

Automated Notes Maker from Audio Recordings



Guided By:
Mr. Ajay Tiwari

Team ID:
CSE-I-53

Team Members:

- 1. Aditya Mehta
(35114802718)**
- 2. Lakshay Chawla
(40814802718)**
- 3. Mehul Rekhi
(35814802718)**

Problem Statement

In the present industry, communication is the key element to progress. Passing on information, to the right person, and in the right manner is very important, not just on a corporate level, but also on a personal level. The world is moving towards digitization, so are the means of communication. Phone calls, emails, text messages etc. have become an integral part of message conveyance in this tech-savvy world. In order to serve the purpose of effective communication between two parties without hindrances, many applications have come to picture, which acts as a mediator and help in effectively carrying messages in form of text, or speech signals over miles of networks. Most of these applications find the use of conversion from speech signals to text. Devise a way for the conversion of voice-based recording of online lessons to text. Compare the accuracy of what is spoken in the voice-based recording the text . The project aims to provide students with text document of their online classes given they are voice based in nature. The accuracy must of the proposed solution must be taken into account.



Abstract



Ever wish you could just speak your thoughts into a document instead of writing or typing them or convert your online lectures to transcripts? We aim to provide a one stop solution for Speech to text conversion using Deep Learning, Convolutional Neural Networks.

Introduction

Scope

Communication among human beings is dominated by spoken language, therefore it is natural for people to expect voice interfaces with computers. This can be accomplished by developing a voice recognition system - speech-to-text which allows the computer to translate voice requests and dictation into text.

Motivation

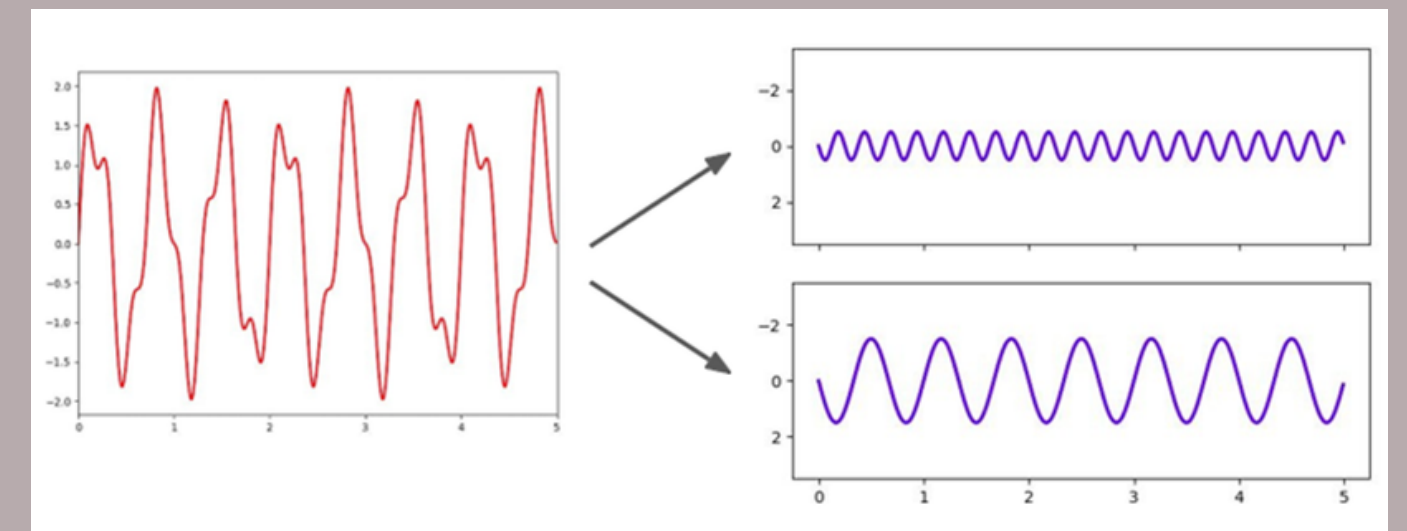
Moving ahead from our last project on Traditional Machine Learning algorithms, we wanted to enhance our knowledge in the field of Deep Learning. Audio domain in deep learning is fairly recent and unexplored. So, we decided to tap in this domain and make a project for the benefit of the society.

Challenges

Audio Domain in itself is a challenging topic. To come up with ideas to process audio signals was the first hurdle. Continuing to that , DEep learning (and audio processing for that matter) tends to be a lot more computationally expensive which adds another layer of challenge.

Audio Processing

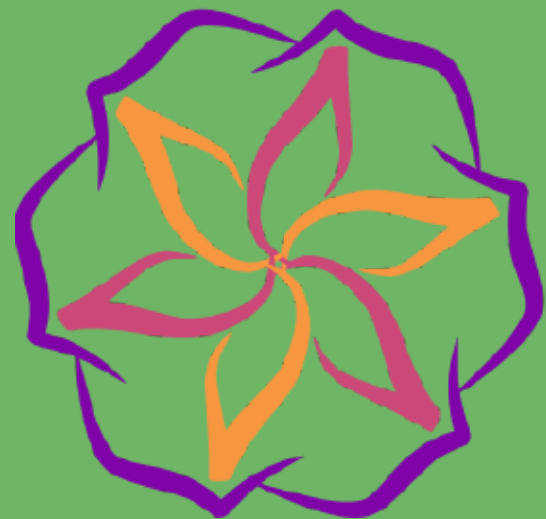
- Perform FFT (Fast Fourier Transform)
- Perform STFT (Short Time Fourier Transform) with Python
- Extract MFCCs



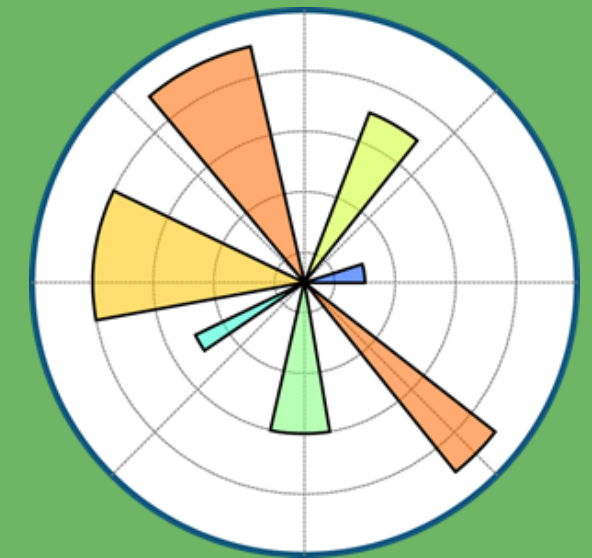
Tech Stack



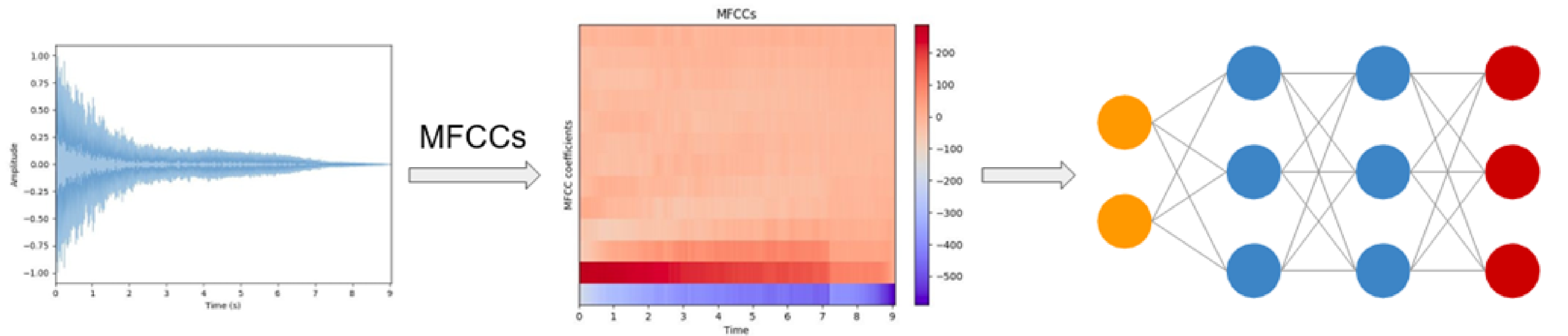
- Python language used to create deep learning models.
- Libraries Used –
 - Tensorflow (Keras)
 - Librosa
 - Numpy
 - Matplotlib



NumPy



Architecture



Results

Dataset	Number of Samples
Training set	17049
Validation Set	4263

Dataset	Accuracy
Training set	89.24%
Validation Set	83.2%

Statement	Time taken by project	Time Taken by google API
One dog on the tree house	0:00:00.001013	0:00:00.010130
Stop, go right	0:00:00.000993	0:00:00.014895

References



[1] Jingdong Chen, Member, Yiteng (Arden) Huang, Qi Li, Kuldip K. Paliwal, "Recognition of Noisy Speech using Dynamic Spectral Subband Centroids" IEEE SIGNAL PROCESSING LETTERS, Vol. 11, Number 2, February 2004.

[2] Hakan Erdogan, Ruhi Sarikaya, Yuqing Gao, "Using semantic analysis to improve speech recognition performance" Computer Speech and Language, ELSEVIER 2005.

[3] Chadawan Ittichaichareon, Patiyuth Pramkeaw, "Improving MFCC-based Speech Classification with FIR Filter" International Conference on Computer Graphics, Simulation and Modelling (ICGSM"2012) July 28-29, 2012 Pattaya(Thailand).

[4] Bhupinder Singh, Neha Kapur, Puneet Kaur "Speech Recognition with Hidden Markov Model:A Review" International Journal of Advanced Research in Computer and Software Engineering, Vol. 2, Issue 3, March 2012.

THANK YOU