



An-Najah National University
Electrical and Telecommunications Engineering
Departments

GRADUATION PROJECT 1

"Your Whisper Key-Sound ID Unlock"

Submitted in partial fulfillment of the requirements
for Bachelor degree in Electrical and
Telecommunications Engineering

Supervisor: Dr. Falah Hasan.

Students Name:

Jana Nazem Hanini (12011091)

Noor AlDeen Moneer Mahameed (12010248)

Academic year: 2024/2025

❖ DEDICATION

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ "وَأَنْ لَيْسَ لِلْإِنْسَانِ إِلَّا مَا سَعَىٰ وَأَنَّ سَعْيَهُ سَوْفَ يُرَىٰ" صدق الله العظيم

لى هذا الوطن الحبيب...

لَكَ الْحُبُّ الْأَوَّلُ، وَالْإِنْتِمَاءُ الْأَبَدِيُّ، وَلَكَ هَذَا الْجُهْدُ ثَمَرَةٌ عِلْمٍ وَسَعْيٍ، نَصَعُهُ بَيْنَ يَدَيْكَ بِكُلِّ فُرْ وَوَلَاءٍ.

لى الأهل الأحرار...

أَنْتُمْ الدِّعَامَةُ الْأُولَى، وَالِدَعَاءُ الصَّادِقُ فِي كُلِّ حُلَّةٍ، هَذَا الْإِنجَازُ مِنْكُمْ وَإِيَّاكُمْ، فَأَنْتُمْ الْأَصْدُ وَالْأَمْتِدَادُ.

لى الأصدقاء ورفاق الذئب...

كُنْتُمْ النُّورَ فِي عَمَمَةِ الطَّرِيقِ، وَالْيَدَ الَّتِي مَسَكَتْ بِنَا حِينَ كِدْنَا أَنْ نَتَعَثَّرَ، شَكَرُوا لِقُلُوبِكُمْ الَّتِي كَانَتْ مَعَنَا.

لى الأساتذة الكرام...

مَنْ حَمَلُوا شُعْلَةَ الْعِلْمِ، وَمَشَوْا بِهَا نَحُونَا بِكُلِّ صَبْرٍ وَتَبَدُّلٍ، فَكُنْتُمْ لِحَرْفِ الْأَوَّلِ فِي كُلِّ إِنجَازٍ نَكْتَبُهُ الْيَوْمَ.

هَذَا الْعَمَلُ لَكُمْ يُحْتَمُّ بِأَسْمَائِنَا وَخَدَنَّا، بَلَى نَحْمَدُ بِصَمَاتٍ مِنْ رَحْمَتِنَا، وَمَنْ آمَنَ بِنَا.

لَكُمْ مِنْ كُلِّ الْإِمْتِنَانِ وَالْتَفَادِ،

المهندسة جنى ناظم حنيني

المهندس نور الدين منير محاميد

❖ **DISCLAIMER**

This report was written by students at the Electrical and Telecommunications Engineering Department, Faculty of Engineering, An-Najah National University. It has not been altered or corrected, other than editorial corrections, as a result of assessment and it may contain language as well as content errors. The views expressed in it together with any outcomes and recommendations are solely those of the students. An-Najah National University accepts no responsibility or liability for the consequences of this report being used for a purpose other than the purpose for which it was commissioned.

❖ ACKNOWLEDGMENT

- First and foremost, we thank God for the strength, patience, and success that have brought us to this point and enabled us to complete this project to the fullest.
- We extend our deepest gratitude to our dear families, who have been a key supporter every step of the way. Without them, this achievement would not have been possible.
- We express our gratitude to our supervisor, Dr. Falah Hassan, for his time, effort, and guidance.
- We also thank all our esteemed teachers for answering our questions whenever we needed them.
- Finally, to our dear friends, who have been our greatest support. Thank you for everything.

❖ TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION.....	4
1.1 Statement of the problem.....	4
1.2 Objectives of the work.....	4
1.3 Scope of the work.....	4
1.4 Significance or importance of your work.....	5
1.5 Organization of the report.....	5
CHAPTER 2: LITERATURE REVIEW	6
CHAPTER 3:Methodology.....	6
3.1 intelligent control systems	7
3.1.1 Definition of intelligent control systems:	7
3.1.2 Features of intelligent control systems:.....	7
3.2 Biometric systems.....	8
3.2.1 Definition of Biometric systems:.....	8
3.2.2 Types of Biometric Features:	8
3.2.3 Comparison of Biometric Systems: Fingerprint, Face, and Voice Recognition.....	8
3.2.4 Why Voice Recognition is the Best Option?	9
3.2.5 Speech Recognition vs Speaker Recognition:	10
3.2.6 Applications of Speech Recognition	10
3.2.7 Applications of Speaker Recognition:.....	11
3.2.8 Components of the Speaker Recognition System :	11
3.3 Formants.....	12
3.3.1 Definition of the Formants:.....	12
3.3.2 Basic Definition:	12
3.3.3 Acoustic & Anatomical Basis.....	12
3.3.4 The number and positions of formants depend on:.....	12
3.3.5 Key Formants (F1, F2, F3...).....	13
3.4 Mel-Frequency Cepstral Coefficients (MFCCs)	14
3.4.1 Definition of MFCC:.....	14
3.4.2 Concept Summary:	14

3.4.3 MFCC Extraction Steps:	14
3.4.4 MFCCs vs Formants	15
3.5 Speaker Recognition as a Biometric Identity System: Training and Testing Phases...	15
3.6 Gaussian Mixture Model (GMM)	17
3.6.1 Definition of GMM:	17
3.6.2 Advantages of GMM:	17
3.6.3 Parameters	17
3.6.4 The GMM work in voice Recognition	20
3.7 The pitch period	21
3.7.1 Definition of the pitch period:	21
3.7.2 Technical Definition:	21
3.7.3 The Importance of Pitch in Voice Recognition:	21
3.7.4 Approximate Pitch Period Values for Different Groups:	22
3.7.5 Explanation:	22
3.8 Noise Handling and Signal Processing Techniques	23
3.8.1 Band-Pass Filtering	23
3.8.2 Hamming Windowing	24
3.8.3 Signal Normalization	24
3.8.4 Summary Table of Noise Handling Techniques	25
Chapter 4: Results and Analysis	26
4.1 The principle of code operation	26
4.2 Code Result for an 11-year-old child	26
4.3 Code Result for Female	29
4.4 Code Result for Male	31
Chapter 5: Discussion	34
Chapter 6: Conclusions and Recommendation	35
6.1 Conclusions	35
6.2 Recommendation	36
6.3 Future works	36
Swot Analysis	37
References	38

❖LIST OF FIGURES

Figure 1-3: Speaker Recognition (Tranining and Testing Phases).	16
Figure 2-3: The GMM work in voice Recognition.	20
Figure 3-3: Speaker Recognition Process	25
Figure 3-4: The child's Original Audio in the three Attempts.	26
Figure 4-4: The child's Normalized Audio in the three Attempts.	27
Figure 5-4: The child's MFCC comparison across in the three Attempts.	27
Figure 6-4: The child's Pitch Period and Fundamental Frequency value in the three Attempts.	28
Figure 7-4: The child's Fomant Analysis Result in the three Attempts.	28
Figure 8-4: Female Original Audio in the three Attempts.	29
Figure 9-4: Female Normalized Audio in the three Attempts.	29
Figure 10-4: Female MFCC comparison across in the three Attempts.	30
Figure 11-4: Female pitch Period and Fundamental Frequency value in the three Attempts.	30
Figure 12-4: Female Formant Analysis Results in the three Attempts.	31
Figure 13-4: Male original Audio in the three Attempts.	31
Figure 14-4: Male Normalized Audio in the three Attempts.	32
Figure 15-4: Male MFCC comparison across in the three Attempts.	32
Figure 16-4: Male pitch Period and Fundamental Frequency value in the three Attempts.	33
Figure 17-4: Male Formant Analysis Results in the three Attempts.	33

❖ LIST OF TABLES

Table 1-3: Comparison of Biometric System.	8
Table 2-3: Speech Recognition vs Speaker Recognition.	10
Table 3-3: Key Formants.	13
Table 4-3: MFCCs vs Formants.	15
Table 5-3: Approximate Pitch Period Values for Different Groups.	22
Table 6-3: Summary of Noise Handling Techniques.	25

❖ LIST OF ABBREVIATIONS

MFCC	Mel Frequency Cepstral Coefficient
CNN	Convolutional Neural Network
RNN	Recursive Neural Networks
SVM	Support Vector Machine
IoT	Internet of Things
PID	Proportional Integral Derivative
3D	Three-Dimensional
ASR	Automatic Speech Recognition
NLP	Natural Language Processing
FFT	Fast Fourier Transform
DCT	Discrete Cosine Transform
ID	Identification
GMM	Gaussian Mixture Model
EM	Expectation-Maximization
MAP	Maximum A Posteriori
RAP	Relative Average Perturbation
APQ	Amplitude Perturbation Quotient
NHR	Noise-to-Harmonics Ratio
SPI	Soft Phonation Index
SNR	Signal to Noise Ratio

❖ ABSTRACT

This project follows an intelligent design for an access control system based on voice recognition of specific individuals, enhancing security and facilitating access by relying on voice prints instead of traditional passwords. The system analyzes the unique voice data of authorized individuals, which is then recorded in real time. The recording is then recorded, allowing the door to open immediately if a match is found.

The system boasts stellar voice recognition accuracy, even in noisy environments. It relies on advanced algorithms that are remarkably capable of distinguishing the authorized person's voice from a range of different voices, making it suitable for use in real-world situations requiring a high level of security.

The project reflects the integration of voice control technologies and embedded systems, highlighting the significant potential of voice recognition technologies in building smart solutions and controlling devices.

❖ CHAPTER 1: INTRODUCTION

1.1 Statement of the problem.

Traditional door security systems, such as physical keys, keypads, and access cards, are vulnerable to loss, theft, duplication, or unauthorized access. These methods also lack flexibility and may not provide adequate security in modern smart environments. Therefore, there is a need for a more intelligent and secure access control system that uses biometric features specifically, voice recognition to ensure that only authorized individuals can unlock and access secured areas. This project aims to address this issue by designing a voice-controlled door lock system that offers improved security, convenience, and reliability.

1.2 Objectives of the work.

The main objective of this project is to develop a secure and user-friendly voice-controlled door lock system that can identify and authenticate specific authorized users based on their voice. The detailed objectives include:

- Designing and implementing a voice recognition module capable of identifying pre-registered voices.
- Integrating the voice module with a microcontroller to control the door lock mechanism.
- Ensuring system reliability and accuracy under various environmental conditions.
- Enhancing user convenience by enabling hands-free, keyless access.

1.3 Scope of the work.

The scope of this project includes the design and implementation of a voice-controlled door locking system that recognizes specific, pre-recorded voice inputs. The system will be built using a microcontroller, a voice recognition module, and an electronic locking mechanism. The project will be limited to recognizing a limited number of authorized users, and the impact of environmental factors such as noise, echo, and polyphony will be tested within pre-defined limits.

1.4 Significance or importance of your work.

The proposed voice lock system provides a modern, secure, and contactless alternative to traditional entry methods such as keys, cards, and passwords. The project's significance lies in its reliance on voice recognition, a unique biometric feature that is difficult to replicate, enhancing security. The project also aligns with the growing demand for smart and automated systems in homes and workplaces. Additionally, this system offers a practical and easy to use solution for people with disabilities who may encounter difficulty with traditional locks. By reducing physical interaction and improving access control, the system contributes to a safer and smarter living environment.

1.5 Organization of the report.

This report is organized into six main chapters to provide a clear and systematic presentation of the project.

Chapter 1 introduces the project by outlining the problem statement, objectives, scope, significance, and the structure of the report.

Chapter 2 presents a comprehensive literature review, discussing previous work related to voice recognition, biometric systems, and the technologies used.

Chapter 3 explains the methodology, including the system design, biometric techniques, feature extraction processes, and noise handling methods.

Chapter 4 details the results and analysis, showcasing the system's performance with different users and discussing the extracted features.

Chapter 5 provides a discussion of the key findings, challenges faced, and the effectiveness of the proposed system.

Chapter 6 concludes the report with a summary of the outcomes, recommendations for improvement, and suggestions for future work. Additionally, a **SWOT analysis** is included to evaluate the strengths, weaknesses, opportunities, and threats of the project. Finally, **References** are provided to acknowledge the sources used throughout the research.

❖ CHAPTER 2: LITERATURE REVIEW

voice recognition technology has gained significant attention in recent years due to its growing applications in security, smart systems, and human-machine interaction. Researchers have continuously sought to improve the accuracy, reliability, and noise tolerance of voice-based systems, particularly for access control. Traditional security methods, such as physical keys, PIN codes, and fingerprint sensors, have shown limitations in terms of convenience, vulnerability to theft, and environmental constraints. As a result, biometric systems, especially those based on voice recognition, have emerged as a promising alternative due to their non-contact nature and user-friendliness.

Previous studies have demonstrated that voice recognition can serve as a reliable biometric identifier when combined with appropriate signal processing techniques. For example, the use of Mel-Frequency Cepstral Coefficients (MFCCs) and Gaussian Mixture Models (GMMs) has been widely adopted to enhance the precision of speaker identification. However, early implementations struggled with noise interference and variations in voice due to environmental factors or user behavior. To address these issues, modern approaches integrate intelligent control systems and machine learning algorithms, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), to improve system adaptability and performance.

Moreover, the comparison between different biometric modalities such as fingerprint, facial, and voice recognition reveals that voice recognition offers unique advantages, particularly in accessibility for people with disabilities and its compatibility with hands-free smart environments. Researchers have also distinguished between speech recognition, which focuses on understanding spoken commands, and speaker recognition, which identifies the person speaking. The latter is the core focus of this project.

This project builds upon these studies by designing a practical, real-time, and speaker-specific voice-controlled door lock system, incorporating advanced noise handling, signal processing, and speaker verification techniques to ensure secure and efficient access control in both personal and professional settings.

❖ CHAPTER 3: METHODOLOGY

3.1 intelligent control systems

3.1.1 Definition of intelligent control systems:

An intelligent system uses artificial intelligence and intelligent learning control techniques without direct human intervention. Unlike traditional control systems (such as PID), which rely on mathematical models, intelligent control systems learn from data and adapt to the surrounding environment, even if they are subject to data uncertainty or inaccuracy.

3.1.2 Features of intelligent control systems:

- ✓ **Adaptiveness:** The ability to modify the system's behavior based on changes in the environment or inputs.
- ✓ **Flexibility:** The ability to handle nonlinear and complex systems that are difficult to model mathematically.
- ✓ **Control under uncertainty:** The ability to operate even in the presence of noise or incomplete data.
- ✓ **Self-learning:** Improves performance over time by learning from past experiences.

3.2 Biometric systems

3.2.1 Definition of Biometric systems:

are a common commercial or behavioral technology used to identify people or verify their original identity. This process can be used for periods such as access to protected areas, smart devices, and even system centers.

3.2.2 Types of Biometric Features:

- ✓ Physical: Fingerprint, Face, Iris, DNA.
- ✓ Behavioral: Voiceprint, Signature, Gait, Keyboarding Style.

3.2.3 Comparison of Biometric Systems: Fingerprint, Face, and Voice Recognition.

Criterion	Fingerprint Recognition	Face Recognition	Voice Recognition
Accuracy	Very high	High, but affected by lighting and angle	Good, improving with modern technologies
Reliability	Reliable, but affected by cuts or dirt	Moderate, can be affected by facial changes	Moderate to high with voice filtering
Ease of Use	Requires touching a sensor	Requires camera alignment	No touch or alignment needed – just speak
Privacy	Can be faked using printed fingerprints	Can be spoofed with 3D images	More secure using unique voice frequencies
Accessibility	Limited for certain physical disabilities	Limited for visually impaired	Very suitable for people with disabilities
System Integration	Needs dedicated sensor hardware	Needs high-quality camera	Can use standard microphones
Cost	Medium	Relatively high	Low – no special hardware required
Environmental Adaptability	Affected by sweat or dust	Affected by lighting or shadows	Works in dark and varied conditions

Table 1-3: Comparison of Biometric System.

3.2.4 Why Voice Recognition is the Best Option?

- ✓ **No Specialized Hardware Required:** Voice systems work with standard microphones in phones or computers.
- ✓ **User-Friendly:** The user simply needs to speak – no physical contact or precise positioning needed.
- ✓ **Strong Privacy:** Can use techniques like frequency matching or spoken passphrases to enhance security.
- ✓ **Highly Flexible:** Easy to integrate into smart systems, such as smart locks, voice assistants, and more.
- ✓ **Ideal for Smart Environments:** Perfect for hands-free access in smart homes, offices, and public systems.
- ✓ **Upgradable:** Can be improved using machine learning and signal processing.

3.2.5 Speech Recognition vs Speaker Recognition:

Feature	Speech Recognition	Speaker Recognition
Also Called	Automatic Speech Recognition (ASR)	Voice Biometrics or Speaker Identification
Purpose	Understand what is being said	Identify or verify who is speaking
Focus	Converts spoken words into text	Matches voice patterns to a specific person
Main Output	Text (e.g., "Open the door")	Identity (e.g., "This is Jana speaking")
Technology	Natural Language Processing (NLP), speech-to-text models	Machine learning on vocal features (tone, pitch.)
Sensitive to	Language, accent, grammar	Voice characteristics, emotion, background noise
Used With	Commands, dictation, transcription	Authentication and security systems

Table 2-3: Speech Recognition vs Speaker Recognition.

3.2.6 Applications of Speech Recognition

1. **Virtual Assistants:** Siri, Alexa, Google Assistant.
2. **Dictation & Transcription:** Turning spoken lectures or medical notes into text.
3. **Voice Interfaces:** Smart TVs, cars, or appliances responding to voice commands.
4. **Real-time Translation:** Translating spoken language into another in real time.
5. **Voice-Controlled Apps:** Hands-free texting or controlling apps while driving.

3.2.7 Applications of Speaker Recognition:

1. Security & Authentication

- Voice-based login to banking apps or secure systems.
- Smart locks that open only to specific people's voices.

2. Forensics & Law Enforcement

- Identifying speakers in recorded phone calls or surveillance.

3. Personalized Services

- In smart homes: the assistant recognizes who is speaking and adjusts preferences (music, lighting, temperature).
- In cars: auto-sets driving profiles (seat, mirrors) based on the driver's voice.

4. Call Centers & Customer Support

- Voice ID to verify customers instead of security questions.
- Fraud prevention in financial services.

3.2.8 Components of the speaker Recognition System:

1. Front-end processing - the "signal processing" part, which converts the sampled speech signal into set of feature vectors, which characterize the properties of speech that can separate different speakers. Frontend processing is performed both in training and testing phases.

2. Speaker modeling - this part performs a reduction of feature data by modeling the distributions of the feature vectors.

3. Speaker database - the speaker models are stored here.

4. Decision logic - makes the final decision about the identity of the speaker by comparing unknown feature vectors to all models in the database and selecting the best matching model.

3.3 Formants

3.3.1 Definition of the Formants:

Formants are the resonant frequencies of the human vocal tract. They appear as peaks in the speech spectrum and are critical in shaping the distinct sounds of vowels and some consonants.

3.3.2 Basic Definition:

A formant is a frequency band where acoustic energy is concentrated due to resonance in the vocal tract during speech production.

3.3.3 Acoustic & Anatomical Basis

1. The lungs push air through the vocal folds, which may vibrate to produce a glottal sound.
2. The sound then travels through the vocal tract (throat, mouth, nasal cavity).
3. The shape of the vocal tract selectively amplifies certain frequencies—these are formants.

3.3.4 The number and positions of formants depend on:

- Vocal tract length and shape
- Tongue and lip position
- Jaw opening
- Nasal cavity involvement

3.3.5 Key Formants (F1, F2, F3...)

Formant	Frequency Range (Typical Male Voice)	Acoustic Meaning
F1	300–900 Hz	Inversely related to tongue height. Low F1 = high tongue.
F2	850–2500 Hz	Related to tongue frontness High F2 = front tongue.
F3	1800–3000 Hz	Influenced by lip rounding and tongue tip position.
F4	3000+ Hz	Less often used, but helpful in speaker identification.

Table 3-3: Key Formants.

3.4 Mel-Frequency Cepstral Coefficients (MFCCs)

3.4.1 Definition of MFCC:

The human speech contains numerous discriminative features that can be used to identify speakers. Speech contains significant energy from zero frequency up to around 5 kHz. The objective of automatic speaker recognition is to extract, characterize and recognize the information about speaker identity. The property of speech signal changes markedly as a function of time. To study the spectral properties of speech, signal the concept of time varying Fourier representation is used. However, the temporal properties of speech signal such, as energy, zero crossing, correlation are assumed constant over a short period. That is its characteristics are short-time stationary. Therefore, using hamming window, Speech signal is divided into a number of blocks of short duration so that normal Fourier transform can be used.

3.4.2 Concept Summary:

- ✓ **Mel-frequency:** Mimics how humans perceive pitch (nonlinear frequency scale).
- ✓ **Cepstral:** Represents the rate of change in the spectral envelope (the shape of the spectrum).
- ✓ **Coefficients:** The actual numbers that describe the shape of the spectrum.

3.4.3 MFCC Extraction Steps:

- ✓ Convert the audio signal to the frequency spectrum using FFT.
- ✓ Map the frequency bands to the Mel scale (human auditory scale).
- ✓ Apply the logarithm to compress large variations.
- ✓ Use the Discrete Cosine Transform (DCT) to generate the final coefficients.

3.4.4 MFCCs vs Formants

Feature	MFCCs (Mel-Frequency Cepstral Coefficients)	Formants (F1, F2, F3...)
What They Represent	Overall spectral envelope of the speech signal	Resonant frequencies of the vocal tract
Derived From	Log Mel-spectrum → Discrete Cosine Transform (DCT)	Peaks in the speech spectrum
Biological Basis	Indirect: reflects vocal tract via spectrum shape	Direct: tied to anatomical vocal tract properties
Interpretability	Not directly interpretable	Highly interpretable (linked to articulation)
Used In	ASR, speaker ID, emotion recognition	Forensics, phonetics, speech pathology
Robustness	More robust to noise, distortion, and channel effects	Sensitive to recording quality and noise
Language Dependency	Generally, language-independent	More influenced by phoneme/language characteristics
Feature Dimensions	Typically, 13–39 (with deltas and acceleration)	Usually 3–5 (F1–F5)
Extraction Tools	Easy: Librosa, Kaldi, Python, MATLAB	Requires spectral analysis, LPC, or Praat
Application Speed	Fast, low computational cost	Slower, sometimes requires manual verification

Table 4-3: MFCCs vs Formants.

3.5 Speaker Recognition as a Biometric Identity System: Training and Testing Phases.

Anatomical structure of the vocal tract is unique for every person and hence the voice information available in the speech signal can be used to identify the speaker. Recognizing a person by her/his voice is known as speaker recognition. Since differences in the anatomical structure are an intrinsic property of the speaker, voice comes under the category of biometric identity. Using voice for identity has several advantages. One of the major advantages is remote person authentication. Like any other pattern recognition systems, speaker recognition systems also involve two phases namely, training and testing. Training is the process of familiarizing the system with the voice characteristics of the speakers registering. Testing is the actual recognition task. The block diagram of training phase is shown in Figure 1-3. Feature vectors representing the voice characteristics of the speaker are extracted from the training utterances and are used for building the reference models. During testing, similar feature vectors are extracted from the test utterance, and the degree of their match with the reference is obtained using some matching technique. The level of match is used to arrive at the decision. The block diagram of the testing phase is given in Figure 1-3.

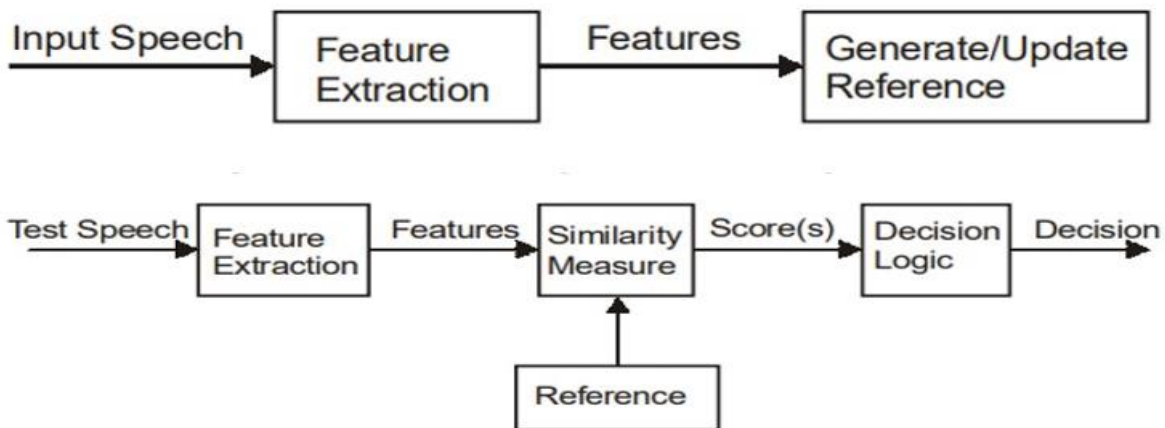


Figure 1-3: Speaker Recognition (Training and Testing Phases).

3.6 Gaussian Mixture Model (GMM)

3.6.1 Definition of GMM:

is a parametric probability density function represented as a weighted sum of Gaussian component densities. GMMs are commonly used as a parametric model of the probability distribution of continuous measurements or features in a biometric system, such as vocal-tract related spectral features in a speaker recognition system. GMM parameters are estimated from training data using the iterative Expectation-Maximization (EM) algorithm or Maximum a Posteriori (MAP) estimation from a well-trained prior model.

3.6.2 Advantages of GMM:

- ✓ Simple and easy to implement.
- ✓ Very effective for tasks that do not depend on text content (text-independent).
- ✓ Does not require massive data growth or high computational power.
- ✓ Very suitable for embedded systems such as the Arduino or Raspberry Pi.

3.6.3 Parameters:

- 1- **Jitter:** is relative evaluation of the period-to-period (very short-term) variability of the pitch within the analyzed voice sample.

Voice break areas are excluded.

$$\text{Jitt} = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |T_o^{(i)} - T_o^{(i+1)}|}{\frac{1}{N} \sum_{i=1}^N T_o^{(i)}} \quad (1)$$

where $T_o^{(i)}$, $i = 1, 2 \dots N$ is the extracted pitch period data and $N = PER$ is the number of extracted pitch periods.

2- RAP: (Relative Average Perturbation) is the relative evaluation of the period-to-period variability of the pitch within the analyzed voice sample with smoothing factor of 3 periods.

Voice break areas are excluded.

$$\text{RAP} = \frac{\frac{1}{N-2} \sum_{i=2}^{N-1} \left| \frac{To^{(i-1)} + To^{(i)} + To^{(i+1)}}{3} - To^{(i)} \right|}{\frac{1}{N} \sum_{i=1}^N To^{(i)}} \quad (2)$$

where $To^{(i)}$, $i = 1, 2 \dots N$ is the extracted pitch period data and $N = PER$ is the number of extracted pitch periods.

3- Shimmer: Shimmer Percent [%] is relative evaluation of the period-to-period (very short term) variability of the peak-to-peak amplitude within the analyzed voice sample.

Voice break areas are excluded.

$$\text{Shim} = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |A^{(i)} - A^{(i+1)}|}{\frac{1}{N} \sum_{i=1}^N A^{(i)}} \quad (3)$$

where $A^{(i)}$, $i = 1, 2 \dots N$ is the extracted peak-to-peak amplitude data and $N = PER$ is the number of extracted impulses.

- 4- APQ:** (Amplitude Perturbation Quotient) /%/ is relative evaluation of the period-to-period variability of the peak-to- peak amplitude within the analyzed voice sample at smoothing of 11 periods.

Voice break areas are excluded.

$$APQ = \frac{\frac{1}{N-4} \sum_{i=1}^{N-4} \left| \frac{1}{5} \sum_{r=0}^4 A^{(i+r)} - A^{(i+2)} \right|}{\frac{1}{N} \sum_{i=1}^N A^{(i)}} \quad (4)$$

where $A^{(i)}$, $i = 1, 2 \dots N$ is the extracted peak-to-peak amplitude data and $N = PER$ is the number of extracted impulses.

- 5- NHR:** (Noise-to-Harmonic Ratio) is the average ratio of the inharmonic spectral energy the frequency range 1500-4500 Hz to the harmonic spectral energy in the frequency range 70- 4500 Hz. This is a general evaluation of noise present in the analyzed signal.
- 6- SPI:** (Soft Phonation Index) is the average ratio of the lower- frequency harmonic energy in the range 70-1600 Hz to the higher-frequency harmonic energy in the range 1600-4500 Hz.

3.6.4 The GMM work in voice Recognition:

1. Training

- Take voice samples from each person (e.g., MFCC).
- Create a GMM model for them, i.e., a system that "learns" the characteristics of the voice in terms of frequency, pitch, and speed.

2. Validation or Recognition (Testing)

- When a person becomes a person, their voice requests are obtained (e.g., MFCC).
- The probability (probability) of this decision is calculated for each existing GMM.
- A person is accepted if their probability of belonging to the GMM is the highest (and exceeds a specified threshold).

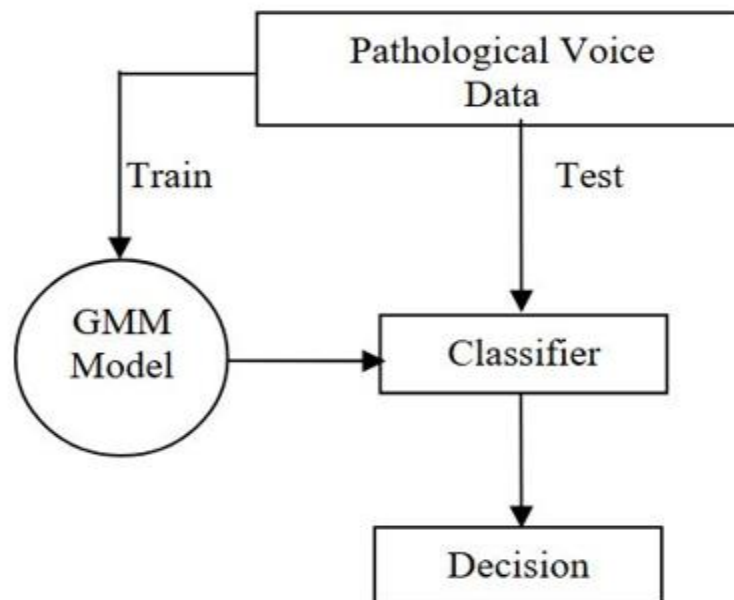


Figure 2-3: The GMM work in voice Recognition.

3.7 The pitch period

3.7.1 Definition of the pitch period:

Pitch period is the time interval between two consecutive repetitions of a sound wave produced by the vocal cords (often when pronouncing a vowel or tone), it determines the pitch of a speaker.

3.7.2 Technical Definition:

It is the time interval (in seconds or samples) between two consecutive peaks in a periodic audio signal produced by the vibration of the vocal cords.

3.7.3 The Importance of Pitch in Voice Recognition:

➤ Speaker recognition:

Each person has a distinct pitch. Men typically have a fundamental frequency between 85–180 Hz, and women between 165–255 Hz.

➤ Feature extraction:

Pitch period is used in some biometric systems as an additional feature with MFCC or LPC.

➤ Signal classification:

It helps distinguish between:

- **Voiced signals:** have a clear pitch period (such as vowels).
- **Unvoiced signals:** do not have a regular pitch (such as f and s).

3.7.4 Approximate Pitch Period Values for Different Groups:

Group	Typical Fundamental Frequency (F_0)	Pitch Period (T_0) = $1 / F_0$
Males	85 – 180 Hz	5.6 – 11.7 milliseconds
Females	165 – 255 Hz	3.9 – 6.1 milliseconds
Children	250 – 400 Hz	2.5 – 4 milliseconds
Elderly	100 – 200 Hz	5 – 10 milliseconds

Table 5-3: Approximate Pitch Period Values for Different Groups.

3.7.5 Explanation:

1. Males:

- Have longer and thicker vocal cords → vibrate slower → lower frequency → longer pitch period.

2. Females:

- Shorter and thinner vocal cords → vibrate faster → higher frequency → shorter pitch period.

3. Children:

- Very small vocal folds → very high frequency → shortest pitch period.
- Their voices tend to be higher-pitched.

4. Elderly:

- Changes in muscles and nerves affect vocal folds.
- Frequency often decreases in males and slightly increases in females, reducing the gap between genders.

3.8 Noise Handling and Signal Processing Techniques

In real-world environments, voice recognition systems are often exposed to various sources of signal distortion such as background noise (side conversations, device sounds, traffic, etc.), variations in voice intensity, and changes in microphone sensitivity. To minimize the impact of these factors and improve system performance, a set of noise handling and signal processing techniques were applied.

3.8.1 Band-Pass Filtering

3.8.1.1 What is Band-Pass Filtering?

Band-pass filtering is a technique that allows frequencies within a specific range to pass through while attenuating frequencies outside this range.

3.8.1.2 Why is it Used?

- ✓ The critical frequency range for human speech typically lies between 300 Hz and 3400 Hz.
- ✓ Frequencies below 300 Hz often represent noise such as vibrations or air conditioning sounds.
- ✓ Frequencies above 3400 Hz are typically sharp noises or electrical interference.

3.8.1.3 Benefits:

- ✓ Isolates the relevant speech signals.
- ✓ Improves the signal-to-noise ratio (SNR).
- ✓ Makes the system more accurate and less sensitive to unwanted sounds.

3.8.2 Hamming Windowing

3.8.2.1 What is Hamming Windowing?

It is a mathematical function applied to small segments (frames) of the audio signal to reduce distortion when converting the signal to the frequency domain.

3.8.2.2 Why is it Used?

- ✓ The speech signal changes rapidly and cannot be accurately analyzed as a long continuous signal.
- ✓ The signal is divided into short-time frames (20 to 30 milliseconds), which can be considered stationary.
- ✓ Without windowing, abrupt changes at the frame edges cause "spectral leakage" during Fourier Transform analysis.

3.8.2.3 Benefits:

- ✓ Minimizes edge discontinuities between frames.
- ✓ Provides more accurate spectral analysis for feature extraction.

3.8.3 Signal Normalization

3.8.3.1 What is Signal Normalization?

Signal normalization is the process of adjusting the amplitude of the signal so that all samples have a consistent energy level.

3.8.3.2 Why is it Used?

- ✓ Speakers may talk closer or farther from the microphone.
- ✓ Speakers may talk loudly or softly.

3.8.3.3 Benefits:

- ✓ Ensures fair comparison across all samples regardless of voice intensity.
- ✓ Improves the performance of the recognition algorithm by focusing on the unique features of the voice rather than its volume.

3.8.4 Summary Table of Noise Handling Techniques

Technique	Purpose	Benefit
Band-Pass Filtering	Isolate human speech frequencies.	Reduces external noise.
Hamming Windowing	Prevent spectral leakage.	Accurate frequency analysis.
Signal Normalization	Standardize signal amplitude.	Fair comparison across samples.
MFCC + GMM	Extract and classify voice features.	Improves recognition accuracy.

Table 6-3: Summary of Noise Handling Techniques.

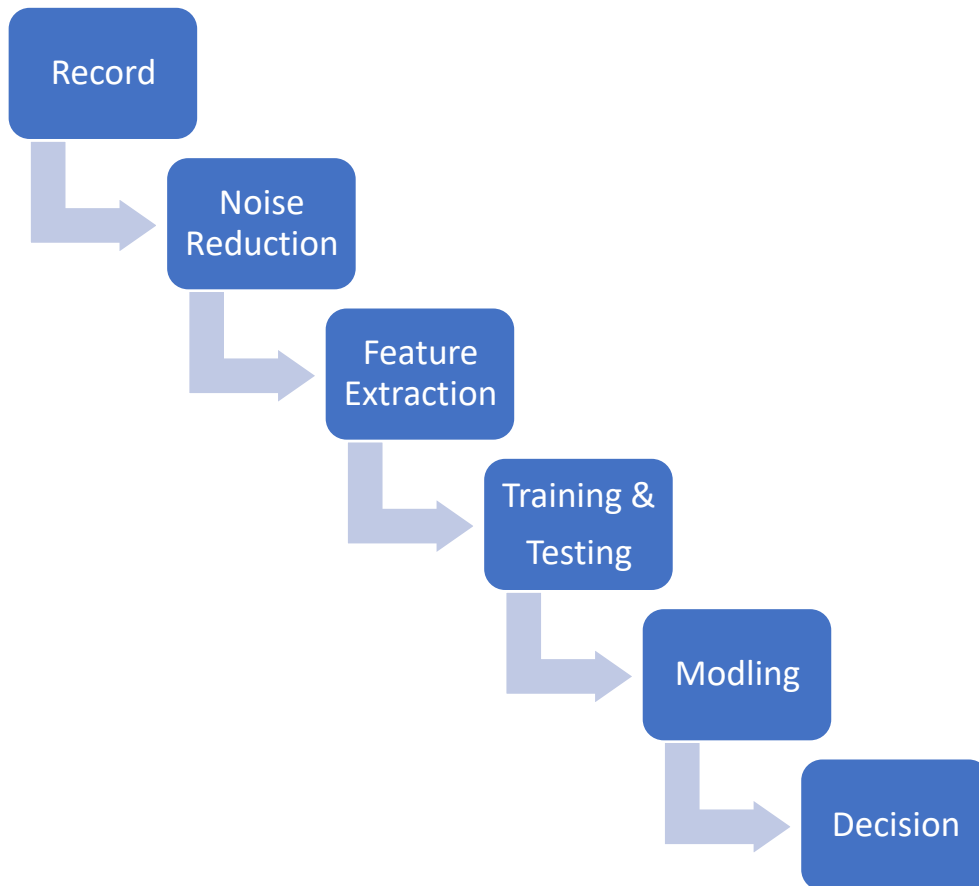


Figure 3-3: Speaker Recognition Process.

❖ Chapter 4: Results and Analysis

4.1 The principle of Code Operation

The code's function is to implement what was explained previously in the report regarding the topic of feature extraction. The code records 3 attempts of the speaker, then purifies them as much as possible from noise by raising the sound Amplitude for high-value sounds, and reducing the Amplitude of the sounds with low values. Then it performs extraction operations for the mentioned values (MFCC, PP, FF), prints their values, and draws curves representing them using ready-made libraries.

4.2 Code Result for an 11-year-old child

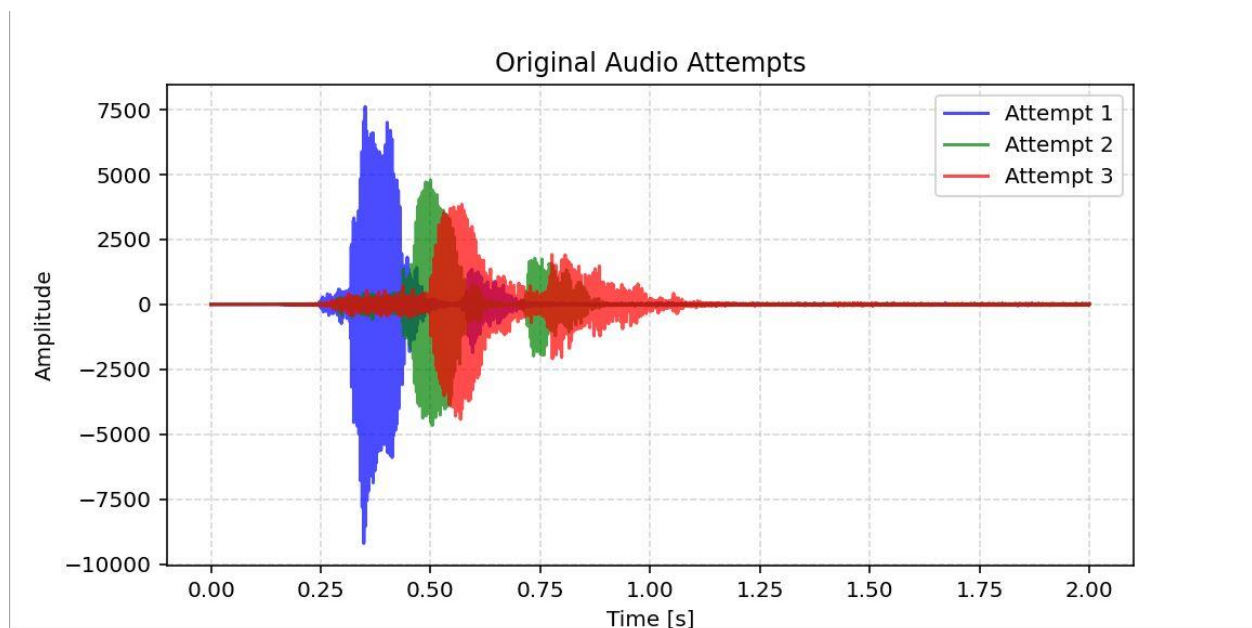


Figure 3-4: The child's Original Audio in the three Attempts.

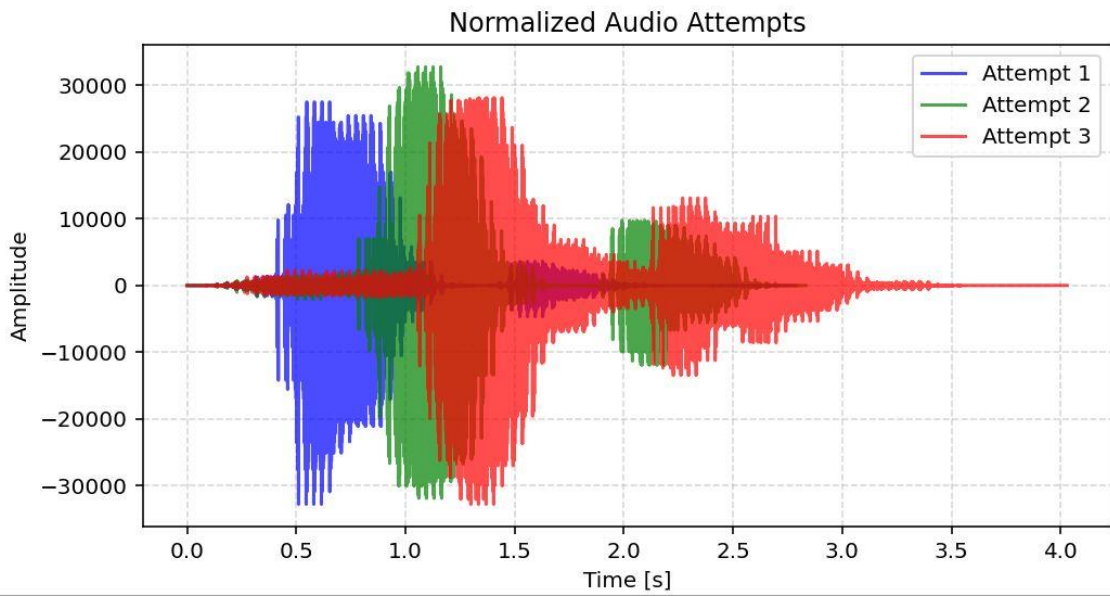


Figure 4-4: The child's Normalized Audio in the three Attempts.

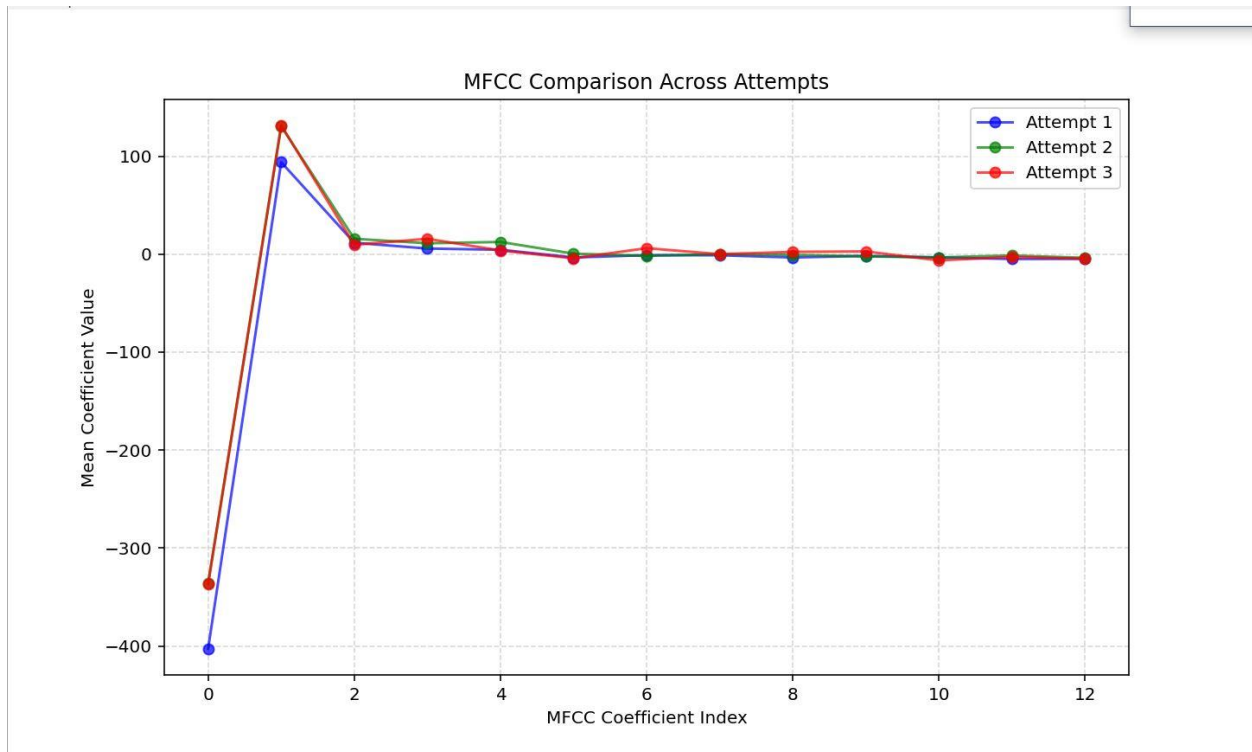
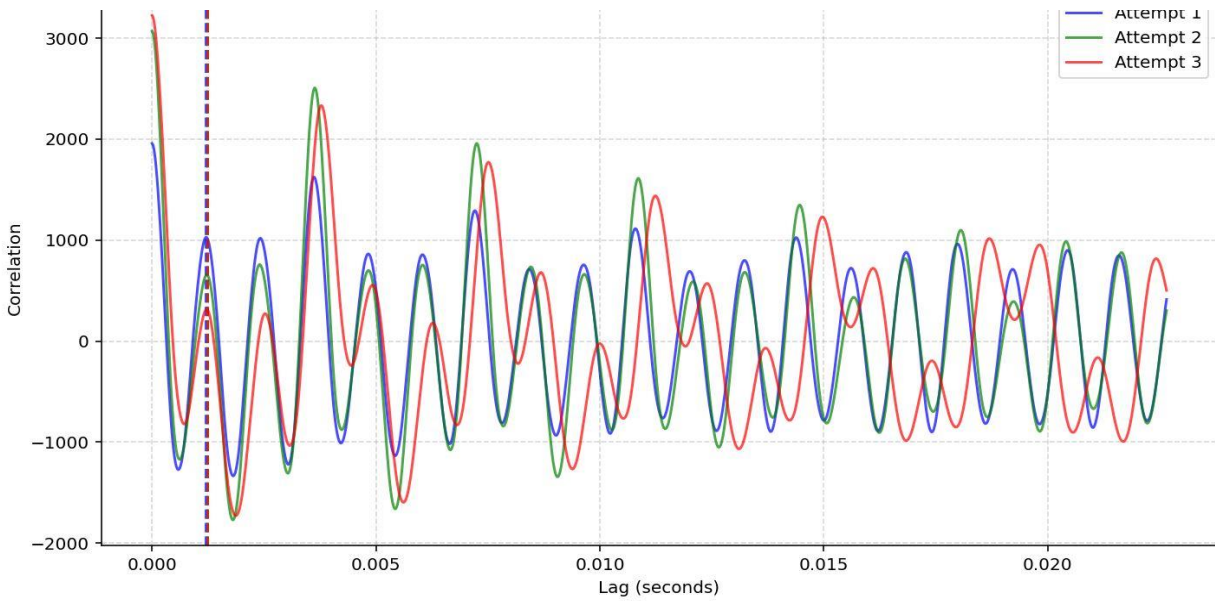


Figure 5-4: The child's MFCC comparison across in the three Attempts.



Pitch Analysis Results:

- Attempt 1:
 - Pitch Period: 0.0012 s
 - Fundamental Freq: 832.08 Hz
- Attempt 2:
 - Pitch Period: 0.0012 s
 - Fundamental Freq: 816.67 Hz
- Attempt 3:
 - Pitch Period: 0.0012 s
 - Fundamental Freq: 816.67 Hz

Figure 6-4: The child's Pitch Period and Fundamental Frequency value in the three Attempts.

Time (s)	
Formant Analysis Results:	
Attempt 1 Median Formants:	
F1:	503.9 Hz
F2:	1430.8 Hz
F3:	2247.2 Hz
F4:	3235.4 Hz
Attempt 2 Median Formants:	
F1:	472.9 Hz
F2:	1399.1 Hz
F3:	2238.4 Hz
F4:	3167.8 Hz
Attempt 3 Median Formants:	
F1:	502.5 Hz
F2:	1262.9 Hz
F3:	2170.9 Hz
F4:	2998.0 Hz

Figure 7-4: The child's Fomant Analysis Result in the three Attempts.

4.3 Code Result for Female

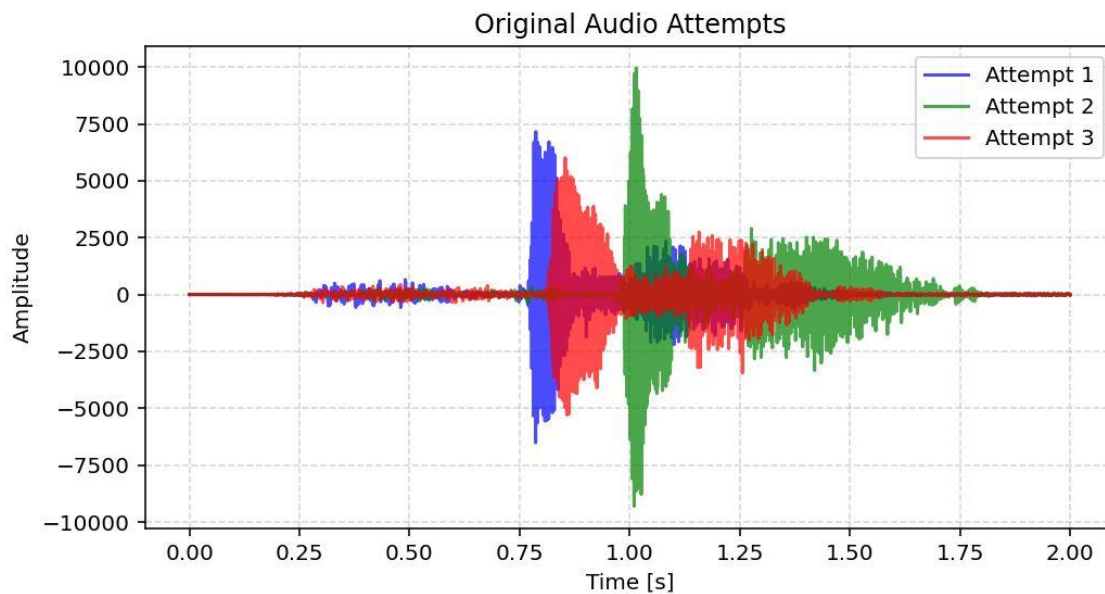


Figure 8-4: Female Original Audio in the three Attempts.

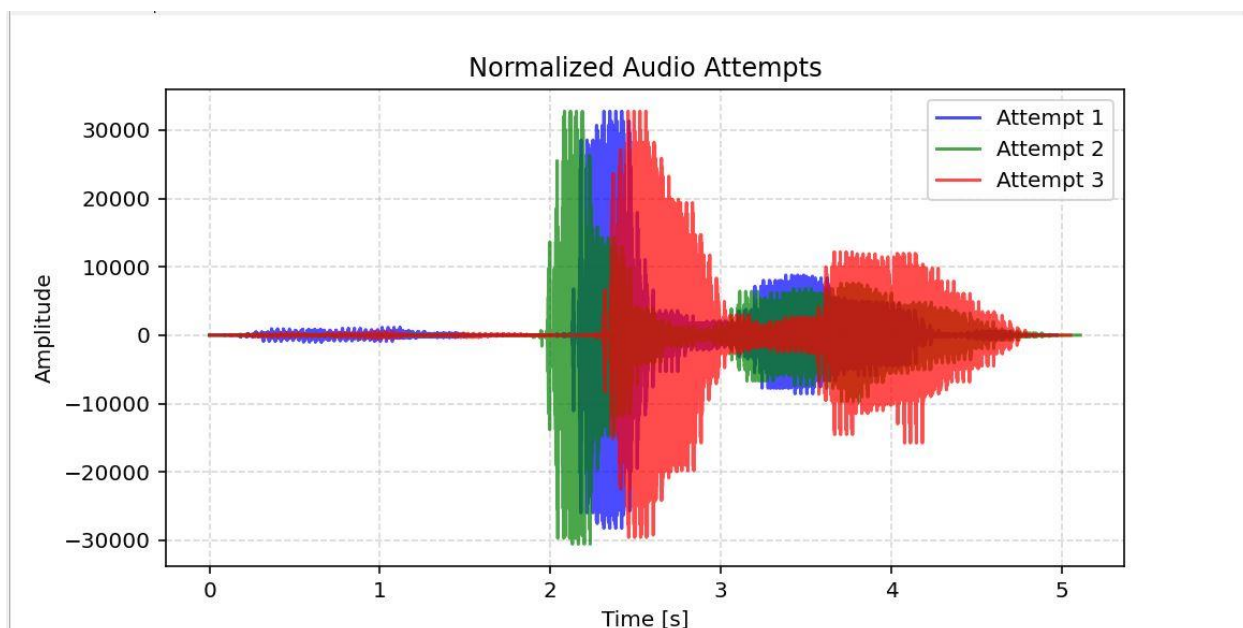


Figure 9-4: Female Normalized Audio in the three Attempts.

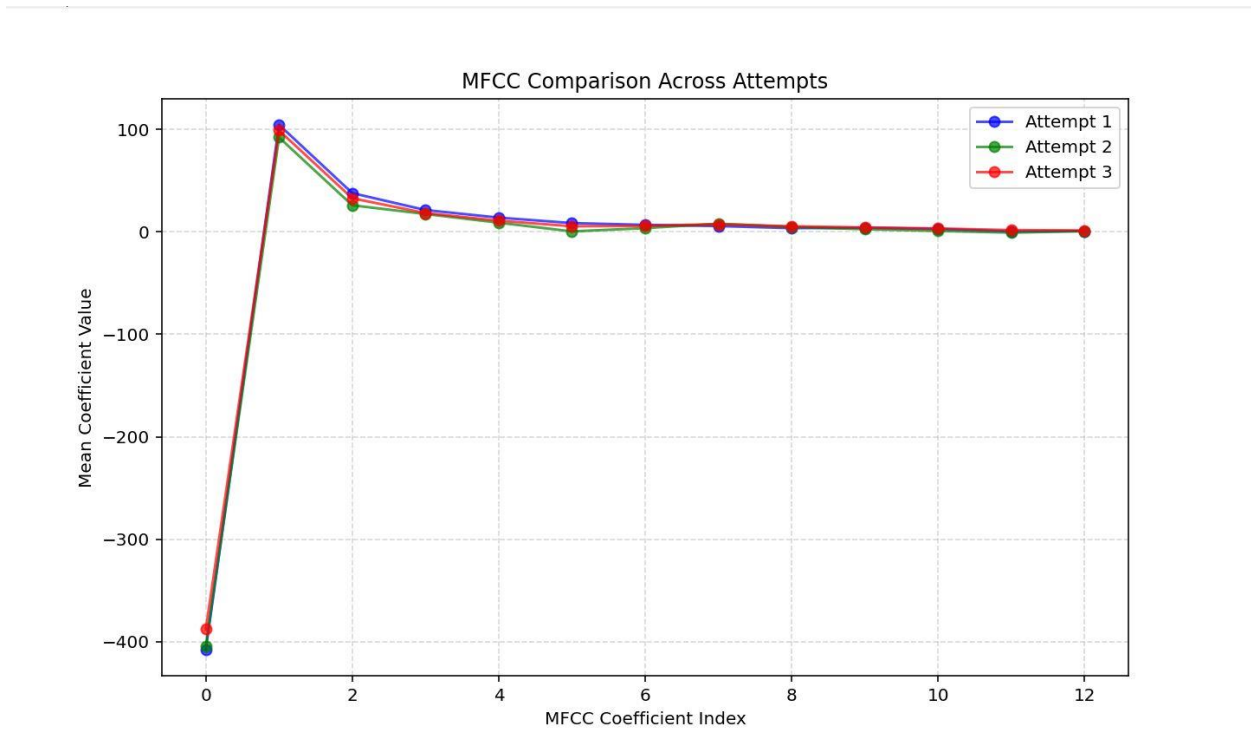


Figure 10-4: Female MFCC comparison across in the three Attempts.

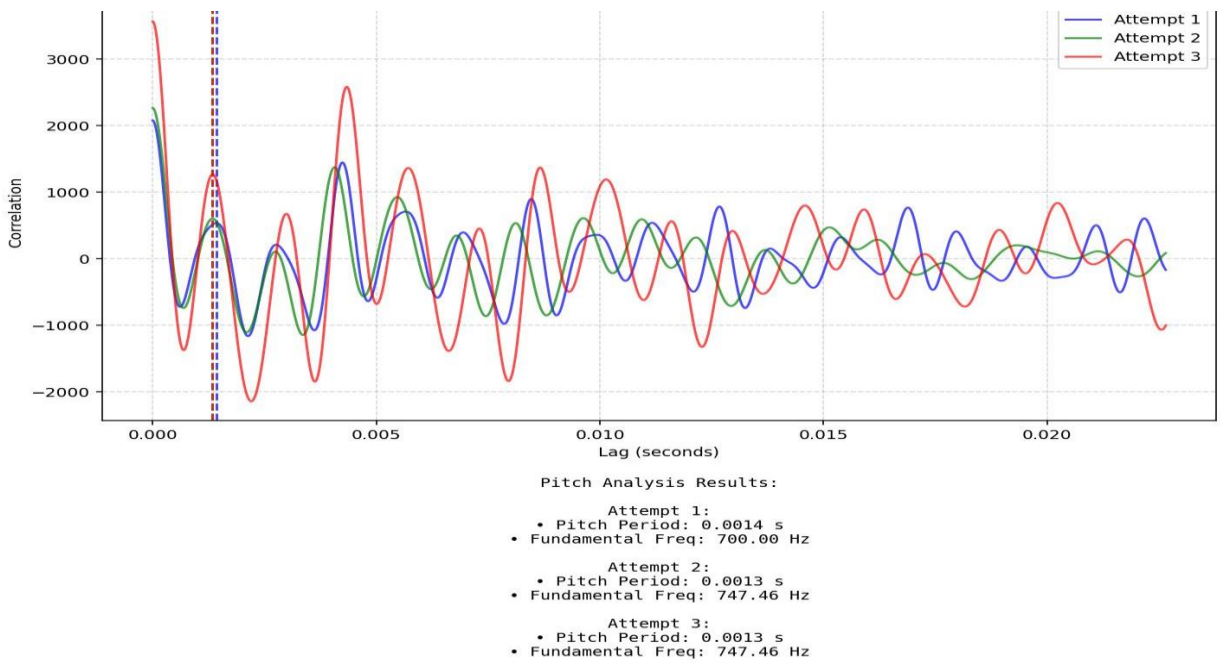


Figure 11-4: Female pitch Period and Fundamental Frequency value in the three Attempts.

```

Time (s)
Formant Analysis Results:

Attempt 1 Median Formants:
F1: 266.0 Hz
F2: 1341.1 Hz
F3: 2243.8 Hz
F4: 3084.0 Hz

Attempt 2 Median Formants:
F1: 300.9 Hz
F2: 1298.7 Hz
F3: 2247.7 Hz
F4: 3122.5 Hz

Attempt 3 Median Formants:
F1: 275.1 Hz
F2: 1366.0 Hz
F3: 2220.2 Hz
F4: 3088.7 Hz

```

Figure 12-4: Female Formant Analysis Results in the three Attempts.

4.4 Code Result for Male

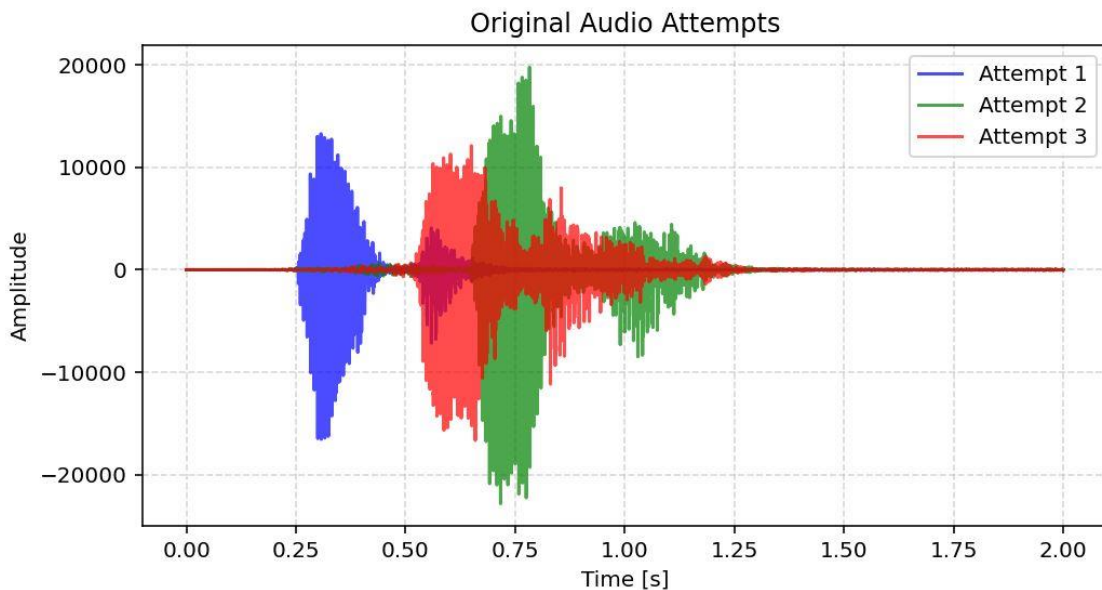


Figure 13-4: Male Original Audio in the three Attempts.

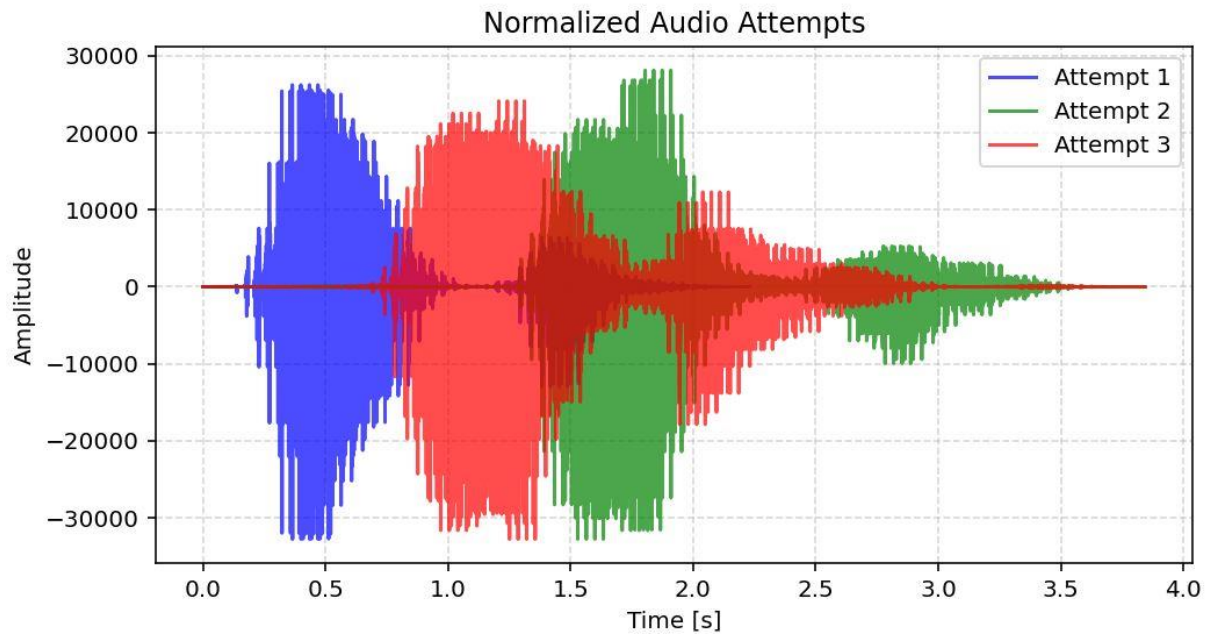


Figure 14-4: Male Normalized Audio in the three Attempts.

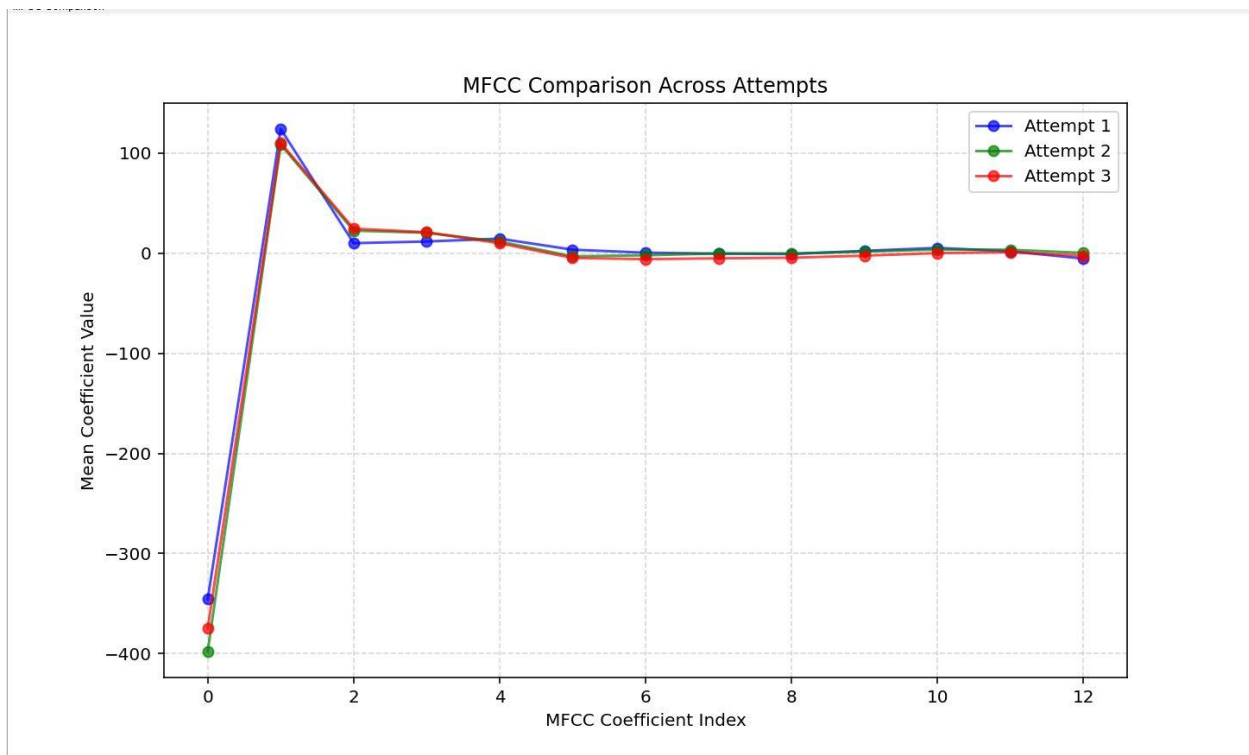


Figure 15-4: Male MFCC comparison across in the three Attempts.

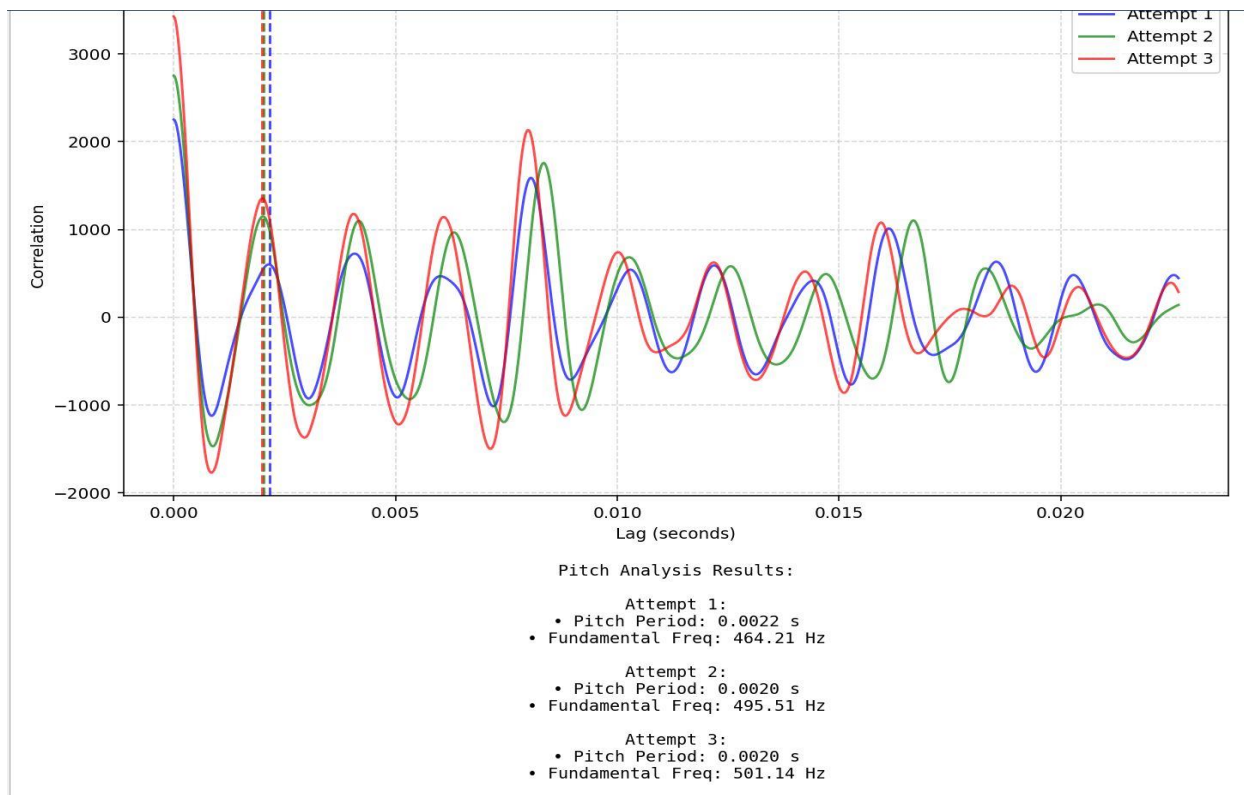


Figure 16-4: Male pitch Period and Fundamental Frequency value in the three Attempts.

Time (s)

Formant Analysis Results:

Attempt 1 Median Formants:

F1: 504.3 Hz
 F2: 1524.8 Hz
 F3: 2247.1 Hz
 F4: 3280.2 Hz

Attempt 2 Median Formants:

F1: 540.8 Hz
 F2: 1404.6 Hz
 F3: 2492.9 Hz
 F4: 3409.3 Hz

Attempt 3 Median Formants:

F1: 496.8 Hz
 F2: 1309.6 Hz
 F3: 2352.7 Hz
 F4: 3386.9 Hz

Figure 17-4: Male Formant Analysis Results in the three Attempts.

❖ Chapter 5: Discussion

The results obtained from the system's practical implementation and testing provide clear evidence of the effectiveness of the proposed voice-controlled door lock system. By analyzing the performance across different user groups — a child, a female, and a male — the system successfully demonstrated its ability to extract reliable voice features such as Mel-Frequency Cepstral Coefficients (MFCC), Pitch Period (PP), and Formants, which are critical for accurate speaker identification.

One key observation from the results is the consistency of the extracted features within the same speaker across multiple attempts. This confirms the system's stability and accuracy in recognizing the correct user, even when slight variations in voice tone, speed, or distance from the microphone occur. The figures presented in Chapter 4 clearly show that the MFCC patterns, pitch periods, and format distributions are unique to each user and remain stable across their three attempts.

The system also showed strong capability in noise handling. Techniques such as band-pass filtering effectively isolated the human voice frequency range (300 Hz to 3400 Hz), while Hamming windowing minimized spectral leakage, and signal normalization reduced the influence of varying voice amplitudes. These methods collectively improved the system's robustness, as the results remained valid even when the recordings included moderate background noise.

Another important discussion point is the system's ability to distinguish between the typical pitch periods and forms of different user categories. For instance, the pitch period for the child user was consistently shorter than that of the male and female users, which aligns with the theoretical pitch values explained in Chapter 3. This strengthens the reliability of the biometric identification process.

However, despite these positive outcomes, the system faced some limitations in very high noise environments where competing sounds could partially affect the recognition process. Also, the current version of the system is limited in its capacity to store a large number of authorized users, which could become a scalability issue in more complex applications.

In summary, the system proved to be a practical and reliable voice-controlled access solution, with satisfactory performance across different genders and age groups. It demonstrated robustness in handling noise and voice variations.

❖Chapter 6: Conclusions and Recommendation

6.1 Conclusions

This project successfully achieved its main objective of designing and implementing a voice-controlled door lock system capable of identifying and authenticating specific authorized users based on their unique voice features. By utilizing advanced techniques such as Mel-Frequency Cepstral Coefficients (MFCC), pitch period analysis, and formant extraction, the system demonstrated a high level of accuracy in recognizing different speakers and distinguishing them from unauthorized individuals.

The system's ability to process and adapt to various environmental conditions, supported by noise-handling methods like band-pass filtering, Hamming windowing, and signal normalization, further enhanced its reliability in real-world scenarios. The results confirmed that voice recognition can serve as a practical and secure alternative to traditional security methods such as physical keys and password-based systems.

Additionally, the project highlights the advantages of using voice as a biometric feature, especially in terms of convenience, accessibility, and low-cost implementation, as no special hardware is required. The system is particularly beneficial for people with disabilities and in smart environments where touchless access is essential.

In conclusion, the project provides a solid foundation for the development of intelligent, voice-controlled access systems and proves the viability of integrating biometric voice recognition in everyday security applications.

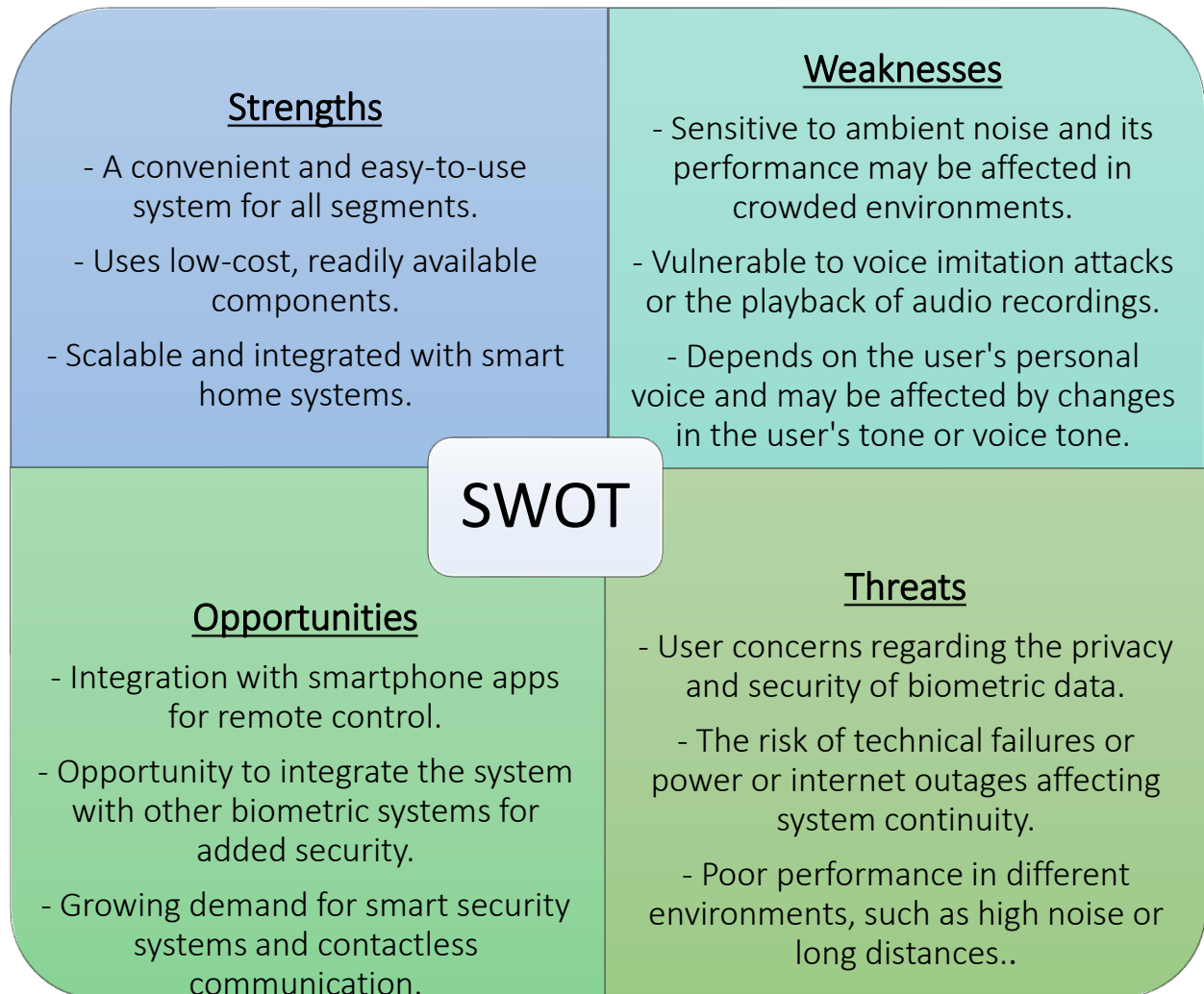
6.2 Recommendation

It is recommended to improve the system by focusing on increasing its resistance to various environmental challenges, particularly high levels of background noise and echo, by applying more sophisticated filtering and noise suppression algorithms. Additionally, expanding the voice database to support a larger number of users while maintaining quick response time would make the system more scalable and practical for real-life applications. The integration of real-time processing units is also encouraged to ensure faster decision-making and smoother system performance. To further enhance security, it is advisable to develop a multi-factor authentication system that combines voice recognition with other biometric techniques. Developing an intuitive user training module would also help users provide clearer and more consistent voice samples during registration. Finally, it is recommended to design a mobile application that offers users the ability to remotely control and monitor the system, providing greater convenience and flexibility.

6.3 Future works

In the future, this system can be further developed by platforms to enable remote control, real-time monitoring, and access management through mobile applications. Expanding the system to support multilingual voice commands and adaptive learning would also enhance its flexibility and usability in diverse environments. Moreover, future improvements could include the development of a more sophisticated noise cancellation module to allow the system to perform reliably in highly crowded or industrial areas. Additionally, integrating artificial intelligence and deep learning techniques, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), could improve the system's ability to differentiate between similar voices and handle more complex voice variations. Ultimately, future versions of this system may target commercial, residential, and institutional applications, offering an even more secure and intelligent access control solution.

❖ Swot Analysis



❖References

1. <https://ieeexplore.ieee.org/abstract/document/8743482>
2. <https://jontallen.ece.illinois.edu/uploads/537.F18/Book/main-all.pdf>
3. <https://citeseerx.ist.psu.edu/document?doi=b4e9c14c67b8aa431a40041cce0a3564144e1a2a&repid=rep1&type=pdf>
4. https://www.ijirset.com/upload/2018/may/8_Speaker.pdf
5. <https://new.eurasip.org/Proceedings/Ext/SPECOM2006/papers/021.pdf>
6. https://web.ece.ucsb.edu/Faculty/Rabiner/ece259/Reprints/016_parallel%20processing%20pitch%20detector.pdf
7. <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=7965c68bedb3f7a7d566a1a1bb1d7a9e72c5a46a>
8. <https://assta.org/proceedings/sst/2006/sst2006-84.pdf>
9. <https://jjem.jnnce.ac.in/journals/SP-2/JJEMSP0231.pdf>
10. <https://www.akademiabaru.com/submit/index.php/ard/article/view/6137>