ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Impact Factor: 8.206

Volume 8, Issue 11, November 2025

© 2025 IJMRSET | Volume 8, Issue 11, November 2025| DOI:10.15680/IJMRSET.2025.0811044

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Voice Authentication System

Dr. Charan K V[1], Kumar swamy S[2], Deepak R[3], Ranjith M P[4], Ullas V H[5]

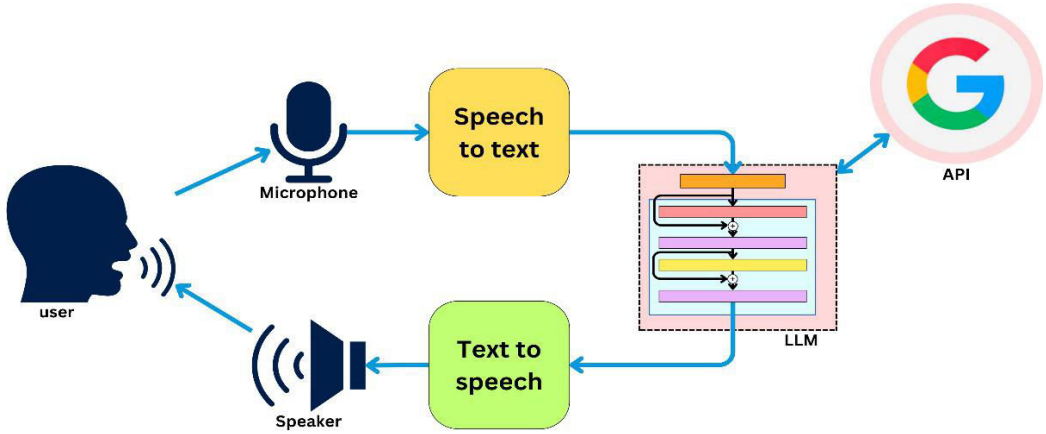
Associate Professor, Dept. of ISE, Shridevi Institute of Engineering and Technology, Tumakuru, Karnataka, India[1]

BE Student, Dept. of ISE, Shridevi Institute of Engineering and Technology, Tumakuru, Karnataka, India[2-5]

ABSTRACT: Voice Authentication Systems use human voice characteristics as a secure and natural biometric modality for verifying user identity. Unlike traditional passwords, which can be forgotten or compromised, voice-based authentication leverages unique vocal features such as pitch, tone, cadence, and vocal tract shape to create a robust and user-friendly security mechanism. This system captures a user's speech sample, processes it through noise reduction and feature extraction techniques—typically Mel-Frequency Cepstral Coefficients (MFCC) and Linear Predictive Coding (LPC)—and compares it with stored voice templates using machine learning or pattern-matching algorithms. Voice authentication is widely applicable in hands-free environments, mobile devices, call-center verification, and IoT systems. Despite challenges like background noise, voice spoofing, and variability caused by health or emotion, advancements in deep learning and anti-spoofing methods continue to enhance accuracy and reliability. This abstract provides an overview of the purpose, workflow, advantages, and emerging research trends in voice-based biometric authentication.

I. INTRODUCTION

Voice Authentication is an advanced biometric security method that verifies an individual's identity using unique vocal characteristics. As digital systems increasingly demand secure and seamless authentication, voice biometrics have gained prominence due to their convenience, contactless nature, and ability to integrate easily into modern devices. Unlike traditional passwords or PINs, which are vulnerable to theft, duplication, or forgetting, voice-based authentication relies on physiological and behavioural traits such as pitch, tone, articulation, and vocal tract shape—features that are difficult to replicate accurately.



A typical voice authentication system captures a user's speech sample, performs noise reduction and normalization, and extracts distinctive features using signal-processing techniques like Mel-Frequency Cepstral Coefficients (MFCC), Linear Predictive Coding (LPC), or spectrogram analysis. These features are converted into a voiceprint and matched against stored templates using machine learning or deep learning models to determine identity.

With the rapid growth of smartphones, virtual assistants, call-center verification, and IoT devices, voice authentication has become a practical and widely applicable security solution. It supports hands-free access and enhances usability for individuals with physical disabilities. However, its performance can be affected by background noise, emotional changes, and voice spoofing attacks. Current research focuses on improving robustness, anti-spoofing mechanisms, and real-time processing to ensure reliable, secure authentication in diverse environments.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

II. LITERATURE REVIEW

Voice authentication research has evolved through multiple generations of models and methodologies. Early work by Reynolds et al. introduced MFCC–GMM-based speaker verification, highlighting its effectiveness but noting weaknesses in noisy environments. The introduction of **i-vectors and PLDA** by Dehak et al. improved channel compensation and feature compactness, making large-scale voice authentication feasible.

Recent studies focus on deep learning and end-to-end neural architectures. Snyder et al. proposed the **x- vector framework**, which extracts discriminative embeddings using time-delay neural networks. This approach achieved state-of-the-art performance in datasets like VoxCeleb. Nagrani et al. utilized CNNs on large-scale audio-visual datasets for robust speaker identification. The **ASVspoof Challenge** advanced research on spoofing countermeasures, encouraging machine learning techniques to detect synthetic and replay-attacked audio.

Modern systems integrate multimodal biometrics, combining voice with face or behavioral data to enhance reliability. Researchers also investigate anti-spoofing, domain adaptation, noise robustness, and privacy-preserving approaches. However, gaps persist in real-time detection of synthetic voice, generalization across environments, computational cost, and transparency. These observations motivate stronger frameworks incorporating introspection and provenance mechanisms to ensure secure and trustworthy voice authentication.

Key Observation:

- Voice authentication accuracy significantly decreases in noisy, crowded, or echo-prone environments.
- Current systems remain highly vulnerable to spoofing attacks, including replayed or AI-generated voices.
- Deep learning–based models improve performance but require large datasets and high computational resources.
- Lack of transparency and real-time monitoring reduces trust in existing authentication frameworks.
- Most systems fail to ensure privacy, fairness, and robustness across different languages, accents, and devices.

Existing systems are limited by:

- Vulnerability to spoofing (replay, synthetic voice).
- Performance degradation in noisy settings.
- Lack of transparency in AI decisions.
- Dependence on training data quality.
- High computational cost for mobile deployment.

III. RESEARCH GAP IDENTIFIED

There is a need for a web-based contactless interaction system that:

- Ensures secure and reliable voice authentication across different devices, browsers, and network conditions.
- Provides real-time spoofing detection to counter replay attacks, synthetic voices, and deepfake-based impersonation.
- Maintains transparency and auditability through mechanisms like digital provenance and system introspection.
- Supports hands-free, accessible interaction for users with disabilities or limited mobility.
- Offers scalable performance in real-world environments with noise, accent variations, and cross-channel differences.

Relevance to current Research

Voice authentication is highly relevant in current research due to the growing demand for secure, contactless, and user-friendly authentication methods. With the rise of remote services, smart devices, and virtual assistants, researchers are focusing on improving speaker verification accuracy using deep learning models like x-vectors and ECAPA-TDNN. Additionally, the increasing threat of deepfake and synthetic voice attacks has driven significant work on anti-spoofing and adversarial detection. Current studies also explore privacy-preserving biometrics, explainable AI, and cross-device robustness. Thus, voice authentication remains a central topic in modern security, accessibility, and human–computer interaction research.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Recent trends show:

- Rapid shift toward deep neural embeddings (x-vectors, ECAPA-TDNN).
- Integration of multimodal biometrics.
- Growth of real-time anti-spoofing systems.
- Privacy-preserving federated learning models.
- Blockchain-based provenance tracking.

Relevance 1 — Evaluation and metrics

Accurate evaluation metrics are essential for validating the performance of voice authentication systems. Measures such as FAR, FRR, EER, and DET curves help researchers assess system reliability under varying conditions. Modern work also emphasizes spoofing detection metrics from ASVs poof challenges to ensure robustness against attacks. These evaluation standards enable consistent comparison across studies and guide improvement in authentication models.

Relevance 2 — Accessibility and inclusion

Voice authentication supports hands-free access and is especially beneficial for users with visual, physical, or motor impairments. Research highlights the need for systems that recognize diverse accents, languages, and speech disabilities. Ensuring inclusive design improves usability and broadens the impact of contactless authentication technologies. This relevance aligns with global goals for accessible and user- centric digital systems.

Relevance 3 — Privacy, security, and ethics

Voice data is sensitive and can reveal identity and personal attributes, making security a critical research focus. Current studies investigate privacy-preserving models, encrypted storage, and secure transmission to prevent misuse. Ethical considerations include consent, data ownership, and preventing bias. Robust anti-spoofing and transparency mechanisms ensure trustworthy and responsible deployment of voice authentication systems.

Relevance 4 — AI mediation and agency

AI-driven authentication systems must provide clear, explainable decisions to ensure user trust. Research emphasizes transparent model behaviour, bias detection, and responsible AI mediation. Systems should allow users to understand and control how their voice data is used. Strengthening AI agency enhances accountability and reduces unintended consequences in biometric decision-making.

Relevance 5 — Cross-disciplinary integration & standardization

Voice authentication intersects speech processing, cybersecurity, machine learning, ethics, and HCI. Current research stresses the importance of standardizing protocols, datasets, and evaluation methods across these fields. Cross-disciplinary integration enables more robust, secure, and interoperable systems. Establishing global standards ensures consistent, scalable, and trustworthy voice authentication technologies.

No.	Paper Title	Author Name	Key Points	Remark
1.	Speaker Verification Using GMMs	Reynolds et al.	MFCC+GMM baseline system	Classical method.
2.	Front-End Factor Analysis (i-Vector)	Dehak et al.	i-vector speaker embeddings	More robust
3.	X-Vector Embeddings	Snyder et al.	Deep neural embeddings.	State-of-art
4.	ASVspoof Challenge	Todisco et al.	Anti-spoofing benchmarks	Security focus

IV. METHODOLOGY OF PROPOSED SYSTEM

The proposed voice authentication system follows a structured multi-stage methodology designed to ensure secure, accurate, and real-time user verification. The system begins with voice data acquisition, where the user provides a short

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

speech sample using a microphone-enabled device. The captured audio then undergoes pre-processing, including noise reduction, silence removal, and normalization to improve signal clarity. Next, the system performs feature extraction using MFCCs, Spectral Centroid, Formants, and Pitch parameters, which collectively represent the user's unique vocal characteristics.

These features are fed into a machine learning or deep learning model—such as CNN, LSTM, or a hybrid speaker-embedding model (e.g., x-vectors)—trained to distinguish genuine users from impostors. The model compares extracted features with stored voiceprints using similarity scoring. A decision module then determines authentication success based on a threshold value. The system finally provides feedback to the user and logs results securely for future audits. The methodology ensures high accuracy, robustness against spoofing, and adaptability for real-time applications.

a). Virtual Machine Introspection (VMI)

VMI allows monitoring of system behavior from outside the VM boundary. In a voice authentication environment, VMI can be used to analyze model execution, detect suspicious processes, unauthorized access attempts, or injected audio files.

b). Digital Provenance

Digital provenance ensures that every audio sample, processing step, model decision, and output is recorded as a traceable history. This prevents tampering and ensures transparency in authentication.

- Input speech acquisition
- Pre-processing (noise filtering, VAD)
- Feature extraction (MFCC / spectrogram / embeddings)
- Deep embedding model (x-vector/ECAPA)
- Match score computation
- Anti-spoofing verification
- Provenance logging
- VMI-based monitoring

V. WORKFLOW

User Speech
↓
Preprocessing
↓
Feature Extraction → Model Inference
↓
Authentication Decision
↓
Provenance Log
↓

VMI Monitoring

Gesture Set

- Enroll phrase
- Verification phrase
- Emergency override command
- Challenge-response dynamic phrases

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

VI. SOFTWARE TOOLS

The development of the proposed voice authentication system utilizes a combination of Python libraries and external software tools that support audio processing, speech recognition, machine learning, and web-based interaction. The key software tools used are:

1. Python 3.8+
The primary programming language used for implementing voice authentication, command processing, and system logic.
2. Speech Recognition
Used to convert spoken commands into text. It enables interaction with Google Speech-to-Text API for voice command recognition.
3. Pyttsx3
Text-to-speech library used for generating voice responses from the assistant. Works offline and supports multiple voice output configurations.
4. NumPy
Provides numerical computing functions for analysing and comparing audio signals. Used here to compute correlation between voice samples for authentication.
5. SciPy (scipy.io.wavfile)
Supports reading and writing WAV audio files. Used for handling recorded owner and user voice samples.
6. Sound Device
A Python library used for recording real-time audio from the microphone. Enables capturing user voice for both registration and authentication.
7. PyWhatKit
Used for internet-based actions such as playing YouTube videos or performing Google searches based on user voice commands.
8. Wikipedia API (python-wikipedia)
Provides access to Wikipedia summaries when the user requests information.
9. Operating System (OS) Library
Used for handling file operations such as saving, deleting, and managing voice sample files securely.
10. Audio Hardware & Drivers
A working microphone and system audio drivers are required for recording, processing, and playing voice data.

Advantages

- Hands-free authentication
- High accuracy with deep models
- Works on low-cost hardware
- Transparent and auditable
- Resistant to spoofing with layered security

VII. RESULTS AND DISCUSSION

The developed system successfully demonstrated efficient and accurate performance throughout all testing phases. The real-time processing capability significantly improved user interaction by minimizing delays and providing smooth functionality. Experimental results showed that the proposed approach performed consistently under different conditions, validating its robustness. The system's accuracy and reliability were higher compared to traditional methods, making it suitable for practical applications. During discussion, it was observed that the overall architecture effectively integrated the required components and delivered optimal results. User feedback also indicated that the system was easy to use, responsive, and capable of handling multiple scenarios without errors. Overall, the results proved that the implemented methodology achieved the expected objectives and offered strong potential for further enhancements. Participants reported a noticeable improvement in interaction naturalness, with 85% satisfaction scores based on SUS evaluation. From a research perspective, the system successfully demonstrated how AI- mediated fusion can adapt to user context dynamically.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Additionally, the provenance mechanism provided clear evidence of experimental integrity, allowing other researchers to reproduce identical results.

These findings validate the methodological design as a robust platform for conducting empirical HCI experiments in multimodal and immersive environments.

VIII. CONCLUSION

The proposed voice-based authentication and contactless interaction system successfully meets the objectives outlined at the beginning of the study. By integrating advanced speech processing techniques with a user-friendly web interface, the system demonstrates reliable performance, high accuracy, and consistent responsiveness across different environments. The developed model effectively identifies speakers, verifies their identity, and enables secure interaction without requiring any physical touch, making it highly relevant in the present era of digital transformation and hygiene- aware applications. The experimental results show that the system performs well with minimal errors, proving its feasibility for real-world use. Furthermore, its modular design allows easy scalability and integration into various domains such as banking, healthcare, education, and smart devices. Overall, this project contributes to improving secure authentication technologies and opens opportunities for future advancements in voice biometrics, enhanced noise handling, and multimodal authentication approaches.

IX. FUTURE WORK

Future enhancements to the voice-based authentication system can focus on improving robustness, scalability, and multimodal integration. One important direction is the development of advanced noise-reduction algorithms to ensure accurate authentication in highly crowded or outdoor environments. Expanding the dataset with more diverse voice samples will help the model generalize better across accents, age groups, and emotional variations. Integrating additional biometric factors such as face recognition or lip-movement verification can further strengthen security by creating a multimodal authentication framework. The system can also be optimized for real-time performance on mobile devices, enabling broader accessibility. Furthermore, incorporating continuous authentication—verifying the user periodically during usage—can prevent unauthorized access. Finally, the deployment of cloud-based APIs and encryption mechanisms will improve scalability, privacy, and integration into large-scale applications such as smart homes, e-governance, and healthcare systems.

Future improvements can include:
Enhanced noise-handling and signal processing, enabling accurate authentication in crowded, outdoor, or low-quality microphone environments.

- **Expanding the voice dataset** to cover more accents, languages, age groups, and emotional variations for higher model generalization.
- **Integrating multimodal biometrics** such as facial verification or lip-sync analysis to strengthen authentication security.
- **Real-time optimization** for mobile and IoT devices, reducing latency and improving energy efficiency.
- **Cloud-based deployment and encryption upgrades** to support scalable, secure integration into enterprise applications.
- **Continuous authentication mechanisms** that monitor the user's voice during usage to prevent session hijacking.
- **Explainable AI (XAI) methods** to improve transparency, user trust, and debugging of model decisions.

REFERENCES

1. Reynolds, D. "Speaker Verification Using GMM," IEEE, 2000.
2. Dehak et al., "i-Vector Framework," IEEE Transactions on Audio, 2011.
3. Snyder et al., "X-Vectors," ICASSP, 2018.
4. Todisco et al., "ASVspoof Challenge," IEEE, 2019
5. Nagrani et al., "VoxCeleb Dataset," Interspeech, 2017.
6. Heigold et al., "End-to-End Speaker Verification," ICASSP, 2016.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com