Visvesvaraya National Institute of Technology (VNIT), Nagpur

Advanced Digital Signal Processing (ECL412)

Project Report

Submitted by :
Allamsetti Jayaram(BT21ECE052)
Semester 6

Submitted to :
Dr. Vipin Kamble
(Course Instructor)
Department of Electronics and Communication Engineering,
VNIT Nagpur

Gender analysis using speech signal

1. Objective

The primary objective of this project is to develop a system that can analyze the speech signal of an individual and determine their gender (male or female). This task is often employed in various applications, such as speaker recognition, speech-based user interfaces, and audio forensics.

2. Concepts/Theory

The project utilizes the following key concepts and techniques:

2.1. Signal Filtering

A low-pass filter is applied to the input audio signal to remove high-frequency components that may not be relevant for gender identification. This helps to focus on the characteristics of the speech signal that are more indicative of gender.

2.2. Feature Extraction

Three main features are extracted from the audio signal:

Fundamental Frequency: The fundamental frequency of the speech signal, which is related to the vibration of the vocal cords and can be used to distinguish between male and female voices.

Zero-Crossing Rate: The rate at which the signal changes sign, which is related to the formant structure of the speech and can provide information about the speaker's gender.

Short-Term Energy: The energy of the speech signal within a short-time window, which can also be used to differentiate between male and female voices.

2.3. Feature-based Classification

The extracted features are used to classify the speaker as male or female. In this project, a simple threshold-based approach is used, where the weighted sum of the normalized feature values is compared to a predefined threshold. If the value is greater than the threshold, the speaker is classified as female; otherwise, the speaker is classified as male.

3. Procedure

The project consists of the following steps:

1. **Audio Data Acquisition:**

The project utilizes a dataset of audio recordings, with each recording representing a single utterance from a speaker.

The audio files are in the WAV format and have a known sampling rate.

2. **Signal Preprocessing:**

A low-pass filter is applied to the input audio signal to remove high-frequency components that may not be relevant for gender identification.

The filtering is performed using a digital filter designed with a cutoff frequency of 1000 Hz.

3. **Windowing and Frame Processing:**

The pre-processed audio signal is divided into overlapping frames, each representing a 30-millisecond segment of the signal.

The frames have a 10-millisecond overlap to capture the temporal dynamics of the speech signal.

Frame Size: num_of_samples = fs * 30 * 0.001 = 0.03 * fs (30 milliseconds)

Frame Overlap: num_over = fs * 10 * 0.001 = 0.01 * fs (10 milliseconds)

Number of New Samples in Each Frame: num_samp = num_of_samples - num_over

Number of Frames: n = ceil((length_samp - num_of_samples) / num_samp)

4. Feature Extraction:

For each frame, three key features are extracted:

Fundamental Frequency: It is calculated from the autocorrelation of the frame, representing the vibration of the vocal cords.

Zero-Crossing Rate: Computed by counting the number of sign changes in the frame, related to the formant structure of the speech.

The average zero-crossing rate is computed as zcr_avg = mean(zcr_sum / 2).

Short-Term Energy: It is calculated by summing the squared values of the samples in the frame, indicative of the overall energy of the speech.

5. **Feature Normalization and Weighting:**

The extracted features are normalized based on predefined threshold values for each feature.

A weighted sum of the normalized features is calculated, with the weights reflecting the relative importance of each feature in the gender classification task.

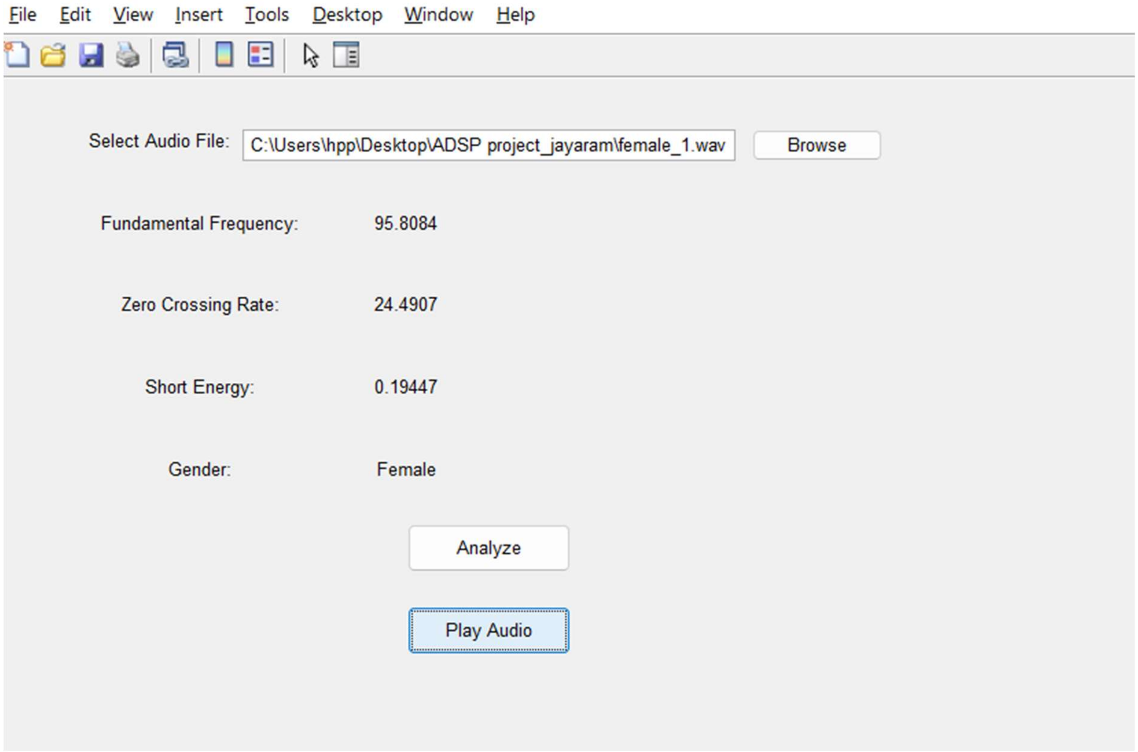
6. **Classification:**

The weighted sum of the normalized features is compared to a predefined threshold value.

If the weighted sum is greater than the threshold, the speaker is classified as female; otherwise, the speaker is classified as male. (i.e. If the weighted sum of the normalized feature value is greater than 1, then it classifies the speaker as female; otherwise, it classifies the speaker as male.)

4. GUI Design

To provide a user-friendly interface for the gender analysis system, a Graphical User Interface (GUI) has been developed using MATLAB's built-in GUI tools. The GUI includes the following components:



Select Audio File: C:\Users\hpp\Desktop\ADSP project_jayaram\female_1.wav

Fundamental Frequency: 95.8084

Zero Crossing Rate: 24.4907

Short Energy: 0.19447

Gender: Female

Fig.1: Figure showing the GUI interface

Audio File Selection: A text box and a "Browse" button allow the user to select the audio file to be analyzed.

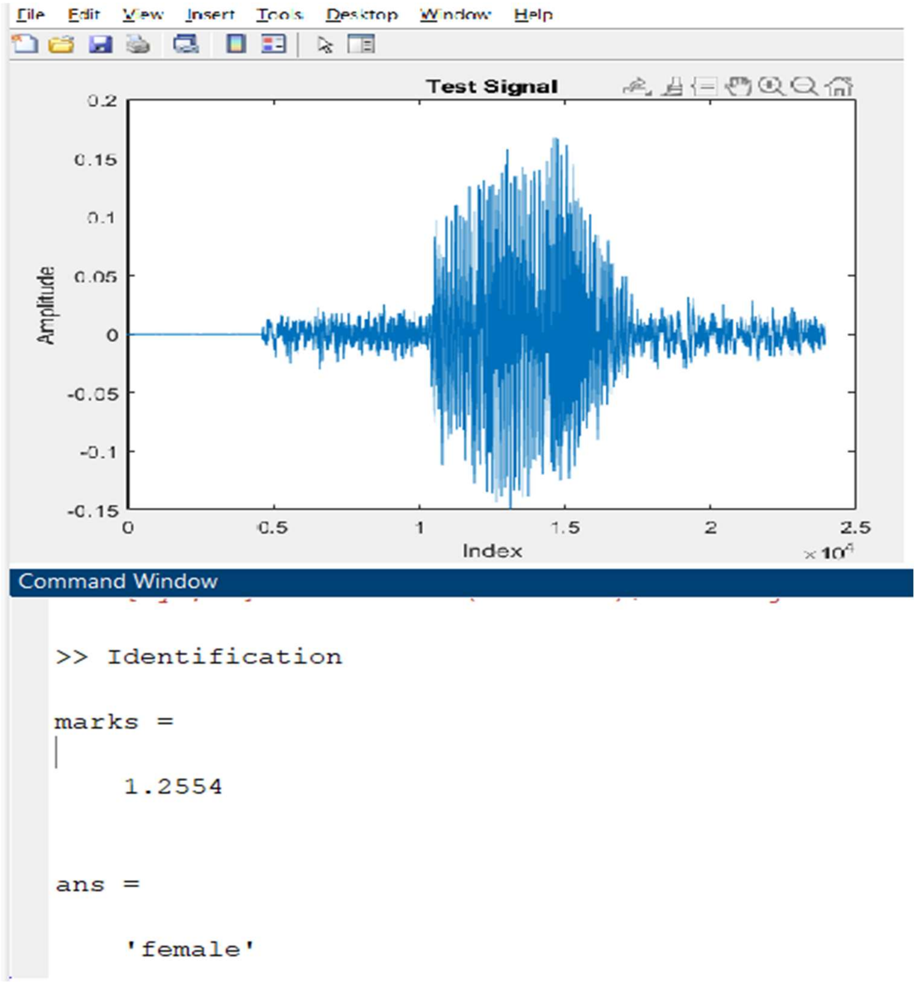
Feature Display: Three text boxes display the extracted features: fundamental frequency, zero-crossing rate, and short-term energy.

Gender Identification: A text box displays the identified gender of the speaker (male or female).

Audio Playback: A "Play Audio" button allows the user to listen to the selected audio file directly from the GUI. All these components can be seen in the figure Fig.1

The GUI provides a seamless and interactive experience for the users, enabling them to easily analyse the gender of speakers and explore the extracted features.

Output:



```
>> Identification

marks =

    1.2554


ans =

    'female'
```

Fig-2: Output plot observed for an audio input of female voice

5. **Discussions and Observations:**

The proposed system has been tested on a limited dataset of audio recordings, and the results have shown promising performance in correctly identifying the gender of the speakers. However, it is important to note that the accuracy of the system may be influenced by various factors, such as the quality of the audio recordings, the diversity of the speakers, and the robustness of the feature extraction and classification algorithms.

The use of a simple threshold-based classification approach in this project is a simplistic approach and may not be able to handle more complex cases, such as speakers with ambiguous or gender-neutral vocal characteristics. More advanced machine learning techniques, such as support vector machines or neural networks, could potentially provide more accurate and robust gender classification results.

6. Future Work

To improve the performance and generalization of the gender analysis system, the following future work can be considered:

1. Expand the Dataset: Collect a larger and more diverse dataset of audio recordings, covering a wider range of speakers, accents, and speaking styles.

2. Investigate Advanced Classification Algorithms : Explore the use of more sophisticated machine learning techniques, such as support vector machines or deep neural networks, to improve the gender classification accuracy.

3. Incorporate Additional Features : Explore the use of additional speech signal features, such as spectral characteristics or prosodic features, to further enhance the gender identification capabilities.

4. Implement Real-time Processing : Develop a real-time version of the system that can perform gender analysis on live audio input, enabling applications in areas such as interactive voice response systems or audio surveillance.

5. Evaluate Performance on Different Domains : Assess the system's performance on different types of audio data, such as telephone conversations, broadcast media, or audio recordings in different languages or environments.

7. References

[1]. Prabhakar, S., Pankanti, S., and Jain, A. “Biometric recognition: security and privacy concerns”IEEE Security and Privacy Magazine 1(2003), 33-42.

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

[3]. Martin, A., and Przybocki, M. Speaker recognition in a multi-speaker environment.In Proc. 7thEuropean Conference on Speech Communication and Technology (Eurospeech 2001) (Aalborg,Denmark, 2001), pp. 787–90.

[4]. Tomi Kinnunen " Spectral Feature for Automatic Voice-independent Speaker Recognition"Depertment of Computer Science, Joensuu University,Finland. December 21, 2003.

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

[6]. Rabiner Lawrence, Juang Bing-Hwang, "Fundamentals of Speech Recognitions", Prentice Hall NewJersey, 1993, ISBN 0-13-015157-2.

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