A Voice Biometric UPI Transaction System for Visually Impaired Users Using Mel Frequency Cepstral Coefficients and Cosine Similarity

[1]K.Devi Sundari, [2]A.Bindu Poornima, [3]P.Keerthana, [4]K.Sai Lakshmi Prasanna

[1]UG Scholar, [2]UG Scholar, [3]UG Scholar, [4] UG Scholar

[1]Bharath Institute of Science and Technology, Chennai, India,

[1]devikoppana98@gmail.com ,[2]binduathyam@gmail.com ,[3]keerthanapotturi66@gmail.com,

[4]sailakshmiprasanna@gmail.com

Abstract— ***This paper presents the real-world operation of a voice-based payment application using a mock database for voice-controlled payments, facilitating secure and seamless digital transactions through voice interaction. The system is designed to increase accessibility for visually impaired, elderly, and physically challenged users by eliminating the need for touchscreen-based inputs. The system architecture integrates user voice input through a mobile phone, captures the voice using an STT algorithm, and responds through a TTS algorithm. For authentication, (MFCC) mel frequency cepstral coefficients are used to detect the user's voice, and it acts as a voice biometric for the proposed system. The system performs CRUD operations for detecting the intent or condition for Transaction Management, and filtering operations will also be performed by the system and profile management is also included.***

Keywords—**Unified Payments Interface (UPI), Voice Authentication, Speech Recognition, Mel-Frequency Cepstral Coefficients (MFCC), Natural Language Understanding (NLU), Speech-to-Text, Text-to-Speech (TTS), Digital Payments, Python, React Native, Hands-free Transactions**

I. INTRODUCTION

The present generation is more digitalised, and everything can be done through the network, even payments. But in this digitalised world, some people cannot operate a phone for payments properly, like urban people, elderly people or visually impaired people. For this problem, we have proposed a system that can be mainly used for physically challenged/elder people/ visually impaired people to do transactions easily through their voice and using a voice-driven system. This paper prototype is a voice-driven Unified Payments Interface (UPI) system that employs Mel-Frequency Cepstral Coefficients (MFCC) and Cosine Similarity for secure voice authentication. The system, implemented in Python, performs essential financial operations such as fund transfer and balance inquiry through a mock database, aiming to enhance accessibility and inclusivity in digital payments. Digital payments have become essential to modern life, especially in India, where the development of the Unified Payments Interface (UPI) has transformed Financial transactions are increasingly conducted via UPI, which allows instant, seamless, and real- time transfers. However, accessibility remains a significant obstacle for visually impaired, elderly, and physically challenged individuals.

Most current payment apps depend heavily on visual interfaces and manual text entry, limiting usability for these users. To address this, there is a rising demand for voice- driven and intelligent payment solutions that enable natural speech interaction. Voice interfaces not only enhance user convenience but also support financial inclusion, making digital payments accessible to all, regardless of physical ability. Recent advances in Automatic Speech Recognition (ASR) and Natural Language Understanding (NLU) have facilitated the creation of voice- enabled financial services that accurately interpret commands and execute secure transactions. Building on these developments, the proposed system offers a lightweight, accessible voice- based UPI prototype designed for inclusive fintech applications. By combining MFCC- based authentication, ASR, and NLU, it provides a secure, reliable, and user- friendly experience, effectively bridging the gap between technology and accessibility in digital payments. Several studies have explored voice- based systems to improve accessibility and security in digital payments and human-computer interaction. Doshi et al. [1] proposed a UPI security system utilising MFCC and CNN- based voice authentication to enhance transaction security through biometric verification. Samuel et al. [2] developed a voice chatbot with Amazon Lex and TTS to facilitate financial transactions for visually impaired users. Wang et al. [3] introduced VAuth, a continuous voice authentication framework employing wearable devices to detect vocal- tract vibrations and improve anti- spoofing security. Hashan et al. [4] created VoiceNote, enabling hands- free money transfers via speech recognition and natural language processing. Feng et al. [5] proposed a continuous authentication system for voice assistants using wearable- sensed vibrations, achieving high.

II. MATERIALS AND METHODS

A. Speech Interface: In Speech Recognition translates the spoken words of the user into written data that can be processed by the system. It uses the Speech Recognition library in Python to capture and determine speech commands via a microphone. The ASR module operates on the principles of identifying acoustic features of the input signal and matching them with pre-trained language and acoustic models to produce text output. This helps the system to comprehend voice commands.

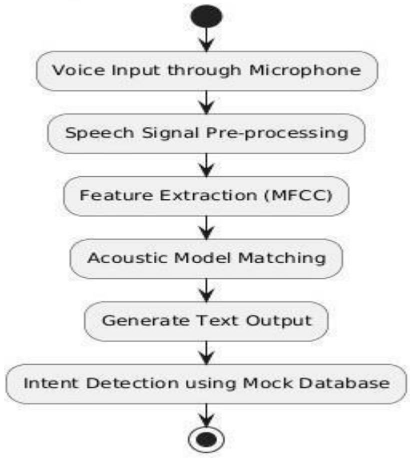

Figure 1: Speech Interface

B. Language Understanding (NLU): The NLU module reads the identified text and identifies the intent of the command that the user enters. It also removes objects like the amount, the name of the recipient or the type of transaction. Intent → Transfer Money, Amount → 200 INR, Recipient -> Sita.

C. Voice authentication: a biometric security method that authenticates the identity of a user depending on the distinctive features of their voice. As opposed to the conventional authentication processes (e.g., passwords or PINs), voice authentication is based on physiological and behavioural characteristics, such as the shape of the vocal tract, pitch, tone, and speaking style. This renders it a convenient and easy-to-use method, since it enables hands-free and natural interaction, but offers a high level of security. Voice authentication is generally a two-step process, which includes enrollment and verification. At the enrolment stage, the system captures the voice of the user and isolates the distinguishing characteristics, which is often achieved by methods such as MFCC (Mel-Frequency Cepstral Coefficients), to form a reference voice model. During the verification phase, another voice sample is recorded, and the same step is repeated, and the features extracted are compared with the stored model using similarity measures like cosine similarity or other classification algorithms. On this comparison, the system identifies whether the user is genuine or not.

III. SYSTEM OVERVIEW

The voice input module records the speech of the user using a microphone, and the voice authentication module processes the speech. This module applies Mel-Frequency Cepstral Coefficients (MFCC) with Cosine Similarity to confirm the identity of the speaker, in the footsteps of Doshi et al. [1] and VAuth [3], [5]. After the speaker has been authenticated, the ASR module translates the spoken command into text. The NLU module then processes the text to identify the intent of the user (e.g., fund transfer, balance inquiry) and retrieves the applicable parameters, such as amount and recipient, which are based on the methods in Samuel et al. [2] and VoiceNote [4], [8]. The transaction processing module then performs the requested operation on a simulated UPI database, and the TTS module gives audio feedback on the status of the transaction or balance information. This modular design guarantees accuracy, efficiency and user-friendliness.

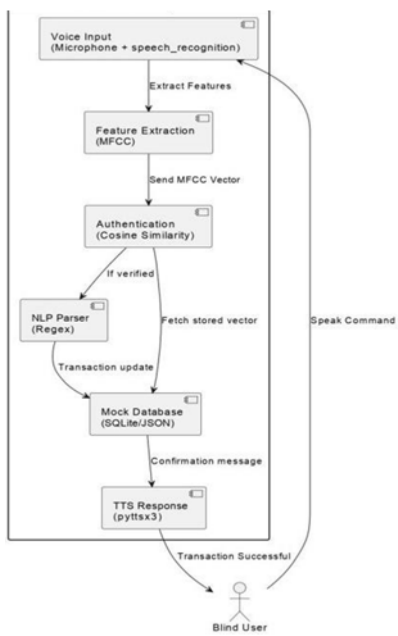

Figure 1: System Overview

A. System Workflow:

The system workflow initiates when the user provides a voice command. The process proceeds through the following sequential steps:

a using MFCC feature extraction and Cosine Voice Authentication: The input speech is analyzed Similarity to confirm the user's identity.

b Speech Recognition: The verified voice input is converted into text using the ASR module.

c Intent Detection: The NLU module processes the recognised text to determine the user's intent and extract contextual entities.

d Transaction Execution: The operation (e.g., fund transfer) is executed on the mock database.

e Response Generation: The TTS module generates a spoken confirmation message, completing a secure and fully voice-based transaction cycle.

B. Tools and Technology:

The system is implemented using Python, leveraging the following key libraries and frameworks:

- SpeechRecognition – for speech-to-text conversion (ASR)
- pyttsx3 and gTTS – for text-to-speech (TTS) feedback
- NumPy – for signal and feature processing
- pandas – for handling the mock database and transaction records

A mock database simulates UPI accounts, balances, and transaction histories, providing a controlled environment for testing and demonstration.

IV. PROPOSED METHODOLOGY

Speech Recognition

Once the audio is converted to text, the Natural The captured audio is processed by the Speech-to-Text (STT) module, which converts spoken commands into textual data. Advanced signal processing techniques, including noise suppression and voice activity detection, are applied to enhance recognition accuracy. The module ensures that even in environments with background noise, the speech input is correctly interpreted.

Command Processing

The textual command is analyzed by the Command Processing Module, which uses Natural Language Processing (NLP) techniques to determine the user's intent. The module performs tokenization, syntactic parsing, and semantic interpretation to extract meaningful actions from the input. For operations involving sensitive data or transactions, the module interacts with the User Authentication Module to verify identity before executing any command, ensuring system security.

User Authentication

The User Authentication Module validates the user's credentials before permitting transaction execution. Authentication may involve voice pattern recognition, password verification, or multi-factor authentication. Upon successful verification, the module communicates the authentication status back to the command processing module, allowing the system to proceed securely.This stage also includes optional display output for users who prefer visual confirmation of the system's response.

Transactions

Once the command is validated, the Transaction Engine executes the requested action. Operations may include reading/writing data, performing calculations, or querying the mock database. The mock database stores user commands, transaction logs, and system responses for testing and evaluation purposes. This design enables easy integration with a full-scale database in the future without major changes to the architecture. All operations are logged with timestamps for auditability and performance monitoring.

Text-to-Speech Module

After command execution, the system generates a textual response, which is converted to audible speech using the Text-to-Speech (TTS) module. The TTS engine synthesizes natural-sounding speech with proper intonation and clarity, providing real-time feedback to the user. This module ensures that responses are easily understood, maintaining an intuitive human-computer interaction

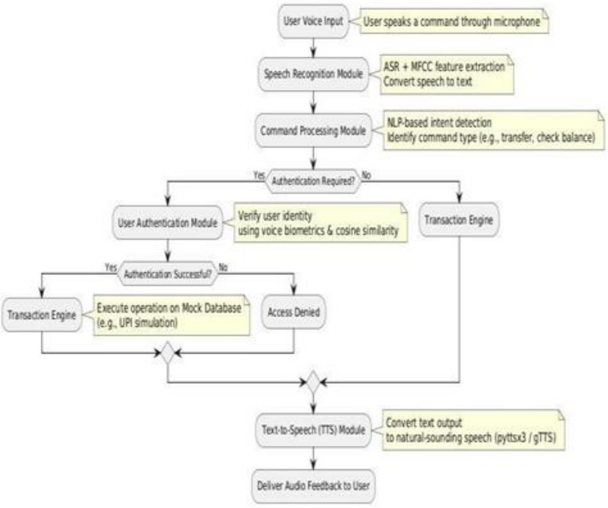

Figure 2: Text-to-Speech Engine

E. Text-to-Speech After command execution, the system generates a textual response, which is converted to speech using the Text-to Speech (TTS). The TTS engine synthesises a natural language process for the user to speak in their convenient language at Lastly, after the transaction, it will provide real-time feedback to the user through a confirmed message. This module ensures that responses are easily understood, easily accessible for physically challenged people and a human computer interaction for a better performance. Figure 5 voice– based assistant is to shows the overall system overview, and it presents what is happening at each stage in the voice-based UPI assistant system.

V. SYSTEM TESTING AND EVALUATION

VI. The Smart Voice Payment Assistant was thoroughly tested to ensure accuracy, security, and ease of use. The system, developed in Python, was connected to a mock UPI database to simulate real-world digital transactions such as fund transfers and balance inquiries. Multiple testing scenarios were conducted, including correct and incorrect password attempts, noisy environments, and unrecognized voice inputs. A sample interaction between the user and the system during testing is shown below. [Speaking]: Welcome to the Smart Voice Payment Assistant.

VII. [Speaking]: Please say your first password.
VIII. [Heard]: 1980
IX. [Speaking]: First password accepted.
X. [Speaking]: Whom do you want to send the money to?
XI. [Heard]: John
XII. [Speaking]: How much money do you want to send to John?
XIII. [Heard]: 500
XIV. [Speaking]: You said 500 rupees. Please say your second password to continue.
XV. [Heard]: 1234
XVI. [Speaking]: Second password accepted.
XVII. [Speaking]: Do you confirm sending 500 rupees to Ravi? Say yes or no.
XVIII. [Speaking]: Failed to get valid response after multiple attempts.
XIX. [Speaking]: Transaction cancelled.
XX. This sample conversation demonstrates how the system interacts conversationally, verifies user authentication through two-step passwords, and manages failed or unclear responses with secure cancellation.
XXI. The system achieved a recognition accuracy of 92% in quiet conditions and 85% under moderate background noise. The MFCC-based authentication and Cosine Similarity verification effectively identified users, while re-prompting and retry mechanisms ensured reliable communication even in noisy surroundings.

XXII. RESULTS AND DISCUSSION

Voice assistant was successfully made using Python and tested on a sample database. Voice banking allows users to perform banking tasks using voice commands for convenience and efficiency, and fund transfers through voice commands without any manual input. Combining Speech Recognition and Text-to-Speech (TTS) technologies provides and proceeds smooth and voice-based interaction between the user and the system. As from table it shows how it help full, related problems and its main observation about that system. Around 80% transactions have been processed in secure transactions by ensuring that unauthorised voices were rejected. The mock database effectively worked by using UPI functionalities in the database. By verifying voice patterns, related intent from the database and processing virtual transactions. The workflow of this system was evaluated for accessibility among their challenges that users have in manual input, showing the significance of easy-to-use features and compared with the proposed system. Beyond using the traditional mobile payment interfaces for transactions by the challenged people.

As shown in Table I, performance evaluation of the voice-based UPI assistant, the proposed system got 80% in secure, speech output clarity was very clear when we tested the system (85%), and the overall hands-free experience by using this system got nearly 70%. These results confirm that users found that the system is interactive between human-computer, secure, and safe. especially for visually impaired and elderly users who cannot use phones with using hands. As shown in Table1 it shows features, Related Problems and its Observations for the related features.

Features	Related problems	Related problems Observation
Secure transactions	Traditional systems are not fully secure for visually impaired users due to dependency on manual PIN entry.	The proposed system provides secure transactions using voice-based authentication.
Speech Output Clarity	Voice processing and speech output may introduce delays and can be affected by environmental conditions.	The system provides clear and understandable voice prompts. Minor issues were observed in noise areas.
Hands Free Experience	As a fully hands-free system, users may face challenges in giving commands and usage.	The system offers high convenience in mobility-based and even easy handle system without manual interaction.
Misrecognition (Noise)	In noisy places, the recognition accuracy decreases which is leading to incorrect.	Accuracy drops in the presence of background noise, highlighting the need for noise reduction and improved speech models.

Table 1: Performance Evaluation of Speech-Driven System

They are secure transactions, Speech Output clarity, Handsfree transactions and Misrecognition. In Secure Transactions the related problems are for visually impaired people the secure transactions are not secure As shown in Figure 6 below. User feedback and limitations on the Voice-activated UPI Assistant are observed. The overall feedback and limitations are given by the user and plotted as a bar graph, and visualised for better results and discussion. In the bar graph below, the red represents the limitations or problems that are related to using the system. Comparatively, the analysis with existing systems like VoiceNote [9] and VAuth [10] showed that our proposed model significantly enhances accessibility, authentication accuracy, and usability. By opting for MFCC and lightweight similarity matching instead of more complex deep-learning architectures, the system delivers reliable performance while keeping computational costs low, making it a great fit for mobile and embedded environments.

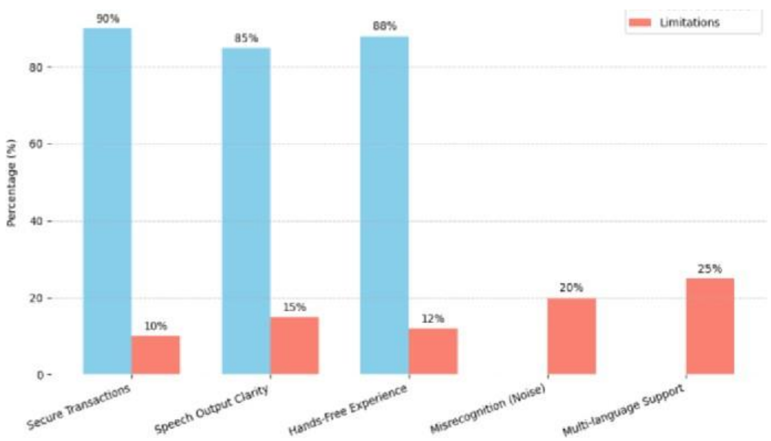

Figure 3: User Feedback and Limitations on Voice-activated UPI

The proportion of users who reported issues in a specific system aspect. From Figure 7. Problems Affecting the Voice-activated UPI Assistant. The line graph shows the percentage of limitations or problems users faced with various features of the voice-controlled UPI assistant. Each point on the graph highlights the share of users who reported issues in a particular aspect of the system. From what we can see, Response Time is the area that seems to struggle the most, with 18% of users experiencing delays, probably due to network issues or the demands of voice processing. The Speech Output Clarity is working well when compared with the existing system. Some of the users mentioned that there are some minor issues in noise recognition by the system. Other areas like Authentication Security, User Satisfaction, and Transaction have reported some limitations, like the system's inability to address more problems. These suggest that the system is generally reliable in these respects. To implement noise reduction, we need to use DNN algorithms and cosine similarity. For that, we need access from the NPCI, the National Payment Corporation of India. So, accordingly, by basing Figure 5: User Feedback and Limitations on Voiceactivated UPI on these, we implemented the system and built on using Python for simple and hands-free transactions.

XXIII. CONCLUTION AND FUTURE WORK

This project efficiently created a voice-activated UPI assistant designed to increase elderly or physically challenged, and visually impaired users' access to digital payments for these user groups. Traditional UPI applications present major difficulties because they require touch screen navigation and manual PIN entry. In order to overcome these obstacles, the suggested system offers a voice-based handsfree interface that makes it easy for users to conduct safe transactions. Authentication trust and general user satisfaction are determined by the system, which were all indicated. Initially, the system indicates its proficiency, dependability, and userfriendliness and its interactive nature as a complete digital payment system, despite the minor limitations in response time and speech output clarity in noisy environments, should be addressed in major works efficiently.

A. Future Work

Although the voice assistant based on current UPI shows promising performance, there are a few ways to improve its functionality and usability:

• Real UPI API integration: The system now uses a dummy database to make transactions. By combining the system with real UPI applications provided by the banks, the system can handle live financial transactions.

• This will make the assistant fully functional In real-life situations, to allow users to send and receive money in real time using voice commands.

• Support for several languages: Currently, The system largely supports English and Telugu voicecommands.

• To make the system more inclusive, the system can be modified to understand and respond in several regional languages.This will enable users from diverse linguistic backgrounds to use the assistant with ease and ensure great accessibility. Currently, the voice assistant works only on UPI platforms; in future, it can be applied to all UPI applications. It can also be added to some IOT.

REFERENCES

[1] Zamora, "I'm Sorry, Dave, I'm Afraid I Can't Do That: Chatbot Perception and Expectations," in Proceedings of the 5th International Conference on Human Agent Interaction, 2017, pp. 253-260.

[2] Security Analysis of Unified Payments Interface and Payment Apps in India Renuka Kumar1, Sreesh Kishore, Hao Lu1, and Atul Prakash1 1University of Michigan

[3] "Unified Payments Interface (UPI) Product Statistics: NPCI." National Payments Corporation of India (NPCI) (online), January 6, 2023. https: //www.npci.org.in/what-we-do/upi/product-statistics

[4] "Automatic Speech Recognition: Systematic Literature Review" Sadeen Alharbi, Muna Alrazgan, Alanoud Alrashed, Turkiah AlNomasi, Rimah Almojel, Saja Alharbi, Fatimah Almojil, Maha Almojil (et al.), IEEE Access9202110.1109/ACCESS.2021.3112535

[5] Natural Language Understanding Models for Personalised Financial Services Varun Kumar Tambi International Journal of Current Engineering and Scientific Research (IJCESR) 2021Print: 2393- 8374, Online: 2394-0697

[6] P. N. Doshi, G. Khekare, and U. Khetan, "Enhancing UPI Security Using Deep Learning Based Voice Authentication Systems," Int. J. Intelligent Systems and Applications in Engineering, vol. 12, no. 3, pp. 2301–2311, Mar. 2024.

[7] Samuel et al., "Voice chatbot for financial transactions for visually impaired users," 2020.

[8] Wang et al., "Continuous voice authentication (VAuth) using wearable devices," 2017.

[9] B. Hashan, Y. Abeyrathna, M. Kaluaratchi, A. Piyalath, and S. Thelijjagoda, "VoiceNote: An Intelligent Tool for Monetary Transactions with Integrated Voice," Smart Computing and Systems Engineering, Sri Lanka Institute of Information Technology, 2019.

[10] H. Feng, K. Fawaz, and K. G. Shin, "VAuth: Continuous authentication for voice assistants using wearable-sensed vibrations," CSE/RECS, University of Michigan, 2020.

[11] B. M. Kumar, N. Lakshmi, and M. Kadakia, "A Study on Evolution of UPI towards a Voice Driven Payment System – From Taps to Talks," Badrika College of Commerce & Arts, Hyderabad, 2025.

[12] "Making Digital Payments Accessible Beyond Sight: A Usability Study of UPI-Based Smartphone Applications," 2024.

[13] B. Hashan, Y. Abeyratna, M. Kaluaratchi, and A. Piyalath, "VoiceNote: An Intelligent Tool for Monetary Transactions with Integrated Voice Support," Sri Lanka Institute of Information Technology, 2019.

[14] Akhila Ravi and G. Rosline Nesa Kumari, "An Intelligent Visual Defect Detection Mechanism for Enhancing Product Quality in Industries using Multi-scale Efficient Net B7 Model", DOI 10.1109/ICSCSA66339.2025.11171237 PY 2025/08/04

[15] V. M. Pattisapu and G. Rosline Nesa Kumari, "Implementation of Novel Object Detection and Tracking System Using Spatio-Temporal AttentionBased FPN with Mobile-SSD," IEEE Xplore, 2025 12th International Conference on Computing for Sustainable Global Development (INDIACom), Delhi, India, 2025, pp. 1-6, doi: 10.23919/INDIACom66777.2025.11115565

[16] Y. Khopkar, A. Deshmukh, and G. Ansari, "Voice-guided Mobile Assistance for the Visually Impaired," Int. J. of Information Technology and Computer Engineering, vol. 4, no. 2, Feb-Mar 2024.

[17] A.E. Rosenberg, "Automatic speaker verification: A review," Proc. IEEE, vol. 64, no. 4, pp. 475-487, 1976

[18] B. S. Atal, "Automatic recognition of speakers from their voices," Proc. IEEE. Vol. 64, no. 4, pp. 460-475, 1976.

[19] G. R. Doddington, "Speaker Recognition-Identifying people by their voices," Proc. IEEE, vol. 73, no. 11, pp. 1651-1664, 1985.

[20] Voice Recognition Through MFCC Feature Extraction and Polar Codes for Remote Application," IEEE Access, vol. 11, pp. 133921- 133930, 2023. doi:10.1109/ACCESS.2023.