Digital Signal Processing course project

Gender Recognition by Voice

[1] Amir Rusrus 1190987
[2] Osama Rihami 1190560
1190987@student.birzeit.edu, 1190560@student.birzeit.edu

Abstract

This project aims to build a technique for identification of males and females (maybe further adults and children) from their speech analysis , using MATLAB we want to recognize the gender of a speaker from a voice sample and analyzing parameters of this voice sample to determine the gender of the speaker.

1. Introduction

Gender Recognition of a speaker has various potential applications such as speaker identification for securing access to confidential information or virtual spaces. As a human it is easy to differentiate between male and female voices , but we need to find implement this method using MATLAB..

2. Problem Specification

We can know the difference between genders by finding the vibration rate of the vocal folds of the speaker, which is usually high for the children, female speakers, and relativity low for the male speakers, so after we read the voice sample from our dataset, we can get a simple short-time pitch frequency (or fundamental frequency F0) which can be estimated from the short frames (20-30ms) using auto-correlation method. The fundamental frequency here represents the vibration rate we need to find, then we compare the value with a threshold frequency which for a male voice lies between 85-155 Hz whereas for a female, it lies between 165 to 255 Hz.

3. Data

We used a dataset of voice samples for a different male and female speakers, we downloaded the dataset from ‘VoxCeleb’ website (available in the references), it had a 32-voice sample (16 for men, 16 for women) in m4a format, then we used an online web converter to convert the files to wav format with 32000 HZ sampling frequency.

4. Evaluation Criteria

we measured the performance of our program by testing the result of each dataset sample we have.

5. Approach

We used two approaches of voice recognition for our program to recognize the gender of the speaker in the dataset using MATLAB.

5.1 The 1st approach:

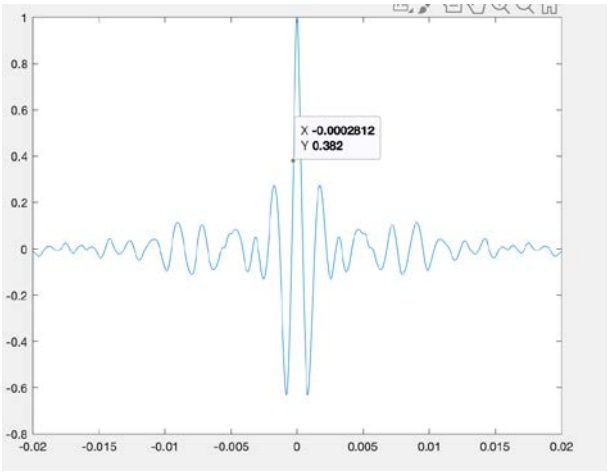
After we read the voice sample data and find the sampling frequency, we take the first 60ms sample and use the ‘pitch’ function to find the pitch frequency (fundamental frequency F0). The ‘pitch’ function is a MATLAB built-in function which returns estimates of the fundamental frequency over time for the audio input, then we test if the pitch frequency was less than 160 then It’s more likely to be a male’s voice. Otherwise it’s either a child or a female’s voice.

5.2 The 2nd approach:

After we read the voice sample data and find the sampling frequency, we find the autocorrelation value of the voice sample for (20ms) as shown:

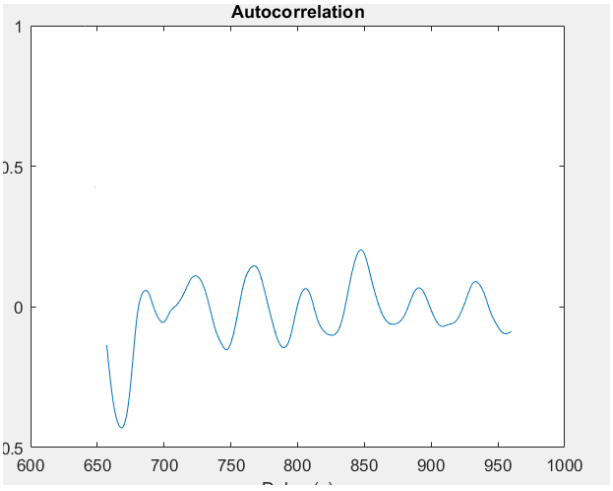
$$R(k) = \frac{1}{N}\sum_{n=0}^{N-1} s[n]s[n-k]$$

Where, R(k) is the autocorrelation function at k. s[n] is the short frame samples with length N samples.



(Fig 1)

, then we find the second peak value on the right side of the function (the first peak is at 0) located between (50 + 20ms, and 30ms):



(Fig2)

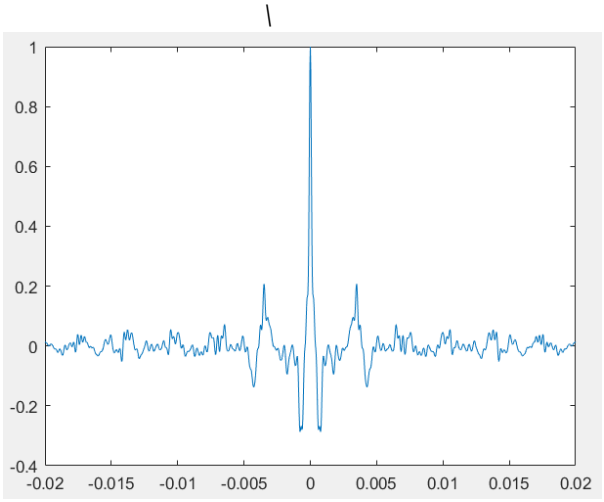
, next we can calculate the pitch period at that point, then from the pitch period we find the pitch frequency, and if it's less than 160 then it's likely to be a male's voice otherwise it's a female's or a child's voice.

6 Results and Analysis

Using the "female/0.wav" as a female data sample and "male/0.wav" as a male data sample.

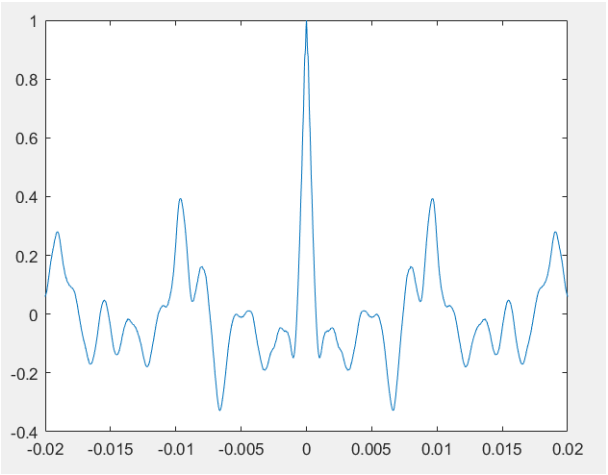
the results of the 1st approach are pitch frequency = 247 for the female and 105 for the male, and we can notice that the female sample had a pretty much obvious higher pitch frequency than the male's sample.

the results of the 2nd approach are pitch frequency = 283.1 for the female and an autocorrelation plot:



(Fig3)

And for the male a pitch frequency = 103.2 and an autocorrelation plot:



(Fig4)

, and we can notice again that the female sample had higher pitch frequency than the male's sample.

The 1st approach succeeded to recognize the gender in 25 out of 32 samples with a success rate of 78.12%

The 2nd approach succeeded to recognize the gender in 28 out of 32 samples with a success rate of 87.50% which is higher than the first approach.

7 Development

We can improve our project to overcome the shortcomings by finding a more accurate threshold value, or by taking the pitch frequency on a different time periods and compute the average value of them.

8 Conclusions

From this approach now we understand how voice recognition works and how the smart devices nowadays can recognize my voice over other people like virtual assistants like Siri and Alexa, and so it can be used for identify a person by their voice so it can be used for security wise and many other applications.

9 References

[1] https://drive.google.com/file/d/1HRbWocxwClGy9Fj1MQeugpR4vOaL9ebO/view

[2] https://cloudconvert.com/m4a-to-wav

[3] https://www.mathworks.com/help/audio/ref/pitch.html