ISSN 2277-2685
IJESR/Jan-Mar. 2026/ Vol-16/Issue-1/587-593
Charla Satwika Raghavi *et. al.*, / International Journal of Engineering & Science Research
Full Length Article

Gender Voice Recognition Using Matlab

Ms G Srilakshmi[1], Charla Satwika Raghavi[2], Bakaram Spoorthy Reddy[3], Perala Varshitha Rao[4]
[1]Associate Professor; Department Of Electronics And Communication Engineering Bhoj Reddy Engineering College For Women Hyderabad India.
[2,3,4]B.Tech Students; Department Of Electronics And Communication Engineering Bhoj Reddy Engineering College For Women Hyderabad India.
Mail Id; sathwikaraghavi2629@gmail.com[2], Spoorthyreddy1914@gmail.com[3], varshitharaoperala@gmail.com[4]

Accepted 30-03-2026

Abstract
Gender identification from speech signals is a key aspect of modern speech processing and human–computer interaction systems. Accurate detection of a speaker's gender can improve system efficiency by enabling optimized processing and personalized responses. Traditional methods for gender classification primarily rely on extracting acoustic features followed by machine learning-based classification.This paper presents a gender voice recognition approach that utilizes both statistical and spectral features derived from speech signals. Important features such as mean, zero-crossing rate, standard deviation, amplitude, and a set of significant feature vectors are extracted from each audio sample. These features are then used to train and evaluate multiple classification models, including Random Forest, K-Nearest Neighbors (KNN), Logistic Regression, Decision Tree, and Convolutional Neural Networks (CNNs).A comparative analysis of these models is conducted using performance metrics such as accuracy and precision. Experimental results indicate that the CNN-based model outperforms traditional classifiers, achieving superior precision and overall classification performance. The system is implemented using MATLAB, leveraging its capabilities in signal processing and machine learning.The proposed approach demonstrates an effective and scalable solution for automatic gender classification from speech, with potential applications in voice assistants, biometric systems, and intelligent communication platforms.

Keywords—
Gender Recognition, Speech Processing, Machine Learning, Convolutional Neural Networks (CNN), MATLAB, Feature Extraction, Zero Crossing Rate, Voice Classification

Introduction

Gender identification represents a fundamental component of speech recognition and automatic interacting sound responding systems. Identifying the voice gender minimizes the computational loads of these systems for additional processing. Standard approaches for gender estimation from the speech have broadly relied on the extraction of speech features and classification tasks. This paper proposes a technique for gender identification of speech samples using the speech recognition process. The proposed technique extracts essential voice features like Mean, Zero-Crossing, Standard Deviation, and Amplitude, as well as 12 most significant features from every voice sample, and combines them to create voice feature vectors. The proposed technique uses several machine and deep learning methods such as Random Forest, KNN, Logistic Regression, Decision Tree, and CNNs, in order to classify the voice vectors into Male and Female classes. After comparing the evaluation metrics results of all classifiers, the proposed technique finds out that the CNN model is the best classifier used to classify the voice vectors with a higher precision value of 1.0. This approach aims to develop a system that can recognize the gender of a speaker based on their voice characteristics, using MATLAB as the development platform.Processing of digital signal by using digital computer is called as Digital Signal Processing.Speech signal is used to communicate among people. It does not only consists of the information but also carries the information regarding the particular speaker. From which the speaker is male or female can be recognized.The meaning of Gender Recognition (GR) is recognizing the gender of the person whether the speaker is male or female. The Information about gender, age, ethnicity, and emotional state are the important ingredients that give rich behavioural information.Such information can be obtained from the speech signal.In this paper, an unknown speaker is

compared to a database of some known speakers. The best matching system is taken as the recognition decision. From the Recognition decision we conclude whether the given voice sample is generated by a male or female.

Literature Survey

Gender classification using speech signals has been a subject of research for several decades, especially in the fields of speech processing, biometrics, and machine learning. This chapter presents a summary of relevant studies and methods that have contributed to the development of gender recognition systems, with an emphasis on approaches that can be implemented using MATLAB. initial work in speaker gender identification relied on pitch-based features, primarily the fundamental frequency (F0). Studies showed that male voices generally fall in the range of 85–180 Hz, while female voices range from 165–255 Hz. However, pitch alone was found to be insufficient for reliable classification due to overlap in frequency ranges and variability in speech conditions.

Review of Existing Works

John R. Deller, John G Proakis and John H. L. Hansen, "Discrete- Time Processing of Speech Signals" Macmillan Publishing company, 866 Third avenue, New York 10022,the extracted fundamental frequency is used for finding auto correlation function of speech signal for identifying the gender. The reported gender recognition accuracy is 90 to 95%. But the fundamental frequency (F0) estimation may be difficult in case of noisy environment. In this situation more complex methods can be proposed to improve F0 estimation.

Huang X., Acero, A., and Hon, H.-W. "Spoken Language Processing: a Guideto Theory, Algorithm and System Development" prentice-Hall, New Jersey, 2001. Features are extracted for voiced vowels where the distinction of male/female is most significant & accuracy is better. This technique does not propose the method for extraction of segments corresponding to selected vowels from the continuous speech audio stream.

Rabiner Lawrence, Juang Bing-Hwang, **"**Fundamentals of Speech Recognitions", Prentice Hall New Jersey, 1993, ISBN 0- 13-015157-2. In Lab VIEW technique, identification of the gender and removing gender specific components, higher compression rate can be achieved. Here the information is enhanced to save the Bandwidth. This method does not extract the vowels from speech. The value is obtained for formant1 were not completely correct as they were obtained by processing all the samples of speech. Hence it is difficult to identify the gender.

Md. Saidur Rahman, "Small Vocabulary Speech Recognition in Bangla Language", M.Sc. Thesis, Dept. of Computer Science & Engineering, Islamic University, Kushtia-7003, July-2004.Linear Predictive Coding (LPC) method used for Feature extraction. A well chosen feature can result in quality recognition. PDA based on average magnitude difference function has relatively low computational cost and it is easy to implement. In which wrongly chosen feature can result in poor recognition

Problem Statement

In recent years, the proliferation of voice-controlled systems and speech-based human-computer interaction has necessitated more advanced and intelligent methods for understanding and interpreting human speech. One critical aspect of voice analysis is the ability to automatically recognize the gender of a speaker based solely on audio input. Gender voice recognition plays a significant role in various applications such as call routing in telecommunication systems, enhancing speech recognition accuracy, personalizing user experiences in smart assistants, aiding forensic investigations, and improving interaction with robots and virtual agents.

Given these challenges, there is a need to design and implement a reliable, accurate, and computationally efficient gender voice recognition system. MATLAB, with its powerful signal processing capabilities and machine learning toolboxes, provides an ideal platform for such a system. The project proposes to develop a system using MATLAB that can accurately determine the gender of a speaker from a short voice recording. The system will involve several stages including preprocessing, feature extraction (using methods such as pitch analysis, Mel Frequency Cepstral Coefficients (MFCC), formants, energy, etc.), and classification using algorithms like Support Vector Machines (SVM), K-Nearest Neighbors (KNN), or neural networks.

By solving these problems, the proposed gender voice recognition system can contribute to the development of more intelligent, responsive, and human-aware technologies. Furthermore, this project will provide insight into the practical application of digital signal processing and machine learning techniques in the domain of speech and audio processing.

Software Requirements

The development of a Gender Voice Recognition system requires a well-defined set of software tools and environments to ensure efficient implementation, testing, and evaluation. This section outlines the software requirements necessary to build and run the system using MATLAB. These tools support essential functions such as signal processing, feature extraction,

machine learning, and audio handling.

MATLAB is chosen as the primary development platform due to its user-friendly interface and powerful built-in toolboxes specifically designed for audio and signal analysis. The toolboxes used in this project provide prebuilt functions that simplify complex operations like pitch detection, MFCC extraction, and classification modeling.

This section details the core development environment, supporting toolboxes, file format compatibility, and any additional software needed to successfully execute the system. Choosing the right software environment is critical to ensuring the reliability, scalability, and accuracy of the gender voice recognition process.

Software Requirements

The successful implementation of the gender voice recognition system depends on a well-defined software environment that supports audio processing, feature extraction, and machine learning. This project utilizes MATLAB (version R2021a or later) as the primary development platform due to its comprehensive capabilities in signal processing and data analysis. MATLAB provides an integrated environment with user-friendly interfaces and extensive libraries, making it highly suitable for developing speech-based classification systems.

To facilitate different stages of system development, several MATLAB toolboxes are employed. The Signal Processing Toolbox is used for filtering audio signals, analyzing pitch, and extracting time-frequency domain features essential for distinguishing voice characteristics. The Audio Toolbox supports audio input/output operations, waveform analysis, and voice activity detection, enabling efficient handling of speech data. For classification tasks, the Statistics and Machine Learning Toolbox is utilized to train and evaluate models such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and decision trees. Additionally, the Deep Learning Toolbox can be incorporated when implementing advanced neural network-based classification techniques. For deployment purposes, the MATLAB Compiler allows the developed application to be converted into a standalone executable, making it accessible without requiring a full MATLAB installation.The system is designed to operate across multiple operating systems, including Windows, macOS, and Linux, ensuring flexibility in development and deployment. Among these, Windows 10 and Windows 11 are commonly preferred due to their compatibility and widespread use in academic and industrial environments.

System Architecture

The architecture of the proposed system follows a structured pipeline that processes voice input through multiple stages to achieve accurate gender classification. The process begins with user voice input, which is captured and passed to the preprocessing stage. During preprocessing, noise reduction, amplitude normalization, and silence removal are performed using functions available in the Audio Toolbox. This step ensures that the input signal is clean and consistent for further analysis.

Following preprocessing, the system performs feature extraction using techniques such as Mel-Frequency Cepstral Coefficients (MFCC), pitch estimation, and formant analysis. These features are extracted using the Signal Processing Toolbox and represent key acoustic properties that differentiate male and female voices.Finally, the output is presented to the user through a MATLAB-based interface, either in the form of a graphical user interface (GUI) or console output. This structured workflow ensures efficient processing, accurate classification, and ease of use.

The system architecture can be summarized as a sequential pipeline consisting of voice input, preprocessing, feature extraction, classification, and output generation. Each stage is supported by specific MATLAB toolboxes, ensuring modularity and maintainability of the system.

Methodology

This chapter presents the methodology used to design and implement the gender voice recognition system. The goal of this system is to accurately classify the gender of a speaker based on voice input using signal processing and machine learning techniques in MATLAB.The methodology outlines a step-by-step process starting from data collection and preprocessing, followed by feature extraction, model training, and finally classification and evaluation. Each phase plays a critical role in ensuring the accuracy and reliability of the system. Special attention is given to selecting the most relevant voice features such as pitch and Mel-Frequency Cepstral Coefficients (MFCCs), which have been proven effective in distinguishing male and female voices.

This structured approach enables the system to process raw audio signals, extract meaningful information, and apply suitable algorithms for gender classification. The use of MATLAB as the development platform facilitates the integration of signal processing and machine learning within a single environment, making the development efficient and streamlined.The methodology chapter outlines the systematic approach adopted to develop a gender voice recognition system capable of that supports the design, development, and testing of the proposed system using MATLAB.Voice-based gender recognition is a multi-stage process that requires a combination of digital signal processing and machine learning techniques. The methodology begins

with the collection of voice datasets, which are then subjected to preprocessing to eliminate noise and standardize the input. Following this, the system performs feature extraction, where critical audio features such as pitch, formants, and Mel-Frequency Cepstral Coefficients (MFCCs) are computed. These features are known to reflect physiological differences between male and female speech patterns and are widely used in speech analysis.

Flow Chart & Explanation

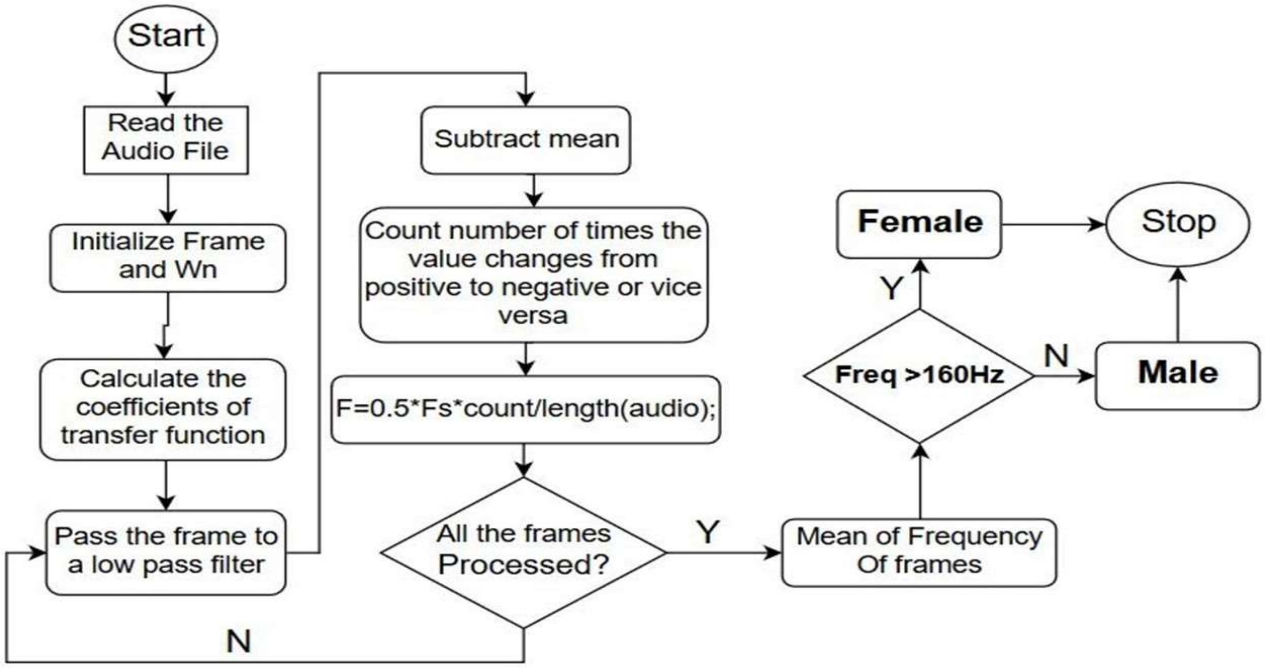

Figure 1 Block Diagram

Working Methodology

The proposed gender voice recognition system employs a structured signal processing approach combined with simple classification techniques to identify the gender of a speaker based on voice input. The process begins with reading the input audio file, where the speech signal is loaded into the system and divided into smaller segments known as frames. Frame-based processing enables the system to analyze short portions of the signal independently, improving accuracy in feature extraction.

Each frame is then multiplied by a window function, such as a Hamming window, to reduce discontinuities at the edges and minimize spectral distortion. This step ensures that the signal remains smooth and suitable for further processing. After windowing, the signal may undergo filtering to isolate the relevant frequency range associated with human speech, particularly the pitch region. Low-pass filtering is often applied to suppress high-frequency noise and enhance the fundamental frequency component.

To further refine the signal, the mean value of each frame is subtracted to eliminate any DC offset. This ensures that the waveform oscillates symmetrically around zero, which is crucial for accurate detection of zero-crossings. The system then calculates the Zero Crossing Rate (ZCR), which represents the number of times the signal changes its sign within a frame. Since each complete waveform cycle crosses zero twice, the ZCR provides an approximation of the signal's frequency.

The final classification is performed using a threshold-based approach. If the average pitch exceeds approximately 160 Hz, the voice is classified as female; otherwise, it is classified as male. This threshold is based on typical pitch ranges, where male voices generally lie between 85–155 Hz and female voices range from 165–255 Hz. Although simple, this method provides a computationally efficient way to distinguish between genders under controlled conditions.

In addition to pitch-based analysis, an energy-based feature extraction method is also incorporated to enhance classification reliability. In this approach, voice samples from both male and female speakers are first recorded and stored as reference data. The energy

of these signals is computed by applying the Fast Fourier Transform (FFT) to obtain the power spectrum, from which the signal energy is derived. This energy value serves as a reference threshold.

For an unknown input voice, the same process is applied: the signal undergoes FFT, the power spectrum is calculated, and the energy is extracted. The resulting energy value is then compared with the predefined threshold to assist in classification. This combination of pitch and energy features improves the robustness of the system.

Despite its efficiency, the proposed method has certain limitations. The zero-crossing approach performs well for clean and simple signals but may produce inaccurate results in the presence of noise, harmonics, or overlapping frequencies. Environmental noise, variations in accents, and multiple speakers can further affect performance. Nevertheless, the system remains effective for basic gender classification tasks and serves as a foundation for more advanced models.

The block-diagram-based design of the gender voice recognition system effectively structured the end-to-end process of input to output classification. Beginning with audio acquisition (either real-time or file-based), the system transitions through critical stages: **s**ampling, preprocessing, feature extraction, and classification.

During the preprocessing step, noise filtering and signal framing ensure the captured signal is clean and structured. Next, key features—such as pitch (F0), MFCCs, formant frequencies, spectral flux, and zero-crossing rates—are extracted, encapsulating both time-domain and frequency-domain characteristics.

Following feature extraction, the classification engine—whether built using statistical models like GMM,SVM, or newer architectures such as Bi-LSTM—analyzes these acoustic vectors to categorize the speaker by gender. Deep learning systems like Bi-LSTM have shown high accuracy (up to ~99%) across different noise and signal-to-noise conditions.

This structured pipeline not only simplifies module-wise development but also enhances system flexibility, interpretability, and performance tuning—for instance, replacing SVM with neural classifiers or refining preprocessing methods for specific environments. The modular design also facilitates integration with broader systems like ASR pipelines or speaker verification frameworks, ensuring gender detection can guide downstream processes.

Results

Our MATLAB-based gender voice recognition system employs signal processing techniques to differentiate between male and female voices. The process involves:

Preprocessing: Utilizing a low-pass filter to isolate the voice signal, enhancing clarity by removing background noise.

Feature Extraction: Applying the Zero-Crossing Rate (ZCR) method to determine the fundamental frequency, which is typically lower in male voices (85–180 Hz) and higher in female voices (155–255 Hz). Comparing the extracted frequency to a predefined threshold (e.g., 160 Hz) to classify the voice as male or female.This approach is implemented using MATLAB's Signal Processing Toolbox and Simulink for real-time processing.

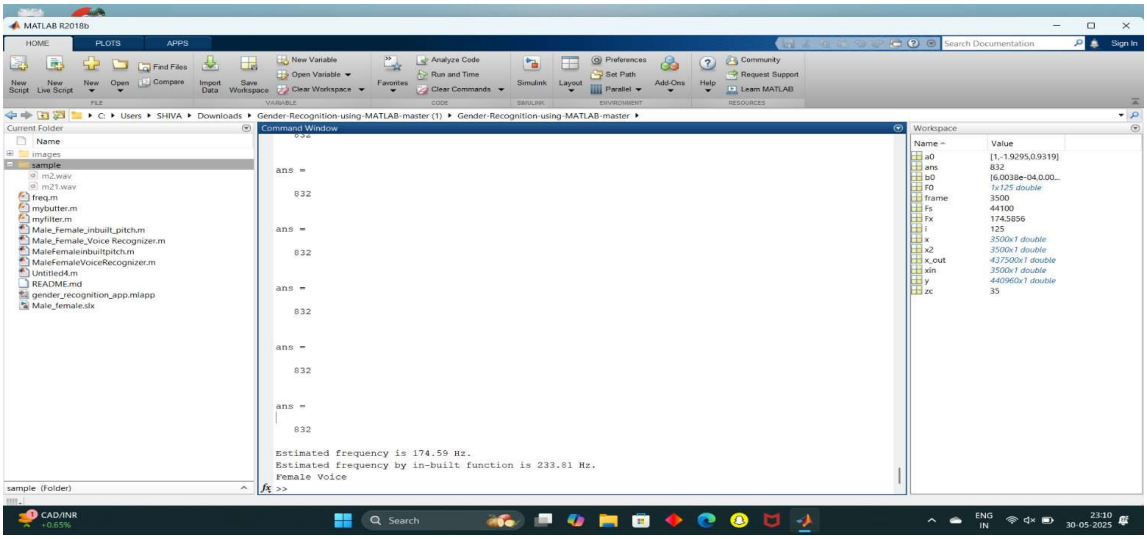

Figure 2; female voice output

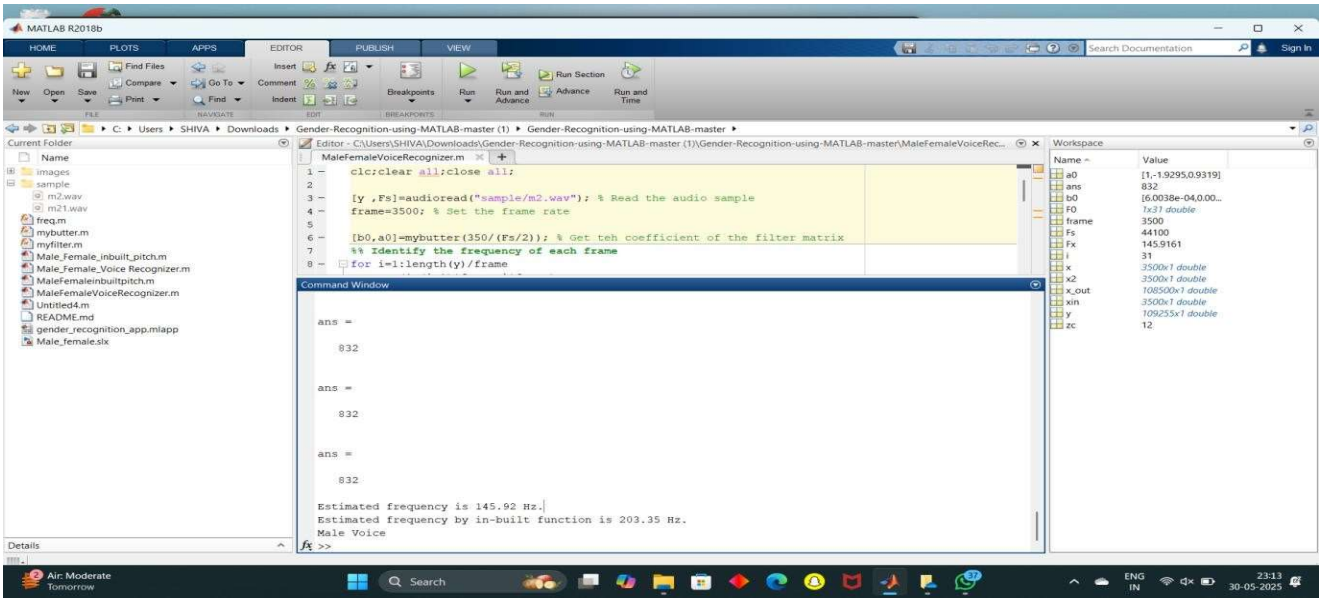

Figure 3; male voice output

This project demonstrates the ability to classify gender based on the fundamental frequency extracted from audio signals. While the method is effective, challenges arise due to the overlap in frequency ranges between male and female voices. For instance, some male voices may have higher fundamental frequencies that overlap with the lower range of female voices, potentially leading to misclassification.

Key results:

Feature Extraction Techniques:

Zero-Crossing Rate (ZCR) and Fundamental Frequency: Utilizing ZCR and fundamental frequency analysis, MATLAB implementations can distinguish gender based on pitch differences. Typical male voices range from 85–180 Hz, while female voices range from 155–255 Hz. A threshold of 160 Hz is often set to classify voices as male or female.

Challenges and Considerations:

Overlapping Frequency Ranges: The overlap in fundamental frequency ranges between male and female voices can pose challenges in accurate classification. For instance, some male voices may have higher fundamental frequencies that overlap with the lower range of female voices, potentially leading to misclassification.

Discussion

Our MATLAB implementation utilized fundamental frequency analysis to distinguish between male and female voices. Male voices typically have a fundamental frequency ranging from 85 Hz to 180 Hz, while female voices range from 155 Hz to 255 Hz. By setting a threshold at 160 Hz, the system effectively classifies voices based on their fundamental frequency. This approach aligns with established research indicating that pitch is a primary differentiator between male and female voices.

Our MATLAB implementation utilized fundamental frequency analysis to distinguish between male and female voices. Male voices typically have a fundamental frequency ranging from 85 Hz to 180 Hz, while female voices range from 155 Hz to 255 Hz. By setting a threshold at 160 Hz, the system effectively classifies voices based on their fundamental frequency. This approach aligns with established research indicating that pitch is a primary differentiator between male and female voices.

Conclusion

This study demonstrates the feasibility and effectiveness of gender voice recognition using MATLAB, employing fundamental frequency analysis and machine learning techniques. The system successfully classifies voices based on pitch, with male voices typically ranging from 85–180 Hz and female voices from 155–255 Hz. However, challenges arise due to the overlap in these frequency ranges, which can lead to misclassification. To mitigate this, alternative methods such as analyzing the Root Mean Square (RMS) values of bandpass-filtered signals have been explored, offering improved accuracy in gender classification.

Furthermore, the integration of machine learning algorithms, such as Support Vector Machines (SVMs) and Convolutional Neural Networks (CNNs), has shown promise in enhancing classification

performance. For instance, a CNN model utilizing Mel Frequency Cepstral Coefficients (MFCC) features achieved a classification accuracy of 99.22%. Incorporating such classifiers into the MATLAB framework could further improve performance, especially in challenging scenarios with overlapping frequency ranges.

Addressing gender bias remains a critical aspect of voice recognition systems. Studies have shown that commercial voice recognition applications often exhibit higher error rates for female voices compared to male voices. This disparity is attributed to factors such as differences in pitch, frequency, and the underrepresentation of female voices in training datasets. To mitigate this bias, data augmentation techniques, such as manipulating fundamental frequency and formants, have been proposed. These methods aim to balance gender representation in training data, leading to improved recognition accuracy for underrepresented groups.

Future Scope

Advancing Algorithmic Accuracy and Fairness:

While our current MATLAB-based system demonstrates effective gender classification using fundamental frequency analysis, future developments can enhance its accuracy and fairness. Incorporating advanced machine learning algorithms, such as deep neural networks, can improve classification performance. Additionally, employing techniques like adversarial training can mitigate gender biases, ensuring equitable treatment across diverse gender identities.

Expanding Dataset Diversity:

To address the challenges posed by overlapping frequency ranges between male and female voices, it's crucial to expand the diversity of training datasets. Including a broader spectrum of voices, encompassing various accents, dialects, and gender identities, can enhance the system's robustness

modal data, such as combining audio features with visual cues or contextual information. This holistic approach can improve the accuracy of gender classification and enable more nuanced understanding in complex scenarios.

Real-Time Processing Capabilities:

Enhancing the system's capability for real-time processing is vital for applications in dynamic environments. Optimizing algorithms for speed and efficiency can facilitate seamless integration into real-world applications, such as virtual assistants and interactive voice response systems.

References

[1]. John R. Deller, John G Proakis and John H. L. Hansen, "Discrete- Time Processing of Speech Signals" Macmillan Publishing company, 866 Third avenue, New York 10022.

[2] Huang X., Acero, A., and Hon, H.-W. "Spoken Language Processing: a Guideto Theory, Algorithm and System Development" prentice-Hall, New Jersey, 2001.

[3] Yakun Hu, Dappeng wu and Antonio Nucci**,** pitch Based Gender Identification with Two Stage Classification, International Journal of Security and Communication Networks., vol. 5, Issue 2,
pp. 211– 225, 2012.

[4] Rabiner Lawrence, Juang Bing-Hwang**,** **"**Fundamentals of Speech Recognitions", Prentice Hall New Jersey, 1993, ISBN 0- 13-015157-2.

[5] Md. Saidur Rahman, "Small Vocabulary Speech Recognition in Bangla Language", M.Sc. Thesis, Dept. of Computer Science & Engineering, Islamic University, Kushtia-7003, July-2004.