EmoTale Datasheet

In line with the proposal on *datasheets for datasets* by Gebru et al. [1], we provide the datasheet for the EmoTale corpus, also available as a standalone document with the dataset.

A. Motivation

For what purpose was the dataset created?
Unavailability of Danish affect data sets not only impedes the development of the technology, but also impacts the validation of existing methods on Danish speakers. The introduction of our corpus is necessary to, at the very least, be able to validate the performance of SER models for the Danish language.
Who created the dataset and on behalf of which entity?
The dataset was created by Maja Jønck Hjuler, Line Katrine Harder Clemmensen, and Sneha Das at the Technical University of Denmark.
Who funded the creation of the dataset?
The dataset creation is funded by the larger WristAngel project which is funded by an exploratory Synergy grant from the Novo Nordisk Foundation and is a collaboration with Copenhagen University Hospital, the Child Psychiatry Research Unit.

B. Composition

What do the instances that comprise the dataset represent?
The instances are audio files of enacted emotional speech in Danish and in English. The speakers enact predefined sentences while expressing a predefined emotions.
How many instances are there in total?
The EmoTale corpus consists of a total of 800 audio instances, comprising 450 emotional speech recordings in Danish and 350 in English. Each recording features one of five different enacted emotions, and the dataset is balanced across these emotions.
What data does each instance consist of?
Each instance consists of raw audio data, in WAV format, captured at a sampling frequency of 48 kHz. Each recording corresponds to one of five enacted emotions; Neutral, Anger, Sadness, Happiness, or Boredom, and is based on predefined sentences that are translations from the German Emo-DB corpus, designed to be emotionally neutral to minimize contextual bias.
Is there a label or target associated with each instance?
In addition to the enacted emotion, three independent annotators provided four labels per instance; one categorical for the emotion chosen from the five possible classes, and three numerical for arousal, valence, and dominance in a range of 1 to 5 with increments of 0.5. The ranges are defined as: Valence [1-negative, 5-positive], activation [1-calm, 5-excited], and dominance [1-weak, 5-strong].
Is any information missing from individual instances?
Everything is included. No data is missing.
Are there recommended data splits?
There are no recommended data splits for training, validation, and testing within the dataset itself. However, it is common practice to create stratified splits across speakers and emotions.
Are there any errors, sources of noise, or redundancies in the dataset?
See preprocessing below.
Does the dataset contain data that might be considered confidential?
The data does not contain any signals reflecting on the state of an individual, minimizing the potential negative impact on the individuals.
Does the dataset identify any subpopulations?
Participants range in age from 9 to 39 years. The dataset includes 18 participants, with 12 females and 6 males.
Is it possible to identify individuals, either directly or indirectly from the dataset?
Individuals can be identified indirectly from the EmoTale corpus due to the unique characteristics of each participant's voice, which can reveal their identity. All participant information has been pseudoanonymized by assigning random IDs.

C. Collection Process & Preprocessing

How was the data associated with each instance acquired?
The data recordings were performed in several sessions in different locations. In each session, the participant was placed in a quiet room and fitted with wireless RØDE microphones paired with the corresponding receiver. Five sentences were enacted with five different emotions, and the participant enacted all sentences for a specific emotion before moving on to the next. The participants were allowed to repeat sentences as often as they liked, but only the last recording was kept in the database. Most often, the recording was made in the first attempt.
Who was involved in the data collection process?
Participants with acting experience and Danish and English language skills were recruited through physical flyers and posts on social media, and theater schools in the Greater Copenhagen area were contacted by email and phone.
Over what timeframe was the data collected?
The data was collected as part of master thesis work lasting 5 months.
Were any ethical review processes conducted?
Ethical approval was obtained from the institutional review board prior to the study [2].
Did the individuals in question consent to the collection and use of their data?
Abiding by GDPR requirements, written consent was obtained from participants or their guardians prior to data collection.
Was any preprocessing/cleaning/labeling of the data done?
Some instances were cropped to exclude silent 'clicks' from the experimenter pressing the keyboard at the beginning or end of recordings. The audio files are named according to the same template including information about the language, speaker ID, emotion, and sentence. For example, the

file *DK_004_A_5.wav* contains the fifth sentence spoken by speaker 004 in Danish, with angry affect.

Was the "raw" data saved in addition to the preprocessed/cleaned/labeled data?

Yes. The authors can provide the raw data upon request.

D. Uses

Has the dataset been used for any tasks already?

The dataset paper investigates the dataset's capacity in predicting speech emotions through the development of speech emotion recognition models using Self-Supervised Speech Model embeddings and the openSMILE feature extractor. Furthermore, cross-sorpus transferability of the models was investigated.

What (other) tasks could the dataset be used for?

The dataset can also be used for ASR, due to the availability of speech and the corresponding transcription. The enacted English speech in addition to Danish will aid research and investigation into speech systems, for instance when the speaker remains identical, but language changes, hence towards more universal speech emotion models.

Are there tasks for which the dataset should not be used?

Given the size of the dataset, it should not be used for tasks that require large-scale training of complex machine learning models. Additionally, it is not suitable for tasks that require spontaneous emotional speech, as the recordings consist of enacted emotions rather than natural emotional expressions.

E. Distribution & Maintenance

How will the dataset will be distributed?

The dataset can be accessed at https://github.com/snehadas/EmoTale.

Will the dataset be distributed under a copyright or other intellectual property (IP) license?

The data will be distributed under a copyright. There is no license, but there is a request to cite the corresponding paper if the dataset is used.

Who will be maintaining the dataset and how can they be contacted?

The dataset will be maintained by the corresponding author Sneha Das (sned@dtu.dk).

Will the dataset be updated?

This dataset will not be updated in terms of the number of samples or participants.

REFERENCES

[1] T. Gebru, J. Morgenstern, B. Vecchione, J. W. Vaughan, H. Wallach, H. D. III, and K. Crawford, "Datasheets for datasets," 2021. [Online]. Available: https://arxiv.org/abs/1803.09010

[2] "Ethical approval application to the IRB for DanskEmoTale: a pilot study," https://github.com/DTUComputeStatisticsAndDataAnalysis/Analysis-of-emotions-using-physiological-signals-a-pilot-study/blob/main/Analysis_plan_modifed.pdf, Sept 2022.