Exploring Affective Experiences Evoked by Music: Study Plan for a Neurophysiological Deep Dataset "ManyMusic🎶"

Seung-Goo Kim[1], Pablo Alonso-Jiménez[2], Dmitry Dogdanov[2], Xavier Serra[2], Daniela Sammler[1,3]

[1] Research Group Neurocognition of Music and Language, Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, Germany
[2] Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain
[3] Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

upf. Music Technology Group

MAX PLANCK INSTITUTE FOR EMPIRICAL AESTHETICS

Rationale

Background. Despite considerable advances in understanding music-evoked emotions [1], our understanding of the underlying mechanisms has remained fragmental. Major limitations come from the enormous dynamics and complexity of emotional experiences during natural music listening [2], the variability of the music people hear, know, and like [3], and the uniqueness of individual neural functioning which we only begin to sense [4]. Therefore, it is most certainly impossible to attain a more comprehensive understanding of music-evoked emotions with traditional experiments that are based on a small number of short and often artificially manipulated pieces of selected musical genres (mainly classical), and rely on group statistics of brain activity.

Vision. We aim to establish a massive multi-modal dataset for music-evoked emotions, entitled "ManyMusic🎶." This dataset will be open-access and will become a workbench for researchers from computational and affective neurosciences and artificial intelligence to test a vast number of questions related to music perception and music-evoked emotions — Ultimately, this dataset will help to understand how abstract musical sounds transform into vivid emotional experiences via the brain.

Novelties. We will build a unique dataset (**Figure 1**) of behavioural, physiological and multi-modal brain responses to a massive amount of naturalistic music pieces from an open-access music dataset, in an approach known as "deep phenotyping." Specifically, we aim to simultaneously collect time-resolved ratings of emotional experiences (emotion tracking) and neural and physiological recordings (i.e., fMRI and EXG, including EEG, facial EMG, heart rate, respiration, EGG, and EDA) while 6 carefully screened participants are listening to real-world music from a wide range of genres (1,021 audio tracks; 5 min each). A large-scale behavioural online study prior to the acquisition of the "ManyMusic🎶" data will map the space of emotions evoked by the music and will validate the choice of the musical pieces used in the fMRI-EXG experiments. Here, we focus on the fMRI and EXG dataset.

Advances in deep neural networks and cutting-edge machine learning algorithms that are able to trace human experiences during real-world events [5-7] will be used to comprehensively model how human affective experiences emerge from the representation of expressive and structural musical features in the brain.

Implications for music industry, clinical research, and our daily lives. Firstly, the deep phenotyping of musical emotions can offer vital insights into artificial intelligence systems, including music streaming algorithms and music generation. Secondly, the comprehensive profiling of neural and physiological activity can help to build a normative model which aids the early detection of affective disorders such as depression. Finally, when combined with wearable devices like smart watches, real-time music recommendation systems will become feasible, potentially changing our daily music listening behaviours.

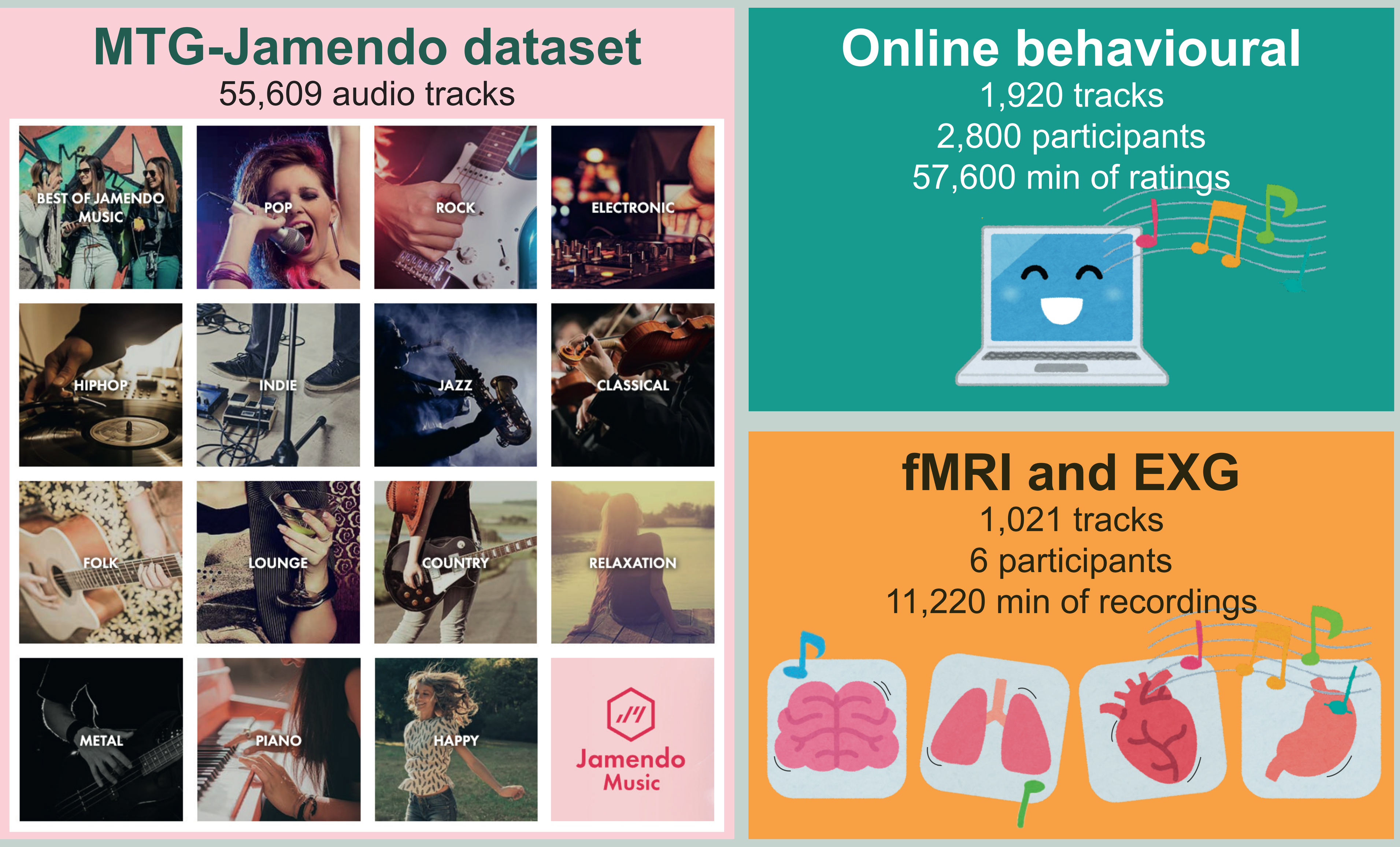

Figure 1. Overview of the project. Music from the massive, open-access MTG-Jamendo music dataset [8] will be used in large-scale online experiments and deep-phenotyping fMRI (functional magnetic resonance imaging) & EXG (electroencephalography and other physiological recordings such as breathing and heart rate) experiments.

Experiment Design

Participants

- Six participants, musically experienced, emotionally responsive and willing to listen to a massive amount of music (170 tracks x 2 modalities [fMRI and EXG]) across 34 sessions.

Stimulus selection

- **MTG-Jamendo music:** From the MTG-Jamendo Dataset (55,609 tracks) [9], 960 tracks in various musical styles (e.g., electronica, classical, rock, jazz) will be selected to evoke a range of musical emotions based on DNN models' predictions of valence, arousal and associated moods [9].
- **Participant-selected music:** 10 tracks per participant (total 60) to estimate the maximal emotional/neural response level in each individual.
- **Reference music:** Queen's "Bohemian Rhapsody" at the end of all sessions to estimate noise ceiling and facilitate aggregation of multi-session/-subject data

fMRI acquisition – 17+1 sessions

- **BOLD imaging**: multi-band multi-echo EPI at 3-T
- **Functional localizers**: Tonotopy mapper (Da Costa) | Voice localizer (Belin) | Music localizer (Kim & Overath)

EXG acquisition – 17+1 sessions

- **Electroencephalogram:** actiCHamp 32 channels
- **BrainAmp ExG:** 8 bipolar channels (DC), 8 auxiliary channels
  - 6 bipolar channels: Facial EMG: 2 ch. (corrugator & zygomaticus major) | Electrogastrogram: 4 ch. [10]
  - 7 auxiliary channels: Heart rate: pulse oximeter: 1 ch. | Electrodermal activity: 1 ch. | Respiratory belt: 1 ch. | Stimulus mono copy: 1 ch. | 3-D accelerometer: 3 ch.
- **Artifact localizers**: muscle contraction [facial, jaw, neck, abdominal], body movement [head, foot], trackball control

Emotion tracking – 34 sessions [fMRI and EXG]

- Continuous ratings of emotional experiences using a trackball
- Two 2-D rating planes (counterbalanced across participants and fMRI-EXG sessions)
  - Core affect [Valence, Arousal]
  - Appraisal [Engagement, Liking]

Planned Analyses

Musical feature extraction via computational models

- Embeddings of a convolutional neural network [9] and an audio-domain music transformer [11] across multiple levels.

Information theoretical analysis

- Partial information decomposition will be used to decompose multivariate data of musical features, fMRI or EXG data, and behavioural ratings into unique, redundant, and synergistic components.

Data aggregation

- fMRI-EXG data will be transformed using the reference music to maximize intra-/inter-subject correspondence ("hyper-alignment" [12]).

Predictive modeling

- Ridge regression will be used to predict neural/physiological responses from the CNN/transformer embeddings.
  - The CNN is expected to model the expressive (timbral) features in music.
  - The transformer is expected to model the temporal structures in music.
- Predicted neural/physiological responses will be used to further predict behavioural responses (emotional ratings).

Seung-Goo Kim | seung-goo.kim@ae.mpg.de
www.ae.mpg.de | +49 69 8300 479 851

Acknowledgement We thank Mr. Harin Lee for advice on music datasets and irasutoya.com for image sources. This research is supported by the Max Planck Society.

References [1] Koelsch, 2014, doi:10.1038/nrn3666 [2] Cowen et al., 2020, doi:10.1073/pnas.1910704117 [3] Park et al., 2019, doi:10.1038/s41562-018-0508-z [4] Westlin et al., 2023, doi:10.1016/j.tics.2022.12.015 [5] Huth et al., 2016, doi:10.1038/nature17637 [6] Allen et al., 2022, doi:10.1038/s41593-021-00962-x [7] Hebert et al., 2023, doi:10.7554/eLife. 82580 [8] Bogdanov et al., 2019, http://hdl.handle.net/10230/42015 [9] Alonso-Jiménez et al., 2022, doi:10.5281/zenodo.7316790 [10] Wolpert et al., 2019, doi:10.1111/psyp.13599 [11] Alonso-Jiménez et al., 2023, doi:10.48550/arXiv.2309.16418 [12] Haxby et al., doi:10.7554/eLife.56601