**Dataset Description: Moods and activities in music**

**Grant**: ES/K00753X/1

**Date**: 16/10/2015

**Authors**: T. Eerola, P. Saari

**Affliation**: Durham University, UK

**Location**: UK DataService ReShare

**Citation**: University of Durham. Department of Music, Tagging music for emotions [computer file]. 1st Edition. Durham, Country Durham: UK Data Archive [distributor], October 2015, http://dx.doi.org/ [to be added]

**Description**: This dataset contains tagged data concerning moods and activities in music, as well as music preferences and generic demographic information of the taggers as well and brief music metadata. All data is related to an ESRC funded research project called 'Tagging music for emotions' (ES/K00753X/1). The data has been obtained through a series of three tasks (1. mood and activity tagging, 2. track search and tagging, 3. track tagging for moods and activities) designed to map various moods and activities related to music. These tasks have been run in Crowdflower (a crowdsourcing service) using workers of a specific quality. 2508 workers contributed to the dataset with varying number of rounds (which contains all tasks and consistency check) completed. No personal information was recorded in the initial data collection and the demographics reveal gender, age, country and a raw index of musical preferences, language skill, and musical expertise. The mood terms (88) and activities (9) and genres (113) are the results of pilot experiments, and there is a validation stage of the data as well. A full description of the tasks will be available in the scientific reports, which are currently in preparation (15-12-2015).

**Methodology**: This data was collected using crowdsourcing paradigm (Paolacci & Chandler, 2014) that utilised Crowdflower (a company running one of the online crowdsourcing services) as the source of participants. Technically the tasks were carried out within a server hosted at Durham University to which the participants were directed from the Crowdflower service. It is worth noting that the participants can live anywhere in the world and we do not have access to any personal information (name, address, etc.) of the participants since the third party (Crowdflower) handles the accounts and the transactions.

**Summary**: 2508 participants (round 2) from 95 countries (see country coverage in the file countries.xls) providing 18559 responses to mood-to-activity tagging task, 8681 responses to track search task yielding 5129 unique tracks, and 28550 responses to a track tagging task. The precise structure of the data is detailed within *data structure description* files. The findings suggested that there is a consistent structure of moods that accounts for a large number of different mood terms. Moreover, types of activities are linked with the mood structures, which can also be further linked with music tracks.

**Data structures and formats**

**Relational data structures**

As the three tasks within the data collection are fairly complex and relational, the dataset reflects the relational nature of the original database used in the data collection. Hence there are several descriptive files, which give label to the index integer values used in the table. For example, 'users.csv' contains for each user a unique userid that is used in the other files as a pointer to the users. Files 'bg\_activities\_general.csv', 'bg\_activities\_musical.csv', 'bg\_genres.csv', 'searches.csv', 'tags.csv' etc. relate to these indices.

**The formatting of the data**

All data is stored in comma-delimited csv files, with headers in the first row.

The values in the files are either integers or strings, depending on the type of data collected. For a more specific explanation of the different files within the files, see 'index.md' which lists the length (number of rows and columns) of the files, the variable types and a general explanation of the variables.

**Other metadata**

‘Consent\_Form.docx’ gives an account of the consent issues related to data collection.

‘Information\_Sheet.docx’ provides a summary of the purposes of the study communicated to the participants.

‘Questionnaire\_Form.docx’ shows the actual wordings for each question and task.

‘countries.xls’ summarises the countries represented by the participants.

‘ReadMe.txt’ contains technical information about the fields in the databases (\*.csv files).