Notes\_on\_datasets

I use a text editor (Sublime) to help me count occurrences within datasets.

# Aesdd

From the readme

For the creation of v.1 of the database, 5 (3 female and 2 male) professional actors were recorded. 19 utterances of ambiguous out of context emotional content were chosen. The actors acted these 19 utterances in every one of the 5 chosen emotions. One extra improvised utterance was added for every actor and emotion. The guidance of the actors and the choice of the final recordings were supervised by a scientific expert in dramatology. For some of the utterances, more that one takes were qualified. Consequently, around 500 utterances occured in the final database.

No gender labels, but I manually determined that speakers 1, 2, and 5 are female-sounding and speakers 3, 4, and 6 are male-sounding. Contrary to the documentation, the dataset actually contains 6 different speakers.

# Anad

There’s something up with the segmented files; they all sound the same?

Labels at the discourse level? Low face validity

# BAUM1

There are misspellings in the emotion labels. Also, “Subject Label”, “Clip”, and “Clip Name” don’t always match.

First pass cleaning in Excel (data\_dir.xslx). Record counts didn’t match, so I redid it in Python.

From the paper

The data was collected from 31 subjects, 17 of which are female, which are shown in Fig. 2. All subjects are native speakers of Turkish, and have an age range of 19-65.

…



The paper is inconsistent on the actual number of participants coded female.

S015\_007 spontaneous emo label does not match emo code in annotations!

Some mp4s no audio? Check downstream

Boredom is mapped to negative valence (van Tilburg & Igou, 2017; Kort, Reilly, & Picard, 2001); Regan, Mandryk, & Atkins (2007)

Interest is mapped to positive valence (Kort, Reilly, & Picard, 2001)

Contempt mapped to negative (Trnka, Mana, & Kuška, 2021; Ferran 2017; Melwani, Mueller, & Overbeck, 2012)

Surprise is negative (Noordewier & Breugelmans 2013; Koch, Alves, & Krüger 2016; Noordewier, Topolinski, & Van Dijk 2016)

# BAUM2

Extensive manual Pre-cleaning done in data\_dir.xlsx; stages of processing go from leftmost sheet to rightmost sheet

Recoded to valence and recounted majority vote

Turkish and English

# Cafe

Surprise is now negative

“This dataset includes six different sentences, pronounced by twelve actors, in six basic emotions plus one neutral emotion. The basic emotions are acted in two different intensities. This represents a total of 936 different audio samples.”

All samples accounted for

Québec French

# CREMA-D

“There are 91 actors, 48 male and 43 female (51 actors worked with one director, 40 with the another). The actors were between the ages of 20 and 74 with a mean age of 36. Table 2 provides detailed age information. Several racial and ethnic backgrounds were represented in the actor group: Caucasian, African American, Hispanic, and Asian. Table 3 provides a detailed breakdown of the racial and ethnic groups.”

FROM README:   
CREMA-D is a data set of 7,442 original clips from 91 actors. These clips were from 48 male and 43 female actors between the ages of 20 and 74 coming from a variety of races and ethnicities (African America, Asian, Caucasian, Hispanic, and Unspecified).

Actors spoke from a selection of 12 sentences. The sentences were presented using one of six different emotions (Anger, Disgust, Fear, Happy, Neutral, and Sad) and four different emotion levels (Low, Medium, High, and Unspecified).

Participants rated the emotion and emotion levels based on the combined audiovisual presentation, the video alone, and the audio alone.

Crowd-sourced ratings

“Binomial majority is used to define majority recognition. Unlike traditional majority, which is defined as more than 50% of raters having selected the specific emotion, binomial majority is achieved when a binomial test would reject at the 95% confidence level the null hypothesis that the most commonly chosen label is selected randomly from the six possible labels.”

* Recode intended emotions to valence
* Recode votes to valence
  + A, D, F, S -> -1, N -> 0, H -> 1
  + H:N -> 1
  + Remaining -> -1 if the vote string doesn’t contain “N”
* Retain the record if any of VoiceVote valence, FaceVoteValence, or MultiModalVote valence matches the intended valence

Discarded 569, kept 6873

Intended emotion used for final valence label

# dzafic

Just 6 samples, so I created the tsv manually

# ekorpus

The corpus contains 1,234 Estonian sentences that express anger, joy and sadness, or are neutral. [867 retained]

Female voice, 44.1 KHz, 16Bit, Mono;

wav, textgrid: phonemes, words, sentences.

The audio-recordings and text of sentences can be downloaded and saved.

^ <https://metashare.ut.ee/repository/browse/estonian-emotional-speech-corpus/4d42d7a8463411e2a6e4005056b40024a19021a316b54b7fb707757d43d1a889/>

Elicited spontaneously

Used <http://peeter.eki.ee:5000/reports/valence>

To filter only samples with >51% rater recognition for positive, negative, and neutral valence

[Text influence: all

Emotion: all

Min. recognition %: 51]

Filter op returned list of sample IDs

Used the corresponding textgrid field for emotion category label

69 Sample IDs missing from dataset: 466, 468, 475, 481, 485, 493, 495, 499, 509, 511, 521, 550, 552, 554, 556, 558, 564, 570, 572, 578, 582, 586, 588, 592, 594, 598, 608, 612, 626, 638, 642, 469, 471, 473, 507, 515, 519, 525, 527, 529, 531, 533, 535, 537, 541, 545, 547, 548, 560, 562, 566, 568, 574, 584, 590, 596, 600, 602, 604, 606, 610, 614, 620, 622, 624, 632, 634, 640, 644, 646

253 sample IDs discarded since perceived valence did not match intended valence:  
120325, 13875, 13701, 13935, 13737, 120706, 120732, 13673, 14039, 121322, 121324, 173, 217, 229, 279, 285, 13435, 13473, 13525, 13661, 13677, 13683, 13687, 13743, 13799, 13877, 13879, 13903, 13917, 13923, 13929, 13967, 13979, 120151, 120161, 120171, 120173, 120203, 120221, 120225, 120227, 120233, 120317, 120365, 120526, 120532, 120534, 120536, 120572, 120590, 120692, 120712, 120718, 120744, 120901, 120909, 120961, 121141, 121188, 121208, 121286, 121312, 171, 121330, 13639, 14131, 14163, 14167, 14179, 14353, 120263, 120311, 120403, 120568, 120596, 120598, 120600, 120616, 120676, 120738, 120819, 120857, 120943, 120951, 120955, 120963, 120971, 121003, 121123, 121145, 121156, 121194, 121206, 57, 91, 175, 269, 395, 13249, 13311, 13335, 13341, 13627, 43, 120279, 120281, 120845, 65, 69, 120323, 121113, 120337, 127, 13583, 121214, 120367, 120369, 121222, 167, 185, 195, 121121, 121266, 120401, 121232, 13643, 13649, 13773, 13949, 13983, 120923, 120445, 120451, 120457, 120540, 14125, 14129, 121137, 14137, 121139, 120586, 120588, 121234, 14165, 121150, 121268, 120967, 120975, 14201, 120979, 120634, 120642, 120995, 14249, 120662, 14259, 14261, 121174, 13001, 121007, 120670, 14281, 14283, 13033, 13071, 120680, 121182, 121051, 121184, 13091, 121186, 120742, 14375, 120750, 13179, 120754, 121276, 121067, 121071, 121073, 14451, 120768, 14463, 120770, 14477, 13255, 121288, 120776, 121198, 120780, 14543, 14547, 13331, 121202, 120814, 120163, 120175, 120177, 120189, 120829, 120831, 13451, 120833, 13459, 120835, 120839, 120696, 13009, 120782, 14295, 13251, 121075, 120219, 13679, 120877, 120881, 120895, 120931, 121041, 121152, 121244, 121310, 305, 13077, 13085, 13181, 13199, 13257, 13267, 13433, 13675, 13721, 13839, 13843, 13859, 13965, 14035, 14109, 14205, 14209, 120139, 120147, 120215, 120297, 120315, 120347, 120355, 120359, 120381, 120538, 120644, 120690, 120788

# EmoDB

From abstract…

Ten actors (5 female and 5 male) simulated the emotions, producing 10 German utterances (5 short and 5 longer sentences) which could be used in everyday communication and are interpretable in all applied emotions.

The recordings were taken in an anechoic chamber with high-quality recording equipment. In addition to the sound electro-glottograms were recorded. The speech material comprises about 800 sentences (seven emotions \* ten actors \* ten sentences + some second versions).

The complete database was evaluated in a perception test regarding the recognisability of emotions and their naturalness. Utterances recognised better than 80% and judged as natural by more than 60% of the listeners were phonetically labelled in a narrow transcription with special markers for voice-quality, phonatory and articulatory settings and articulatory features.

The database can be accessed by the public via the internet (<http://www.expressive-speech.net/emodb/>). 🡨 link is no good

From <https://www.kaggle.com/piyushagni5/berlin-database-of-emotional-speech-emodb> :

The EMODB database is the freely available German emotional database. The database is created by the Institute of Communication Science, Technical University, Berlin, Germany. Ten professional speakers (five males and five females) participated in data recording. The database contains a total of 535 utterances. The EMODB database comprises of seven emotions: 1) anger; 2) boredom; 3) anxiety; 4) happiness; 5) sadness; 6) disgust; and 7) neutral. The data was recorded at a 48-kHz sampling rate and then down-sampled to 16-kHz.

Every utterance is named according to the same scheme:

* Positions 1-2: number of speaker
* Positions 3-5: code for text
* Position 6: emotion (sorry, letter stands for german emotion word)
* Position 7: if there are more than two versions these are numbered a, b, c ....

Example: 03a01Fa.wav is the audio file from Speaker 03 speaking text a01 with the emotion "Freude" (Happiness).

Table

Description automatically generated

# EmoReact\_V\_1.0

Train-val-test split provided, keeping speakers separated between splits

Discard the split in favor of re-splitting, stratifying by language and dataset (and gender?); speaker gender was unbalanced by split too

multimodal emotion dataset of children between the ages of four and fourteen years old. The dataset contains 1102 audio-visual clips annotated for 17 different emotional states: six basic emotions, neutral, valence and nine complex emotions including curiosity, uncertainty and frustration.

Regan, Mandryk, & Atkins (2007) associate excitement with positive valence and frustration with negative valence and boredom is negative

Lang (1995); Russell, Weiss, & Mendelsohn (1989) – excitement is positive

Carleton (2016); Anderson, Carleton, Diefenbach, & Han (2019): uncertainty is negative

Curiosity gets mixed reviews: van Lieshout, Traast, de Lange, Cools (2019); Shin & Kim (2019); Noordewier & van Dijk (2017)

Nojavanasghari, Baltrusaitis, Hughes, & Morency (2016) looked at EmoReact and found that curiosity was associated with positive valence. Hill, Fombelle, & Sirianni (2016) also treat it as positive. So for EmoReact, curiosity could be treated as positively valenced.

Not all emotions in the paper are in the dataset labels… because of v1?

Present: curiosity, excitement, happiness, uncertainty, surprise, disgust, fear, frustration

missing: exploration, confusion, anxiety, attentiveness, anger, sadness, embarrassment

regarding annotating speaker identity, the paper cites Florian et al. as reference #47, but the 47th reference is Schroff, Kalenichenko, & Philbin 2015: “A unified embedding for face recognition and

clustering.”, which doesn’t sound like it’s about annotating speaker identity/gender

Anyways, I’m manually annotating gender of the child subjects myself, mostly using the thumbnails and listening to the audio when I’m not as sure

Sometimes it’s the adult interviewer’s voice, not the child subject; sometimes both. It is unclear whether the label votes are for the interviewer or child subject in these cases, so where I identified it, I manually omit the observation: VCR107\_2.mp4, BULLYING27\_2.mp4, GAMEBOY19\_2.mp4. I didn’t check every single one, however. The interviewer is usually (always?) male sounding

Some are non-verbal vocalizations

Every sample has a valence rating

Multiple ethnicities, NA English

# Emotional\_EMA

From readme…

‘’’

This Electromagnetic Articulography (EMA) database includes articulatory motions recorded by an EMA system.

Talkers produced simulated (acted) emotional speech.

A set of 10 sentences was commonly used for speech recording of a male (AB) and two females (JN, LS), who are native speakers of American English.

On top of the 10 sentences, there are 4 additional sentences used for recording by only AB.

Each sentence was produced five times for four different emotions, such as neutrality, anger, sadness and happiness.

In totol, AB produced 280 utterances (14 sentences x 5 repetitions x 4 emotions), and JR and JN produced 200 utterances (10 sentences x 5 repetitions x 4 emotions).

Each utterance was digitalized in 12-bit amplitude resolution with 16kHz sampling rate.

Speech was recorded simultaneously by the EMA system so that speech and corresponding articulatory movements are aligned in time.

‘’’

Created valence\_scores\_per\_sample from DocumentationEma.txt

In DocumentationEma.txt, there were two filenames misspelled:

|  |  |  |
| --- | --- | --- |
| 4EMO\_~43.WAV | = | 4emo\_ls\_angry\_41\_041.wav |
| 4EMO\_~86.WAV | = | 4emo\_ls\_happy\_32\_032.wav |

For the emotion category votes, each sample obtained a majority vote for one category (3/4 or 4/4 votes). All of these matched the intended emotion. These were from the best\_xxx\_files.txt files.

Samples were also rated on valence separately (different evaluators). These ratings didn’t always match the valence of the intended emotion. I kept the samples where either the majority valence vote (if present) or the average valence rating matched the valence of the intended emotion. I allowed both criteria to increase the number of samples retained. 32 samples were discarded because of perceived-intended mismatch.

The best\_xxx\_files.txt files do not contain all the files listed in DocumentationEma.txt! The leftovers were assessed by valence only.

Discarded 58

# EmoV-DB\_sorted

Belgian French and North American English

elicitation prompts based on CMU Arctic (en) and SIWIS (fr)