Data Report

1. Introduction

IEMOCAP(Interactive Emotional Dyadic Motion Capture Database) collected by the Speech Analysis and Interpretation Laboratory (SAIL) at the University of Southern California (USC).

- Identifier: USC_IEMOCAP
- Modalities: audio, video, Motion Capture
- Emotional content: angry, happy, sad, neutral, frustrated, excited, fearful, disgusted, excited, other
- Emotion elicitation methods: scripts and improvisation
- Size: 10 subjects, ~12 hours of audiovisual data
- Number of splitted audio based on sentence: 10039 samples
- Nature of material: dyadic interactions
- Language: English

2. Structure of dataset

Dataset includes documentation, 5 sessions and file readme.txt:

IEMOCAP_full_release
|------ Documentation
|------|----- corpus.dic
|------|----- Five_face_markers2.png
|------|----- HumaineInfo.txt
|------|----- phonemes.txt
|------|----- timeinfo.txt
|------ Session1 (all Sessions is same)
|------|---- dialog
|------|-----|---- avi
|------|-----|---- EmoEvaluation
|------|-----|---- lab
|------|-----|---- MOCAP_hand
|------|-----|---- MOCAP_head
|------|-----|---- MOCAP_rotated
|------|-----|---- transcriptions
|------|-----|---- wav
|------|---- sentences
|------ Session2
|------ Session3
|------ Session4

|------ Session5
|------ README_backup.txt
|------ README.txt

- The documentation comprises guidelines for the entire dataset, including the corpus, phonemes, time information, and visual representations of facial markers.
- Five folders "Session" contains two subfolders, "dialog" and "sentences."
+ Within the "dialog" folder, there are videos, labels for each sentence in every audio file, recorded data on hand gestures, head movements, dialogues, and audio files for each recording.
+ The "sentences" directory also contains similar data as the "dialog" directory but has been processed and segmented into individual sentences.
- File “README.txt” includes instructions to use this dataset and lincense.

3. Quality of dataset

- The data was recorded in a professional recording studio and encompasses all necessary facets, including audio, video, and transcriptions, making it highly suitable for current research and applications.
- The recordings were performed by multiple individuals, both following predefined scripts and engaging in spontaneous dialogues, ensuring diverse linguistic expressions.
- The data has been processed to segment it into individual sentences.
- The label assessment methodology for the dataset is also rigorous, incorporating multiple factors and perspectives, allowing various emotion experts to assess and select appropriate labels.

=> This dataset is comprehensive and highly suitable for the project involving emotion analysis from speech.

4. Conclusion

The process of dataset selection has undergone meticulous filtering and analysis. The IEMOCAP dataset exhibits a high level of completeness, featuring diverse data types and a substantial volume. Its objectivity ensures a robust foundation for preparing an effective model for emotion analysis based on audio and text.