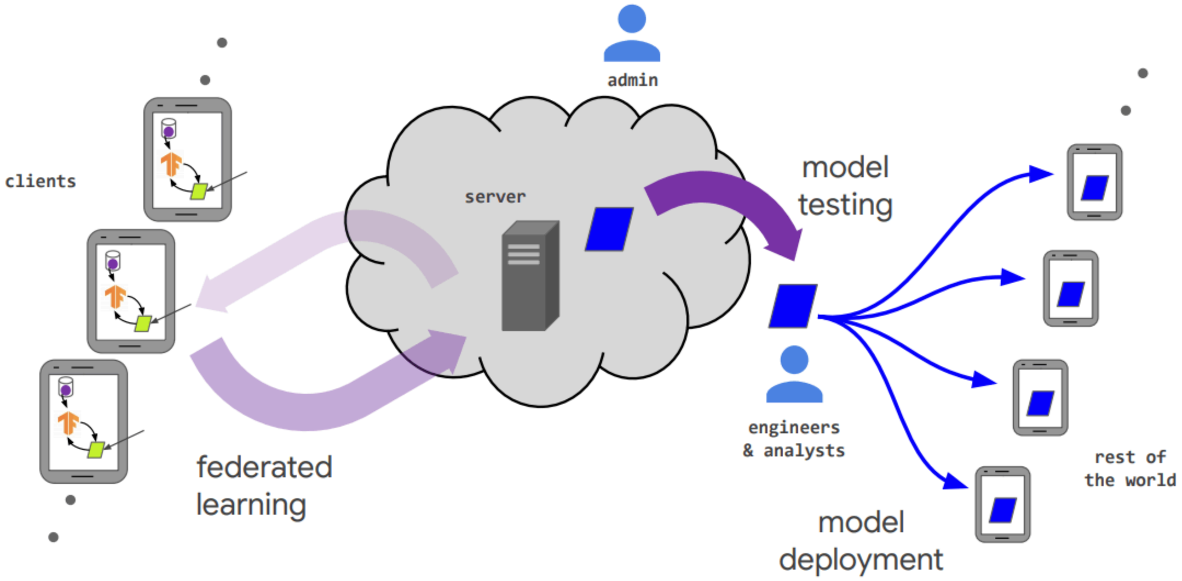
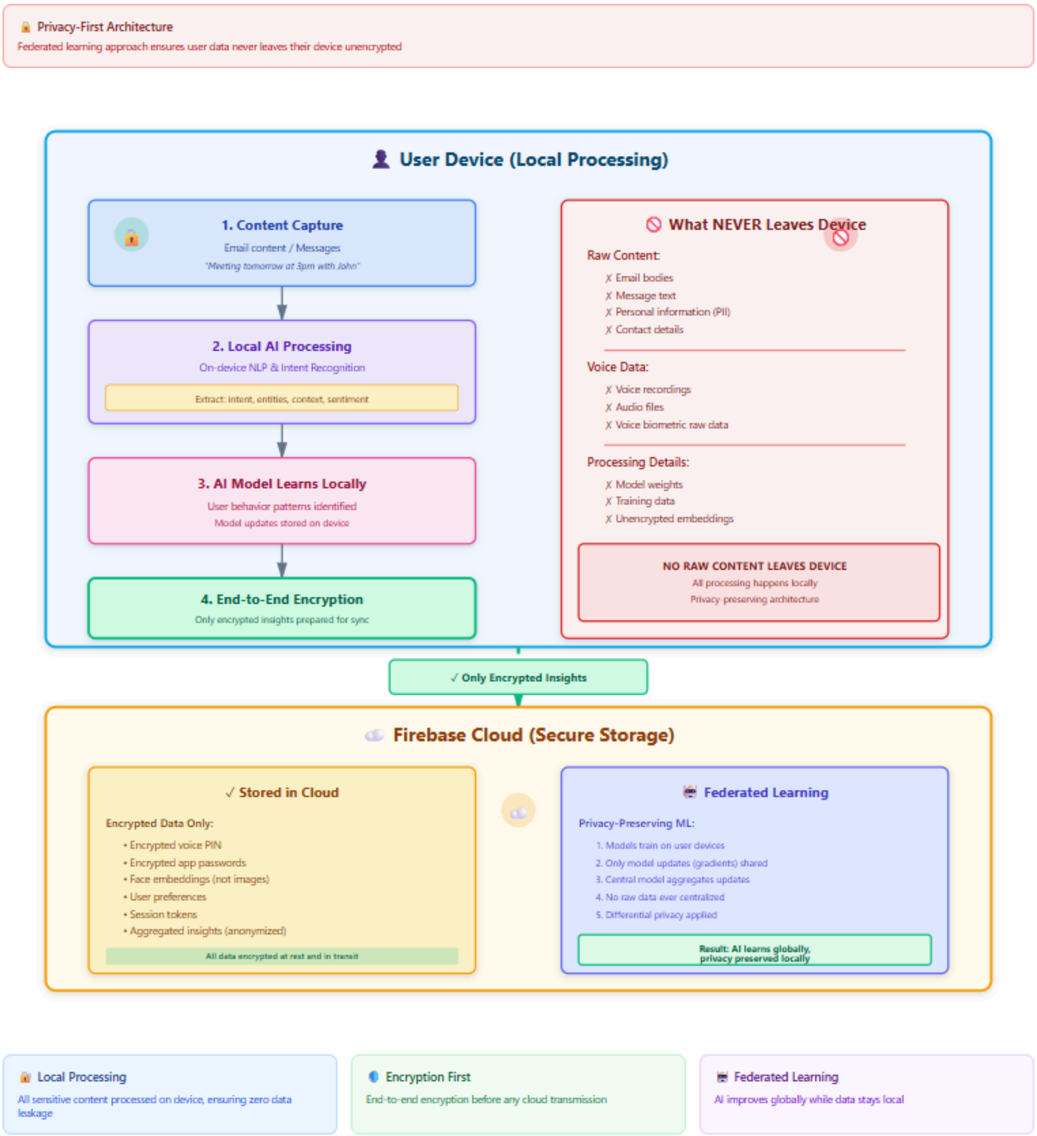


Image source: Link

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



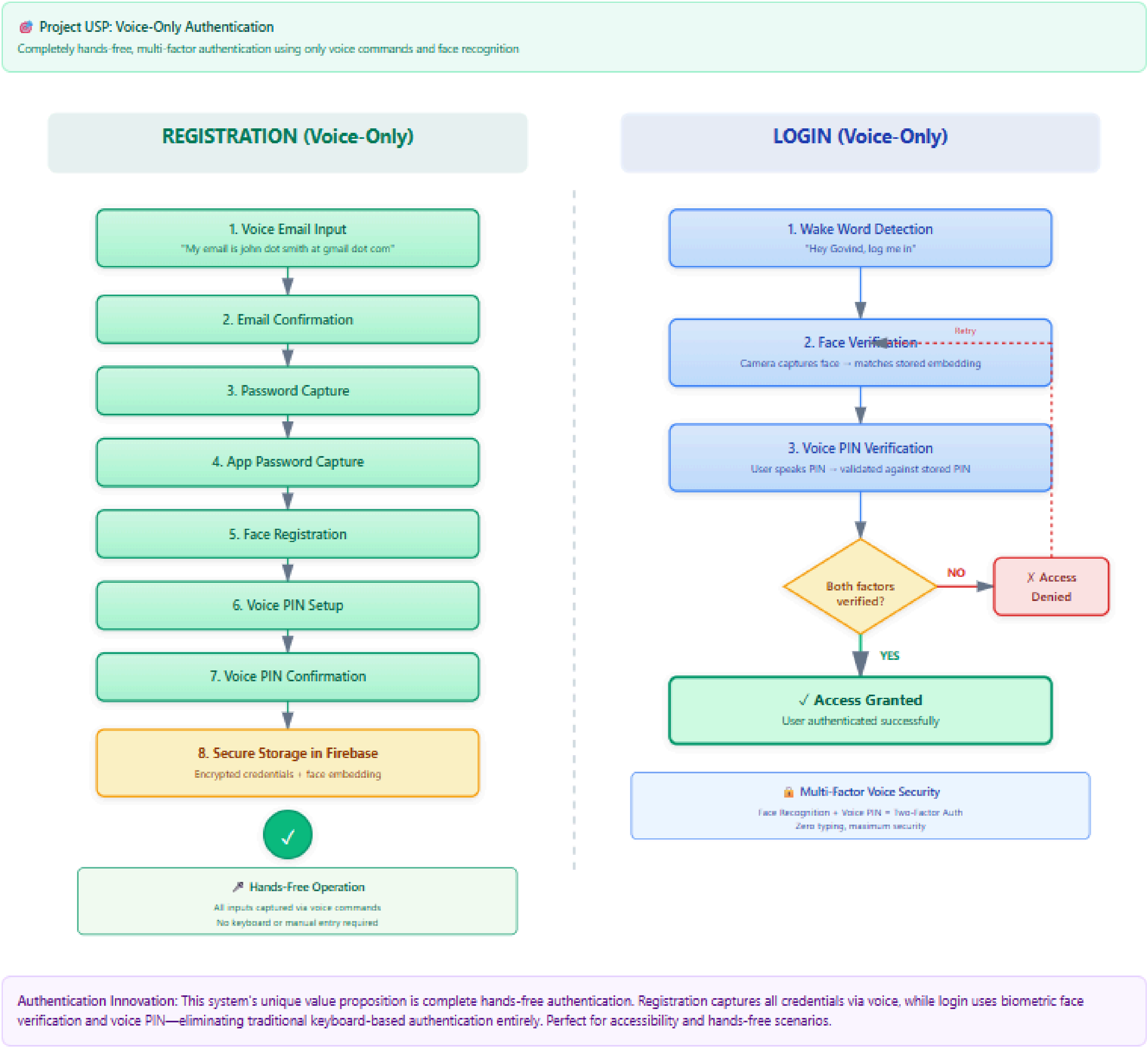
Privacy-First Architecture
Federated learning approach ensures user data never leaves their device unencrypted
User Device (Local Processing)
1. Content Capture
Email content / Messages
"Meeting tomorrow at 3pm with John"
2. Local AI Processing
On-device NLP & Intent Recognition
Extract: intent, entities, context, sentiment
3. AI Model Learns Locally
User behavior patterns identified
Model updates stored on device
4. End-to-End Encryption
Only encrypted insights prepared for sync
What NEVER Leaves Device
Raw Content:
✗ Email bodies
✗ Message text
✗ Personal information (PII)
✗ Contact details
Voice Data:
✗ Voice recordings
✗ Audio files
✗ Voice biometric raw data
Processing Details:
✗ Model weights
✗ Training data
✗ Unencrypted embeddings
NO RAW CONTENT LEAVES DEVICE
All processing happens locally
Privacy-preserving architecture
✓ Only Encrypted Insights
Firebase Cloud (Secure Storage)
✓ Stored in Cloud
Encrypted Data Only:
• Encrypted voice PIN
• Encrypted app passwords
• Face embeddings (not images)
• User preferences
• Session tokens
• Aggregated insights (anonymized)
All data encrypted at rest and in transit
Federated Learning
Privacy-Preserving ML:
1. Models train on user devices
2. Only model updates (gradients) shared
3. Central model aggregates updates
4. No raw data ever centralized
5. Differential privacy applied
Result: AI learns globally, privacy preserved locally
Local Processing
All sensitive content processed on device, ensuring zero data leakage
Encryption First
End-to-end encryption before any cloud transmission
Federated Learning
AI improves globally while data stays local

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



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

REGISTRATION (Voice-Only)
1. Voice Email Input
"My email is john dot smith at gmail dot com"
2. Email Confirmation
3. Password Capture
4. App Password Capture
5. Face Registration
6. Voice PIN Setup
7. Voice PIN Confirmation
8. Secure Storage in Firebase
Encrypted credentials + face embedding
✓
Hands-Free Operation
All inputs captured via voice commands
No keyboard or manual entry required
LOGIN (Voice-Only)
1. Wake Word Detection
"Hey Govind, log me in"
2. Face Verification
Camera captures face → matches stored embedding
Retry
3. Voice PIN Verification
User speaks PIN → validated against stored PIN
Both factors verified?
NO
✗ Access Denied
YES
✓ Access Granted
User authenticated successfully
Multi-Factor Voice Security
Face Recognition + Voice PIN = Two-Factor Auth
Zero typing, maximum security

Authentication Innovation: This system's unique value proposition is complete hands-free authentication. Registration captures all credentials via voice, while login uses biometric face verification and voice PIN—eliminating traditional keyboard-based authentication entirely. Perfect for accessibility and hands-free scenarios.

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

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.

The Agile methodology continued into Milestone-2 with a focus on **feature integration, robustness, and security hardening**.

Milestone-2 sprints focused on:

- Integrating Gmail services
- Persisting user authentication states
- Implementing controlled UI flows for voice actions
- Logging and debugging cloud interactions

E. Each iteration involved:

- Voice-command testing
- Firebase rule validation
- UI behavior verification under partial and failed authentication scenarios

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

VII. Implementation Details

A. Implementation Details (Milestone-1)

Milestone-1 focuses on establishing a **secure, voice-first authentication foundation** for the system. All interactions are designed to be hands-free, accessible, and resistant to unauthorized access.

1. Speech Recognition

Browser-native Speech-to-Text (STT) capabilities available on Windows platforms were utilized instead of cloud-based APIs.

- This approach eliminates dependency on third-party STT services, reducing latency and avoiding API quota limitations or downtime.
- Local processing improves reliability during authentication flows where accuracy and responsiveness are critical.
- Using browser-native STT also enhances privacy, as voice data is not transmitted to external servers during initial authentication.

B. Voice-Based Registration

User registration is conducted entirely through voice interaction, ensuring accessibility and eliminating the need for keyboard or visual input.

- Users provide essential credentials such as email ID and password through spoken input.
- Facial data is captured via the device camera and securely stored as a reference for future verification.
- A voice PIN is recorded and associated with the user profile, enabling an additional authentication factor tailored for voice-based systems.
- The registration flow is sequentially guided using spoken prompts and confirmations, minimizing input errors.

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

C. Voice-Based Login

The login process is designed as a **multi-factor, voice-driven authentication flow**.

- Face verification is performed first to establish visual identity.
- Upon successful facial recognition, voice PIN verification is initiated to confirm the user’s presence and intent.
- This layered approach reduces the risk of impersonation and strengthens overall system security, especially in hands-free environments.

D. State Control

A strict state-control mechanism governs all authentication actions.

- Commands are processed only if they align with the user’s current authentication state.
- Invalid or out-of-order commands are rejected automatically.
- This prevents unauthorized state transitions, such as bypassing verification steps or accessing protected features prematurely.
- The state machine ensures deterministic behavior, improving system reliability and auditability.

B. Implementation Details (Milestone-2)

Milestone-2 extends the authentication framework by introducing **secure service integration and session-aware command execution**, enabling real-world email interaction.

E. Gmail Integration

Gmail functionality is integrated using OAuth 2.0 to ensure secure, permission-based access.

- Users explicitly authorize Gmail access through Google’s OAuth consent flow.
- Access tokens are generated and stored securely without exposing user credentials.
- The system supports core email actions triggered via voice commands, including:

  - Reading inbox emails
  - Composing emails
  - Sending emails after confirmation

Email actions are executed only when:

- The user session is active
- Authentication state is fully verified
- Required permissions are granted

This gating mechanism ensures safe and intentional email operations.

F. Firebase Firestore Enhancements

Firestore data models were extended to support advanced session and integration requirements.

Additional collections and fields include:

- Connected application and service metadata
- Authentication and verification timestamps
- Status flags for face and voice verification

These enhancements enable:

- Seamless session continuity across interactions
- Faster authentication checks without redundant verification
- Structured, audit-friendly data storage for security reviews and debugging

G. Firebase Storage Usage

Firebase Storage is used for securely managing biometric assets.

- Facial reference images captured during registration are stored in encrypted storage buckets.
- Verification-related assets are isolated from general application data.
- Access is governed by strict Firebase security rules, ensuring:
    - Only authenticated users can access their own assets
    - No public or unauthorized access to sensitive biometric data

This separation enhances privacy and compliance readiness.

H. UI & State-Synchronized Pages

Dedicated user interface states were introduced to reflect backend authentication and session status accurately.

Key UI states include:

- Login and verification screen
- Gmail OAuth authorization screen
- Email interaction dashboard
- Each UI page is tightly synchronized with backend authentication states.
- UI transitions occur only after backend validation is complete.
- This design prevents UI-level bypass attacks where users might attempt to access protected views directly.

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.

Milestone-2 achieved the following outcomes:

- Successful Gmail integration without compromising security
- Reliable session persistence across voice interactions
- Accurate command execution bound to authentication state
- Elimination of accidental email actions

User testing indicated:

- Improved confidence in voice-driven actions
- Reduced command misfires
- Seamless transition from authentication to communication tasks

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

Future enhancements now extend beyond authentication to full automation:

- Multi-platform messaging (WhatsApp, Slack)
- AI-driven email summarization
- Context-aware reply generation
- Workflow automation using n8n
- Enterprise policy enforcement layers
- Federated on-device learning for personalization

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

◆ 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,"
   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**

◆ I. Email Integration, Cloud Backend & Secure Sessions

21. **Google Developers,- *“Gmail API Overview,”***
https://developers.google.com/gmail/api
22. **OAuth 2.0 Authorization Framework, RFC 6749, IETFFirebase Documentation,**
“Managing User Sessions & Security Rules,”
https://firebase.google.com/docs
23. **N. Carlini et al., *“Hidden Voice Commands,”* USENIX Security Symposium**