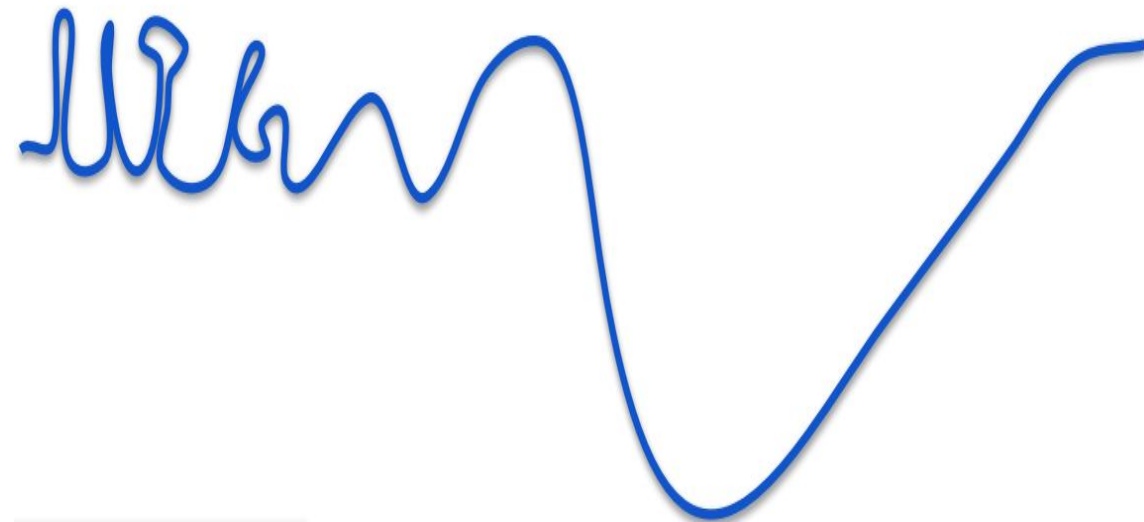


# Accent Transfer Techniques for Speech Synthesis



**SignalMinded**

Pradnesh Prasad Kalkar - 190101103

Mesharya M Choudhary - 190101053

Saket Kumar Singh - 190101081

Anirudh Phukan - 190101104

Korada Pavan Kumar - 190102093

# Topics covered in class

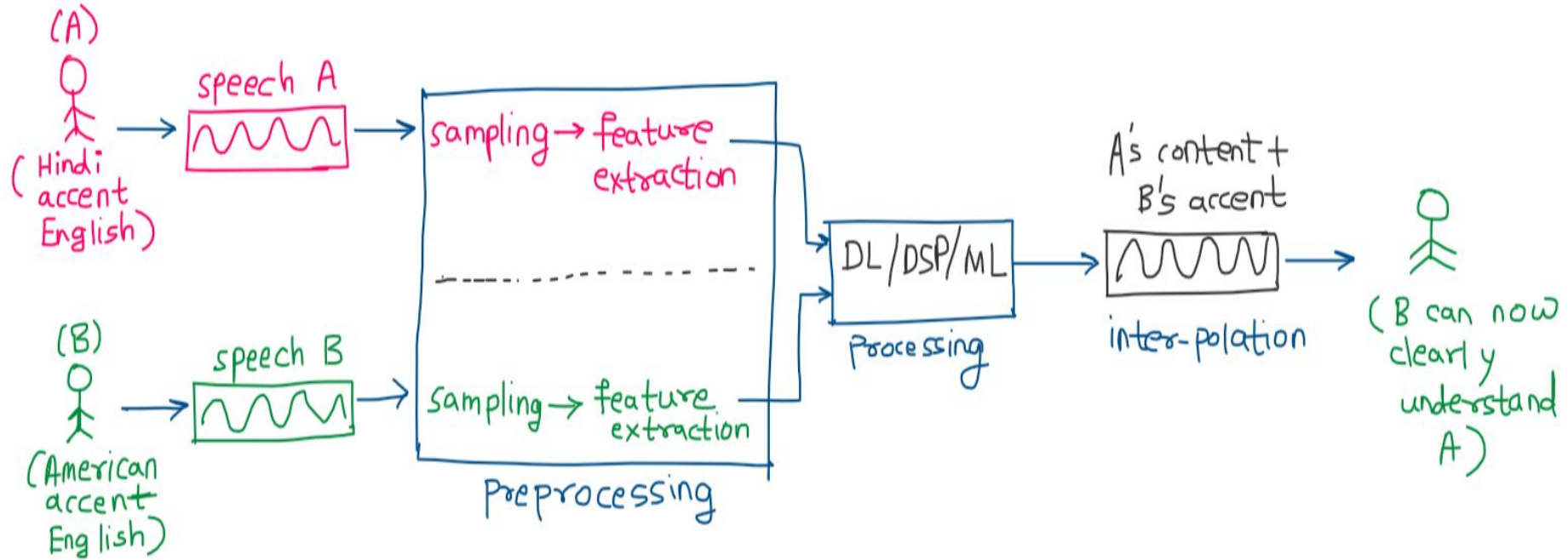
- Signals in our surroundings - audio, temperature, heart signals ...
- Representation of signals - taylor series, fourier series, polynomial...
- Signal processing - convolution, scaling, addition, filtering
- Signal sampling - Nyquist rate, Shannon sampling theorem
- Interpolation - sinc interpolation, lagrange interpolation
- Moving from continuous signals to discrete - Discrete Fourier Transform
- Efficient computation of DFT - FFT

# Hypothesis

The project aims to test the hypothesis that accent transfer techniques, such as deep neural networks, signal processing, or machine learning algorithms, can effectively extract the accent of one speaker and superimpose it on the content of another speaker.

The results of the study will provide insights into the effectiveness of accent transfer techniques and their potential applications in speech synthesis, language learning, and accent modification

# Block Diagram



# Dataset

- *Speech Accent Archive* dataset on Kaggle
- 2140 speech samples, distinct speakers, each speaking in English
- Speakers from 177 countries and 214 diverse native languages



Hindi Accent



American Accent

# Challenges

- Extract the accent features from one's speech without mixing the person's voice features in it
- Combining the above accent features with other person's content features without disturbing his/her voice features and generating a new speech

# Approaches to consider

- **Fourier Transform**
  - Take Fourier Transform of both input signals
  - Use a mathematical aggregation function to combine content of one and the style of other
  - Convert the aggregated signal back to time domain and then interpolate
- **Neural audio style transfer**
  - The process of audio neural style transfer involves using deep convolutional neural networks to extract the complex features from the style clip and then apply them to the content clip
  - A frozen pre-trained CNN is used to extract features from the spectrogram of the style and content clip. The output spectrogram is iteratively updated to minimize style and content loss. Finally the spectrogram is converted back to an audio signal to get a new audio clip
- **ML Methods**
  - Use of ICA (Independent Component Analysis) for feature extraction
  - A powerful tool for separating a mixed signal into independent components

# Validation of Hypothesis

- The output obtained from the model using the two speech signals would then be fed to a classifier and the performance of our model would be evaluated.
- If we are able to achieve a decent accuracy then that would mean that we have successfully validated the hypothesis.
- Otherwise it would mean that the methods explored by us failed to validate the hypothesis.



THANK YOU!