

MSDS 631 Deep Learning

Task 1

Adam Ansari

Daniel Tinoco

Yusong Wang

Project Proposal

Goal:

Make [Samuel Jackson](#) sing the team members Happy Birthday.

Approach:

Create audio embedding using an audio to vector model ([speechbrain](#)), then fine tune a Speech-to-Speech ([speecho5_vc](#)) model from HuggingFace. More specifically train on audio data generated by the team members and create audio embeddings for the speaker (Samuel Jackson).

Resources:

[Example script for generating audio embeddings](#)

[Colab for Loading and using the T5 model](#)

[Original Paper on the T5 model](#)

[Original Paper for Audio to Vector model](#)

[Blog post about Hugging Face model](#)

Data Selection

Embeddings:

Data Format: .wav and .flac files

Size: 30 mb

Attributes: Passionate speech by Samuel Jackson from movies

Description:

[Audio files](#) of Samuel Jackson. Starting with 118 quotes. There are other sources online that have the exact format needed (.wav) as well.

Challenges: Cleaning the data of background noise and any additional speakers that may be in the audio clip.

Input Data:

Data Format: .wav and .flac files

Size: 65 mb

Attributes: Monotone, emotionless speech of the team members

Description:

MSDS 631 Deep Learning

Task 1

Adam Ansari

Daniel Tinoco

Yusong Wang

The team will record audio and use [Audacity](#) to remove background noise and generate clean recordings. The team anticipates having to record around a thousand 5-15 second audio clips.

Challenges: Recording audio is time consuming and the data needs to be monotone and without much emotion for the purposes of training. The recordings might also contain background noise which will require cleaning.