

# **Bengali speech recognition: Identifying the Writing Style of Bangla Language Using Natural Language Processing**

## **Introduction:**

Speech recognition may be an intuitive process for humans, but it turns out to be intimidating to make computers automatically recognize speeches. Although recent progress in speech recognition has been very promising in other languages, Bengali lacks such progress. There are very few research works published for Bengali speech recognizers. In this paper, we have investigated long short-term memory (LSTM), a recurrent neural network, approach to recognize individual Bengali words. We divided each word into a number of frames each containing 13 mel-frequency cepstral coefficients (MFCC), providing us with a useful set of distinctive features. We trained a deep LSTM model with the frames to recognize the most plausible phonemes. The final layer of our deep model is a softmax layer having an equal number of units to the number of phonemes. We picked the most probable phonemes for each time frame. Finally, we passed these phonemes through a filter where we got individual words as the output. Our system achieves a word detection error rate of 13.2% and a phoneme detection error rate of 28.7% on the Bangla-Real-Number audio dataset.

## **Challenges:**

- The inherent complexity of the language.
- Bengali words are intricate, and there is a shortage of comprehensive data to train effective recognition models.

## **LSTM Approach:**

- Long Short-Term Memory (LSTM), is a specialized type of recurrent neural network.
- Particularly well-suited for handling variable-length sequences.

## **Limitations:**

- HMMs also effectively modeled acoustic features and a small set of phonemes that make up words in various languages.
- Traditional models like HMM have practical limitations.
- Gaussian Mixture Models (GMM) were combined with HMM to address this issue.

### **Future Prospects:**

- Deep neural networks are likely to perform exceptionally well in Bengali speech recognition.
- In this study, the Bengali words are divided into frames, each containing 13 mel-frequency cepstral coefficients (MFCC).
- Trained a deep LSTM model to recognize the most probable phonemes.

### **Result Analysis:**

- Achieved a word detection error rate of 13.2%.
- Phoneme detection error rate of 28.7%.
- Results are based on the Bangla-Real-Number audio dataset.

### **Conclusion:**

- The work opens the door to more efficient and accurate Bengali speech recognition.
- Reduces the communication gap between humans and machines in this complex and linguistically rich language.
- Ability to harness the power of Long Short-Term Memory (LSTM) networks, which have proven to be exceptionally well-suited for recognizing individual Bengali words.