Dysarthric Speech Recognition Using a Deep Bidirectional LSTM

Jeremy Tate Campbell (jcamp12@stanford.edu) Stanford University – CS 230

Introduction

- **Goal:** Build a speech recognition system for people with dysarthria, a motor speech disorder caused by muscle weakness in the face, lips, tongue, or throat.
- **Input:** A single, isolated word from a speaker with dysarthria
- **Output:** The phonetic transcription of the word

Dataset & Features

- From UA-Speech Database (~120 hrs)
- Each file was a one-word utterance
- Words chosen using greedy algorithm to maximize uncommon phonemes

Transcriptions → phonemes

- command | k əh əm əe ən əd
- pajamas | ə p əh əj əe əm əh əz
- observation | əə ə b əv əv əy ə əh ən

Audio (WAV) file → filter banks

Each input feature vector \(x_i\) had 123 features from a 10ms window of the normalized filter bank

Model

- **Encoder:** Bidirectional LSTM
- **Decoder:** 1 fully-connected layer to a CTC decoder that condenses phonemes that aren’t separated by a “blank”
- **Loss:** Negative log-likelihood of true phoneme sequence given the softmax probabilities.
  \[ \text{Loss} = - \sum_{(x,y) \in B} \log p(y|x) \]

Results

- **Test Set PER:** 44.68% (~10,000 words)

Discussion

A PER of 45% is about what we expect – other state-of-the-art models on dysarthric speech have PERs of ~35%, and the subjects in our test set had below average speech intelligibility scores. Phonemes that were similar to another, such as “r” (as in their) and “er” (as in her), generally were less successful. Certain phonemes that are more stressful on speech muscles, such as “m” and “er” did particularly worse on dysarthric speakers. Overall, we believe our DBLSTM sufficiently fits the training data and successfully learns the inconsistent temporal acoustic cues present in dysarthric speech.

Future Work

- Incorporate prior knowledge of phonetic relationships
- Conduct a more thorough search of hyperparameter space
- Combine DBLSTM with beam search decoder
- Transfer learning using large corpus of non-dysarthric speech

References

Special thanks to Professor Mark Hasegawa-Johnson of the University of Illinois for allowing us access to the UK-Speech database he helped to create.