SSRG International Journal of Electrical and Electronics Engineering
ISSN: 2348-8379/ https://doi.org/10.14445/23488379/IJEEE-V11I6P135

Volume 11 Issue 6, 351-358, June 2024

Original Article

Development of a Low-Cost Security System Based on Voice Recognition Using Artificial Intelligence

Willians Jeremy Luna Condori[1], Emily Juliana Mamani Macedo[2], Alex Leon Ppacco Huamani[3], Jesús Talavera S.[4], Jarelh Galdos[5]

[1,2,3,4,5]*Department of Electronic Engineering, School of Production and Services, Universidad Nacional de San Agustín de Arequipa, Arequipa, Peru.*

[5]*Corresponding Author : jgaldosb@unsa.edu.pe*

Received: 16 April 2024 Revised: 19 May 2024 Accepted: 16 June 2024 Published: 29 June 2024

Abstract *- Voice recognition has been widely used in various applications, especially in the field of security. In this paper, we propose the development of a low-cost security system based on voice recognition using artificial intelligence. The system utilizes a Raspberry Pi 4B as a microcontroller and Python as a programming language. The system works with a pre-recorded database of voices from 20 people, and the new user's voice is matched against the pre-recorded voices using Gaussian Mixture Model (GMM). We extracted Mel-Frequency Cepstral Coefficients (MFCC) from the recorded voices, which were used to train the GMM. The system achieved an accuracy rate of 95.42%, with an equal error rate of 4.57%. The proposed system is low-cost and easy to use, making it accessible to a wider audience. However, it has some limitations, such as only being able to work with a pre-recorded database of voices.*

Keywords *- Voice recognition, Security system, Gaussian mixture model, Mel-frequency cepstral coefficients, Low-cost biometric-systems.*

1. Introduction

In recent years, home burglary has become an increasing problem in Peru and many other Latin American countries. According to data from the National Institute of Statistics and Informatics (INEI), the number of home burglaries has risen in recent years. Between March and August of 2023 alone, in cities with more than 20,000 inhabitants, 13.4% of homes in the country were affected by burglary or attempted burglary [1].

Home security is a fundamental concern for Peruvian citizens, and many are seeking effective and affordable solutions to protect their homes. In this context, biometric-based security systems have gained traction worldwide, as they offer a high level of security and user identity authentication.

However, despite the many benefits of biometric-based security systems, their high cost has been a significant obstacle to their widespread adoption in developing countries like Peru. This has led to the search for more affordable and accessible solutions for Peruvian citizens. An interesting solution is the use of Raspberry Pi, a low-cost microcontroller, along with programming languages like Python, to develop biometric-based security systems. Biometric-based security systems have proven to be effective in protecting homes and businesses, and their popularity is on the rise. These systems use unique physical or behavioral characteristics of individuals to authenticate their identity and grant them access to a location or device. However, these systems can be expensive and out of reach for many people in developing countries like Peru.

Fortunately, advances in technology and the growing popularity of low-cost microcontrollers like Raspberry Pi have enabled the development of biometric-based security systems at more affordable prices. Both individuals and businesses can use these systems which can help protect properties and assets from the threat of burglaries.

In particular, voice recognition-based security systems are an interesting option for Peruvian citizens, as they may be more affordable than other biometric-based security systems such as fingerprint or facial recognition. Additionally, voice recognition is a unique feature of each individual, making it especially effective for identity authentication.

In this context, the present article introduces the development of a low-cost security system based on voice recognition using artificial intelligence. The system utilizes Raspberry Pi as the microcontroller and the Python programming language for its development. The system

operates by comparing the recorded voice of the user with a pre-recorded database of authorized users' voices. If a match is detected, an electromagnetic lock is unlocked, allowing access to the property. This article also provides a comparison of the system's cost with other biometric-based security systems, such as fingerprint and facial recognition. Additionally, it discusses the results obtained in the system tests and analyzes its potential future applications.

In summary, this article presents an affordable and effective solution for the protection of homes and properties in Peru and other developing countries. The voice recognition-based security system developed in this work represents a viable and accessible alternative to traditional biometric-based security systems, with the potential to enhance security and safeguard the assets of Peruvian citizens.

2. Literature Review

In the development of a low-cost security system based on voice recognition using artificial intelligence, it is essential to review existing literature to understand the advances and methodologies employed in similar domains. This section provides an overview of related work in the fields of home automation, voice-controlled smart home systems, and speaker verification.

Several studies have explored the integration of voice recognition technology into home automation systems. Bharathi et al. [2] introduced a methodology for remote appliance control using Raspberry Pi and Android mobile phones. Similarly, Obaid et al. [3] investigated the use of ZigBee and voice-controlled wireless methodologies for smart homes. Alexakis [4] proposed a speech recognition-based methodology for smart homes. These studies highlight the various approaches and technologies used in the development of smart home automation systems.

Speaker verification is a crucial component of voice recognition systems. Shah et al. [5] proposed a biometric voice recognition system for security applications, where they used MATLAB to determine whether a user is accepted or rejected, highlighting its potential to enhance security measures. Ossama et al. [6] explored the use of Convolutional Neural Networks (CNNs) for speech recognition, demonstrating the effectiveness of deep learning techniques in improving verification accuracy.

The literature review reveals a series of significant advancements and persistent challenges. The adoption of Gaussian Mixture Models (GMMs) and Mel-Frequency Cepstral Coefficients (MFCCs) for voice identification has been established as a robust approach due to its effectiveness in capturing the distinctive characteristics of voice signals. In the work of Liu et al. [7], MFCCs and GMMs were utilized for the development of an access control system, yielding favorable results but emphasizing the importance of the training database. It was noted that better results are obtained when the test voices are included in the training database.

In the study by Ali [8], an automatic voice disorder detection system based on continuous speech was presented, focusing on MFCCs and the GMM model. The system achieved a detection rate of 91.66% with continuous speech, demonstrating its effectiveness in distinguishing between normal and pathological voices.

In Paulose [9], the author focuses on implementing a speaker recognition system where it is demonstrated that GMM outperforms the i-vector method in terms of performance, especially when the duration of test signals is increased. It was further observed that the recognition accuracy increases when specific speaker information, such as pitch, is added to MFCC and IHC features.

In Hanilçi [10], a comparison is made in identity spoofing detection when a person speaks using i-vector and GMM, where the initial experimental results yielded Equal Error Rates (EER) of 4.624% and 2.391%, respectively. This highlights that under simple conditions, GMM surpasses i-vector.

Furthermore, in El Ayadi et al. [11], a performance comparison of classifiers used for speech emotion recognition is conducted, highlighting the average accuracy achieved by GMM compared to other classifiers such as Hidden Markov Model (HMM), Artificial Neural Network (ANN), and Support Vector Machine (SVM). It is also noted that GMM has the shortest training time in comparison.

The framework of our study relies on the demonstrated effectiveness of MFCC and GMM for voice processing and recognition, considering their advantages in modeling unique features of the human voice. Additionally, we want to emphasize the importance of developing robust defenses against unwanted attacks, a critical aspect of ensuring the security and reliability of AI-based systems.

Through this work, we aim to contribute to the existing literature by offering an innovative solution that is not only technically and economically viable but also effectively addresses security concerns and is accessible to users in development contexts.

3. Materials and Methods

The development of the voice recognition-based security system was carried out using various specific components and methodologies. A Raspberry Pi 4B 8GB was employed as the central microcontroller of the system. This device provides the necessary processing power and flexibility to run the voice recognition software.

Additionally, a high-quality microphone connected via the USB ports of the Raspberry Pi was used for voice capture. This microphone ensured clear and precise audio input, which is fundamental for the proper operation of the voice recognition system. An electromagnetic lock was integrated into the system to provide physical access control.

This lock is activated or deactivated based on the identification of the user's voice by the system. The controller will send a signal to a 5v single-channel relay, which will activate the lock. A 12v power source powers the electromagnetic lock. The experimental stage was conducted in two phases: the training phase and the real-time recognition phase.

3.1. Phase 1 - Training

Voice data were collected from a group of 20 voluntary participants. Each person uttered a series of predefined phrases and keywords in a controlled and quiet environment. Multiple recordings were made for each individual to obtain a representative sample of their voice.

High-quality microphones ensured good audio quality during this process, with a sampling frequency of 16 kHz and a resolution of 16 bits. The acquired voice recordings underwent a preprocessing stage to enhance audio quality and reduce background noise. This included the removal of unwanted noise, volume normalization, and equalization adjustments to improve speech clarity. This step was crucial to ensure that the recordings were consistent and suitable for subsequent analysis.

Once the processed recordings were obtained, the Mel Frequency Cepstral Coefficients (MFCC) were extracted. These coefficients are widely used in numerous research studies and systems related to human speech recognition. In the research field, MFCC feature extraction has been utilized for emotion detection [12-14], disease diagnosis [15], and voice recognition [16, 17]. The Gaussian Mixture Model (GMM) was employed to train the voice recognition system. GMM is a probabilistic model commonly used in voice recognition due to its ability to model voice feature distributions [18, 19].

Both MFCC and GMM have been combined for voice recognition in biometric systems [7, 20], demonstrating good results that support their usage. At this stage, the parameters of the GMM were adjusted, such as the number of components and covariances.

With the trained GMM models, a voice database was constructed containing the information of each person registered in the system. This database consisted of GMM models corresponding to the voice features of each individual. Each GMM model was associated with a unique label identifying a person. The process during phase 1 is illustrated in Figure 1.

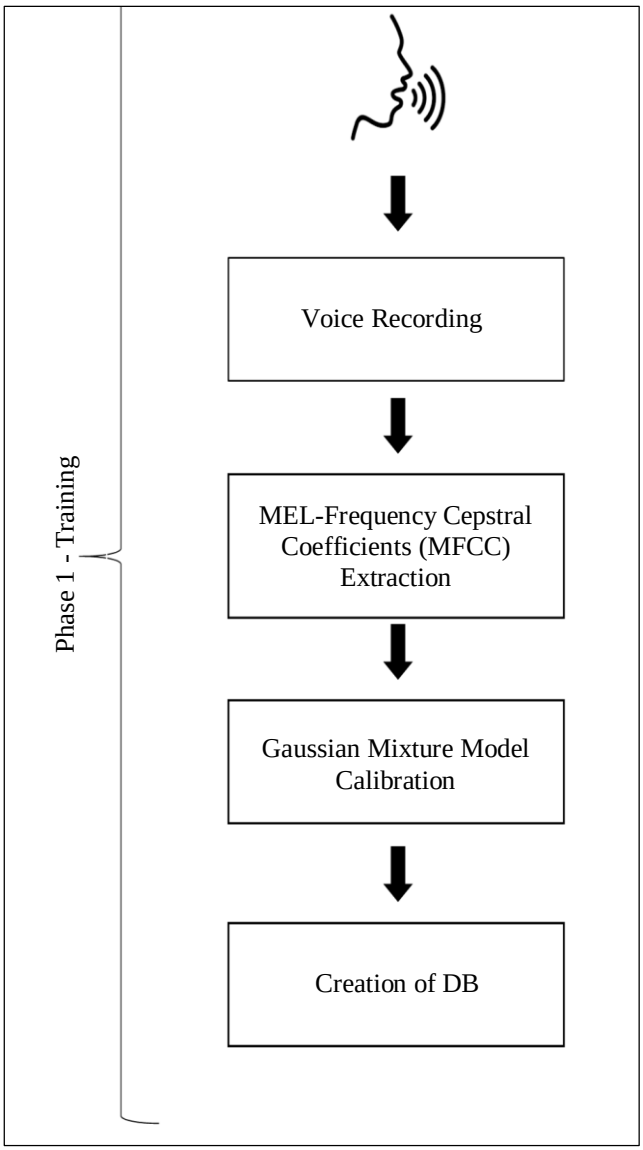

Fig. 1 Process in phase 1

3.2. Phase 2 - Real-Time Recognition

The voice recognition system was implemented in a real-time environment using a Raspberry Pi 4B as the microcontroller. The Raspberry Pi 4B was connected to a microphone to capture the user's voice and perform feature extraction using MFCC analysis. Subsequently, the extracted features were compared with the GMM models stored in the voice database using probabilistic classification techniques.

When an unknown user attempts to access the system, their voice is captured, and features are extracted using MFCC analysis. Then, a comparison is made with the GMM models stored in the voice database. If a high probability match is found, the user's identity is verified, and access to the system is granted. If there is no clear match, the user is considered unauthorized, and access is denied. The process during phase 2 is depicted in Figure 2.

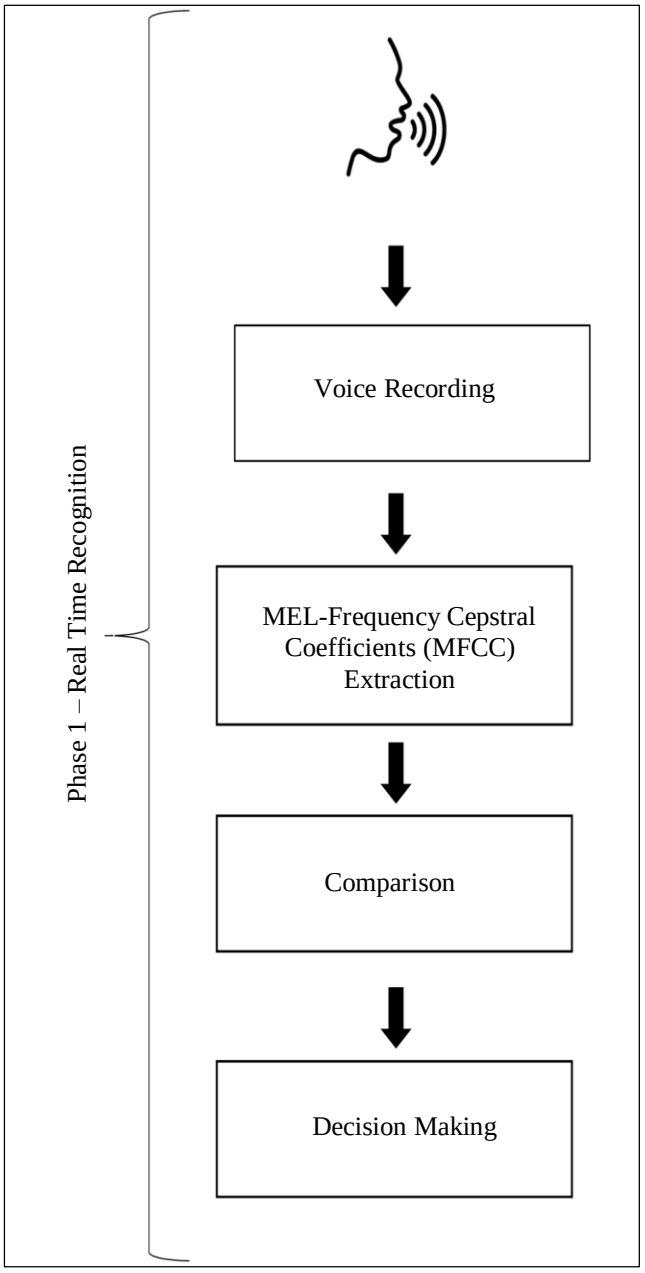

Fig. 2 Process in phase 2

The system operates as detailed in the flowchart depicted in Figure 3. The acceptable threshold θ was empirically defined to determine whether to accept a user. Upon system initialization, the "Tries" variable, referring to the number of user attempts, is created. The new user approaches the microphone and speaks two sentences.

A comparison is then made, and if the user's voice value α does not surpass the threshold for any label, the entry is denied. The "Tries" variable increases by 1, and the user is asked if they would like t'o try again. This process repeats up to 3 times; after that, the program terminates, definitively denying entry.

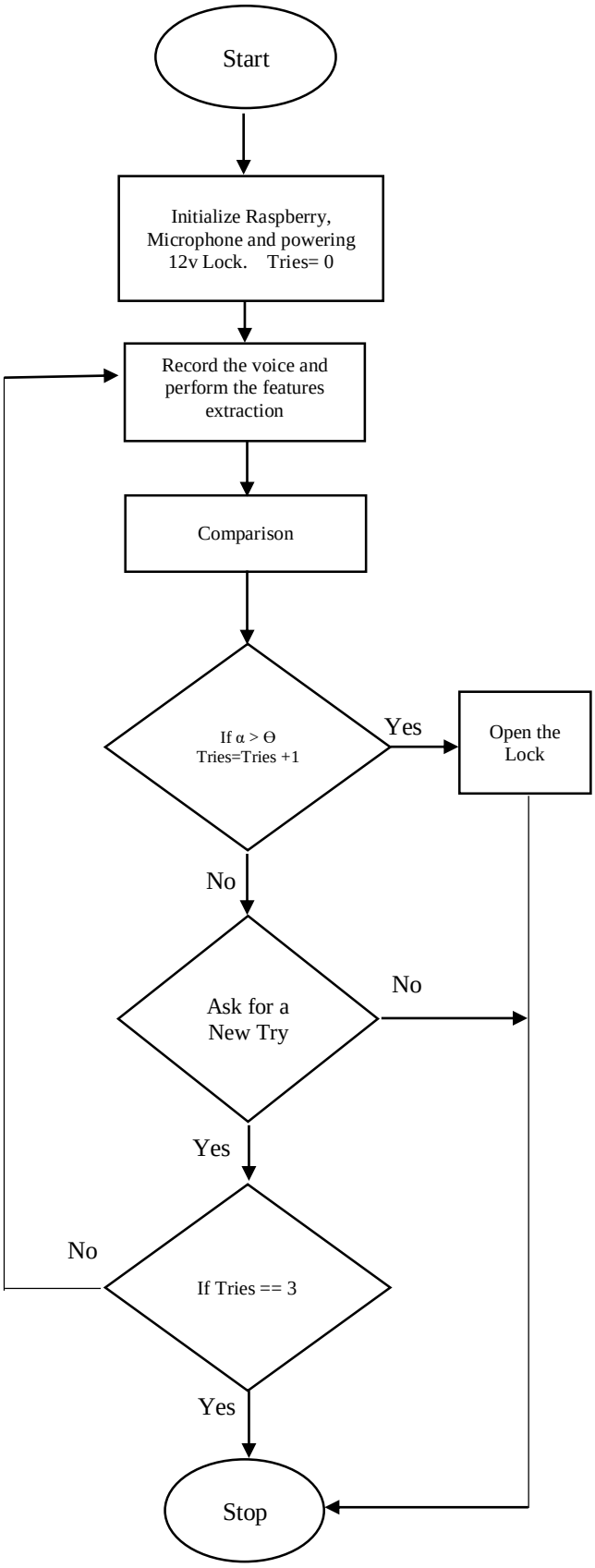

Fig. 3 Flowchart of the System

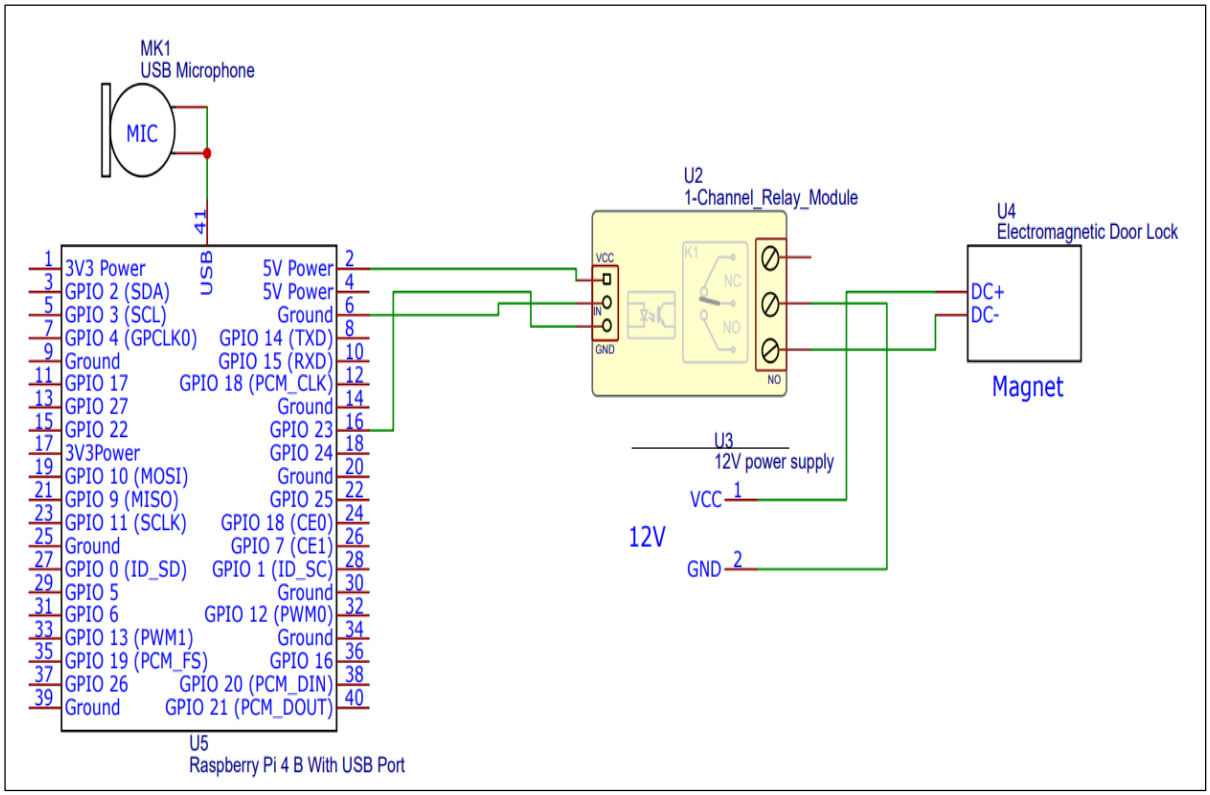

Fig. 4 System's schematic

The connections between the electronic components are detailed in Figure 4, corresponding to the schematic diagram of the system. To evaluate the performance of the voice recognition system, comprehensive tests were conducted using a separate dataset of voice recordings not used during training. These tests allowed for the measurement of the system's accuracy and robustness in real-world situations. Parameters such as false positive rate, false negative rate, true positive rate, and true negative rate were evaluated to obtain the Equal Error Rate (EER). Based on the test results, improvements and adjustments were made to the system to optimize its performance. This included optimizing the parameters of the GMM and adjusting the decision thresholds. Additionally, additional tests were conducted to ensure that the system is reliable and accurate in different environmental conditions and with users of different vocal characteristics. The results are presented and discussed in the following section.

4. Results and Discussion

The voice recognition security system using Gaussian Mixture Model (GMM) has been thoroughly evaluated to measure its performance and accuracy. Below are the results obtained during the conducted tests:

4.1. Evaluation of the Recognition Rate

During the voice recognition tests, recordings of both authorized and unauthorized individuals were utilized. In total, 285 different voice recordings were collected from the 20 voluntary participants, varying in intonation, pronunciation, and speech rate.

The recognition rate was measured to determine the system's ability to identify authorized individuals and reject unauthorized ones correctly. The success rate in identifying authorized individuals was 95.42%. This indicates that the system accurately recognized the majority of individuals registered in the voice database.

This high recognition rate demonstrates the system's effectiveness in verifying the identity of authorized users, suggesting a low probability of granting access to unauthorized individuals. In Table 1 are the results for different thresholds, depicting true positives and false positives. The Equal Error Rate (EER) value is determined by the interpolation of the False Acceptance Rate (FAR) and False Rejection Rate (FRR) as defined by Equations (1) and (2):

$$FRR = \frac{FN}{TP+FN} \quad (1)$$

$$FRR = \frac{FP}{FP+VN} \quad (2)$$

With the obtained data, the Equal Error Rate (EER) was calculated, resulting in a value of approximately EER ≈ 4.57%.

Table 1. True positive and false positive values for different threshold values

Threshold	True Positive (%)	False Positive (%)
0	100	0
0.2	98.12%	1.87%
0.4	95.42%	4.57%
0.6	87.82%	11.18%
1	47.05%	46.94%

4.2. Performance Evaluation under Adverse Conditions

Additional tests were conducted to evaluate the system's performance under adverse environmental conditions, such as background noise and variations in volume and intonation of the voice. The goal was to determine the system's robustness and its ability to maintain a high level of accuracy in real-world situations.

The system demonstrated good tolerance to background noise, being able to recognize the voice of an authorized person even in noisy environments. However, a slight decrease in the recognition rate was observed in the presence of intense and persistent noise. Regarding variations in volume and intonation of the voice, the system showed acceptable ability to adapt to these changes. It was observed that the system correctly recognized authorized individuals even when the voice recording exhibited differences in volume and intonation compared to the training recordings.

4.3. Response Time and System Efficiency

The voice recognition system based on the GMM model demonstrated a fast response time and remarkable efficiency. The processing of voice features and comparison with the GMM models stored in the database was performed in real-time, enabling a rapid response to the user. The average processing time for a voice recording and comparison with the GMM models was approximately 1 second. This response speed contributes to a smooth user experience and ensures convenient access to the security system.

4.4. Limitations and Areas for Improvement

- Although the voice recognition security system using the GMM model showed promising results, several limitations and areas for improvement were identified that could be addressed in future research. While the system demonstrated good tolerance to background noise, it is important to note that extreme noise conditions could affect recognition accuracy. Future improvements could explore the incorporation of advanced noise suppression techniques to enhance performance in noisy environments.
- The system may encounter difficulties in identifying authorized individuals due to variations in pronunciation. This could occur when a person pronounces a word or phrase slightly differently from the recording used during training. To overcome this challenge, techniques for pronunciation adaptation and normalization could be explored.
- In this study, a voice database with recordings from a group of 20 individuals and 285 voice recordings was used. To improve the system's generalization and ensure a higher level of accuracy, it is recommended to expand the database with recordings from a larger and more diverse group of individuals. This would allow for better representation of vocal variations and increased robustness.

4.5. Prices Comparison

To perform the price comparison, the cost of components will be summed up in US Dollars (USD), converting their price in Peru using the current exchange rate.

- Raspberry pi 4B+: 99.95 USD (S/374.50)
- USB Microphone: 16.00 USD (S/60.00)
- Relay 1-channel 5V :3.20 USD (S/12.00)
- Electromagnetic Lock 12V 600lbs: 35.99 USD (S/134.80)
- Power Supply 12v: 9.34 USD (S/35.00)

The total price for all electronic components is 164.48 USD (S/616.30). To this amount, the cost of building and designing a casing to house all the components should be added, which would be approximately 15 USD. This makes the total cost 179.48 USD (S/672.46). Here are the prices of commercial devices and their specifications, again given in USD with the current exchange rate.

In Table 2, you can see the prices for various products sold in Peru is shown in Appendix. All the products significantly exceed the total price of the voice recognition system. In the context of the economy in Peru, many families do not have sufficient funds to afford one of these locks.

5. Conclusion

The present study developed a low-cost security system based on voice recognition using artificial intelligence and a Raspberry Pi microcontroller. The system exhibited a success rate of 95.42% in identifying authorized users and achieved an Equal Error Rate (ERR) of approximately 4.57%. These results underscore the system's effectiveness in identity verification and its potential to protect homes in low-resource environments.

The obtained results indicate that the developed security system is a viable and economical solution to enhance home protection. The high recognition rate and low ERR

demonstrate that the combination of Mel Frequency Cepstral Coefficients (MFCC) and Gaussian Mixture Models (GMM) is effective for voice recognition in security applications. To improve the system's robustness, future research could explore advanced noise suppression techniques and pronunciation adaptation. Expanding the voice database with a more diverse sample of users would also be beneficial. Furthermore, other voice recognition technologies and models could be investigated, along with integrations with other security systems, to create more comprehensive and secure solutions. The implementation of similar systems can significantly contribute to reducing the risk of burglaries and improving the quality of life for citizens in developing countries without having a negative impact on their economies.

Acknowledgments

The authors would like to thank Universidad Nacional de San Agustin de Arequipa.

References

[1] National Institute of Statistics and Informatics, Citizen Security Statistics, 2023. [Online] Available. https://www-gob-pe.translate.goog/institucion/inei/colecciones/6094-estadisticas-de-seguridad-ciudadana?_x_tr_sl=es&_x_tr_tl=en&_x_tr_hl=en&_x_tr_pto=sc

[2] H. Bharathi et al., "Home Automation by Using Raspberry Pi and Android Application," *2017 International Conference of Electronics, Communication and Aerospace Technology (ICECA)*, pp. 687-689, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[3] Thoraya Obaid et al., "Zigbee-Based Voice-Controlled Wireless Smart Home System," *International Journal of Wireless & Mobile Networks*, vol. 6, no. 1, pp. 47-59, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[4] George Alexakis et al., "Control of Smart Home Operations Using Natural Language Processing, Voice Recognition and IoT Technologies in a Multi-Tier Architecture," *Designs*, vol. 3, no. 3, pp. 1-18, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[5] Hairol Nizam Mohd. Shah et al., "Biometric Voice Recognition in Security System," *Indian Journal of Science and Technology*, vol. 7, no. 2, pp. 104-112, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[6] Ossama Abdel-Hamid et al., "Convolutional Neural Networks for Speech Recognition," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 10, pp. 1533-1545, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[7] Jung-Chun Liu et al., "An MFCC-Based Text-Independent Speaker Identification System for Access Control," *Concurrency and Computation: Practice and Experience*, vol. 30, no. 2, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[8] Zulfiqar Ali et al., "Vocal Fold Disorder Detection Based on Continuous Speech by Using MFCC and GMM," *2013 7th IEEE GCC Conference and Exhibition (GCC)*, pp. 292-297, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[9] Suma Paulose, Dominic Mathew, Abraham Thomas, "Performance Evaluation of Different Modeling Methods and Classifiers with MFCC and IHC Features for Speaker Recognition," *Procedia Computer Science*, vol. 115, pp. 55-62, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[10] Cemal Hanilçi, "Data Selection for i-Vector Based Automatic Speaker Verification Anti-Spoofing," *Digital Signal Processing*, vol. 72, pp. 171-180, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[11] Moataz El Ayad, Mohamed S. Kamel, and Fakhri Karray, "Survey on Speech Emotion Recognition: Features, Classification Schemes, and Databases," *Pattern Recognition*, vol. 44, pp. 572-587, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[12] Kunxia Wang et al., "Speech Emotion Recognition Using Fourier Parameters," *IEEE Transactions on Affective Computing*, vol. 6, no. 1, pp. 69-75, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[13] Yongjin Wang, and Ling Guan, "Recognizing Human Emotional State from Audiovisual Signals*," *IEEE Transactions on Multimedia*, vol. 10, no. 5, pp. 936-946, 2008. [CrossRef] [Google Scholar] [Publisher Link]

[14] Peipei Shen, Zhou Changjun, and Xiong Chen, "Automatic Speech Emotion Recognition Using Support Vector Machine," *Proceedings of 2011 International Conference on Electronic & Mechanical Engineering and Information Technology*, pp. 621-625, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[15] S. Sharanyaa, and M. Sambath, "Optimized Hybrid Model for Enhanced Parkinson's Disease Classification Using Feature Fused Voice Signal," *SSRG International Journal of Electronics and Communication Engineering*, vol. 10, no. 11, pp. 11-26, 2023. [CrossRef] [Publisher Link]

[16] Zhizheng Wu et al., "Synthetic Speech Detection Using Temporal Modulation Feature," *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 7234-7238, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[17] Jorge Martinez et al., "Speaker Recognition Using Mel Frequency Cepstral Coefficients (MFCC) and Vector Quantization (VQ) Techniques," *2012 22nd International Conference on Electrical Communications and Computers (CONIELECOMP)*, Puebla, Mexico, pp. 248-251, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[18] M. Shamim Hossain, Ghulam Muhammad, and Atif Alamri, "Smart Healthcare Monitoring: A Voice Pathology Detection Paradigm for Smart Cities," *Multimedia Systems*, vol. 25, pp. 565-575, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[19] Stefan Billeb et al., "Biometric Template Protection for Speaker Recognition Based on Universal Background Models," *IET Biometrics*, vol. 4, no. 2, pp. 116-126, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[20] Fang-Yie Leu, and Guan-Liang Lin, "An MFCC-Based Speaker Identification System," *2017 IEEE 31st International Conference on Advanced Information Networking and Applications (AINA)*, pp. 1055-1062, 2017. [CrossRef] [Google Scholar] [Publisher Link]

Appendix

Table 2. Prices of biometric devices for security in Peru

Product	Price	Characteristics
Samsung DP609	718.25USD	A fingerprint biometric reader capable of storing up to 100 distinct fingerprints.
Yale YM40	592 USD	A fingerprint biometric reader capable of storing up to 100 distinct fingerprints. It operates using 4 AA batteries.
FC3D	373.35 USD	Biometric fingerprint and facial recognition reader.
ZK ci-I700	282USD	A fingerprint biometric reader It operates using 4 AA batteries.