



# Progress Update

Mahir Morshed

15 October 2018

# Data Compilation

- Obtained as much Bengali audio and text as possible from All India Radio website
  - Requires matching of audio and text files due to inconsistencies on said website
- Performing optical character recognition (due to PDF issues) using Tesseract
  - Requires some cleanup to separate long phrases and sentences
- Developing corpus and dictionary based on derived lexicon
  - (At this point, might resort to heuristics to generate phonetic mapping)
  - Alternatively may consider using already-established language model

# Feature Investigation

- Exploring encodings of articulatory information alongside extracted features
  - Similar to approach described in [doi:10.21437/Interspeech.2018-2529](https://doi.org/10.21437/Interspeech.2018-2529) at first (simple determination via deep nets), but ideally closer to [doi:10.21437/Interspeech.2018-2122](https://doi.org/10.21437/Interspeech.2018-2122) later (basis on physical measurements for articulation)
- Determining best method for substituting MFCCs with wavelet coefficients
  - Initial thoughts based on [doi:10.21437/Interspeech.2018-2083](https://doi.org/10.21437/Interspeech.2018-2083) (mel-frequency discrete wavelet coefficients), but with different wavelet bases

# Implementation Logistics

- Weighing benefits of AWS, Azure, (Grainger machines? building a GPU-enabled PC?) for eventually executing experiments
- Resolved to perform initial experiments with Kaldi—wrapped with Python if possible—and later transition to fully PyTorch-based recipes