**SPEECH PROCESSING**

**PROJECT PROPOSAL**

**TOWARDS END-TO-END SPEECH RECOGNITION WITH TRANSFER LEARNING**

**GROUP MEMBERS:**

**ADEELA SAMAR (FA19-MSCS-0023)**

**SAMI-UR-REHMAN (FA19-MSCS-0022)**

**INSTRUCTOR:**

**DR TAFSEER AHMED KHAN**

Towards end-to-end speech recognition

with transfer learning

**Towards end-to-end speech recognition with transfer learning**

End-to-end training of automated speech recognition (ASR) systems requires massive data and compute resources. We explore transfer learning based on model adaptation as an approach for training ASR models under constrained GPU memory, throughput and training data. We show that this technique allows faster training on consumer-grade resources while requiring less training data in order to achieve the same accuracy, thereby lowering the cost of training ASR models in other languages. Model introspection revealed that small adaptations to the network's weights were sufficient for good performance, especially for inner layers.

Transfer learning (TL) is widely used in conventional hybrid automatic speech recognition (ASR) system, to transfer the knowledge from source to target language. TL can be applied to end-to-end (E2E) ASR system such as recurrent neural network transducer (RNN-T) models, by initializing the encoder and/or prediction network of the target language with the pre-trained models from source language. In the hybrid ASR system, transfer learning is typically done by initializing the target language acoustic model (AM) with source language AM. Several transfer learning strategies exist in the case of the RNN-T framework, depending upon the choice of the initialization model for encoder and prediction networks. This report presents a comparative study of four different TL methods for RNN-T framework.

Towards end-to-end speech recognition

with transfer learning