* 1 Introduction: Chris
  + Question explored – best way to classify emotion in audio and can we apply it to political data
  + 2 Literature review: what is SER? Types of models that can be applied?
  + 3 Dataset: Crema-d. Size? Type of audio files? Classes?
* 4 Methodology & Results:
  + (2) CNN: Samiksha
  + (3) LSTM : Mengfei
  + (4) AST: Chris
  + (4) Pretrained model :Faisal
  + (5) Multi-model & Audio-text: Chris
* 6 Conclusion: Samiksha
* \

P1.

Introduction:

Waveform images

Spectrogram images

Examples audio snippets

P2.

CNN LSTM

Architecture (image) Architecture (image)

Results (image) results (image)

AST Transformer

Architecture Architecture

Results (image) Results (image)

Multimodal

Architecture

Results (image)

Summary table of all accuracy

P3.

Demonstration:

Input file

Play audio

Choose model?

Classify with model