IDENTIFICATION OF HUMAN MALE AND FEMALE BY SPEECH RECOGNITION USING PYTHON

Name: Richard Tony

Enroll: BT18GEC091

INTRODUCTION:

The oldest musical instrument is the human voice. The voice produces sound when air from the lungs vibrates the vocal cords in the throat. The air in the hollow spaces of the chest, throat and mouth vibrates and amplifies the sound of the voice. The vibrations of the vocal cords resonate in the cavities of both the chest (in the lower register) and the head (in the upper register).

A voice type is a particular kind of human singing voice perceived as having certain identifying qualities or characteristics.

• Women are typically divided into three groups:

soprano, mezzo-soprano, and contralto.

• Men are usually divided into four groups:

countertenor, tenor, baritone, and bass.

Fundamentals and Harmonics of Human Voice Frequency Range:

- A Female voice frequency range covers fairly up to 350 Hz to 17KHz.Its fundamental frequency is 350Hz to 3KHz and Harmonics is 3KHz to 17KHz.
- Male voice covers a Frequency range of 100Hz to 8KHz. The fundamental is 100Hz to 900Hz and Harmonics is 900Hz to 8KHz.

Sound production, in animals, the initiation of sound as a means of information transmission. Sounds are termed vocal when produced

in the respiratory system and mechanical when produced by mutual contact of body parts or by contact with some element in the environment. Vocal sounds are restricted to vertebrate animals; non vocal sounds are produced by many invertebrates and by some members of all vertebrate classes.

Many animals possess special structures for producing mechanical sounds. Crickets and grasshoppers produce sound by rubbing together rasp like structures on their wings. Cicadas, which emit the loudest sounds known from insects, do so by means of a pair of membranous organs (timbal organs) at the base of the abdomen. A special muscle deadens the hearing apparatus of the insect when it is calling.

This project aims to classify or identify the human gender and different types of animals based on the speech sample, animal sounds using python.

APPLICATION SCENARIO:

- Attendance system for recording the presence of student based on voice
- Home security system to recognize the voice of the recorded family members.
- And many more

METHODOLOGY (PROPOSED):

- Collection of speech samples
- Determine its frequency components using fft commands

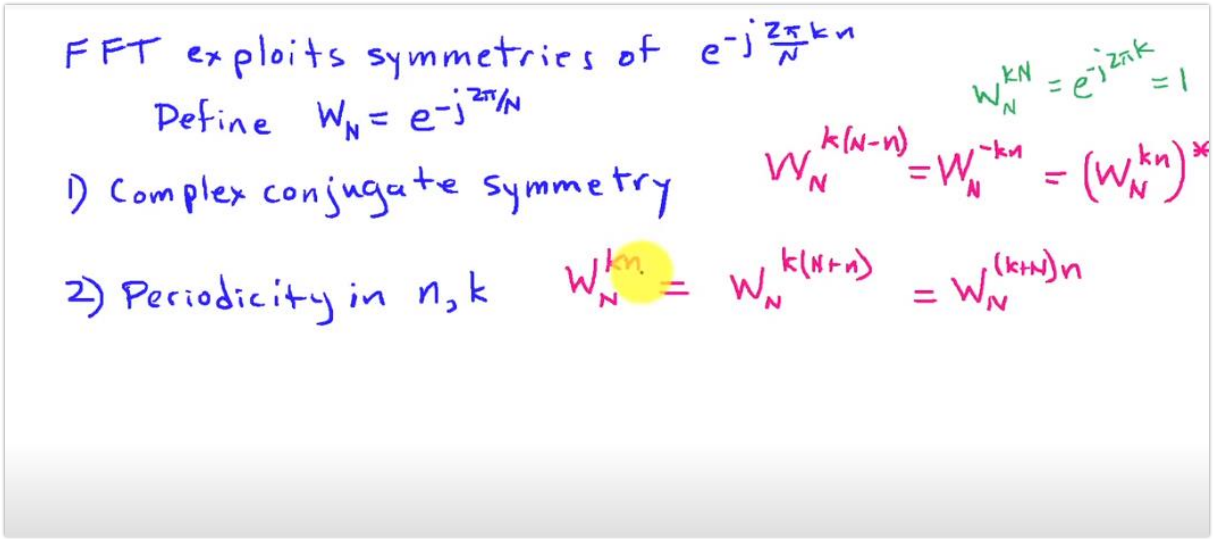
- Compare the frequencies with the determined frequency range of male and female voice

CONCEPTS:

Fast Fourier Transform:

The Fast Fourier Transform (FFT) is an algorithm that determines Discrete Fourier Transform of an input significantly faster than computing it directly.

$$x[k] = \sum_{n=0}^{N-1} x[n] e^{\frac{-j2\pi kn}{N}}$$



FFT exploits symmetries of $e^{-j\frac{2\pi}{N}kn}$

Define $W_N = e^{-j2\pi/N}$

1) Complex conjugate symmetry $W_N^{k(N-n)} = W_N^{-kn} = (W_N^{kn})^*$

2) Periodicity in n, k $W_N^{kn} = W_N^{k(N+n)} = W_N^{(k+N)n}$

Decimation in Time FFT (one of many)

- build a big DFT from smaller ones
- Assume $N = 2^m$

separate $x[n]$ into even and odd-indexed sub sequences

$$X[k] = \sum_{n=0}^{N-1} x[n] W_N^{kn} = \sum_{n \text{ even}} x[n] W_N^{kr} + \sum_{n \text{ odd}} x[n] W_N^{kr}$$

even, odd indices $\begin{cases} n = 2r \\ n = 2r+1 \end{cases}$, $r = 0, 1, \ldots \frac{N}{2} - 1$

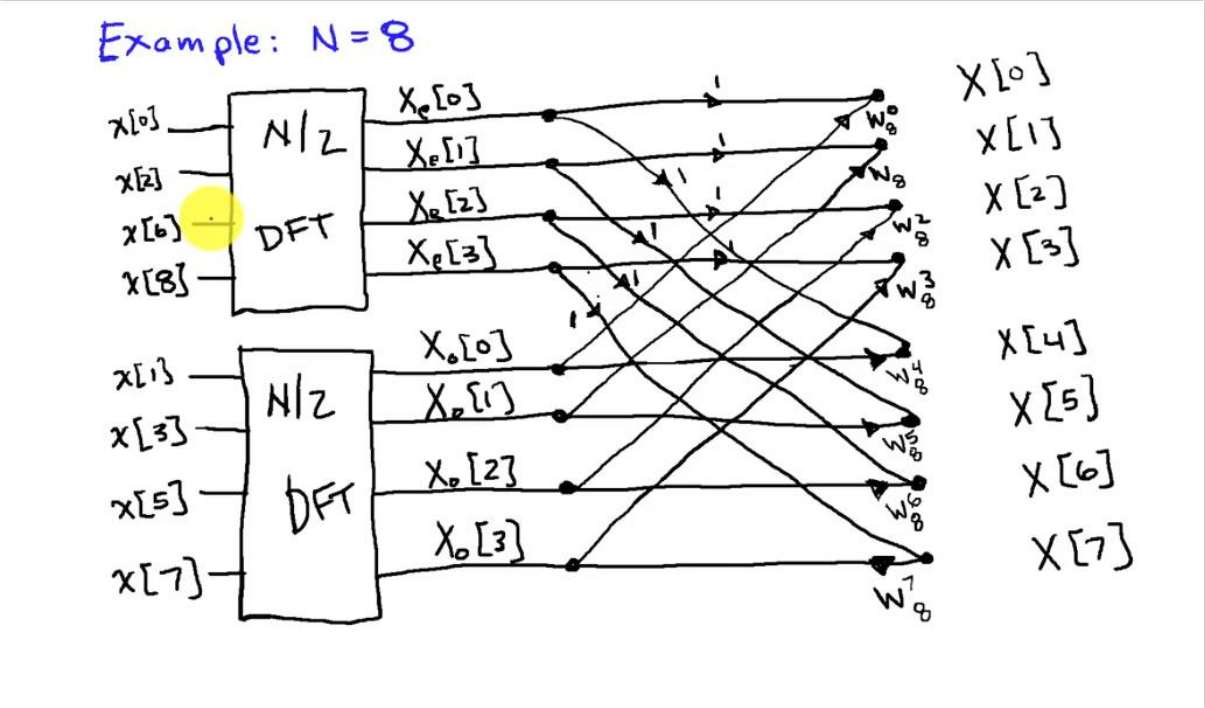
$$X[k] = \sum_{r=0}^{\frac{N}{2}-1} x[2r] W_N^{k2r} + \sum_{r=0}^{\frac{N}{2}-1} x[2r+1] W_N^{(2r+1)k}$$

$$= \sum_{r=0}^{\frac{N}{2}-1} x[2r] (W_N^2)^{kr} + W_N^k \sum_{r=0}^{\frac{N}{2}-1} x[2r+1] (W_N^2)^{kr}$$

But $W_N^2 = e^{-j\frac{2\pi}{N}2} = e^{-j2\pi/(N/2)} = W_{N/2}$

$$X[k] = \underbrace{\sum_{r=0}^{N/2-1} x[2r] W_{N/2}^{kr}}_{\text{N/2 DFT of even samples } X_e[k]} + W_N^k \underbrace{\sum_{r=0}^{N/2-1} x[2r+1] W_{N/2}^{kr}}_{\text{N/2 DFT of odd samples } X_o[k]}$$

$X[k] = X_e[k] + W_N^k X_o[k]$ — sum of 2 $N/2$ point DFTs



PROCEDURE:

- The audio is collected in the .wav format using the wav.read() function.
- The sampling rate and the data which contains the frequency information is collected.
- FFT is performed on the data which gives an array of complex values.
- The average is taken of all these samples by adding each frequency components and then dividing by total number of samples.
- A function called voice is created which takes this average or mean as its argument
- If the mean value is less than 100.0Hz or 0.1KHz then the gender identified will be male, else it will be a female. 100.0Hz or 0.1KHz is taken as a value which is the mean of the fft of the defined range of the human male voice.
- The plot is of the real part of the fft components.

CODE SNIPPET:

```python
#library for plotting the graph
import matplotlib.pyplot as plt
#library for reading from .wav file
from scipy.io import wavfile as wav
#library to perform fft functions
from scipy.fftpack import fft
import numpy as np
#reading from wav file
sampling_rate, data = wav.read('C:\\Users\\Richard Tony\\Downloads\\Music\\Emon.wav')
print('Sampling frequency: '+str(sampling_rate))
#using fft function to get frequency components
fft_out = fft(data)
combined = fft(data).ravel()
meanfunfreeq = sum(combined)/combined.size
print("Sum of frequency components"+str(sum(combined)))
print('Total number of frequency components: '+str(combined.size))
print("Mean frequency: "+str(meanfunfreeq))
#function which determines the gender
def voice(meanfun):
#checks if the mean value of the freqency components less than 0.1Khz then it is a male else it is a female
  if meanfun<=0.1:
    return("The gender identified is: Male")
  else:
    return ("The gender identified is: Female")
#prints the gender recognised in speech
print(voice(meanfunfreeq))


#plot the data of fft
plt.plot(data, np.abs(fft_out))
plt.xlabel
plt.show()

```

OUTPUT SCREENSHOTS:

1. FEMALE VOICE

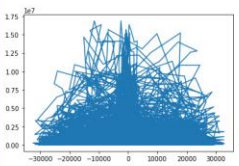
2. MALE VOICE

```python
#library for plotting the graph
import matplotlib.pyplot as plt
#library for reading from .wav file
from scipy.io import wavfile as wav
#library to perform fft functions
from scipy.fftpack import fft
import numpy as np
#reading from wav file
sampling_rate, data = wav.read('C:\\Users\\Richard Tony\\Downloads\\Music\\Man_scream.wav')
print('Sampling frequency: '+str(sampling_rate))
#using fft function to get frequency components
fft_out = fft(data)
combined = fft(data).ravel()
meanfunfreeq = sum(combined)/combined.size
print("Sum of frequency components"+str(sum(combined)))
print('Total number of frequency components: '+str(combined.size))
print("Mean frequency: "+str(meanfunfreeq))
#function which determines the gender
def voice(meanfun):
#checks if the mean value of the freqency components less than 0.1Khz then it is a male else it is a female
  if meanfun<=0.1:
    return("The gender identified is: Male")
  else:
    return ("The gender identified is: Female")
#prints the gender recognised in speech
print(voice(meanfunfreeq))


#plot the data of fft
plt.plot(data, np.abs(fft_out))
plt.xlabel
plt.show()

```

```
In [33]: runfile('C:/Users/Richard Tony/genderrecognition.py', wdir='C:/
Users/Richard Tony')
Sampling frequency: 11025
Sum of frequency components(-7754739.99999996+7.712515071034431e-10j)
Total number of frequency components: 20090
Mean frequency: (-385.9999999999998+3.838982115995237e-14j)
The gender identified is: Male
```




```
In [34]:
```

FURTURE ASPECTS:

- Can be used to identify people based on the speech in case of crimes by police department and other govt agencies.
- Controlling of appliances and devices based on voice using voice locks.
- And many more

CONCLUSION:

The project deals with digital signal processing techniques such as Fast Fourier Transform (FFT) to create a system which can identify the gender based on voice and further can be exploited to create different application scenarios in the digital signal processing fields.

ONLINE MATERIALS AND LINKS:

- https://www.seaindia.in/human-voice-frequency-range/
- https://www.britannica.com/science/sound-production
- https://www.slideshare.net/musicolga26/classification-of-human-voice
- https://towardsdatascience.com/understanding-audio-data-fourier-transform-fft-spectrogram-and-speech-recognition-a4072d228520