AJEMA GLOBAL RESEARCH NETWORK

AMERICAN Journal of Engineering, Mechanics and Architecture
Volume 3, Issue 6, 2025 ISSN (E): 2993-2637

A Systematic Review of Voice Encryption Techniques for Secure Audio Transmission

Jahanvi Pragati
P.G. Student, Department of CSE, Sat Kabir Institute of Technology and Management, Haryana, India

Ritu Dagar
Assistant Professor, of CSE, Sat Kabir Institute of Technology and Management, Haryana, India

Abstract: As voice communication via digital networks becomes more and more common, protecting the integrity and secrecy of audio communications has become crucial. In order to protect recorded and live audio data from eavesdropping, manipulation, and unwanted access, this systematic review investigates and assesses a number of voice encryption methods. The study examines the efficacy of encryption techniques in terms of security strength, computing cost, latency, and resistance to cryptographic assaults, classifying them into time-domain, frequency-domain, hybrid, and machine learning-based approaches. Lightweight encryption techniques that work well in resource-constrained settings, including embedded and mobile systems, are given particular consideration. The paper also emphasizes new developments in adaptive and quantum-resistant voice encryption methods. This book serves as a thorough reference for researchers and practitioners interested in creating reliable and effective speech encryption systems by combining recent developments, difficulties, and prospective study directions.

Keywords: Secure Audio Transmission, Real-time Communication.

INTRODUCTION: Voice communication, which includes technologies like VoIP (Voice over Internet Protocol), secured conferencing, and voice assistants, has become an essential part of contemporary telecommunication networks in the digital age. The potential of security breaches, such as eavesdropping, impersonation, and altering voice data while it is being transmitted, is increasing along with the use of IP-based communication services. Strong voice encryption methods are required to protect audio communication systems' integrity, confidentiality, and authenticity in light of these risks.

To prevent unwanted access to data while it is being transmitted or stored, voice encryption transforms spoken audio signals into unreadable formats. Voice encryption has distinct difficulties compared to typical data encryption, including real-time processing, large bandwidth needs, and sensitivity to distortions and delays. Therefore, the creation of effective and portable encryption techniques is essential for safe voice-based services, particularly in mobile and Internet of Things contexts.

A number of methods have been put out to deal with these issues. To hide the content of the communication, time-domain and frequency-domain encryption techniques either directly alter the spoken signal or its spectral components [1]. To strike a compromise between security and performance, hybrid approaches combine signal processing techniques with cryptographic

algorithms [2]. In order to further improve the durability of voice encryption techniques, recent developments have included the utilization of chaotic systems [3], machine learning [4], and quantum cryptography [5]. The objective of this paper is to provide a systematic review of existing voice encryption techniques, analyze their strengths and limitations, and identify potential directions for future research. This review categorizes voice encryption methods based on algorithmic complexity, application domains, and suitability for real-time systems, providing insights into the evolving landscape of secure audio transmission.

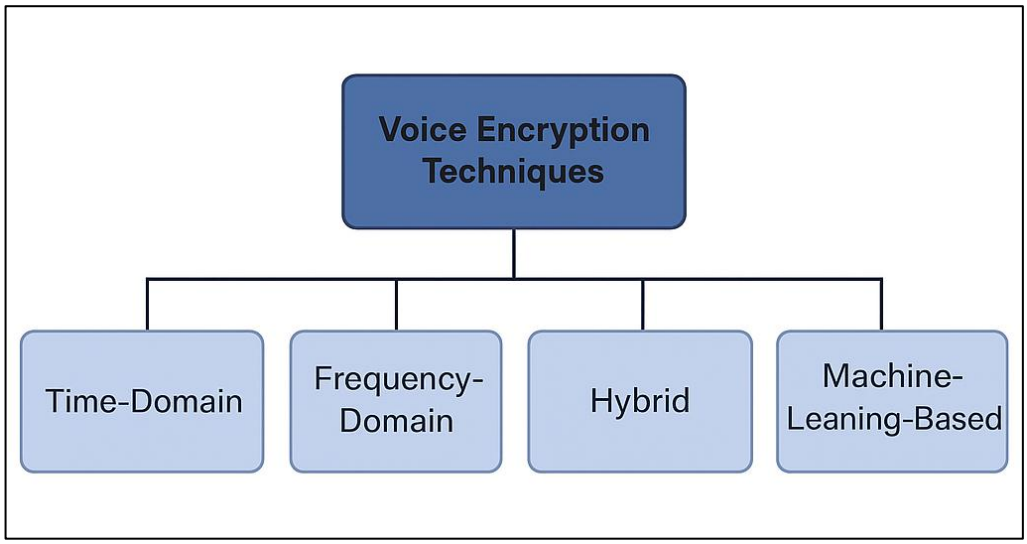

Figure 1: Some Popular Voice Encryption Methods

Research Background

The authors of [6] provide a thorough analysis of speech encryption strategies, emphasizing the different approaches and algorithms created to ensure voice communication security. It is suggested to use multilayer cryptosystems to secure audio conversations [7]. By continually merging the audio signal with a speech signal without silent intervals, these cryptosystems integrate audio signals with other active concealed signals, such speech signals. Preventing other people from listening to encrypted audio conversations is the aim of these cryptosystems. Before they are joined, the speech and audio signals are preprocessed because this is required to prepare the signals for fusion.

The cryptosystems depend on the values of audio samples rather than encoding and decoding techniques, which saves time and makes them more resilient to hackers and noisy surroundings. The primary characteristic of the suggested method is its consideration for all three encryption levels: permutation, substitution, and fusion, where different combinations are taken into account. The resultant cryptosystems are contrasted with other cutting-edge approaches and one-dimensional logistic map-based encryption algorithms. The signal-to-noise ratio (SNR), structural similarity index, histogram, and other metrics are used to assess the performance of the proposed cryptosystems.

Sharma et al. [8] created a technique that uses the RSA algorithm to encrypt audio recordings. A number of methods, including those described in [9], used various shuffling strategies to encrypt text files and images. As explained in [10], speech files were encrypted using the RSA technique, with each word being retrieved and converted to text.

TIME-DOMAIN VOICE ENCRYPTION TECHNIQUES:

Direct manipulation of the raw speech signal in its original temporal form is a component of time-domain voice encryption schemes. Usually, these techniques work with audio samples that haven't been converted to the frequency domain. Their simplicity, reduced computing complexity, and adaptability for real-time applications with little delay are some of their advantages.

a. Bit-level Manipulation:

This involves altering individual bits of the audio signal's digital representation, such as through XOR operations with a pseudo-random sequence (stream cipher), or applying permutation strategies [11].

b. Sample Shuffling (Permutation):

A key-dependent permutation function is used to alter the audio sample order, rendering the signal incomprehensible in the absence of the key. It is suitable for low-bandwidth and embedded systems [12].

c. Amplitude Masking:

A secret signal (mask) is added to or modulates the amplitude of the voice signal. The receiver subtracts the mask using a shared key [13].

Table1: Comparison of Popular Time Domain Voice Encryption Methods

Technique	Domain	Encryption Method	Key Features	Advantages	Limitations
Stream Cipher with PRNG [11]	Time-Domain	XOR encryption using pseudo-random number generator	Real-time stream cipher; keystream generated using PRNG	Fast and suitable for real-time communication	Security depends on PRNG quality; susceptible to attacks if PRNG is weak
Sample Rearrangement + Amplitude Transformation [12]	Time-Domain	Permutation of samples and amplitude alteration	Dual-layer security using sample reordering and value transformation	Lightweight; increases resistance to statistical analysis	May introduce audio distortion; sensitive to timing mismatches
Amplitude Masking [13]	Time-Domain	Adding/subtracting a secret signal (mask)	Masks voice signal by embedding it within another amplitude-based signal	Simple, low-latency encryption method	Vulnerable to noise and compression; not robust against attacks

FREQUENCY-DOMAIN VOICE ENCRYPTION TECHNIQUES:

In order for frequency-domain voice encryption techniques to function, the speech signal must first be broken down into its frequency components, usually using the Fourier Transform or other spectrum methods. These components must then undergo encryption processes. These techniques take advantage of the fact that speech is a useful domain for safe modification because a large portion of its information and intelligibility is contained in particular frequency ranges.

a. Fast Fourier Transform (FFT)-Based Encryption:

This technique converts time-domain signals into frequency spectra using FFT, encrypts the spectral coefficients, and then reconstructs the time-domain signal via inverse FFT (IFFT) [14].

b. Discrete Cosine Transform (DCT)-Based Encryption:

DCT is used to concentrate signal energy into fewer coefficients. Selected coefficients are encrypted using symmetric or asymmetric cryptography [15].

c. Wavelet Transform Encryption:

Wavelet Transform provides multi-resolution analysis. Different frequency bands (low-pass, high-pass) can be encrypted selectively based on their perceptual significance. It has better time-frequency localization compared to FFT [16].

d. Cepstrum-Based Techniques:

The cepstral domain (logarithm of frequency spectrum) allows robust voice feature representation. Encryption of cepstral coefficients prevents speaker recognition or speech content decoding [17].

In these methods altering spectral components affects intelligibility significantly. It allows partial encryption (e.g., only high-energy frequencies), saving computation. It often integrates well with audio codecs like MP3 or AAC.

Table 2: Comparison of Popular Frequency Domain Voice Encryption Methods:

Technique	Domain	Encryption Method	Key Features	Advantages	Limitations
FFT with Chaos-based Key Generation [14]	Frequency (FFT)	Spectral encryption using chaotic maps	Uses Fast Fourier Transform and pseudo-chaotic sequences for encryption	Simple FFT structure; enhances unpredictability with chaos	Sensitive to synchronization errors; low time localization
DCT-based Speech Encryption [15]	Frequency (DCT)	Coefficient encryption using symmetric cipher	Compresses signal using DCT, encrypts significant coefficients	Efficient storage; compression-friendly	Less precise for time-variant signals
Wavelet Transform + RSA Encryption [16]	Frequency (Wavelet)	Multilevel decomposition + RSA encryption	Performs multi-resolution analysis and applies RSA on selected bands	Robust to noise; strong cryptographic foundation	RSA increases computational complexity; harder to implement
Cepstral Domain Scrambling [17]	Cepstral Domain	Coefficient scrambling for identity masking	Applies scrambling to MFCC-like features to anonymize speaker identity	Protects speaker info without affecting intelligibility	Does not fully encrypt speech content; may not ensure privacy

Machine Learning (ML) Based Voice Encryption Methods:

Voice encryption is one of the many cybersecurity domains where machine learning (ML) has emerged as a potent weapon. It is employed not just to protect audio but also to modify encryption tactics according to threat levels, signal characteristics, or context. In contrast to conventional cryptographic techniques, machine learning (ML)-based encryption has the ability to learn patterns, modify keys dynamically, and even optimize encoding to improve efficiency and security. Machine learning in voice encryption can be applied in multiple roles: (i) Key Generation: Generating strong, non-repetitive encryption keys using deep neural networks or generative models. (ii) Feature-Based Encryption: Learning sensitive speech features (e.g., phonemes, MFCCs) and selectively encrypting or masking them. (iii) Adaptive Encryption Schemes: Dynamically changing the encryption technique based on signal complexity or external factors. (iv) Adversarial Obfuscation: Training models to generate audio that fools speech recognizers or biometric systems (speaker obfuscation).

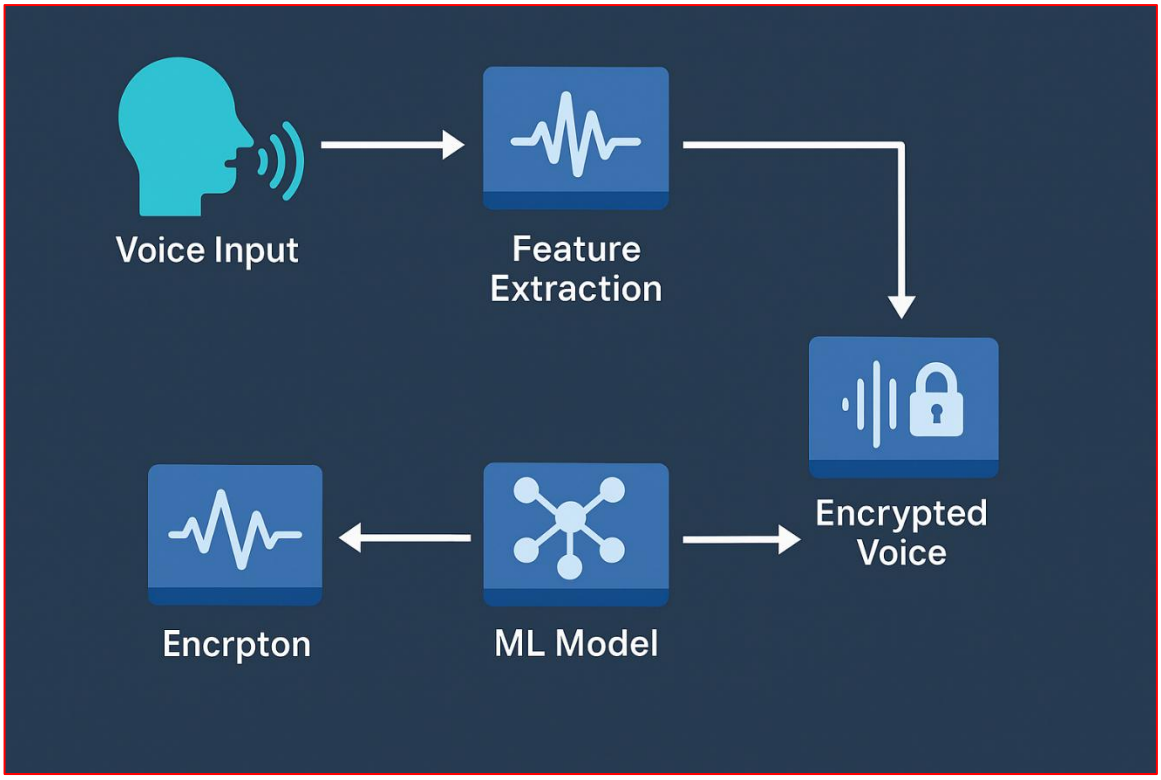

Figure 2: Machine learning Process of Voice Ecryption

a. Neural Key Generation

Neural networks, especially LSTM or GAN-based models, can generate pseudo-random keys that are context-aware and hard to predict. A deep neural network trained on speech patterns generates dynamic keys used to encrypt speech segments [18].

b. Autoencoder-Based Encryption

Autoencoders can be trained to compress and encrypt audio simultaneously. The encoder acts as the encryption mechanism, and the decoder at the receiver's end decrypts it using a learned model. It combines compression and encryption. It requires the model to be synchronized at both ends [19].

c. Feature-Space Masking and Obfuscation

ML models identify key speech features (e.g., pitch, tone, identity markers) and mask or transform them using adversarial techniques to prevent recognition by ASR (automatic speech recognition) or speaker ID systems [20].

d. Reinforcement Learning (RL)-Based Encryption Policy

RL agents can learn encryption strategies that balance security, latency, and quality by observing rewards based on real-time constraints (e.g., channel conditions, energy use).

Table 3: Comparative analysis of the above said ML-based voice encryption methods:

Technique Used	Core Idea	Encryption Strategy	Advantages	Limitations
Deep Neural Networks (DNNs) [18]	Employs DNNs to learn and encode speech features for encryption	DNN encodes raw speech into high-dimensional encrypted vectors	High encryption complexity; hard to break; good generalization	Requires large training data and computation; model sync required
Deep Autoencoders [19]	Uses encoder-decoder architecture to compress and encrypt speech	Encoder performs compression and transformation simultaneously	Combines compression + encryption; efficient for bandwidth	Reconstruction loss can affect quality; decoder must be securely shared
Adversarial Neural Networks (ANNs) [20]	Obfuscates speaker identity while preserving intelligibility	Modifies cepstral/spectral features to confuse speaker ID systems	High-level privacy and identity masking	Doesn't encrypt message content; only hides identity
Reinforcement Learning (RL) [21]	RL agent adapts encryption policy based on channel/state	Policy dynamically selects optimal encryption based on signal conditions	Adaptive; low overhead in low-threat environments	Complexity of agent design; potential instability in training

CONCLUSION

Time-domain, frequency-domain, and machine learning-based speech encryption algorithms are the three main domains that are compared in this research. Various categories offer distinct benefits and compromises based on the intended level of security, complexity, and application. In general, time-domain encryption methods are quick, simple, and lightweight. Techniques like stream ciphers, amplitude masking, and sample rearrangement are appropriate for low-power devices and real-time applications. These techniques, however, may weaken signal quality in noisy settings and are frequently not resistant to complex cryptanalysis. By converting the voice signal into the spectrum domain, frequency-domain methods such as FFT, DCT, wavelet transformations, and cepstral scrambling provide a more reliable and secure option. These techniques use the frequency characteristics of the broadcast to obfuscate identity and intelligibility. Although they provide better defense against compression and signal processing assaults, they may also add latency and computational expense, which limits their applicability in time-sensitive applications. Secure audio transmission is undergoing a radical change because to machine learning-based encryption techniques. Adaptive, intelligent, and highly secure solutions are offered by methods that use deep neural networks, autoencoders, adversarial models, and reinforcement learning. These systems can conceal the identity of the speaker, adjust encryption according to input properties, and instantly improve encryption tactics. Notwithstanding their complexity, they frequently call for substantial processing power, a large amount of training data, and sender-receiver model synchronization.

To sum up, time-domain approaches work best in contexts with limited resources, frequency-domain approaches balance security and performance, and machine learning-based approaches

are ideal for applications requiring high levels of intelligence, flexibility, and privacy. Future speech encryption systems in sensitive and dynamic contexts would benefit most from a hybrid or layered approach that combines these techniques.

REFERENCES

1. H. Kaur and A. Bansal, "Secure voice communication using frequency domain encryption," International Journal of Computer Applications, vol. 112, no. 15, pp. 6–10, 2015.
2. P. Mishra and S. Mishra, "Hybrid encryption technique for real-time voice communication," Procedia Computer Science, vol. 167, pp. 2163–2170, 2020.
3. S. A. Munir and M. H. Rehmani, "Chaotic-based encryption schemes for real-time speech signals," Multimedia Tools and Applications, vol. 76, pp. 3769–3792, 2017.
4. S. Wang, X. Zhang, and J. Liu, "AI-enhanced speech encryption using deep neural networks," IEEE Access, vol. 8, pp. 127342–127355, 2020.
5. H. Wen, Z. Chen, and Y. Zhang, "Quantum voice encryption for secure communication over classical channels," IEEE Journal on Selected Areas in Communications, vol. 38, no. 3, pp. 478–488, 2020.
6. Ravinder, Sumit Dalal, Sumiran and Rohini Sharma, "A comprehensive review of voice encryption techniques," Synergy: Cross-Disciplinary Journal of Digital Investigation, volume 02, issue 6, 2024.
7. Abdallah, H.A.; Meshoul, S. A Multilayered Audio Signal Encryption Approach for Secure Voice Communication. Electronics 2023, 12, 2.
8. Sheetal, S.; Kumar, L.; Sharma, H. Encryption of an Audio File on Lower Frequency Band for Secure Communication. Int. J. Adv. Res. Comput. Sci. Softw. Eng. 2013, 3, 79–84.
9. Yahya, A.; Abdalla, A. An AES-Based Encryption Algorithm with Shuffling. In Proceedings of the 2009 International Conference on Security & Management, SAM 2009, Las Vegas, NV, USA, 13–16 July 2009; pp. 113–116.
10. Yousif, S.F. Encryption and Decryption of Audio Signal Based on Rsa Algorithn. Int. J. Eng. Technol. Manag. Res. 2020, 5, 57–64.
11. A. Alfalou, N. Khalil, and C. Brosseau, "Real-time audio encryption based on stream cipher and PRNG," Signal Processing, vol. 91, no. 12, pp. 2828–2835, Dec. 2011.
12. S. Sharma and R. Nath, "A new technique for audio encryption using sample rearrangement and amplitude transformation," International Journal of Computer Applications, vol. 82, no. 12, pp. 25–29, 2013.
13. S. K. Singhal and N. Mathur, "Secure voice communication using amplitude masking in time domain," International Journal of Advanced Research in Computer and Communication Engineering, vol. 3, no. 7, pp. 7453–7456, 2014.
14. M. A. Younis and S. A. Bukhari, "Speech signal encryption using FFT and chaos-based key generation," International Journal of Computer Applications, vol. 89, no. 6, pp. 1–6, 2014.
15. A. Al-Haj and A. A. H. El-Khatib, "An efficient DCT-based speech encryption scheme," Digital Signal Processing, vol. 22, no. 2, pp. 428–437, Mar. 2012.
16. R. C. Poonia and D. K. Jain, "Voice signal encryption using discrete wavelet transform and RSA algorithm," Procedia Computer Science, vol. 57, pp. 471–477, 2015.
17. Y. Zhang and J. Li, "Robust speaker identity masking using cepstral domain scrambling," IEEE Transactions on Audio, Speech, and Language Processing, vol. 26, no. 1, pp. 141–150, Jan. 2018.

18. S. Wang et al., “AI-enhanced speech encryption using deep neural networks,” IEEE Access, vol. 8, pp. 127342–127355, 2020.
19. H. Zhang and Y. Luo, “Secure speech transmission using deep autoencoders,” Neurocomputing, vol. 400, pp. 178–188, 2020.
20. J. Qian et al., “Voice anonymization using adversarial neural networks,” Interspeech, 2019.
21. T. Zhang et al., “Reinforcement learning for adaptive encryption in IoT audio communication,” IEEE Internet of Things Journal, vol. 6, no. 5, pp. 8565–8577, 2019.