Self-Supervised Speech Representation Learning

Basic Idea

Acoustic Feature Extraction will be performed on imported speech waveform dataset using Mel-frequency cepstral coefficients(**MFCC**s). Natural Language processing would involve use of RNN based end to end modelling followed by **BERT**, Transformer and LSTM. **GELU** (Gaussian Error Linear Unit) will be passed as the activation function, as it often performs better than ReLU for **NLP**.

Technologies Used

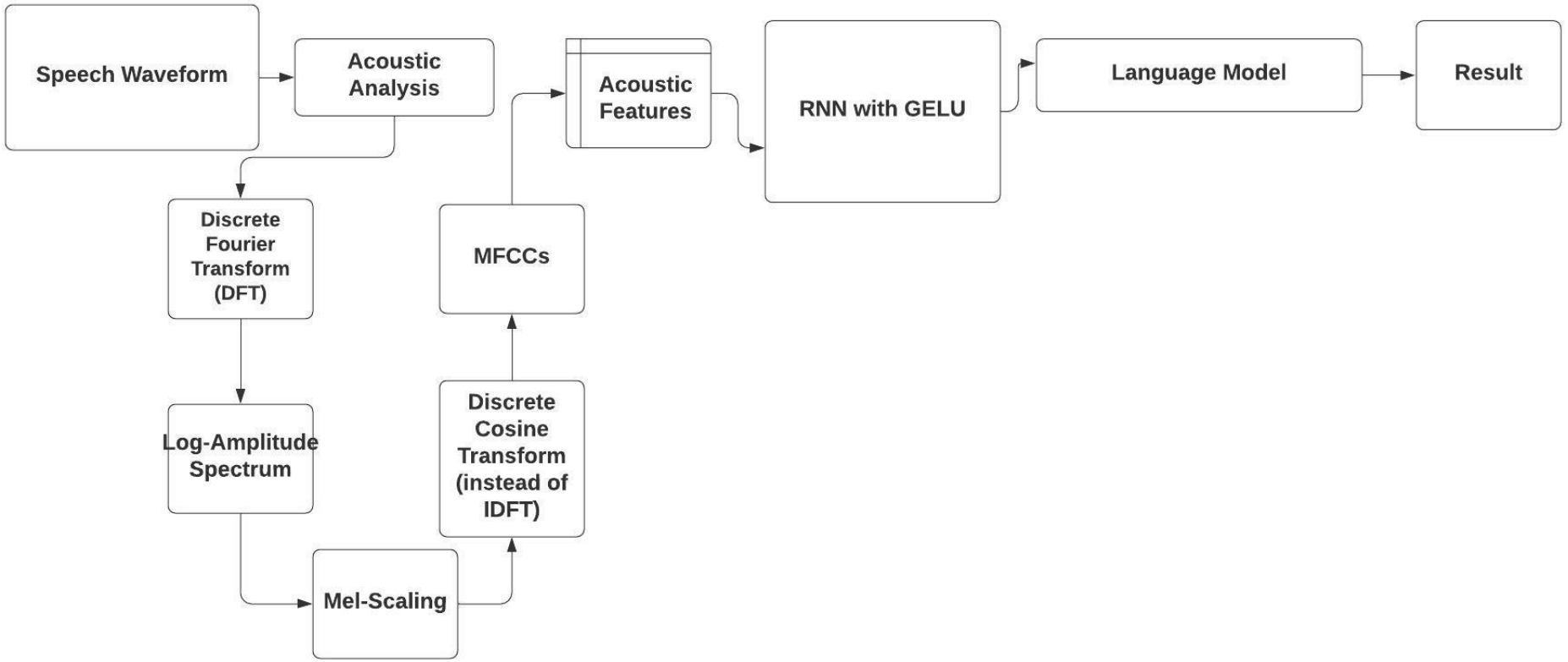
- Deep Learning
- Natural Language Processing(NLP)
- Neural Networks – RNN
- Bi-directional Encoding Representation for Transformers(BERT)
- Tensor Flow, Keras
- Self-Supervised Learning

Platform Identification

- **Google Colab** - Colab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education.

- **Anaconda Navigator** - Anaconda Navigator is a desktop graphical user interface (GUI) included in Anaconda distribution that allows you to launch applications and easily manage conda packages, environments, and channels without using command-line commands. Navigator can search for packages on Anaconda.org or in a local Anaconda Repository.

Architecture Design



References

- https://arxiv.org/pdf/2001.04316v2.pdf
- https://arxiv.org/pdf/1911.03912.pdf
- https://www.researchgate.net/publication/348219901_Hierarchical_Phoneme_Classification_for_Improved_Speech_Recognition
- https://www.researchgate.net/publication/337184235_Effectiveness_of_self-supervised_pre-training_for_speech_recognition
- https://ieeexplore.ieee.org/document/9383575
- https://arxiv.org/abs/2106.07447