International Journal of Convergent Technologies and Future Systems (IJCTFS)
Website: https://www.maulikavyanshufoundation.org/IJCTFS/index.php
Email: maulikavyanshuijctfs@gmail.com | kavyanshufoundation@gmail.com
Research Today, Future Tomorrow Volume 01 Issue-02 April-May 2026 ISSN NO: 3108-1312

Automated Banking Transaction System via Voice Recognition

Akshay Aaba Sonawane
Sinhgad Institute of Computer Application And Management (student)

Dr.Poonam Sawant
Sinhgad Institute of Computer Application And Management (Professor)

Abstract

In today's digital landscape, many individuals struggle with understanding banking terminology and face challenges related to typing. Voice recognition technology has emerged as a practical solution, allowing users to conduct banking transactions seamlessly without relying on traditional keyboards. This study examines the implementation of an automated banking transaction system powered by voice recognition, aiming to enhance accessibility, accuracy, and efficiency in financial services. (Negi et al. (2009))

The integration of voice recognition not only simplifies banking operations but also provides users with a secure and personalized experience. Advanced deep learning models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), play a crucial role in speech recognition, enabling systems to accurately process and interpret user commands. Additionally, natural language processing (NLP) enhances comprehension by understanding contextual meanings, thereby reducing transaction errors.

Security remains a key concern in financial applications, and voice biometric authentication offers a reliable method for verifying users. By analyzing distinct voice patterns, banks can strengthen authentication measures, minimizing the risk of fraud. Furthermore, multi-factor authentication—combining voice recognition with additional security layers such as one-time passwords (OTPs) and facial recognition—adds an extra layer of protection against unauthorized. access .

This paper also explores the potential integration of voice-enabled banking with AI-driven chatbots and virtual assistants, making financial services more accessible to visually impaired individuals and elderly users. Looking ahead, the future of voice banking will rely on continuous advancements to enhance robustness, security, and scalability in real world applications.

Introduction

Speech recognition enables computer systems to identify and convert spoken words into readable text or executable commands. By integrating voice recognition technology, banking applications can provide users with a seamless and hands-free transactional experience. Automatic Speech Recognition (ASR) software, powered by deep learning, allows financial institutions to develop systems capable of accurately understanding and executing banking commands.

This study explores the methodologies, implementation, and benefits of such systems while also addressing key security concerns. The integration of ASR with AI-driven natural language processing (NLP) enhances the precision of voice commands, minimizing errors in financial transactions. Additionally, voice biometrics provide a secure authentication method by analyzing unique speech patterns, reducing the risk of fraud and unauthorized access.

International Journal of Convergent Technologies and Future Systems (IJCTFS)
Website: https://www.maulikavyanshufoundation.org/IJCTFS/index.php
Email: maulikavyanshuijctfs@gmail.com | kavyanshufoundation@gmail.com
Research Today, Future Tomorrow Volume 01 Issue-02 April-May 2026 ISSN NO: 3108-1312

As digital banking continues to expand, voice-enabled transactions improve accessibility, particularly for individuals with disabilities and elderly users who may find traditional interfaces challenging. Real-time speech-to-text conversion, combined with multi-factor authentication, maintains a balance between user convenience and security. Furthermore, cloud-based ASR systems offer scalability, enabling banks to efficiently process large volumes of voice data.

Looking ahead, advancements in voice banking will require robust encryption measures and AI-driven fraud detection mechanisms to strengthen security and enhance user trust.

Need of Project

With the expansion of digital banking, the need for greater accessibility and automation has increased significantly. Many individuals, particularly those who are visually impaired or have difficulty typing, encounter challenges when conducting digital transactions. A voice-enabled banking system can help address these issues by providing a more inclusive and user-friendly experience.

The primary objectives of this project include:

- **Enhancing banking accessibility for individuals with disabilities**: Many users, including the visually impaired and elderly, struggle with traditional banking interfaces. A voice-enabled system removes the need for typing, making banking services more accessible and convenient.
- **Minimizing human errors in transactions**: Mistakes such as incorrect account numbers or transaction amounts can lead to financial losses or failed transactions. Voice recognition technology accurately interprets spoken commands, reducing the likelihood of such errors and improving transaction reliability.
- **Strengthening security with biometric voice authentication**: Traditional authentication methods, such as passwords and PINs, are vulnerable to hacking and identity theft. Biometric voice authentication analyzes unique vocal characteristics, making unauthorized access significantly more difficult and enhancing banking security.
- **Improving efficiency and reducing transaction time**: Traditional banking transactions often require multiple steps, such as logging in, navigating menus, and entering information. A voice-based system streamlines this process, allowing users to complete transactions quickly and with minimal effort, ultimately improving overall banking efficiency.

Objectives

The development of an automated banking transaction system powered by voice recognition aims to achieve several key objectives:

- Integration of Voice Recognition Technology: Enhancing user experience by incorporating voice-based interactions in banking applications.
- Improving Transaction Accuracy and Security: Minimizing errors and preventing unauthorized access to enhance security.
- Enhancing User Convenience: Enabling hands-free banking for greater ease of use.
- Utilizing Deep Learning: Improving speech recognition accuracy through advanced machine learning techniques.

International Journal of Convergent Technologies and Future Systems (IJCTFS)
Website: https://www.maulikavyanshufoundation.org/IJCTFS/index.php
Email: maulikavyanshuijctfs@gmail.com | kavyanshufoundation@gmail.com
Research Today, Future Tomorrow Volume 01 Issue-02 April-May 2026 ISSN NO: 3108-1312

- Implementing Real-Time Speech-to-Text Processing: Ensuring seamless and efficient banking transactions.

Abbreviations and Acronyms

To facilitate understanding, the following abbreviations and acronyms are commonly used in voice recognition and deep learning:
ASR: Automatic Speech Recognition
CNN: Convolutional Neural Network
NLP: Natural Language Processing
AI: Artificial Intelligence
API: Application Programming Interface
HMM: Hidden Markov Model
LSTM: Long Short-Term Memory (a deep learning model)

Methodology

A speech processing system primarily involves three main tasks:

- Speech Recognition – Capturing spoken words, phrases, and sentences.
- Natural Language Processing (NLP) – Understanding the meaning of spoken language.
- Speech Synthesis – Enabling the system to generate spoken responses.
- An ASR system focuses on two critical tasks: phoneme recognition and whole-word decoding.

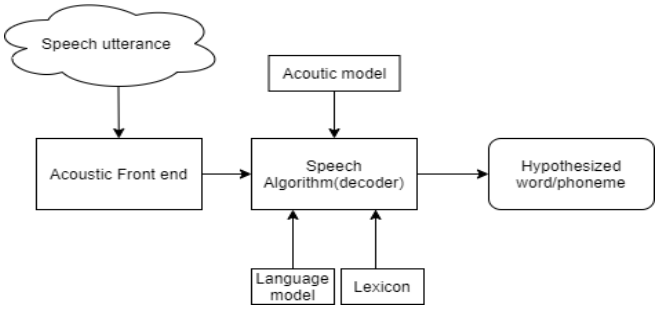

Fig 01:- Voice Recognition Architecture

This process consists of two key steps: **Feature Extraction :** Extracting useful features from speech signals based on prior knowledge. This phase, known as information selection or dimensional reduction, reduces the complexity of speech signals while preserving essential details. Traditional ASR systems commonly use Mel-Frequency Cepstral Coefficients (MFCC) for feature extraction.

Phoneme Prediction and Word Recognition : Discriminative models estimate the probability of each phoneme, followed by a word sequence recognition step using advanced algorithms.

> Deep learning models enhance this process by directly mapping acoustic features to spoken phonemes, facilitating seamless phoneme sequence generation. The three main end-to-end architectures for ASR include: (Srivastava et al. (2017))

Attention-Based Models
Connectionist Temporal Classification (CTC)
CNN-Based Direct Raw Speech Models

Deep Learning Model

Convolutional Neural Network
CNN-based speech recognition models consist of two primary stages:

- Feature Learning Stage – Multiple convolutional layers extract key features from speech signals.
- Classifier Stage – Fully connected layers, including a softmax layer, classify these features using a cost function that minimizes relative entropy.

In this approach, filters in the first convolutional layer extract relevant information, which is then further processed in subsequent layers. The classifier stage categorizes the learned features, achieving performance that is comparable to or even superior to traditional cepstral feature-based systems trained with artificial neural networks (ANNs).

International Journal of Convergent Technologies and Future Systems (IJCTFS)
Website: https://www.maulikavyanshufoundation.org/IJCTFS/index.php
Email: maulikavyanshuijctfs@gmail.com | kavyanshufoundation@gmail.com
Research Today, Future Tomorrow Volume 01 Issue-02 April-May 2026 ISSN NO: 3108-1312

Implementation

Traditional speech recognition models relied on classification algorithms to determine the distribution of phonemes within a speech frame. However, modern speech recognition software integrates Natural Language Processing (NLP) and deep learning neural networks, breaking down speech into interpretable components. These components are then converted into digital format and

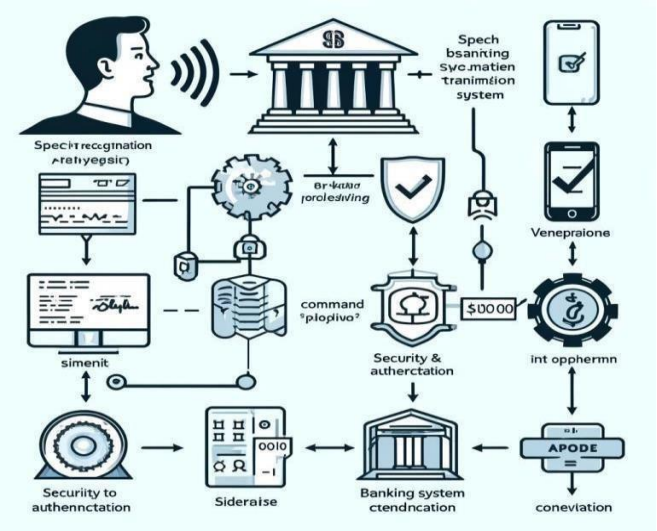

Fig 2

analyzed for segmentation and transcription.

The training process for speech recognition software involves large datasets of known spoken words or phrases, enabling the system to make accurate predictions and convert speech into text. However, independent developers and entrepreneurs often face challenges in building speech detection models due to the limited availability of accessible datasets.

To address this issue, TensorFlow's Speech Commands Dataset provides a large collection of 65,000 one-second utterances of 30 short words, spoken by thousands of individuals. This dataset serves as a valuable resource for training deep learning models in speech recognition.

By leveraging advanced deep learning techniques and high-quality datasets, voice-enabled banking systems can offer a secure, efficient, and user-friendly alternative to traditional banking methods. This innovation makes digital transactions more accessible, reliable, and inclusive, catering to a broader range of users, including those with disabilities

Speech-to-Text Processing

The implementation of speech-to-text recognition relies on **TensorFlow** and **NumPy**, utilizing Python to develop an efficient and accurate system.

TensorFlow plays a crucial role in building neural network models capable of recognizing spoken words. It supports various **Recurrent Neural Network (RNN)** architectures, including:

Static RNN – Processes fixed-length sequences.

Dynamic RNN – Handles inputs of varying lengths.

Static Bidirectional RNN – Considers both past and future contexts, improving accuracy.

NumPy, a widely used Python library for numerical computing, provides **multidimensional array objects** and functions for processing these arrays. It enables mathematical and logical operations essential for efficiently handling speech data in recognition models.

Data Analytics

A combination of **Python** and **Google Cloud** facilitates comprehensive data analysis, supporting the development of an **automated banking transaction system using voice recognition**. (Naik (2016))

Business and Commercial Applications – Python's capabilities extend beyond speech recognition to various domains, including business, research, education, and training.

International Journal of Convergent Technologies and Future Systems (IJCTFS)
Website: https://www.maulikavyanshufoundation.org/IJCTFS/index.php
Email: maulikavyanshuijctfs@gmail.com | kavyanshufoundation@gmail.com
Research Today, Future Tomorrow Volume 01 Issue-02 April-May 2026 ISSN NO: 3108-1312

Machine Learning Pipeline – Python supports all stages of the machine learning workflow, including:

Data Preparation – Cleaning and preprocessing voice data to improve model performance.

Model Training and Validation – Enhancing accuracy and reducing error rates through iterative learning.

Results Visualization – Analyzing model performance and optimizing transactions for better efficiency.

Automated Banking Transactions

Once voice data is processed, it is converted into **structured commands**, enabling users to perform banking tasks securely and efficiently through **speech-based interactions**.

By integrating **speech recognition with**

Fig. 3. Voice Input via Application

banking applications, this system aims to deliver a **seamless, secure, and accessible banking experience**. This innovation is particularly beneficial for individuals with disabilities and those who face challenges using traditional digital interfaces, making banking more **inclusive and user-friendly**.

OPTIMIZATION

To enhance the efficiency of an automated banking transaction system powered by voice recognition, several optimizations have been implemented. The preprocessing of datasets has been streamlined using Pandas, NumPy, and Seaborn, ensuring efficient management of missing values and duplicate entries. Data visualization techniques, such as histograms and counter plots, provide insights into customer demographics, including gender and location, aiding in more effective user analysis. For NLP-based transaction processing, deep learning models like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have been optimized to improve speech recognition speed and keyword extraction accuracy. The system integrates real-time speech-to-text conversion, reducing latency and enhancing response times for users. Security measures have been strengthened by incorporating voice biometric authentication alongside multi-factor authentication (MFA), utilizing methods such as one-time passwords (OTPs) and facial recognition to enhance protection. To ensure faster query execution, the database structure has been optimized using indexing and caching techniques, minimizing retrieval times and improving overall system performance. Additionally, cloud-based architectures like Google Cloud and TensorFlow have been adopted to improve scalability. These technologies enable the system to efficiently manage large datasets and simultaneous user requests, ensuring smooth operations even under high traffic conditions. With these optimizations in place, the system offers greater accuracy, enhanced security, and real-time performance, providing users with a seamless and reliable banking experience

CONCLUSION

The development of an automated banking transaction system powered by voice recognition marks a significant advancement in accessibility, security, and efficiency. This technology enables users, including individuals with disabilities, to conduct transactions effortlessly without relying on traditional interfaces. By integrating deep learning models, Natural Language Processing (NLP), and biometric authentication, the system ensures high accuracy in speech recognition while

International Journal of Convergent Technologies and Future Systems (IJCTFS)
Website: https://www.maulikavyanshufoundation.org/IJCTFS/index.php
Email: maulikavyanshuijctfs@gmail.com | kavyanshufoundation@gmail.com
Research Today, Future Tomorrow Volume 01 Issue-02 April-May 2026 ISSN NO: 3108-1312

enhancing fraud prevention measures. The elimination of manual data entry reduces human errors, while real-time speech processing significantly improves transaction speed and responsiveness. Beyond banking, this technology has the potential to be applied across various industries, including medical bookings, cab services, and food delivery, expanding its impact beyond financial services. The study emphasizes the importance of continuous advancements in AI-driven voice processing and cybersecurity to maintain reliability and security in real-world applications. Looking ahead, future improvements could incorporate enhanced speech emotion recognition and multilingual support, further increasing the system's inclusivity and usability. This innovation paves the way for a smarter, hands-free banking experience, ensuring both security and user convenience in an increasingly digital financial landscape.

ACKNOWLEDGMENT

We express our sincere gratitude to **Dr. Poonam Sawant**, our project guide, for providing continuous support and insightful feedback throughout the development of this system. Additionally, we extend our appreciation to the **faculty members of the IT Department** for their valuable guidance in shaping our research. Special thanks to **our colleagues and industry professionals** who contributed with their inputs on AI, voice recognition, and cybersecurity. Lastly, we acknowledge the **developers of TensorFlow, NumPy, and Pandas**, whose open-source contributions made our research possible. Their efforts in AI and machine learning frameworks have significantly aided the implementation of this voice-based banking system, making it a **secure, scalable, and efficient solution for modern banking applications**.

REFERENCES

1. S. Negi, S., Joshi, S., Chalamalla, A. K., & Subramaniam, L. V. (2009). "Automatically Extracting Dialog Models from Conversation Transcripts." *IEEE International Conference on Data Mining*. doi: 10.1109/ICDM.2009.113.

2. Naik, N. (2016). "Connecting Google Cloud System with Organizational Systems for Effortless Data Analysis." *IEEE International Symposium on Systems Engineering.* doi: 10.1109/SysEng.2016.7753150.

3. Srivastava, S., Soman, S., Rai, A., & Srivastava, P. K. (2017). "Deep Learning for Health Informatics: Trends and Future Directions." *International Conference on Advances in Computing, Communications and Informatics. doi:* 10.1109/ICACCI.2017.8126082.

4. Huang, K., Wu, C., Hong, Q., Su, M., & Chen, Y. (2019). "Speech Emotion Recognition Using Deep Neural Network." *ICASSP - IEEE International Conference on Acoustics, Speech, and Signal Processing.* doi: 10.1109/ICASSP.2019.8682283.

5. Uma, M., Sneha, V., Sneha, G., Bhuvana, J., & Bharathi, B. (2019). "Formation of SQL from Natural Language Query Using NLP." *IEEE International Conference on Computational Intelligence in Data Science.* doi: 10.1109/ICCIDS.2019.8862080.

International Journal of Convergent Technologies and Future Systems (IJCTFS)
Website: https://www.maulikavyanshufoundation.org/IJCTFS/index.php
Email: maulikavyanshuijctfs@gmail.com | kavyanshufoundation@gmail.com
Research Today, Future Tomorrow Volume 01 Issue-02 April-May 2026 ISSN NO: 3108-1312