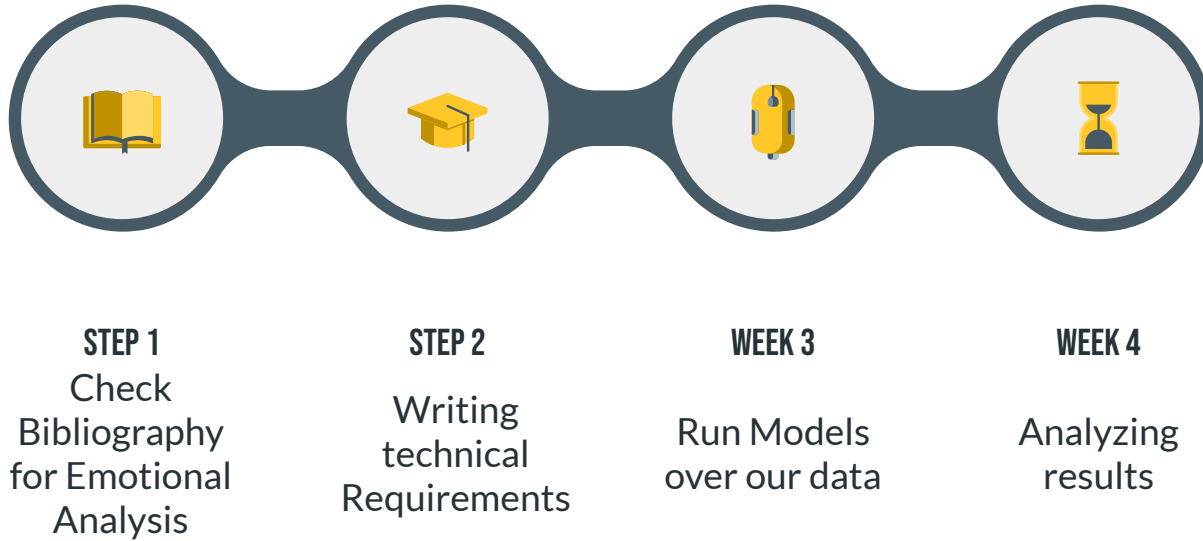


SPEECH EMOTION RECOGNITION

Katerina Xagorari
Marios Diamantopoulos
Christina Borovilou
DSIT - UOA



PROJECT TIMELINE



PROJECT OVERVIEW

At the following presentation we will see:

1. Data:
 - Get Emotion from Podcast Audio data (~113h)
 - Generate Text from audio files
 - Combine data from 2 pipelines
2. Models:
 - Use state-of-the art pretrained models for audio & text
 - Use the models to learn on your data
3. Field-Expert contribution:
 - Analyze emotions into vector-space of arousal & valence

GOALS

1. Find an unbiased real-world dataset to train our model over Sentiment analysis though Speech. The model needs to be annotated with more analyzed voice characteristics like arousal, valence and dominance.
2. Fine-tune a model built for this specific problem using both Speech text content and voice sentiment
3. Evaluate model and its efficacy

01

WAV2VEC

Contextualized speech
representations

02

DISTILROBERTA

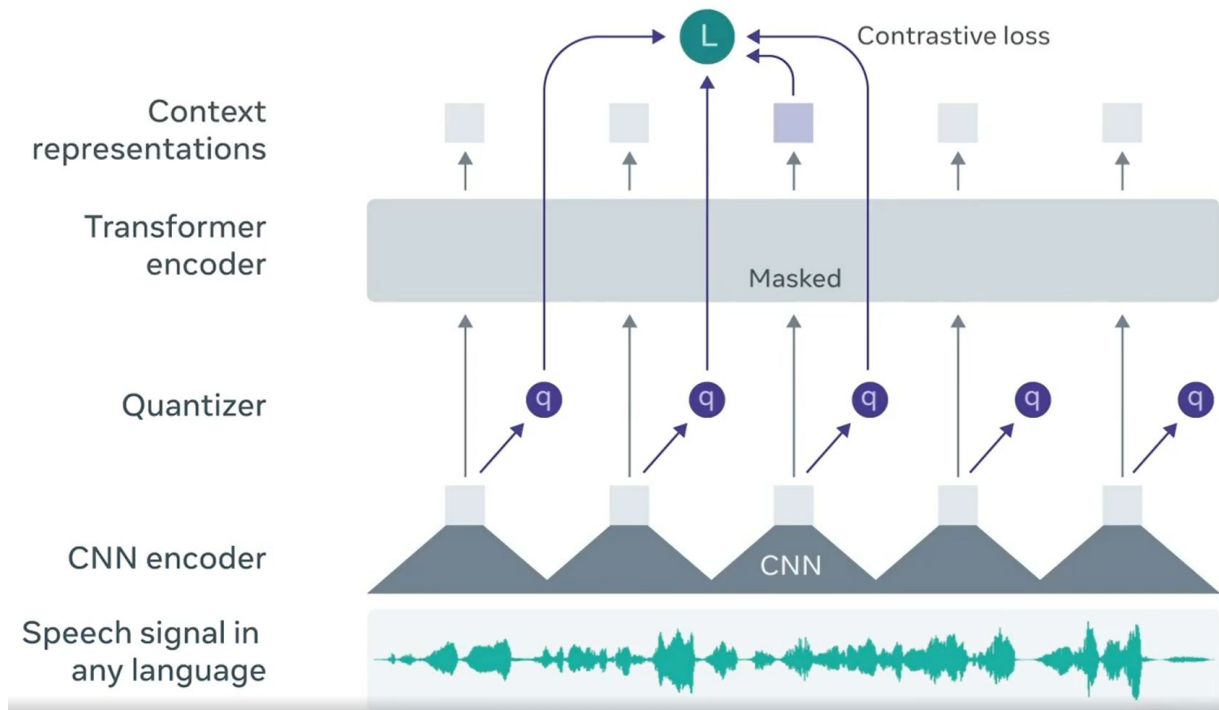
Pretrained on the raw texts only,
without humans labelling

PRETRAINED MODELS



WAV2VEC 2.0

Learns contextualized speech representations by randomly masking feature vectors before passing them to a transformer network



DISTILROBERTA

A distilled version of the RoBERTa-base model.

It was pretrained on the raw texts only, with no humans labelling them.

An automatic process generates inputs and labels

Select model: distilroberta-base

Input Sentence: The girl ran to a local pub to escape the din of her city. Update

Filters: Hide Special Tokens ☒ Show top 70% of att:

Layer: 1 2 3 4 5 6

Selected heads: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

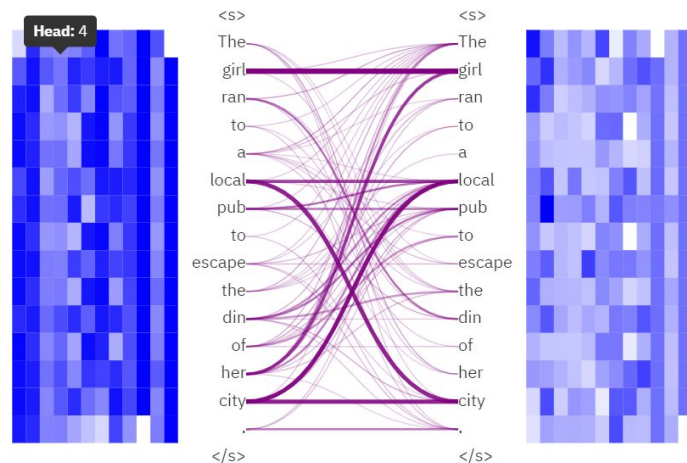
Select all heads

Unselect all heads

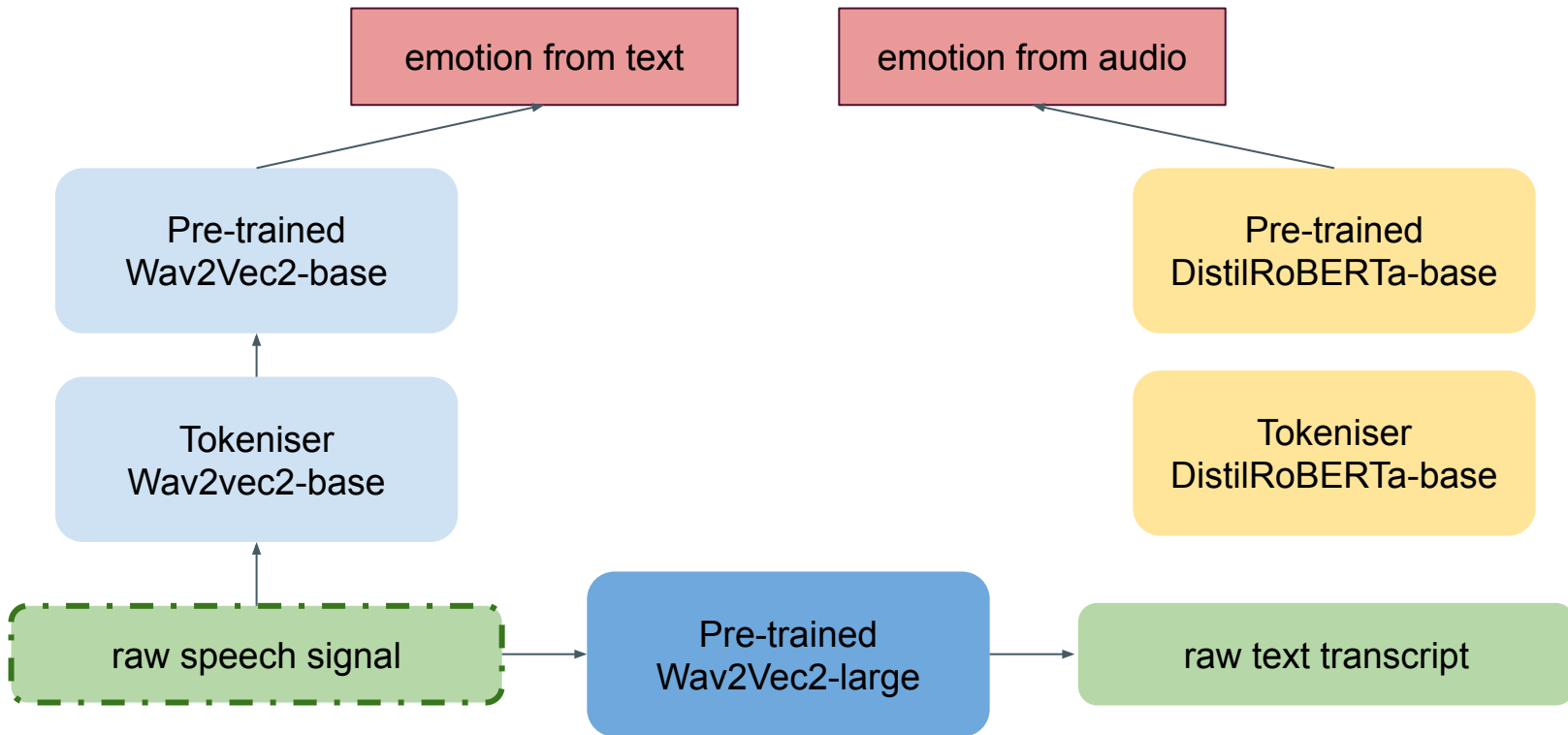
You *focus* on one token by **click**. For bidirectional models, you can *mask* any token by **double click**.

You can *toggle* a head by a **click** on the heatmap columns.

Tokens on the *left* attend to tokens on the *right*.



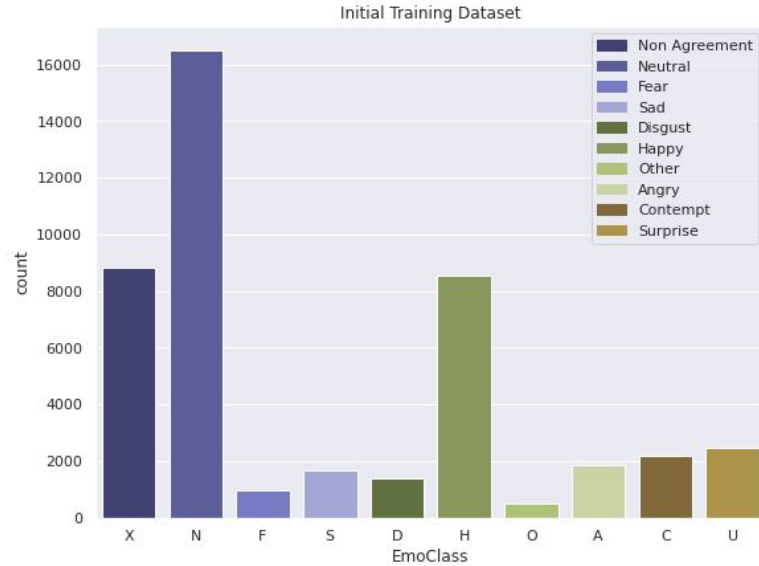
ARCHITECTURE



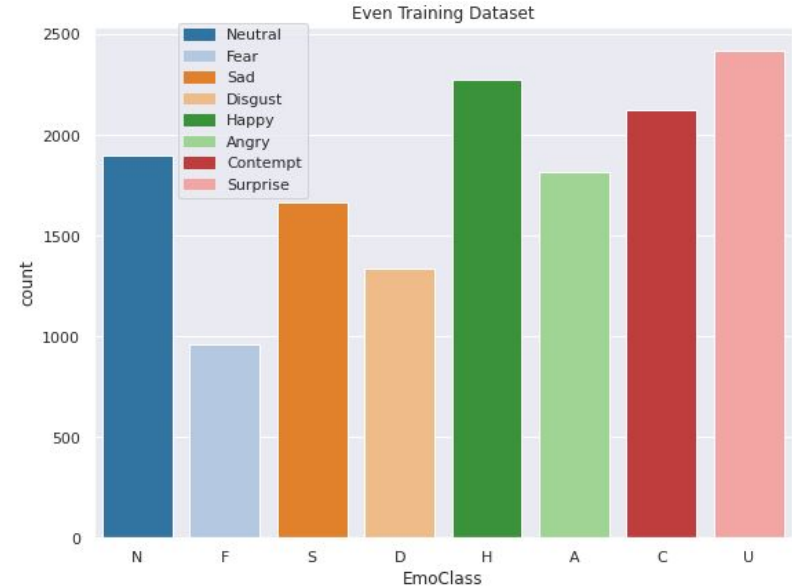


01. HANDS ON DATA

DISTRIBUTION OF DATA AMONG LABELS



ORIGINAL DATASET



RESHAPED DATASET

QUESTIONS?

