Shared Acoustic Codes Underlie Emotional Communication in Music and Speech - Evidence from Deep Transfer Learning (Datasets)

Coutinho, Eduardo

This repository contains the datasets used in the article "Shared Acoustic Codes Underlie Emotional Communication in Music and Speech - Evidence from Deep Transfer Learning" (Coutinho & Schuller, 2017).

In that article four different data sets were used: SEMAINE, RECOLA, ME14 and MP (acronyms and datasets described below). The SEMAINE (speech) and ME14 (music) corpora were used for the unsupervised training of the Denoising Auto-encoders (domain adaptation stage) - only the audio features extracted from the audio files in these corpora were used and are provided in this repository. The RECOLA (speech) and MP (music) corpora were used for the supervised training phase -  both the audio features extracted from the audio files and the Arousal and Valence annotations were used. In this repository, we provide the audio features extracted from the audio files for both corpora, and Arousal and Valence annotations for some of the music datasets (those that the author of this repository is the data curator).

Below, you can find description of the various corpora, the details about the data stored in this repository and information on how to obtain the rest of the data used by Coutinho and Schuller (2017).

**SEMAINE (speech)**

The SEMAINE corpus (McKeown, Valstar, Cowie, Pantic & Schroder, 2012) was developed specifically to address the task of achieving emotion-rich interactions, and it is adequate for this task as it comprises a wide range of emotional speech. It includes video and speech recordings of spontaneous interactions between human and emotionally stereotyped `characters'. Coutinho & Schuller (2017) used a subset of this database (called *Solid-SAL*). The *Solid-SAL* dataset is freely available for scientific research purposes (see http://semaine-db.eu). This repository includes the audio features used in Coutinho & Schuller (2017) (under features/SEMAINE).

**RECOLA (speech)**

The RECOLA database (Ringeval, Sonderegger, Sauer & Lalanne, 2013) consists of multimodal recordings (audio, video, and peripheral physiological activity) of spontaneous dyadic interactions between French adults. Coutinho & Schuller (2017) used the RECOLA-Audio module which consists of the audio recordings of each participant in the dyadic phase of the task. In particular, they used the non-segmented high-quality audio signals (WAV format, 44.1kHz, 16bits), obtained through unidirectional headset microphones, of the first five minutes of each interaction. Annotations consist of time-continuous ratings of the level of Arousal and Valence dimensions of emotion perceived by each rater while seeing and listening the audio-visual recordings of each participant task. The publicly available annotated dataset includes only part of the data which amounts to a total number of 23 instances. The time frame length used by Coutinho & Schuller (2017) is 1s (the original annotations were downsampled). This repository includes the audio features used in Coutinho & Schuller (2017) (under features/RECOLA). To obtain the annotations you should contact the author of the original study (see https://diuf.unifr.ch/diva/recola/download.html for further details).

**ME14 (music)**

The MediaEval ``Emotion in Music'' task is dedicated to the estimation of Arousal and Valence scores continuously in time and value for song excerpts from the Free Music Archive. Coutinho and Schuller (2017) used the whole corpus (development and test sets for the 2014 challenge) which includes 1,744 songs belonging to 11 musical styles -- Soul, Blues, Electronic, Rock, Classical, Hip-Hop, International, Folk, Jazz, Country, and Pop (maximum of five songs per artist). This repository includes the audio features used in Coutinho & Schuller (2017) (under features/ME14). The full dataset (including annotations) can be obtained from http://www.multimediaeval.org/mediaeval2014/emotion2014/.

**MP (music)**

This is a corpus compiled specifically for this work described in Coutinho & Schuller (2017) using data collected in four previous studies. It consists of emotionally diverse full music pieces from a variety of musical styles (Classical and contemporary Western Art, Baroque, Bossa Nova, Rock, Pop, Heavy Metal, and Film Music). Annotations were obtained in controlled laboratory experiments whereby the emotional character of each piece was evaluated time-continuously in terms of levels of Arousal and Valence perceived by listeners (ranging between 35 to 52 in the four studies). In what follows, some details about the various studies are described.

* MPDB1: This subset of the MP corpus consists of the data reported by Korhonen (2004), and gently made available by the author. This dataset includes six full (or long excerpts) music pieces ranging from 151s to 315s in length (only classical music). Each piece was annotated by 35 participants (14 females). The time series correspondents to each music piece were collected at 1Hz. The golden standard for each piece was computed by averaging the individual time series across all raters. This repository includes the audio features used in Coutinho & Schuller (2017) (under features/MP/DB1). To obtain the labels please contact the author of the original study.
* MPDB2: The dataset by Coutinho & Cangelosi (2011) includes 9 full pieces (43s to 240s long) of classical music (romantic repertoire) annotated by 39 subjects (19 females). Values were recorded every time the mouse was moved with a precision of 1 ms. The resultant timeseries were then resampled (moving average) to a synchronous rate of 1 Hz. The golden standard for each piece was computed by averaging the individual time series across all raters. This repository includes the audio features (under features/MP/DB2) and labels (under annotations/MP/DB2) used in Coutinho & Schuller (2017).
* MPDB3: This dataset was collected by Coutinho & Dibben (2012) and it consists of 8 pieces of film music (84s to 130s long) taken from the late 20th century Hollywood film repertoire. Emotion ratings were given by 52 participants (26 females). The annotation procedure, data processing, and golden standard calculations were identical to MPDB2. This repository includes the audio features (under features/MP/DB3) and labels (under annotations/MP/DB3) used in Coutinho & Schuller (2017).
* MPDB4: This dataset was collected by Grewe, Nagel, Kopiez and Altenmüller (2007), and gently made available by the authors. It includes seven music pieces (127s to 502s in length) of heterogeneous styles (e.g., Rock, Pop, Heavy Metal, Classical). Each music piece was annotated by 38 participants (29 females) using an identical methodology to MPDB2 and MPDB3. Data processing and golden standard calculations were also identical. This repository includes the audio features (under features/MP/DB4) used in Coutinho & Schuller (2017). To obtain the labels contact the authors of the original study

**Bibliography**

Coutinho, E., & Cangelosi, A. (2011). Musical emotions: predicting second-by-second subjective feelings of emotion from low-level psychoacoustic features and physiological measurements. *Emotion*, *11*(4), 921.

Coutinho, E., & Dibben, N. (2013). Psychoacoustic cues to emotion in speech prosody and music. *Cognition & Emotion*, *27*(4), 658-684.

Coutinho E, Schuller B (2017) Shared acoustic codes underlie emotional communication in music and speech—Evidence from deep transfer learning. PLoS ONE 12(6): e0179289. https://doi. org/10.1371/journal.pone.0179289.

Grewe, O., Nagel, F., Kopiez, R., Altenmüller, E. (2007). Emotions over time: synchronicity and development of subjective, physiological, and facial affective reactions to music. *Emotion, 7*(4), pp. 774-788. DOI: 10.1037/1528-3542.7.4.774.

Korhonen, M. (2004). Modeling Continuous Emotional Appraisals of Music Using System Identification. Available from: http://hdl.handle.net/10012/879.

McKeown, G., Valstar, M., Cowie, R., Pantic, M., Schroder, M. (2012). The SEMAINE Database: Annotated Multimodal Records of Emotionally Colored Conversations between a Person and a Limited Agent. *IEEE Transactions on Affective Computing*, 3, pp. 5-17. DOI: http://doi.ieeecomputersociety.org/10.1109/T-AFFC.2011.20.

Ringeval, F.,  Sonderegger, A., Sauer, J. & Lalanne, D. (2013). Introducing the RECOLA Multimodal Corpus of Remote Collaborative and Affective Interactions. In *Proceedings of the 2nd International Workshop on Emotion Representation, Analysis and Synthesis in Continuous Time and Space (EmoSPACE 2013)*, Shanghai, China. IEEE