EMOTION RECOGNITON USING TEXT/AUDIO

M.V.B.T. Santhi,

Associate Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, India

Email: santhi_ist@kluniversity.in

Abstract:

The idea of acquiring the human emotional state from one's voice, we've collected the necessary data that helps us understand the principle behind the process. Human emotions can be predicted by his / her facial expressions or voice tone. One of the main tasks involved in image processing is the interpretation of the facial expressions. Computer is well-trained to identify and distinguish human faces, and is also trained to recognize emotions. It is programmed and trained using examples from the real world, since emotions often vary in facial expressions. Each emotion has a different expression to it. It can be trained using various faces in this way. Every emotion likewise holds different tone in one's speech. To measure and evaluate the emotions it needs a particular level of emotional tones. We need to retrieve estimated emotional frequencies. It's the daunting job, as each speaker has different pitch when speaking, and the same person's pitches vary in emotion. Another big problem is the noise in the audio when a person is speaking, because of low quality recordings or surroundings. The emotion list is Happy, angry, sad, bored, shocked.

The prior important principle for this project is speech recognition. The computer must be able to interpret the input as a speech in shape, and must be able to recognize the words. The provided input is converted to text format, and the text is parsed in tokens. Computer must also be capable of fetching the frequencies at the same time.The calculation is carried out using several established methodologies, and we have put into practice another methodology which is capable of recognizing emotions.

Keywords: emotion, human, computer, expression

1. INTRODUCTION

Speech recognition a computer or program's ability to recognize all of the words and phrases in speech and turn them into a computer-readable format. Rudimentary speech recognition code contains a restricted vocabulary of terms and phrases, and if they are spoken terribly clearly it has to define them only.

Speech recognition may also be a subfield of knowledge of linguistics that develops methodologies and technologies that enable the PC models to popularize and translate language into text. This is used as automatic speech recognition (ASR), pc speech recognition or STT (Speech to Text) recognition. Includes knowledge and research in the fields of linguistics, computing and electrical engineering.

Wherever a private speaker reads text or isolated vocabulary into the device, some speech recognition systems want "reading" (also explicit as "enrolment"). The machine analyzes the unique voice of the person and uses it to fine-tune the understanding of the speech of that person, resulting in excessive precision.

Systems not using measurement unit of jobs said as "speaker-independent" systems. Systems using the unit of measurement for jobs said as "speaker based."

Speech recognition systems include speech user interfaces such as

- speech dialing (e.g. 'call home')
- decision routing (e.g. 'I would really like to make a collect call'),
- domestic device management,
- search (e.g. realize a podcast whenever clear words are spoken),
- quick entry of information (e.g. entering a Mastercard number),
- Standardized paper preparation (e.g. a radiology report),
- main speaker features,
- speech-to-text processing (e.g. word processors or emails),
- and crafts (usually called direct voice input).

From the perspective of technical power, speech emphasis consists of a prolonged record of many waves of the foremost developments. Last, developments in deep learning and big data have helped the region. The advance area unit proved not entirely by way of the influx of sector-written tutorial papers at intervals. However, the implementation of a range of deep gathering data on how to spring up with and deploy speech recognition through mistreatment of the worldwide amendment.

In computerized speech focus, an acoustic model is used to reflect the relation between Associate in Nursing Audio Sign and additionally the phonemes or various linguistic gadgets that structure speech. The model is discovered from a collection of audio recordings and the related transcripts thereof. Audio recordings of speech and their text transcriptions and victim code are generated to shape common science representations of the sounds that structure each word.

Voice recognition clashes with the recognition of expressions, voice aims to understand the person or lady saying the words rather than the phrases themselves. But speech recognition technology lacks language. Furthermore, speech cognizance may also be regarded as knowledge of the speaker. Speech sensitivity involves recording victimization of spoken words on a transducer each.

Recognition of speech involves the recording of spoken words, using either a mike or a camera. The audio is then regenerated, and mistreatment 2 factor is assessed with the application of speech awareness: 1. Accuracy (percentage difference in words spoken to digital data in sterilization) 2. Speed (extent to which the computer code can be maintained by the speaker of a personality)

A lengthy list of uses involves speech recognition technology. Computer code package programs for speech recognition are used for general dictation, transcription, hands-free use of a laptop, medical transcription,

2. Methodology

Work into the development of emotions dates back to the nineteenth century. Advancement and operation were extended to the investigation of human correspondence, within Darwin's primary in his 1872 work, The Inclination Function.Charles Robert Henry Martyn Robert Parliamentarian} Charles Robert Darwin discussed the concept of feeling during a shot to assist in his theory of development. He types that discovered terribly like different characteristics, felt like booting created and custom-made after a whileHis research verified not so many physical characteristics in animals, and yet specifically sought to suggest comparisons between behavior in humans and various species.

According to the current theory of the treatment, all entirely different} feelings advanced at different time synonyms / hypernyms (ordered by measurable frequency). Similar to stress, basic feelings open sq. Beat related to antiquated brain parts and likely developed among our premammalian progenitors.

Obedient feelings tend to have formed among early warm-blooded animals, sort of an individual parent affection for her posterity. Nice feelings, close to blame for emotions and abandonment, advanced among friendly primates. Occasionally, as of late developed, the mind moderate companion more formed a {part of} the Darwin convention organized chain rule for the cerebrum, machine-comprehensible content exchange. The first of the trio is that the valuable law which he opened as helpful propensities strengthened the precedent along those lines hereditary by posterity.

He used eyebrow (wrinkling the forehead) as a partner pattern, which he noted was practical to forestall partner degree exorbitant light-weight aggregate from wagging into the eyes. He must boot the same way that eyebrow fostering is used to construct the circle of exteroception. He applies to oldster experiments attempting to recall one problem and lift their eyes, similar to when they were making a shot to recall..

The second of those criteria is the absolute opposite. Whereas some sq propensities. Useful measure, Darwin arranged that some activity in law or propensities territory unit only as an end of it is inverse in nature to a workable propensity, yet do not have all the characteristics of being useful itself.

2.1 Audio clip playing using python:

Using a Clip (python.sound.sampled.Clip) when, for example, a short solid record is needed to play non-constant sound information. Before playing back, the entire record is pre-stacked into memory, and then we have full command over the playback.

2.2 Playing source data file using python

Use a SourceDataLine (python.sound.sampled.SourceDataLine) if you need to play a long, secure record that can not be pre-stacked into memory, or stream ongoing sound content, such as playing back sound as it is captured.

2.3Discrete Fourier transform:

The discrete Fourier change (DFT) transforms over a small grouping of similarly spaced power examples into an equal duration arrangement of similarly distributed examples of a discrete-time Fourier change (DTFT), which is a complex-estimated recurrence power. The interval at which the DTFT is analyzed is the length of the information succession corresponding. A backward DFT is a Fourier arrangement which uses the DTFT tests at the comparing DTFT frequencies as the coefficients of complex sinusoids. It has similar esteems of example as the first grouping of knowledge. The DFT is said to be a common region representing the first grouping of information. Off chance that the first grouping will go through all the non-zero estimates of a capacity, its DTFT will be constant (and occasional), and the DFT will give discrete examples of one cycle. In the off chance that the first succession is an irregular power period, the DFT provides all non-zero estimates of one DTFT period.

$$\int_{-\infty}^{\infty} f(t)\, e^{-2\pi i \nu t}\, dt.$$

$$F_n \equiv \sum_{k=0}^{N-1} f_k\, e^{-2\pi i n k/N}.$$

$$f_k = \frac{1}{N} \sum_{n=0}^{N-1} F_n\, e^{2\pi i k n/N}.$$

The plots below display the genuine (red), fanciful (blue), and complex (green) module of the discrete Fourier changes in capability f(x)=sinx (left) and f(x)=sinx+sin(3x)/2 (right) inspected several times longer than two years. In the left diagram, the right and left balanced spikes are the "positive" and "negative" recurrence sections of the single sine wave. In addition, there are two sets of spikes in the correct figure, with the bigger green spikes relative to the lower-recurrence, more grounded segment sinxand the littler green spikes compared to the more fragile section. A fairly scaled plot of a discreet Fourier shift mind boggling modulus is usually known as a power set.

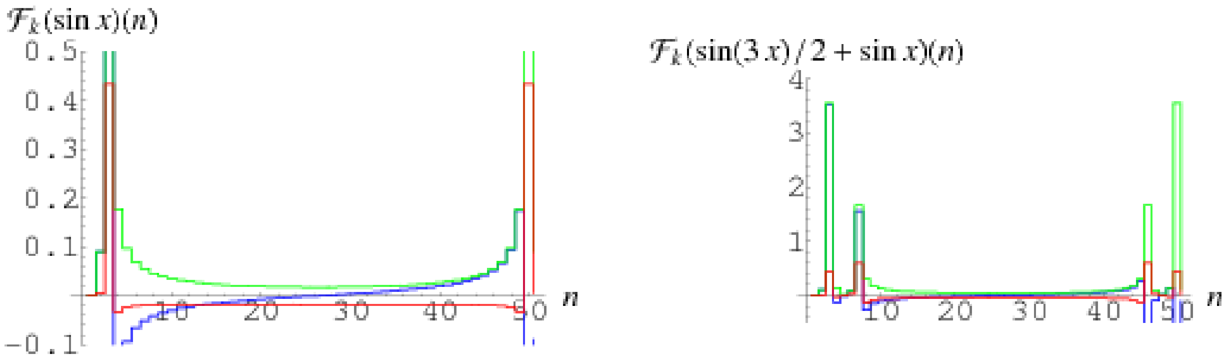
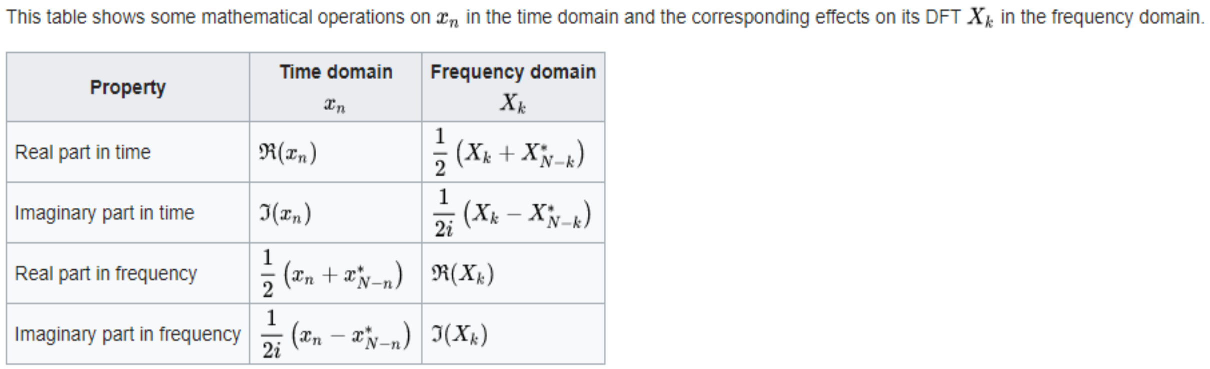


Figure 1



This table shows some mathematical operations on x_n in the time domain and the corresponding effects on its DFT X_k in the frequency domain.

Property	Time domain x_n	Frequency domain X_k
Real part in time	$\mathfrak{R}(x_n)$	$\frac{1}{2}\left(X_k + X^*_{N-k}\right)$
Imaginary part in time	$\mathfrak{I}(x_n)$	$\frac{1}{2i}\left(X_k - X^*_{N-k}\right)$
Real part in frequency	$\frac{1}{2}\left(x_n + x^*_{N-n}\right)$	$\mathfrak{R}(X_k)$
Imaginary part in frequency	$\frac{1}{2i}\left(x_n - x^*_{N-n}\right)$	$\mathfrak{I}(X_k)$

Table 1

2.4Fast Fourier Transform:

A Fast Fourier Change (FFT) is a calculation which records a grouping's or its opposite (IDFT) discrete Fourier Change (DFT). Fourier investigation switches around a sign from its particular area (regularly time or space) to a representation in the field of recurrence and vice versa.

The DFT is obtained by decaying a grouping of qualities into sections of various frequencies. This operation is useful in many fields, but it is always too delayed to actually be down to earth to find out straight from the description.

By factorizing the DFT lattice into a product of (generally zero) variables, FFT easily finds those shifts. Thus, it shows how to diminish the multifaceted nature of DFT registration from {left(N^{2}\right)}{O\left(N^{2}\right)}, which appears in case one essentially applies the definition of DFT, to {O(N\log N)}O(N\log N), where {N}N is the size of the details.

The difference in speed can be enormous, particularly for long information collections where N may be in the thousands or millions. Within the sight of modification error, multiple FFT estimates are considerably more accurate than a true or tacit evaluation of the DFT concept. Various FFT calculations rely on a broad range of distributed speculations, from straight complex-number maths to aggregate hypothesis and hypothesis of numbers.

Quick Fourier changes are commonly used in architecture, music, research, and arithmetics applications. The basic thoughts were advanced in 1965, but a few estimates were as timely as 1805.[1] In 1994, Gilbert Strang described the FFT as "the most important numerical calculation of our lifetime" and the IEEE magazine Computing in Science and Engineering listed it for the Top 10 Twentieth Century Algorithms.

The best established FFT calculations depend on N factorization, but for all N, in any case, for prime N, there are FFTs with multifaceted existence O(N log N). Numerous FFT calculations rely solely on the manner in which {e^{-2\pi I / N}}{e^{-2\pi I / N}} is an N-th crude solidarity base and can thus be extended to undifferentiated changes over any limited area, such as number theoretical changes. Since the backward DFT is equal to the DFT, but with the opposite sign in the form and a factor of 1/N, any FFT measurement can be modified for it without much stretch.

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi kn/N} \qquad k = 0, \ldots, N-1,$$

It is a calculation that takes on a substantial job in the calculation of an arrangement's Discrete Fourier Transform. It switches to flag of the recurrence region over a space or time symbol.

The DFT signal is generated by dispersing significant value arrangements to different part of the recurrence.

It is computationally unnecessarily expensive to work straightforwardly to turn over on Fourier shift. By factorizing the DFT network as a result of meager components, Fast Fourier change is used in this way as it easily registers. Therefore, it reduces the intricacy of DFT calculation from O(n2) to O(N log N). What else.

3. Results and Discussion

Steps for the cycle of understanding emotions through speech:

1) Mp3 to wave length.

2). Frequency of Wavelengths

3) Finding frequency Amplitude.

4) Reinforcement of an audio file.

5) Separation of Framerates.

6) Median level.

7) Use emotional frequency to differentiate the Emotion.

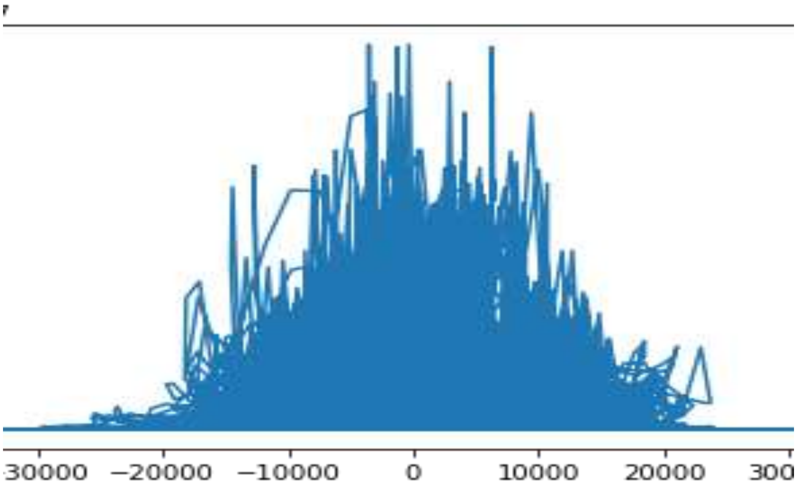


Figure 1 Frequency of Wavelengths

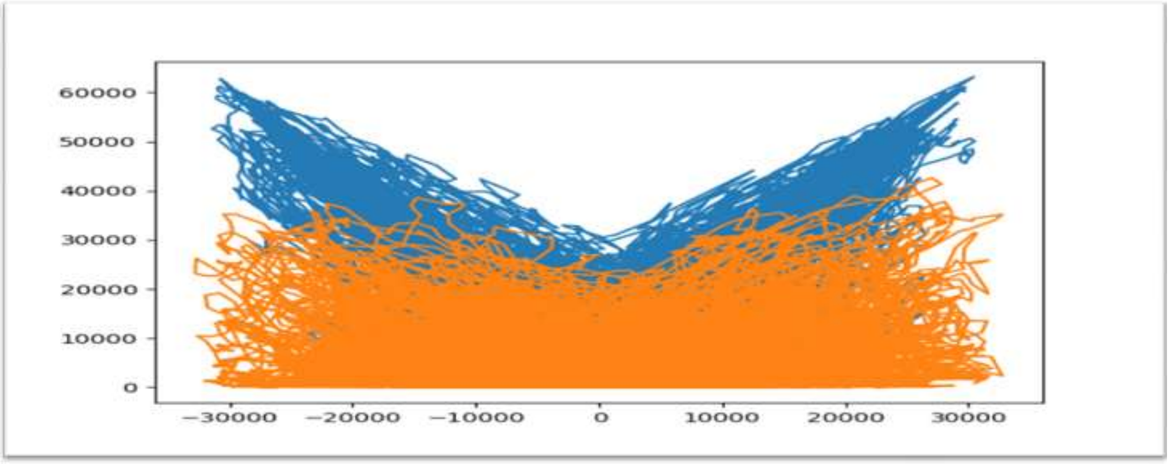


Figure 2 Frequency of Wavelengths

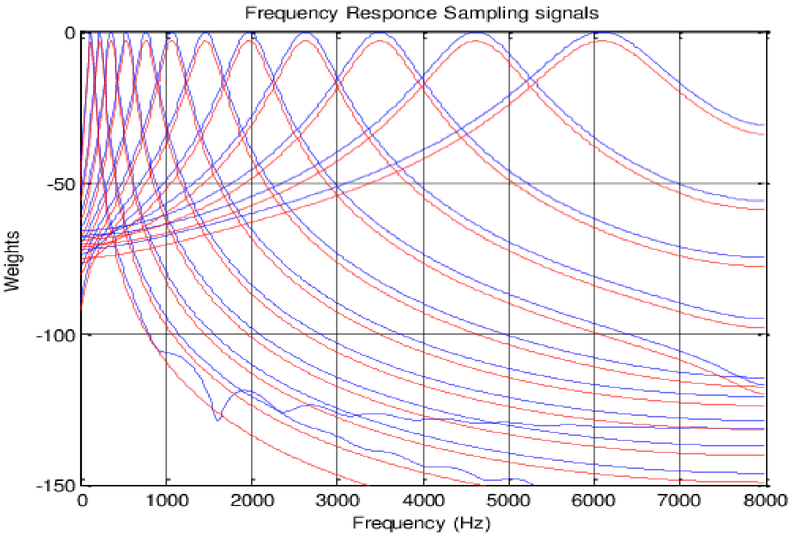


Figure 3 **Discrete Fourier Transform (DFT)**

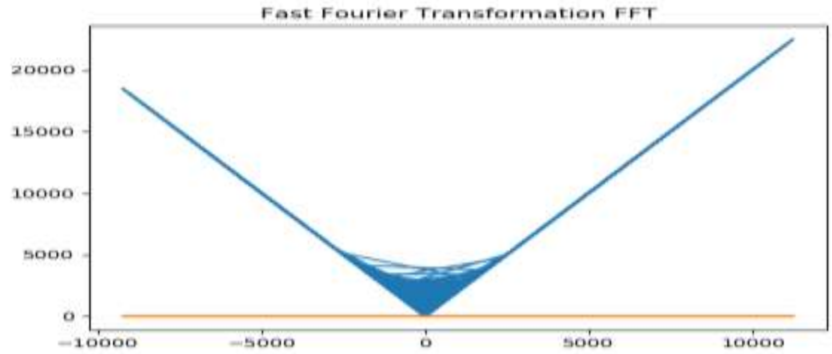
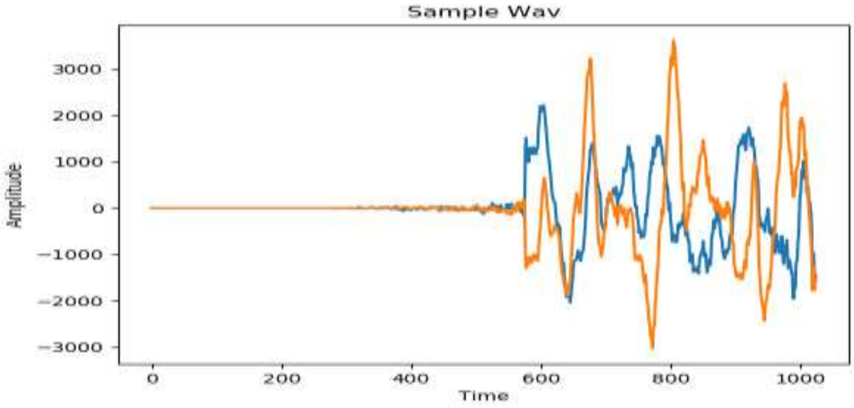
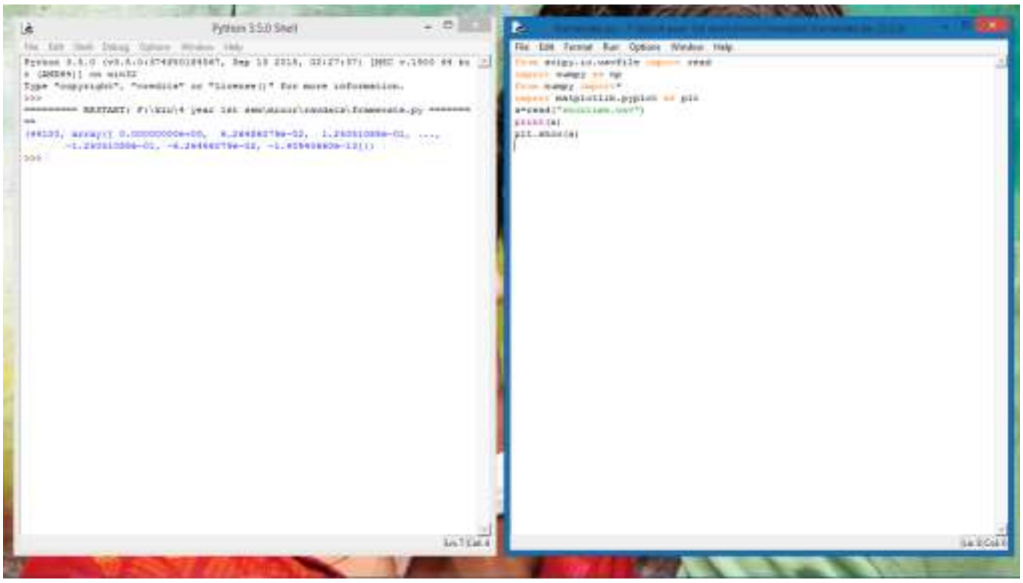


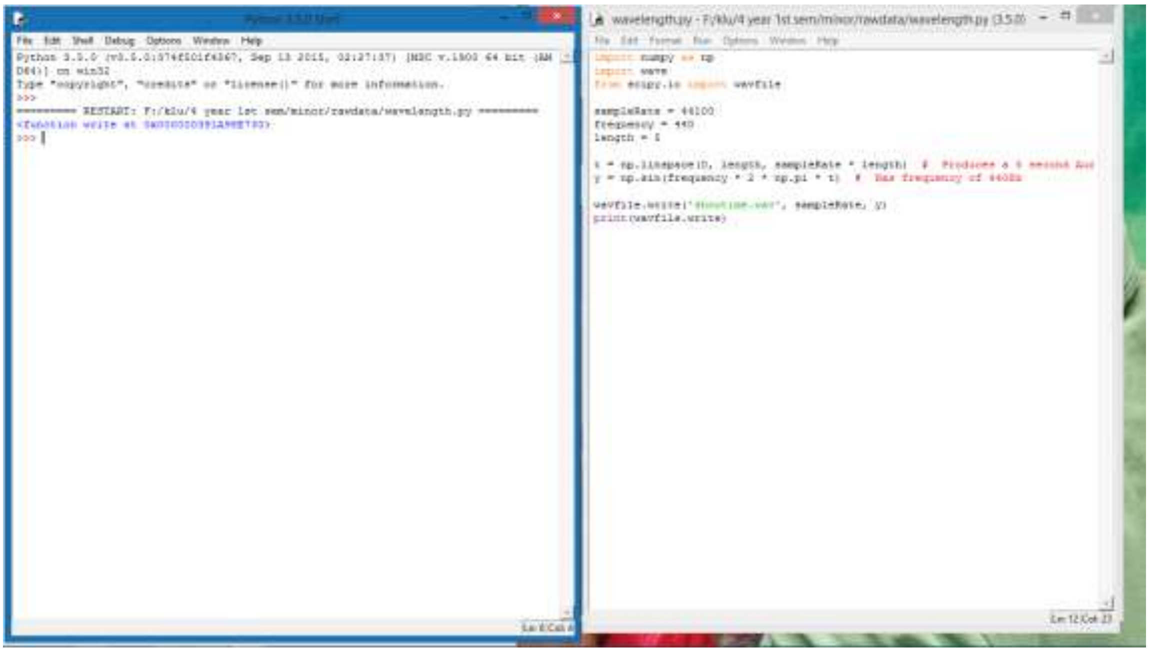
Figure 4 Discrete Fourier Transform (DFT)




Python 3.5.0 Shell
File Edit Shell Debug Options Window Help
File Edit Format Run Options Window Help

```
Python 3.5.0 (v3.5.0:374f501f4567, Sep 13 2015, 02:27:37) [MSC v.1900 64 bit (AMD64)] on win32
Type "copyright", "credits" or "license()" for more information.
>>>
========= RESTART: F:\[illegible]\4 year 1st sem\[illegible]\[illegible].py =======
==
(44100, array([ 0.00000000e+00,  6.26[illegible]e-02,  1.2[illegible]e-01, ...,
       -1.2[illegible]e-01, -6.2[illegible]e-02, -1.[illegible]e-12]))
>>>
```

```
from scipy.io.wavfile import read
import numpy as np
from numpy import*
import matplotlib.pyplot as plt
a=read("[illegible].wav")
print(a)
plt.show(a)
```



```
Python 3.5.0 (v3.5.0:374f501f4567, Sep 13 2015, 02:27:37) [MSC v.1900 64 bit (AM
D64)] on win32
Type "copyright", "credits" or "license()" for more information.
>>>
========= RESTART: F:/klu/4 year 1st sem/minor/rawdata/wavelength.py =========
<function write at 0x[illegible]>
>>> 
```

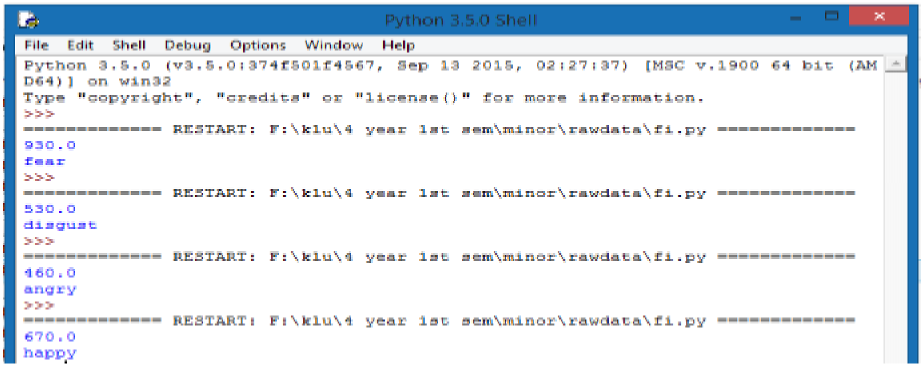
wavelength.py - F:/klu/4 year 1st.sem/minor/rawdata/wavelength.py (3.5.0)

```
import numpy as np
import wave
from scipy.io import wavfile

sampleRate = 44100
frequency = 440
length = 1

t = np.linspace(0, length, sampleRate * length)  #  Produces a 1 second Aud
y = np.sin(frequency * 2 * np.pi * t)  #  Has frequency of 440Hz

wavfile.write('[illegible].wav', sampleRate, y)
print(wavfile.write)
```



```
Python 3.5.0 (v3.5.0:374f501f4567, Sep 13 2015, 02:27:37) [MSC v.1900 64 bit (AM
D64)] on win32
Type "copyright", "credits" or "license()" for more information.
>>>
============= RESTART: F:\klu\4 year 1st sem\minor\rawdata\fi.py =============
930.0
fear
>>>
============= RESTART: F:\klu\4 year 1st sem\minor\rawdata\fi.py =============
530.0
disgust
>>>
============= RESTART: F:\klu\4 year 1st sem\minor\rawdata\fi.py =============
460.0
angry
>>>
============= RESTART: F:\klu\4 year 1st sem\minor\rawdata\fi.py =============
670.0
happy
```

The feelings are understood by the above approaches. The emotions are classified according to the speech frequencies obtained from waveforms. Growing emotion has some different frequencies, using those frequencies that we are classifying. That results in a person's giving of emotions.

4. CONCLUSION

In this step, in extracting the features, we have learned about several concepts and their approach form. Our approach to voice, speech recognition, emotions, forms of emotions, methods of communicating emotions, neural networks, frequency-using extraction and more. We came with feature extraction using frequencies, we used different frequencies to compare and differentiate between emotions and made use of several algorithms to detect and eliminate noise at last we specify that FFT is best among the algorithms to change the audio file. Our future target is to do web scraping from the website for further classification of the text in it.

5. References

[1] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. G. Taylor, “Emotion recognition in human-computer interaction,” IEEE Signal processing magazine, vol. 18, no. 1, pp. 32–80, 2001.

[2] F. Burkhardt, J. Ajmera, R. Englert, J. Stegmann, and W. Burleson, “Detecting anger in automated voice portal dialogs,” in Ninth International Conference on Spoken Language Processing, 2006.

[3] C. Vinola and K. Vimaladevi, “A survey on human emotion recognition approaches, databases and applications,” ELCVIA Electronic Letters on Computer Vision and Image Analysis, vol. 14, no. 2, pp. 24–44, 2015.

[4] M. El Ayadi, M. S. Kamel, and F. Karray, “Survey on speech emotion recognition: Features, classification schemes, and databases,” Pattern Recognition, vol. 44, no. 3, pp. 572–587, 2011.

[5] P. Chandrasekar, S. Chapaneri, and D. Jayaswal, “Automatic speech emotion recognition: A survey,” in Circuits, Systems, Communication and Information Technology Applications (CSCITA), 2014 International Conference on. IEEE, 2014, pp. 341–346.

[6] S. G. Koolagudi and K. S. Rao, “Emotion recognition from speech: a review,” International journal of speech technology, vol. 15, no. 2, pp. 99–117, 2012.

[7] A. Stuhlsatz, C. Meyer, F. Eyben, T. Zielke, G. Meier, and B. Schuller, “Deep neural networks for acoustic emotion recognition: raising the benchmarks,” in Acoustics, speech and signal processing (ICASSP), 2011 IEEE international conference on. IEEE, 2011, pp. 5688–5691

. [8] K. Han, D. Yu, and I. Tashev, “Speech emotion recognition using deep neural network and extreme learning machine,” in Fifteenth Annual Conference of the International Speech Communication Association, 2014.

[9] A. Satt, S. Rozenberg, and R. Hoory, “Efficient emotion recognition from speech using deep learning on spectrograms,” Proc. Interspeech 2017, pp. 1089–1093, 2017.

[10] D. Bahdanau, K. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate,” arXiv preprint arXiv:1409.0473, 2014.

[11] M.-T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," arXiv preprint arXiv:1508.04025, 2015.

[12] J. Zhang, J. Du, S. Zhang, D. Liu, Y. Hu, J. Hu, S. Wei, and L. Dai, "Watch, attend and parse: An end-to-end neural network based approach to handwritten mathematical expression recognition," Pattern Recognition, vol. 71, pp. 196–206, 2017.

[13] Linqin Cai,Yaxin Hu,Jiangong Dong,and Sitong Zhou, "Audio-Textual Emotion Recognition Based on Improved Neural Networks", Mathematical Problems in Engineering 2019.