Gender Recognition of a speaker using MATLAB

Hasan Hamed, Eyab Ghifari, Shorooq Ngjar and Osama Qutait

Department of Computer Engineering
Faculty of Engineering
Birzeit University
Birzeit, Ramallah
Palestine, P-600-P699
E_mail: 1190496@student.birzeit.edu, 1190999@student.birzeit.edu, 1192415@student.birzeit.edu, 1191072@student.birzeit.edu

1. Introduction

This paper describes gender recognition method using autocorrelation method and pitch method involving the preprocessing and the extraction of pitch frequency. It also presents the implementation and the basic experiments and discussions. Voice recognition has several many applications in terms of security and safety.
Pitch detection is very important for many speeches processing algorithm. Speech recognition system of tonal language use pitch tracking for tone recognition, which is important in disambiguating the myriad of homophones. Pitch is also crucial for prosodic variations in text-to-speech systems and spoken language systems. The fundamental frequency (F0) is the main cue of the pitch.

2. Problem Specification

Speaker gender recognition might be used for a variety of purposes, including safeguarding access to secret information or virtual places. It is simple for a person to distinguish between male and female sounds. The aim of the project is to utilize a strategy for identifying males and females (maybe other adults and children) from their voice analysis by examining multiple aspects of the voice sample, this technique seeks to determine the speaker's gender. It contains a simple short-time pitch frequency (or fundamental frequency F0) that may be computed using the auto-correlation approach from short frames (20-30ms).

3. Data

The data that were collected are different samples of different males and females' adult people that are saying different words and sentences which were taken from a previous project in that filed that was taken from GitHub [1]. Data can also be obtained by recording own voices and saving them into the PC. Data can also be obtained using different websites that have saved files of recorded sounds.

4. Evaluation Criteria

The system will be evaluated according to the number of correct recognitions which means that the system will be given the samples that were taken from the different sites as shown in the data section which are females and males voices and then counting the number of correct recognitions which means true recognition of male voices when the input voice is for male and female recognitions when the input voice is a female voice. And then calculating the accuracy of the system by dividing the total number of correct estimations over the whole number of samples. If the system provided a good accuracy, then everything is ok but if the system provided a bad accuracy, then a different approach must be taken or several modifications must be done to the method.

5. Approach

Basically, pitch detection algorithms use short-term analysis techniques. For every frame xm we get a score f(T| xm) that is a function of the candidate pitch periods T. Algorithm determine the optimal pitch by maximizing (1).

$$T_m = \underset{T}{\operatorname{argmax}} f(T \mid x_m) \qquad (1)$$

A commonly used method to estimate pitch is based on detecting the highest value of the autocorrelation function in the region of interest. Given a discrete time signal x(n), defined for all n , the auto-correlation function is generally defined in (2):

$$R(k) = \frac{1}{N}\sum_{n=0}^{N-1} s[n]s[n-k] \ldots\ldots\ldots\ldots (2)$$

The autocorrelation function of a signal is basically a (noninvertible) transformation of the signal that is useful for displaying structure in the waveform. Thus, for pitch detection, if we assume x(n) is exactly periodic with period P, i.e., x(n) = x(n + P) for all n, then it is easily shown that:

$$R_x(m) = R_x(m + P), \qquad (3)$$

i.e., the autocorrelation is also periodic with the same period. Conversely, periodicity in the autocorrelation function indicates periodicity in the signal. For a nonstationary signal, such as speech, the concept of a long-time autocorrelation measurement as given by (2) is not really meaningful. Thus, it is reasonable to define a short-time autocorrelation function, which operates on short segments of the signal as:

$$R_x(m) = \frac{1}{N}\sum_{n=0}^{N'-1}[x(n+l)w(n)][x(n+l+m)w(n+m)],$$
$$o \le m \le M_0 \qquad (4)$$

where w(n) is an appropriate window for analysis, N is the section length being analyzed, N' is the number of signal samples used in the computation of R(m), Mo is the number of autocorrelation points to be computed, and l is the index of the starting sample of the frame. For pitch detection applications N' is generally set to the value in (5):

$$N' = N - m \qquad (5).$$

So that only the N samples in the analysis frame (i.e., x(l), x (l+1), . . ., x (l + N - 1)) are used in the autocorrelation computation. Values of 200 and 300 have generally been used for Mo and N, respectively, it is corresponding to a maximum pitch period of 20 ms (200 samples at a 10 kHz sampling rate) and a 30 ms analysis frame size. [1,3]
For that MATLAB was used following this procedure:
The audio files were read from the data that were saved.
sampling frequency (Fs) was found.
The samples were divided into short frames, e.g., 0.29fs with overlap 0.03fs.
Find autocorrelation values of each short frame, for all k values k:0→N-1.
Find the first max peak of the autocorrelation values, which is corresponding to the pitch period (in samples). And convert the estimated pitch period P into seconds and then find the corresponding pitch frequency F0 using Pitch function.
If F0 is above a specific threshold then the voice is of a male, otherwise it is of a female. If you get incorrect results, try changing the threshold values. Threshold frequency (Fth) of a male voice lies between 85-155 Hz whereas for a female, it lies between 165 to 255 Hz

6. Results and Analysis

6.1. Females Samples

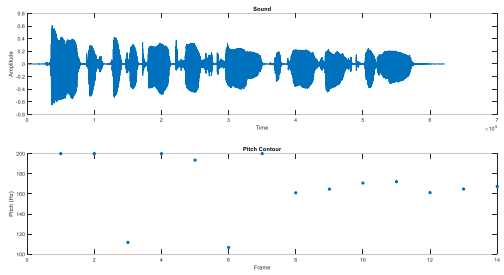

Fig. 1 Sample #1
F0 = 169.5770 HZ

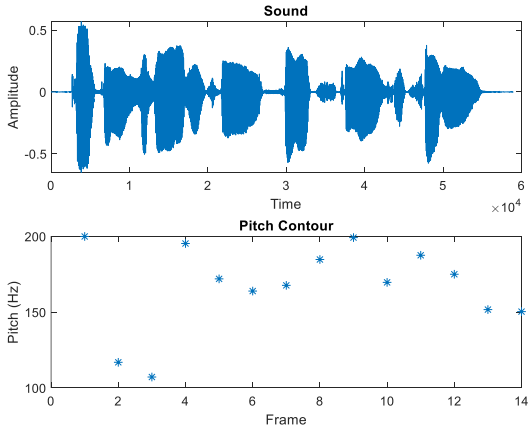

Fig. 2 Sample #2
F0 = 167.2518 HZ

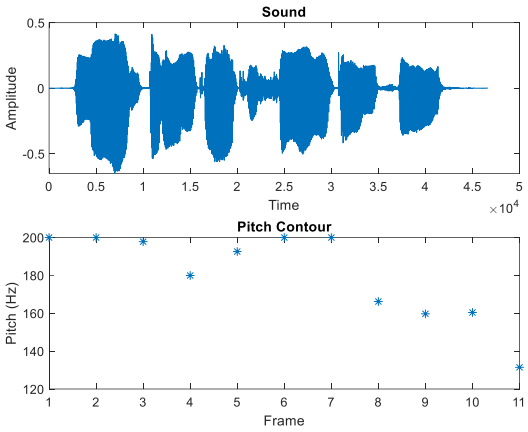

Fig. 3 Sample #3
F0 =180.7117 HZ

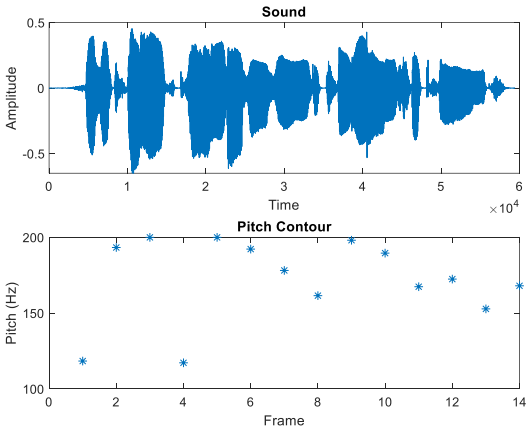

Fig. 4 Sample #4
F0 = 172.0634 HZ

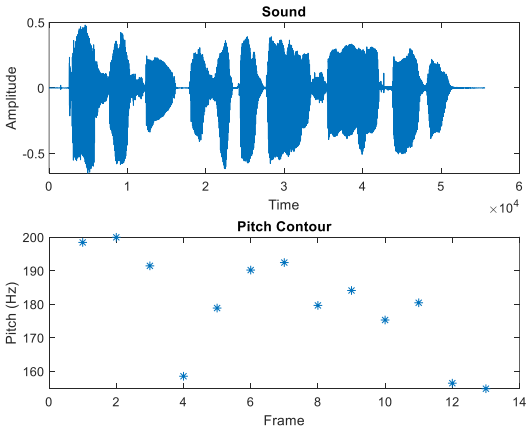

Fig. 5 Sample #5
F0 = 180.0444HZ

As shown in above figures that female pitch frequency for every frame is high and the average of all those frequencies is in the female threshold which lies between 165 to 255 Hz.

6.2. Males Samples

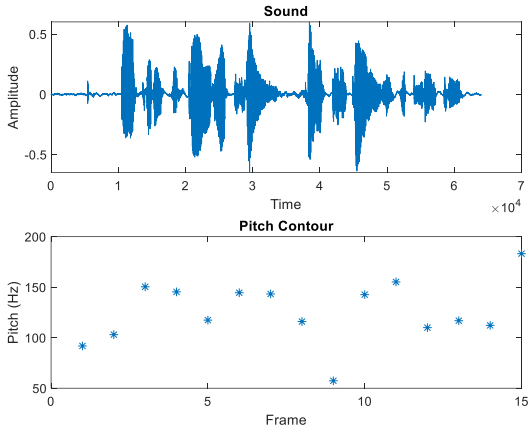

Fig. 6 Sample #1
F0 = 125.9399 HZ

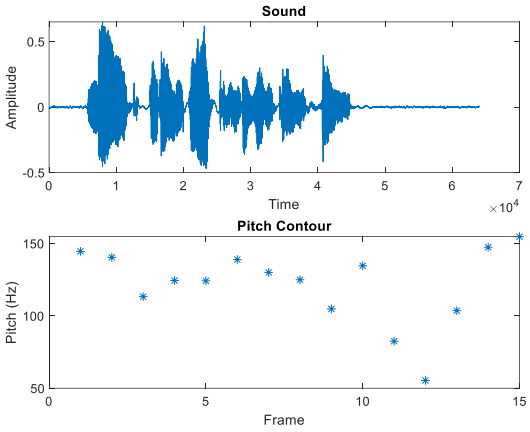

Fig. 9 Sample #4
F0 = 121.5558 HZ

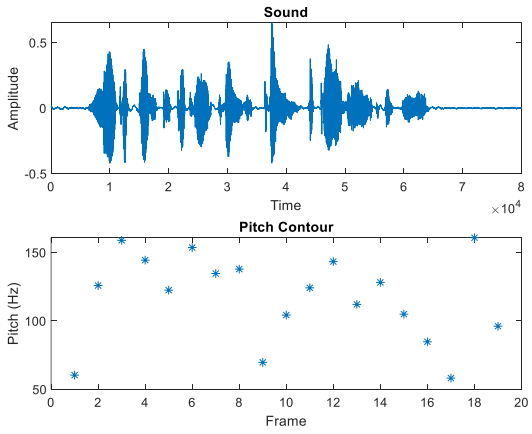

Fig. 7 Sample #2
F0 = 117.0311 HZ

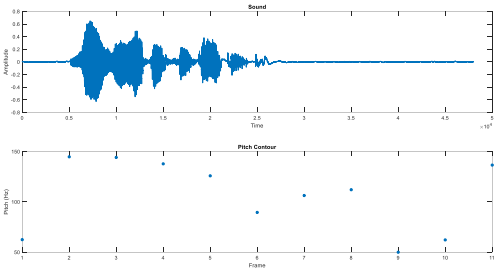

Fig. 10 Sample #5
F0 = 106.5037 HZ

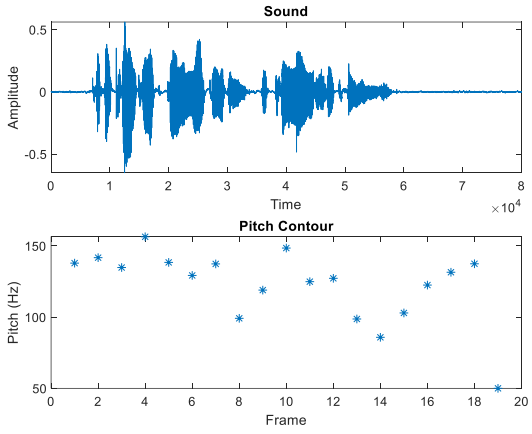

Fig. 8 Sample #3
F0 = 122.2896 HZ

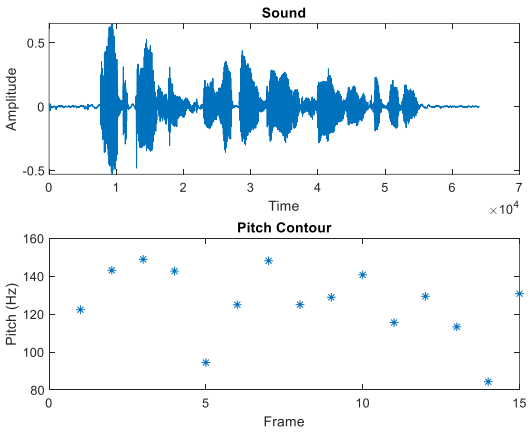

Fig. 11 Sample #6
F0 = 126.1655 HZ

As shown in above figures that the average pitch frequency for all of the frames for the male voices lies between 85-155 Hz which are less than female frequency and thus, we can conclude that female voices are higher than male voices.

7. Development

As shown in all the test cases for both females and males that finding f0 depends heavily on the frequencies of the frames, which means that there is a high error rate in recognition that can be fixed using deep learning which depends on given the systems a lot of test cases that the system can learn from them and decide which is the gender of the voice. The system can be also developed by adding the capabilities of deciding if the voice is for an adult or a kid. The system accuracy is high which from 15 sample test it gave a 14 correct result.

8. Conclusion

In conclusion, we understood how to use pitch frequency for voice recognition which is used for deciding if the voice is a male or female voice, and we understood how the recognition method works which depends on framing the voice signal into frames with a certain length and overlap and then finding autocorrelation of each frame to find the pitch period which is changed into pitch frequency and then finding the average of each frame pitch frequency to decide the voice and all that was done using a single MATLAB function called Pitch().

9. References

[1] https://github.com/AvinashUmmagani/Gender-Recognition-using-Speech-Analysis/tree/master/Gender%20Recognition%20uising%20Speech%20Analysis

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

[3] https://www.researchgate.net/publication/228854783_Pitch_detection_algorithm_autocorrelation_method_and_AMDF

10. Appendix

```
[audioIn,fs] = audioread("1.wav");
sound(audioIn,fs)
windowLength = round(0.29*fs);
overlapLength = round(0.03*fs);
f0 =
pitch(audioIn,fs,WindowLength=windowLength,OverlapLength=overlapLength,Range=[50,200],Method="PEF");
F0=mean(f0)
figure
subplot(2,1,1)
plot(audioIn)
ylabel("Amplitude")
xlabel("Time")
title("Sound")
subplot(2,1,2)
plot(f0,"*")
ylabel("Pitch (Hz)")
xlabel("Frame")
title("Pitch Contour")
if F0>=85 && F0<155
   disp('It is a male voice!')
else if F0>=165 && F0<255
   disp('It is a Female voice!')
   else
       disp('Error!')
   end
end
```