

CS 72: ACCELERATED COMPUTATIONAL LINGUISTICS

DATASET INFORMATION

Carlos Guerrero Alvarez, Amittai Siavava, Aiwei Zhang

03/12/2024

The dataset we used is an aggregation of four different datasets. Because of this reason, we felt that splitting it for submission may result in a non-representative sample since the files from each dataset are quite different.

We have the data in our Google Drive, which can be accessed here:

<https://drive.google.com/drive/folders/1k2FYbPnaTV8RmU8VgW6sQHkum20paS5q?usp=sharing>. However, the unzipped dataset is quite huge. If you prefer to download the compressed version from Kaggle, the relevant link is here: <https://www.kaggle.com/datasets/uldisvalainis/audio-emotions/data>.