WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Survey on Voice Recognition Based Money Transaction System Using Steganography

F. P. Jawalkar[1], A. K. Kamble[2], M. S. Khanolkar[3], M. A. Gangarde[4]

[1]Department of Electronics and Telecommunication Engineering, Savitribai Phule Pune University, Pune, Maharashtra, India
[2]Department of Electronics and Telecommunication Engineering, Savitribai Phule Pune University, Pune, Maharashtra, India
[3]Department of Electronics and Telecommunication Engineering, Savitribai Phule Pune University, Pune, Maharashtra, India
[4]Department of Electronics and Telecommunication Engineering, Savitribai Phule Pune University, Pune, Maharashtra, India

A B S T R A C T

Security has been one of the most crucial challenges in today's environment when insecurity is widespread. Voice biometrics is a developing field in security, particularly for the purpose of authentication. Voice biometric speaker recognition utilizes the distinctive qualities of the human voice, including physiological and behavioral traits. These traits have the ability to identify a person and have distinct and relevant vocal features. This method also makes it possible to verify a user regardless of environment or channel changes. With the use of Artificial Intelligence (AI), software can recognize spoken language and translate it into text. Voice recognition is a part of speech recognition. A software system can match a customer's identification to their voice using voice recognition, a feature of AI. In this system, first the user needs to register using their details (email and voice) and then login in through email and voice. Then the user has to select the person to whom the transaction is to be made. Then the user will be able to make a money transaction by simply saying “transfer Rs.” and then the amount. After that the user will get a prompt to check the amount. After approving it the transaction will be successful. In this project, the main objectives are to use voice recognition for login and recognizing amounts for money transfer, and to use steganography for security purposes.

Keywords: Voice Authentication, Feature Extraction, MFCC, Audio Steganography

1. Introduction

With the rapid growth of mobile internet and smartphones, security shortcomings of mobile software and mobile data communication have shifted the focus to strong authentication. The existing user-id/password system is insufficient for mobile use since it is challenging to enter data on a small form factor device and there is a greater chance that the device may fall into the hands of unauthorized users, even though it works fine for desktops and laptops. Strong voice-based authentication in mobile situations has a lot of potential because of recent advancements in voice biometrics. This is especially relevant to the banking and finance sectors, as financial institutions are searching for solutions to provide flexible and simple authentication to mobile consumers while ensuring security and drastically lowering illegal usage [1].

The voice assistant revolution, led by Amazon's Alexa, Apple's Siri, Google Assistant, and others, has naturally led to voice-based payments [3]. According to recent market statistics from Voicebot study, almost 45 million consumers used voice assistants to shop for products at least once in 2021. That represents a significant increase from 2018, when a similar study revealed that only 20.5 million U.S. people have used voice shopping at least once. It shows a 120% increase in the proportion of adults who feel secure enough to make voice-activated purchases on their gadgets. Retailers must develop a next-level customer experience because more individuals worldwide are using their mobile devices for practically all transactional needs. That is how payments work. Voice technology is changing the way we make purchases. Customers on your brand's online store can enable the voice payment system and select the option to pay using voice due to voice commerce and voice payments. Transactions are made even easier when this option is already pre-selected.

Steganography is an intelligent data hiding technique where the secret data is embedded in a cover media in a way that the media carrying the secret message are undetectable and unnoticed by the intruder or attacker. There are 6 steganography techniques, as follows Text Steganography, Image Steganography, Audio Steganography, Video Steganography, DNA Steganography, and Network Steganography. In our system we are using Audio steganography. Audio steganography is a method of hiding information in an audio signal. When data is incorporated into the signal, it is modified. This modification must be made so that the human ear cannot distinguish it.

2. Review of Literature

According to Mondal and Bours, biometrics is a method of providing identification using detectable physical properties [9]. It employs body traits as a tool to encode, scramble, or descramble data. Currently utilized biometric identification techniques include voice recognition, retina and iris scanning,

palm prints, handwritten signatures, finger vein analysis, and facial anatomy. Biometrics is a practical answer in the struggle against fraud and theft, especially when it comes to the Internet, because the aforementioned characteristics are particular to each individual. Because biometric features are difficult to lose, hack, or duplicate, this sophisticated application is regarded to be superior to employing passwords or PINs. The idea is that you are your own password, based on these features. People misplace cards, misplace documents that have been countersigned, or write down PINs on pieces of paper so that others can access them. Using a part of yourself, a registered biometric identifier, that can be used to verify your identity is one way to protect data.

Mohammad Al Rousan and Benedetto Intrigila concluded that biometrics are able to achieve fast and easy-to-use authentication with high accuracy and a relatively low cost [2]. In this paper, the authors have tabulated a comparison of various biometric techniques based on various parameters. The voice biometric has a medium distinctiveness and a medium performance. But the issue is with the acceptance. People still think voice biometric is somewhat insecure. Biometrics, particularly voice biometric, in few years, use of money, credit cards, and checks will all be replaced by the use of biometrics. But it is important to be cautious about who can use the collected biometric data and for what purposes.

According to Nilu Singh, Alka Agrawal and R. A. Khan, voice based authentication goes through two different processes one after the other [7]. First is feature extraction and second is the model creation. Feature extraction uses techniques like MFCC and LPC. Different types of model creation include GMM, HMM, pattern matching, vector quantization, decision tree and neural networks. In this research paper, features like speaking rate, speaking style, pitch prosody is considered. These features are associated with linguistic components of voice such as syllables and it is noticeable that changes occur in measurable parameters, for example fundamental frequency F0, energy and duration of speech. This solution has many advantages like convenience, increased security and accuracy whereas the cost of implementation is high.

Sonali Goyal and Neera Batra compared the LPC and the MFCC technique for voice authentication systems. They concluded in their paper that the LPC technique is 89.23% accurate and the MFCC technique is 92.7% accurate [10].

According to Rohit Tanwar, Kulvinder Singh and Sona Malhotra [5], different finance firms as well as industry are relying on voice recognition authentication for their security. With the improvement in research for NLP, speech recognition can be used for disable people who are otherwise not able to authenticate themselves using traditional techniques.

Hanlin Liu, Jingju Liu and Xuehu Yan [6] proposed an audio steganography scheme based on the time length of WeChat voice message, which is oriented to social network behavior. After chaotic scrambling and run length encoding, the secret information is encoded into run length, which then is mapped to time length that represents the secret data. The sender sends a voice message with corresponding time length, and the receiver extracts the secret information based on the time length of the voice message.

As per Dr. G. Sundari and Mrs. Alaknanda S. Patil [8], stenography is one of secure methods to transform secret data in wired or wireless communication. Though a variety of embedding methods are available for audio steganography the LSB embedding is the only method with lower computational difficulty but higher security by varying the secret data embedding position. The sturdiness can be again increased by introducing PN sequence generator and secret key.

Enas Wahab Abood et al developed a hybrid approach to protect a plain text message that was encrypted using the Hill encryption method and randomly distributed within an audio file using audio symbol signs to represent message bits [4]. The audio file can only be a *.wav stereo file, and two secret keys are required to decrypt the system and generate hiding spots. The findings of the calculation of PSNR, MSE, and SSIM between the cover before and after embedding reflected the system's imperceptibility criterion. Also, for various text sizes, the time required to secure messages was fairly low, and the encryption process took less time than hiding.

Ali M. Meligy, Mohammed M. Nasef and Fatma T. Eid offered an audio steganography algorithm for embedding text, audio, or images that is based on the Lifting Wavelet Transform (LWT) transform with modification of the Least Significant Bit (LSB) technique and three random keys. These keys are used to increase the robustness of the LSB technique because without them, no one would be able to determine the type of secret message, its length, or its initial position within the LSB. Performance of this algorithm is calculated by comparing the SNR [14].

M. Parthasarathi and T. Shreekala proposed an alternative strategy which focused on minimizing the distortion to the prediction error and the data size overhead [12]. This strategy is based on the prediction error that results from it and the challenge of handling the nonlinear quantization process. The advantage is that they achieve a lower distortion in the quality of audio while the disadvantages are that the data size increases during extracting and embedding. Also there may be a loss in the hidden data. Based on this study, we compared Recognition Rates for various feature extraction techniques for English language, English numerals and for Devanagari [15]

Table 1 – Comparison of Feature Extraction Techniques

Sr. no.	Technique	Language	Recognition Rate
1.	Linear Predictive Coding (LPC)	English	91.4%
		English Numerals	94%
		Devanagari	82.3%
2.	Mel Frequency Cepstral Coefficient (MFCC)	English	99.9%

		English Numerals	92.93%
		Devanagari	85.3%
3.	Zero Crossing with Peak Amplitude (ZCPA)	English	96.67%
		English Numerals	95.4%
		Devanagari	38.5%
4.	Dynamic Time Wraping(DWT)	English	90.5%
		English Numerals	91.1%

Based on previous studies, we compared various audio steganography techniques on various parameters like embedding techniques, strengths, weaknesses and hiding rate. [11]

Table 2 – Comparison of Audio Steganography Techniques

Sr. no	Methods	Embedding Technique	Strengths	Weaknesses	Hiding Rate
	Least Significant Bit	LSB of each sample in the audio is replaced by one bit of hidden information	Simple and easy way of hiding Information with high bit rate	Easy to extract and to destroy	16 kbps
	Echo Hiding	Embeds data by introducing echo in the cover signal	Resilient to lossy data compression algorithms.	Low security and capacity	40-50 bps
	Phase Coding	Modulate the phase of the cover signal.	Robust against signal processing manipulation and data retrieval needs the original signal	Low Capacity	333 bps
	Parity Coding	Break the signal into separate samples and embeds each bit from secret message in sample region parity bit.	Sender has more of a choice in encoding the secret bit.	Not Robust	320 bps
	Spread Spectrum	Spread the data over all signal frequencies.	Provide better robustness.	Vulnerable to time scale modification	20 bps

3. Keywords

1) Voice Authentication: It is a type of security authentication that relies on a person's unique voice patterns for identification in order to gain access. This method of verification requires both a device that can correctly records a person's voice and software that can analyze voice patterns and compare them to known patterns. Voice authentication is also known as voice biometrics, voice ID or speaker recognition.

2) Feature Extraction: By converting the speech waveform to a parametric representation at a significantly lower data rate for further processing and analysis, features are extracted. Typically, this is referred to as front end signal processing.

3) Linear Predictive Coding(LPC): Linear predictive coding (LPC) makes use of the data from a linear predictive model. This is mostly employed in audio signal processing and speech processing to capture the spectral envelope of a digital signal of speech in compressed form.

4) Mel Frequency Cepstral Coefficient (MFCC): MFCC is a feature extraction technique which converts audio in time domain to frequency domain so as to get the information present in the audio signal or the speech signal. Working of this technique can be represented with a block diagram shown below [12].

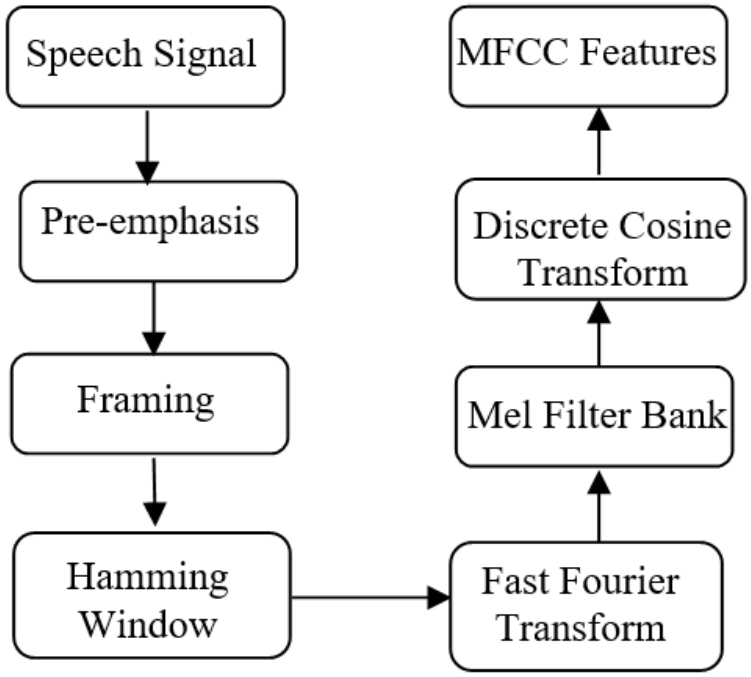

Fig. 1 - Flowchart of MFCC feature extraction technique

To calculate MFCC coefficients, overlapping frames of the signal is separated. Let consecutive frames be spaced by M samples, where M > N, and let each frame contain N samples [10]. A Hamming window is applied to each frame, and its equation is as follows:

$$Wn = 0.54 - 0.46 \cos (2\pi n /(N-1)) \quad (1)$$

In the third step, the signal is transformed using the Fourier Transform from the time domain to the frequency domain. Equation (2) represents a signal's discrete Fourier transform:

$$x_n = \frac{1}{N} \sum_{k=0}^{N-1} (X_k . e^{i2\pi kn/N}) , \quad n \in Z \quad (2)$$

The frequency domain signal is then transformed to Mel frequency scale, which is more in line with human hearing and perceptions, in the following stage. A group of triangular filters are employed to compute a weighted sum of spectral components in order to achieve this goal and produce an output that closely resembles a Mel scale. The magnitude frequency response of each filter is triangular in shape, equal to one at its center frequency, and drops linearly to zero at the centers of two consecutive filters. The formula used to determine the Mel at a particular frequency is given in equation (3):

$$M = 2595 log 10(1 + (f / 700)) \quad (3)$$

The log Mel scale spectrum is then transformed via the discrete cosine transform into the time domain (DCT). The following expression gives the definition of DCT, where is an N-dependent constant:

$$x_k = \alpha \sum_{i=0}^{N-1} x_i . \cos \left(\frac{(2i+1)\pi k}{2N} \right) \quad (4)$$

5) Zero Crossing with Peak Amplitude (ZCPA): A zero-crossing is an instantaneous point at which the sign of a mathematical function changes (e.g. from positive to negative). So the ZCPA method uses time domain zero crossings to extract frequency information and related intensities in a number of psycho-acoustically scaled sub-bands.

6) Dynamic Time Wraping(DWT): Dynamic Time Warping is a technique for aligning sequences and calculating the distance between them. These sequences can be both time based and non-time-based, such as protein sequences, and can be either time-based like audio recordings.

7) Audio Steganography: Audio steganography is a method of hiding information in an audio signal. When data is incorporated into the signal, it is modified. This modification must be made so that the human ear cannot distinguish it. Different audio steganography techniques are Least Significant Bit Coding, Phase Coding, Parity Coding, Echo Data Hiding and Spread Spectrum.

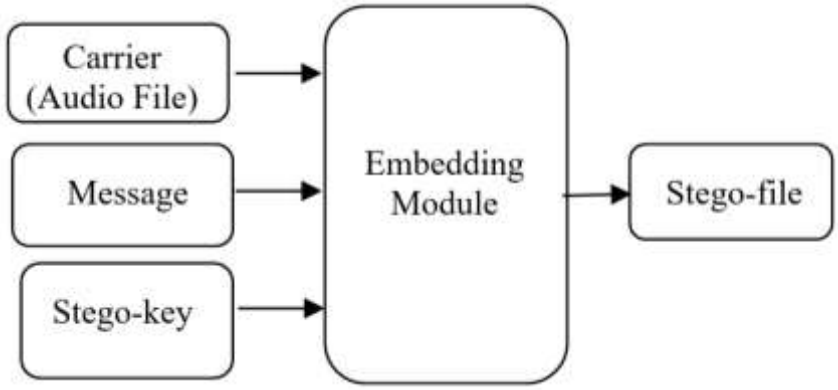

Fig. 2 - Basic Audio Steganography Model

The fundamental components of audio steganography are a carrier file, a message that is provided, and a password, sometimes known as a stego-key. A carrier, or cover-file, is a document that conceals sensitive information. This process can be understood with the help of a block diagram shown [11].

8) Echo Hiding: Echo hiding used to embeds secret data in an audio file by pass an echo into the discrete signal. When compared to other approaches for echo hiding, this technology has the advantages of offering a high data transfer rate and resilience. The original signal is divided into blocks before the encoding process begins, allowing one echo from the original signal to yield one bit of secret data. The blocks are concatenated back together to generate the final signal after the encoding procedure is complete.

9) Phase Coding: The HAS insensitivity to the relative phase of various spectrum components is exploited by phase coding. With this technique, we can substitute concealed information for a few phase components from the original sound signal spectrum. Phase components medication should be kept in a minimal dose due to the difficulty in hearing information. In terms of the SNR ratio, it is a very effective coding technique. Phase dispersion happens when the ratio between each frequency component's phase changes. An inaudible coding can be accomplished if the phase modification is sufficiently small.

10) Parity Coding: One of the reliable methods for audio steganography is parity coding. It divides a signal into separate samples instead of dividing it into individual samples and embeds each bit of the secret message information from a parity bit. The technique inverts the LSB of one of the sections in the area if the of a chosen parity bit region does not match the secret message bit to be encoded. The sender then has a range of choices for encoding the secret bit.

11) Spread spectrum: This method spreads the encoded data over the range of frequencies. This is comparable to a system that randomly distributes the message bits throughout the whole sound file by implementing an LSB coding method. This technique uses a code that is separate from the original signal to disperse the secret data around the audio file frequency spectrum. The final signal uses more bandwidth than is actually necessary for transmission at the end. Chip rate for the communication of sound signals uses sampling.

Spread Spectrum can be used in two different ways:

a) Direct-sequence: Direct-sequence The chip rate, a pseudorandom signal, and an interleaved cover signal are all used by SS to spread out the cover signal and the secret message, respectively.

b) Frequency-hopping schemes Frequency-hopping SS modifies the audio files' frequency spectrum to quickly switch between frequencies. Below are the spread spectrum steps:

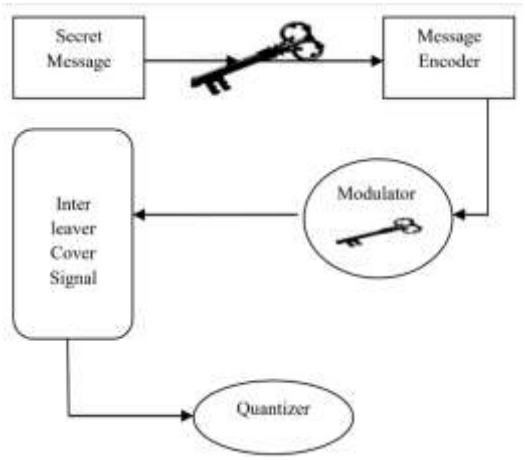

Fig. 3 - Spread spectrum method of audio steganography

4. Applications

Some of the use-cases of voice based money transaction include:

1) Peer-to-peer transfers using online wallets and platforms.
2) Payment gateways
3) Purchasing goods from ecommerce sites or physical stores and paying using voice-activated credit card applications.
4) Utilizing the "reorder" tool to place orders for things you frequently purchase.
5) Perform a variety of financial transactions including paying off card debt and fund transfers.

5. Limitations

1) Security and privacy: A recent study found that users still have trouble believing the voice of virtual or intangible technologies. Customers worry about extraterrestrial technology, but they also worry that the same technology may become overly personal.
2) Accents: Voice assistants still have problems understanding various accents, particularly those that aren't American.
3) Point of service integration: To ensure widespread adoption, voice-activated technology must be compatible with the hardware and software at the POS of retailers. The majority of the costs associated with integrating with Wi-Fi and Bluetooth-enabled systems must, however, be provided by merchants, which poses a significant barrier.
4) Financial institutions: Many well-known companies in the financial industry have begun implementing voice payments. Due to limited roll-outs and the lack of a significant global collaboration with banks, voice payments are still gradually shifting.

6. Conclusion

Different feature extraction algorithms and audio steganography algorithms have been studied and investigated in this paper. The objective is to provide a voice based money transaction and secure it with the help of audio steganography techniques. However, it is quite difficult to secure voice. This analysis revealed that the recognition rate of the English-language database is higher than that of other languages. For languages other than English, the automatic speech recognition system has a recognition rate in between 80% to 90%. Mel Frequency Cepstral Coefficient (MFCC) is one of the most useful methods for encoding good quality speech at a low bit rate and provides extremely accurate estimates of speech parameters. The analysis of different techniques of audio steganography based on some parameters like strengths, weaknesses, embedding technique, hiding rate have been discussed in this paper. The study showed that the Least Significant Bit technique can be easily broken, while the phase encoding method is better in terms of security and signal manipulation.

Acknowledgements

We would like to express our sincere gratitude to our college especially our E&TC department for providing an opportunity to work on the project. We would like to convey our heartfelt gratitude to Dr. M. A. Gangarde for his tremendous support and assistance in the completion of survey paper of our project and constantly encouraging and guiding us throughout the semester without which completing out required project work in short span could not be possible. His initial guidance regarding the study of several research papers related to our project helped us a lot while completing our project.

References

[1] Amjad Hassan Khan, M. K., & Aithal, P. S. (2022): “Voice Biometric Systems for User Identification and Authentication – A Literature Review”. International Journal of Applied Engineering and Management Letters (IJAEML), 6(1), 198-209. DOI: https://doi.org/10.5281/zenodo.6471040

[2] Mohammad Al Rousan and Benedetto Intrigila (2020): “A Comparative Analysis of Biometrics Types: Literature Review”. Journal of Computer Science, 16 (12): 1778.1788 DOI: 10.3844/jcssp.2020.1778.1788

[3] G. Yu. Peshkova, O. V. Zlobina (2020): “Digital transformation of banking with speech technologies”. -ISSN: 2357-1330. ICEST. DOI: 10.15405/epsbs.2020.10.03.34

[4] Enas Wahab Abood et al (2020): “Securing Hill encrypted information With Audio steganography: a New Substitution Method”, J. Phys.: Conf. Ser. 1591 012021

[5] Rohit Tanwar, Kulvinder Singh, Sona Malhotra (2019): “An Approach to Ensure Security using Voice Authentication System” International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878

[6] Hanlin Liu, Jingju Liu, Xuehu Yan (2018): “Social Network Behavior-Oriented Audio Steganography Scheme” Eighth International Conference on Instrumentation and Measurement, Computer, Communication and Control

[7] Nilu Singh, Alka Agrawal, and R. A. Khan.(2018): "Voice Biometric: A Technology for Voice Based Authentication". Article in Advanced Science, Engineering and Medicine. DOI: 10.1166/asem.2018.2219

[8] Mrs. Alaknanda S. Patil, Dr. G. Sundari (2018): "An Embedding of Secret Message in Audio Signal" 3rd International Conference for Convergence in Technology (I2CT)

[9] Mondal, S., & Bours, P. (2017): "A study on continuous authentication using a combination of keystroke and mouse biometrics".

[10] Sonali Goyal and Neera Batra (2017): "Issues and Challenges of Voice Recognition in Pervasive Environment". Indian Journal of Science and Technology, Vol 10(30), DOI: 10.17485/ijst/2017/v10i30/115518.

[11] Palwinder Singh (2016): "A Comparative Study of Audio Steganography Techniques" (IRJET)

[12] M. Parthasarathi and T. Shreekala (2017): "Secured Data Hiding in Audio Files Using Audio Steganography Algorithm", International Journal of Pure and Applied Mathematics Volume 116.

[13] Shweta Vinayakarao Jadhav, Prof. A.M Rawate (2016): "A New Audio Steganography with Enhanced Security based on Location Selection Scheme", IJESC.

[14] Ali M. Meligy, Mohammed M. Nasef and Fatma T. Eid (2015): "An Efficient Method to Audio Steganography based on Modification of Least Significant Bit Technique using Random Keys", I.J. Computer Network and Information Security.

[15] Pratik K. Kurzekar, Ratnadeep R. Deshmukh, Vishal B. Waghmare, Pukhraj P. Shrishrimal (2014): "A Comparative Study of Feature Extraction Techniques for Speech Recognition System". International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2319-8753, DOI: 10.15680/IJIRSET.2014.0312034