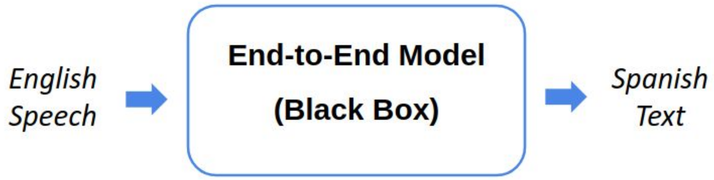
Combining ASR and NMT as *speech2texto*

Nick Hinke and Katie Brandegee

JOHNS HOPKINS WHITING SCHOOL of ENGINEERING

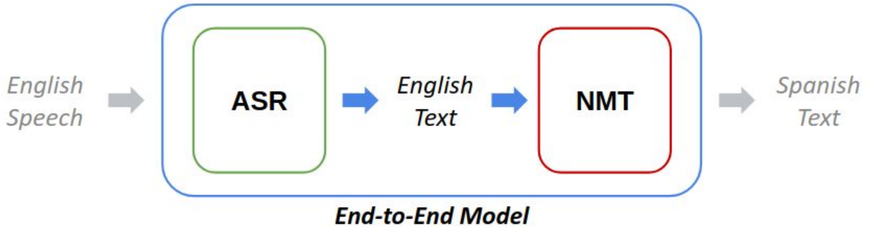
Problem Statement

- **Project Goal:** Convert audio samples of natural language (English) into translated text (Spanish)



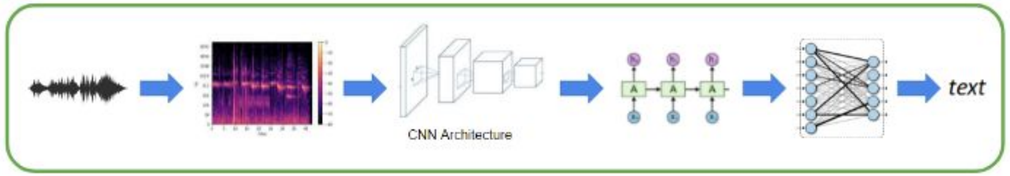
Overall Approach

- **Project Approach:** Construct end-to-end model using two-stage pipeline where each stage can be individually designed, trained, and evaluated before chaining them together
  - *Stage 1:* Automatic Speech Recognition (ASR)
  - *Stage 2:* Neural Machine Translation (NMT)

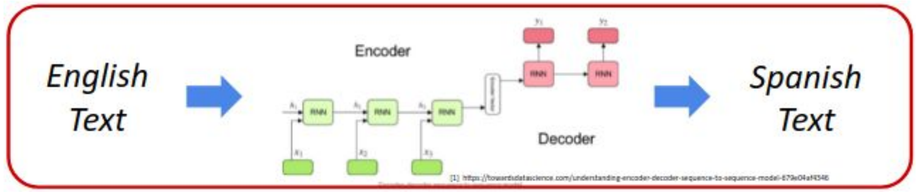


Solution

- **ASR Stage:** Audio Signal --> Spectrogram --> CNN --> RNN --> FC with Softmax --> Text



- **NMT Stage:** English Text --> Sequence to Sequence Model --> Spanish Text



Milestones

- **Preliminary Project Calendar:**

3/20 - 3/26	3/27 - 3/2	4/3 - 4/9	4/10 - 4/16	4/17 - 4/23	4/24 - 4/30
Preliminary Planning and Research	Convert input audio to images, develop an optimal CNN architecture	Develop an optimal RNN architecture, establish a smooth connection between networks, and tune hyperparameters	Develop the sequence to sequence model used in the NMT stage	Train and tune whole model using cross-validation	Address any issues that may arise from language context issues