NSTC 國家科學及技術委員會
National Science and Technology Council

ISMIR 2023

A DATASET AND BASELINES FOR MEASURING AND PREDICTING THE MUSIC PIECE MEMORABILITY

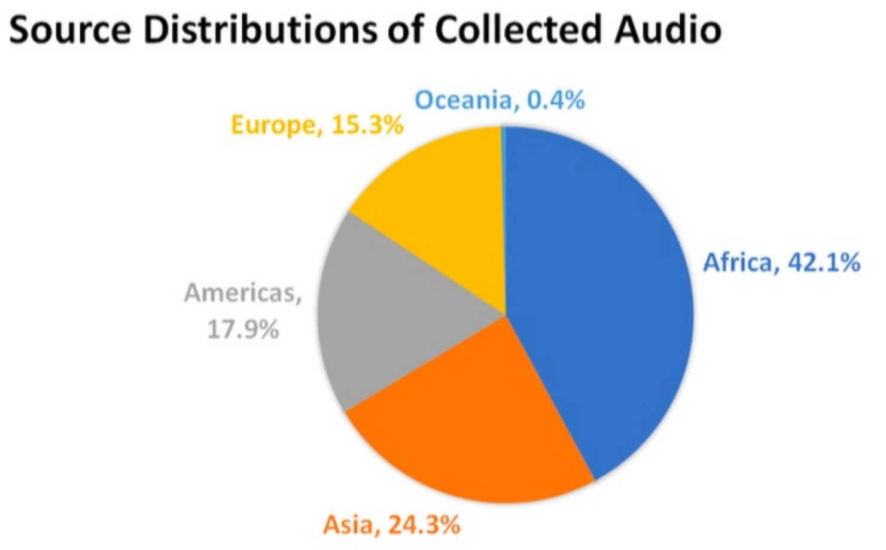
Li-Yang Tseng, Tzu-Ling Lin, Hong-Han Shuai, Jen-Wei Huang, Wen-Whei Chang
National Yang Ming Chiao Tung University, Hsinchu, Taiwan
{liyangtseng.ee10, tzulinglin.11 ,hhshuai, admsd.ee10, wwchang} @ nycu.edu.tw

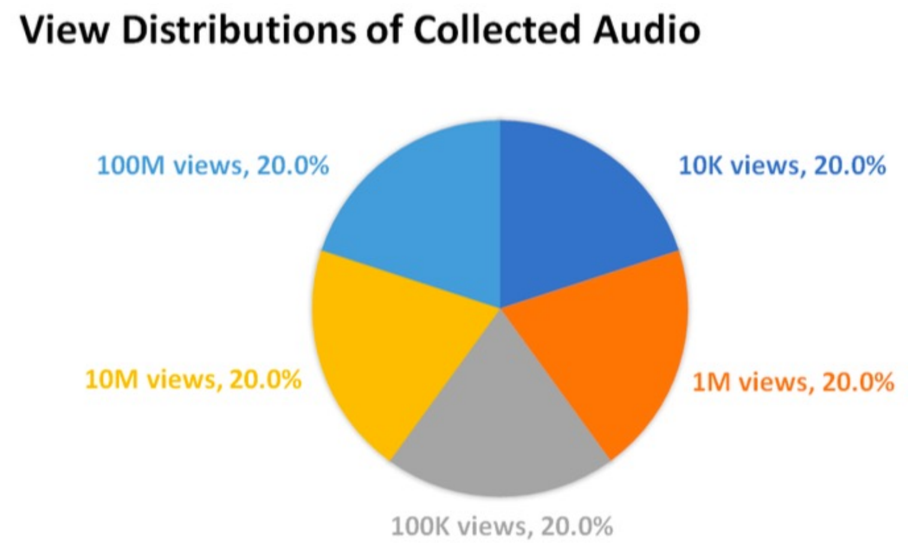
Abstract Despite the abundance of music in this music-saturated world, some pieces stand out and become more memorable, prompting our interest in measuring and predicting music memorability. We create a dataset with memorability labels through an interactive process and employ deep learning techniques. Our findings suggest that music features like valence, arousal, and tempo influence memorability, offering potential applications in music recommendations and style transfer. **To the best of our knowledge, we are the first to explore music memorability regression (MMR) using data-driven deep learning-based methods.**

Dataset

Data Selection

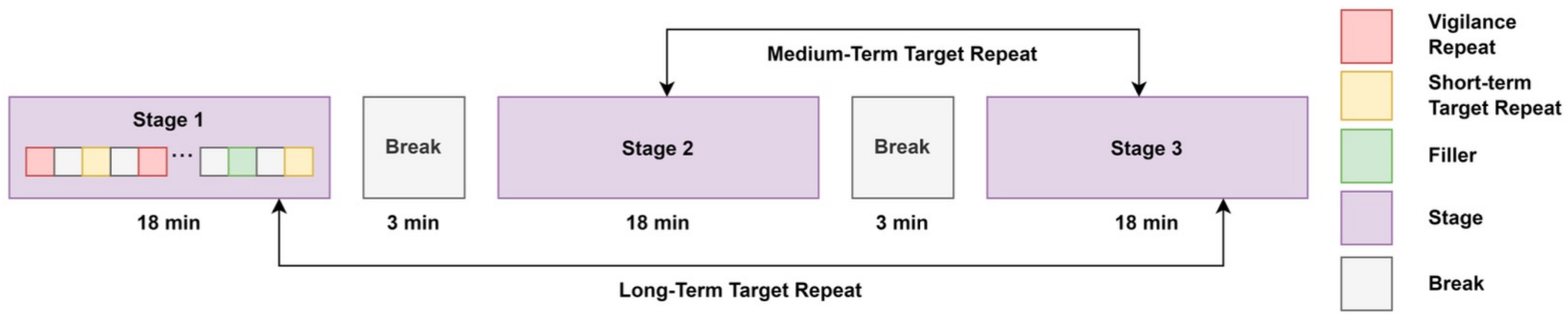
Music pieces are randomly selected from YouTube API. Manual filtering are applied to the music to confirm the queried videos contain pure music content. Pilot study was also implemented to ensure the unfamiliarity to target annotator group.


Source Distributions of Collected Audio


View Distributions of Collected Audio

The Music Memory Game

We to design a novel music listening experiment.



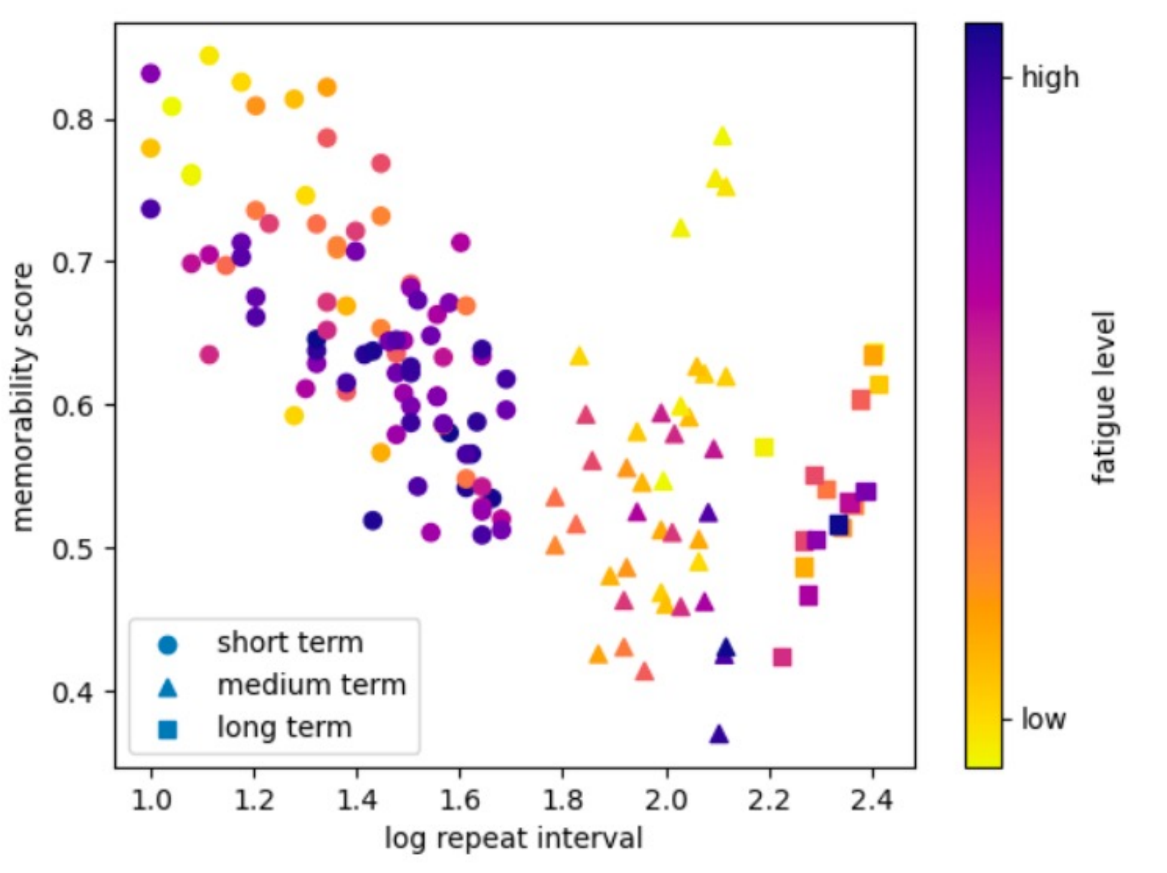
Music memorability is measured as the tendency to correctly recognize a music piece when encountering it again in the experiment among all users. The memorability score of music i, denoted by $m^{(i)}$

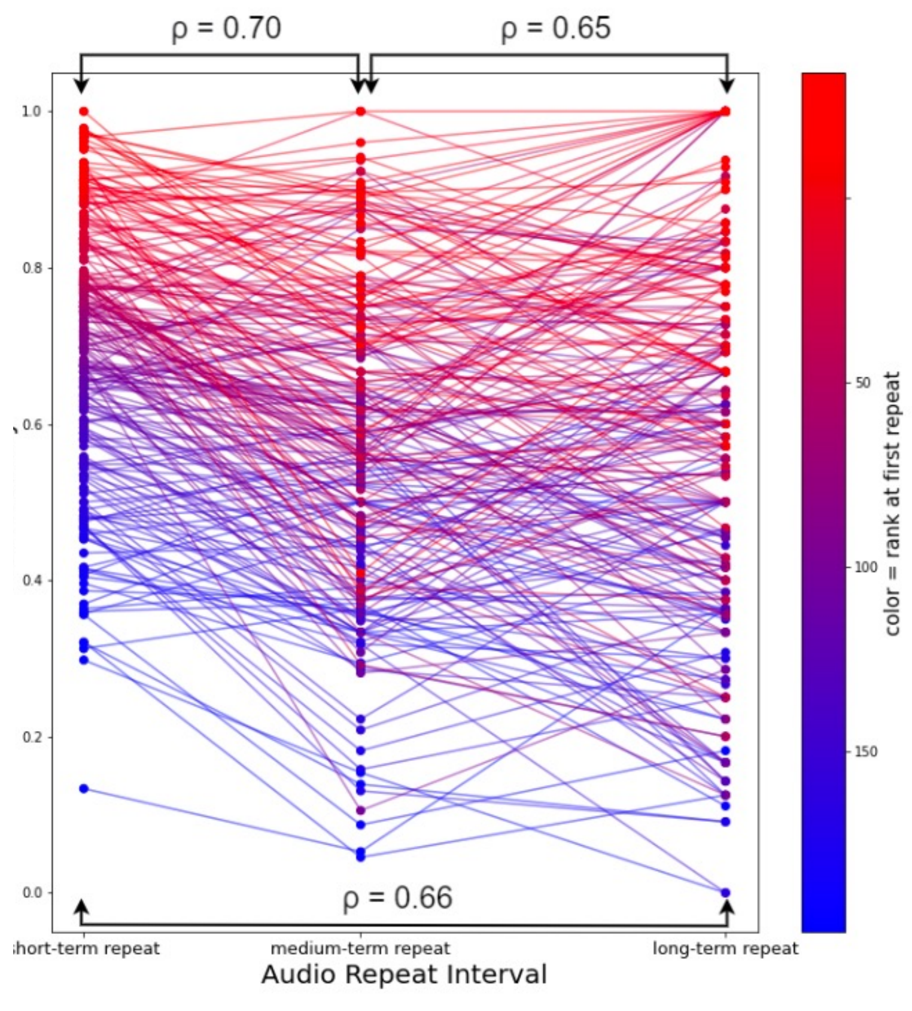
$$m^{(i)} = \frac{1}{n_i} \sum_j x_j^{(i)}$$

where n_i is the total number of target music pairs, and $x_j^{(i)}$ denote whether the i-th music piece can be recalled by the j-th data annotator.

Consistency Analysis

Human consistency is measured in terms of Spearman's rank correlation. The average Spearman's rank correlation coefficient ρ is 0.83 after 25 random splitting, indicating the consistency of the collected data. Music is found to possess a linear relation between memorability score and log-scaled repeat interval. The final collected dataset can be found in https://lin-tzuling.github.io/YTMM_dataset/.





Memorability Prediction

Handcrafted Features

Explainable Handcrafted (EHC) features provide interpretable information for more application insights. As a result of this, low-level features like the chromagram (harmony), beats per minute (rhythm), source separation magnitudes (timbre), and zero crossing rate are selected, while the predicted valence and arousal (mood), music and instrument (genre) are chosen as the high level features.

End-to-End Deep Learning

Evaluation metrices:

- Spearman's Rank Correlation for relative ranking.
- Mean Squared Error (MSE) loss for absolute memorability score.

Baselines:

- Chroma/MFCCs along with their respective derivative + MLP
- Convnet (CNN pretrained on music tagging) features + MLP
- Mel-spectrograms + SSAST (transformer-based model pretrained on multiple audio tasks)

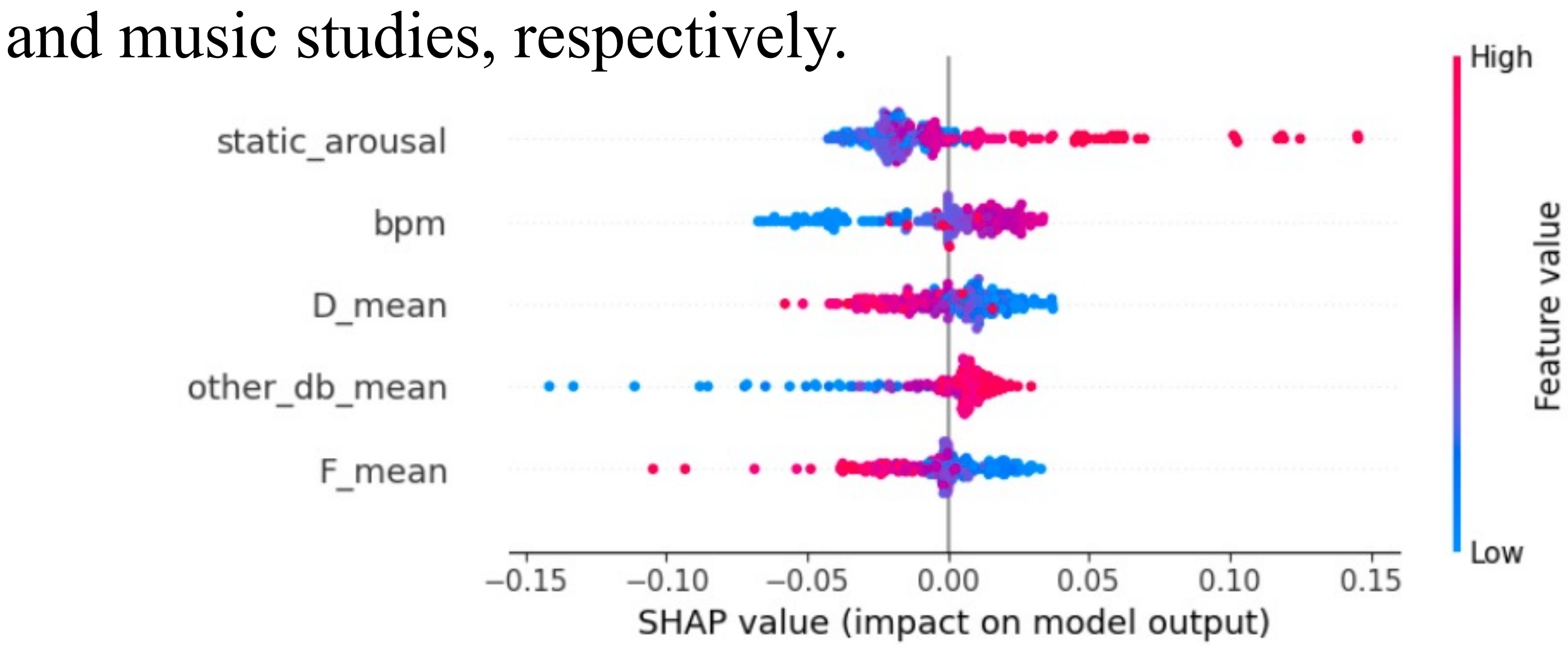
Method	Corr.	MSE	MSE STD
chroma + MLP	0.1740	0.0326	-
MFCCs + MLP	0.1179	0.0353	-
convnet features [38] + MLP	0.1889	0.0314	-
EHC features + SVR	**0.2988**	0.0339	0.0128
EHC features + SVR + PS	0.2084	0.0391	0.0129
EHC features + MLP	0.2656	0.0263	0.0058
EHC features + MLP + PS	0.2388	0.0275	0.0059
mel-spectrograms + SSAST	0.0124	0.0298	0.0061
mel-spectrograms + SSAST + PS	0.2658	0.0265	0.0074

Prediction Results & Ablation Study:

- Chroma and MFCCs are incomplete features for MMR tasks.
- EHC method produces the best correlation results by combining both low and high-level features that help improve MMR.
- Convnet and SSAST outperform because of prior knowledge.
  - *The results manifest that data-driven MIR tasks are notably reliant on huge data quantities to be resilient and general.*
- Pitch shifting (PS) is effective for the models that take sequence information into account.

Interpretability

The results of SHAP (a post-hoc Explainable AI strategy that perturbs a particular instance in the data and examines the impact of these perturbations on the model's output) reveal that greater arousal value and magnitude of "others" (where the main melody is located) lead to better music memorability, which match the psychological and music studies, respectively.


static_arousal
bpm
D_mean
other_db_mean
F_mean
SHAP value (impact on model output)
High
Low
Feature value