Towards Characterising Induced Emotions: Exploiting Physiological Data and Investigating the Effect of Music Familiarity

Ismaël Tankeu[1], Geoffray Bonnin[2,*]

[1]ENIT, Tarbes, France
[2]Loria, Nancy, France

Abstract

Music recommendation aims to suggest songs that align with listeners' preferences. Recent work has shown a strong correlation between induced emotion, i.e., the emotion actually felt during music listening, and music appreciation. However, the potential of leveraging induced emotions for music recommendation still remains unexplored. This paper explores the use of physiological data and music features to predict discrete emotional responses, as defined by the Geneva Emotional Music Scale (GEMS) model. Our results show that integrating physiological data and music familiarity enhances prediction accuracy compared to integrating music features only. Additionally, feature importance analysis revealed that, although music features remained the primary predictor of induced emotions, physiological data contributed substantially to the prediction model.

Keywords

Physiological Data, Music Features, Music Listening, Induced Emotions, Machine Learning

1. Introduction

Music appreciation can be influenced by various psychological phenomena. For instance, music that induces tension is generally associated with lower levels of appreciation, while music that induces joy tends to be positively received [1]. Since the goal of music recommenders is to provide users with music they will enjoy, being able to predict the emotions a piece of music will evoke in a specific user is of major importance. The research field related to this task is known as Music Emotion Recognition. It contains extensive research on perceived emotions, i.e., emotions that listeners believe the music is expressing or conveying, independent of their own emotional state. In contrast, research on induced emotions, i.e., emotions actually felt by listeners as a direct result of the music, remains relatively scarce [2].

Although induced and perceived emotions have been found to often align [3], a same music can induce very different emotions on different listeners. This is demonstrated in Figure 1, which shows the frequency of emotions induced by several tracks in the Music-Mouv' dataset [1]. As can be seen, almost none of these tracks induced a uniform emotional response among all participants. The only exception was "Djadja" from Aya Nakamura, which consistently induced feelings of "Tension" in all participants despite the song's moderatley high perceived emotional positivity[1]. In other words, none of the participants experienced the perceived positive emotion; rather, they seemed to endure discomfort throughout its duration.

Research on the inference of induced emotions usually involves physiological data, such as the heart rate or the pupil size. One of the few works that follow this line of research is [4], which investigates the effectiveness of using heart rate and skin conductance combined with acoustic features to classify induced emotions, as well as users' Big-Five personality traits. Their results show that using physiological features significantly improved the valence classification (the positivity or negativity

✉ ismael.tankeu@enit.fr (I. Tankeu); bonnin@loria.fr (G. Bonnin)

🆔 0009-0009-6386-7078 (I. Tankeu); 0000-0002-9234-3849 (G. Bonnin)

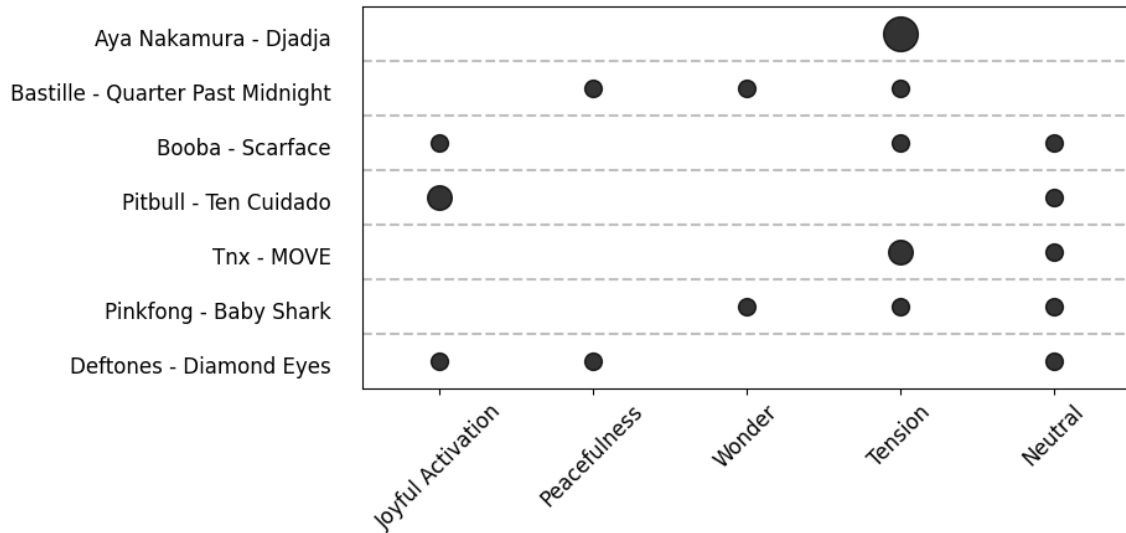[1]The song has a valence value of 0.54 according to Spotify.

Figure 1: Emotions induced by several tracks on different listeners in the Music-Mouv' dataset [1].

of emotions) when combined with acoustic features. However these features offered less benefit for arousal classification (the intensity of emotions), where acoustic features alone were very effective. One limitation of this approach is that it uses a classifier on oversimplified emotion categories derived from the dimensional model of affect. In this work, we will rely on a music-specific discrete emotion framework developed and validated for music, to achieve a more nuanced classification of emotional responses.

Another approach proposed by [5] applied regression models that explicitly differentiate between perceived and induced emotions to predict valence and arousal ratings provided by users during music listening. More precisely their assumption was that listeners assigned these ratings by mixing both types of emotions through a weighted judgment. Their results show a slight improvement compared to models that assume purely induced or purely perceived emotion ratings. However, it is not possible to determine the exact accuracy of their assumption, nor is it possible to know the extent to which the weights used in the weighted judgement models correspond to reality. Our experiments solely focus on induced emotions, i.e., we use subjective data from participants who were specifically asked to the emotions they actually felt through a semi-directed interview approach.

In our previous work [1], we collected a comprehensive dataset that captures the interplay between emotions, physiological responses, physical movement, familiarity and liking judgement. Our aim was to study how the emotional context induced by music listening impacts gait initiation. Our preliminary findings indicated that the induced emotions significantly influenced gait initiation, with certain emotional states facilitating or hindering the movement. Additionally, our results suggested that music familiarity has a major importance and should be taken into account for induced emotion recognition. In this paper, we therefore aim to also investigate the impact of the familiarity factor relatively to the physiological and musical features to further improve the classification accuracy.

Our research questions are therefore the following:

RQ 1: How effectively can physiological data collected during music listening predict discrete emotional responses to music, as defined by a music-specific discrete emotion model?

RQ 2: How do music features compare to physiological data to predict discrete emotional responses?

RQ 3: Does music familiarity further impact the effectiveness of predicting discrete emotional responses?

Table 1
Main physiological features extracted

Type	Name	Description
HRV	SDRMSSD	Standard deviation of root mean square of successive differences
HRV	CVSD	RMSSD divided by the mean of the RR intervals
HRV	MedianNN	Median of the RR intervals
HRV	MadNN	Median absolute deviation of the RR intervals
HRV	IQRNN	Interquartile range of the RR intervals
HRV	Prc80NN	80th percentile of the RR intervals
HRV	Prc20NN	20th percentile of the RR intervals
HRV	MinNN	Minimum of the RR intervals
HRV	HTI	HRV triangular index
HRV	TP	Total spectral power
EDA	Phasic Mean	Mean of the phasic component
EDA	Phasic AUC	AUC of the phasic component
EDA	Phasic Median	Median of the phasic component
EDA	Phasic STD	Standard deviation of the phasic component
EDA	Phasic Number of Peaks	Number of detected peaks
EDA	Phasic Skewness	Skewness of the phasic component
EDA	Phasic Kurtosis	Kurtosis of the phasic component
EDA	Phasic Sum Amplitudes	Sum of amplitudes of the detected peaks
EDA	Phasic Mean Amplitudes	Mean of amplitudes of the detected peaks
EDA	Phasic Max Amplitudes	Maximum of amplitudes of the detected peaks

2. Methodology

2.1. Feature extraction

Our experiments rely on the Music-Mouv' dataset [1]. The most important aspect of this dataset in relation to our research questions is that the induced emotions it contains were collected using a semi-structured interview approach, ensuring that the participants only reported the emotions they actually felt. This dataset contains the physiological responses of 35 participants recorded during music listening. These data were collected using Empatica E4 wristbands and include Blood Volume Pulse (BVP) and electrodermal activity (EDA). Overall, 223 trials where made, and a total of 188 tracks where played, i.e., 27 tracks were played twice or more. Participants also indicated the emotions they experienced during music listening according to a discrete model, the Geneva Emotional Music Scale (GEMS) model [6], and a dimensional model, the usual valence and arousal scales. The dataset also contains binary familiarity ratings and the following track features extracted from the Web API of Spotify[2]: danceability, energy, loudness, key, mode, speechiness, accousticness, instrumentalness, liveness, valence and tempo.

In order to answer our research questions we extracted several features from the raw physiological data of the dataset. We extracted time-domain and frequency-domain Heart Rate Variability (HRV) features [7, 4] by using the neurokit library [8] on the recorded BVP signals of the dataset. We also extracted several EDA related physiological features. EDA signals include two components: the slowly changing tonic component, which reflects a person's general skin conductance level, and the rapidly changing phasic component, which is often elicited by external stimuli [4]. We also relied on the neurokit library to extract several peak-related features from the phasic component. The main extracted features[3] are presented in Table 1.

---

[2] See https://developer.spotify.com/documentation/web-api/.
[3] We only included the 10 main EDA and HRV features for space reasons.

2.2. Experimental protocol

We compared the results of applying Random Forest classification using different types of input attributes: (1) only physiological features, (2) only music features, (3) both types of features, and (4) both types of features along with music familiarity. We used scikit-learn's RandomForestClassifier[4] with a number of estimators of 100.

Although the dataset also contains subjective ratings of valence and arousal for each track, in this paper we only focus on the task of classifying discrete emotions. Approximately 70% of these emotions were represented by "Joyful Activation" and "Tension." Given the corresponding low number of trials of the other emotions in the dataset, we decided to focus only on these two emotions for the remainder of our study, leaving us with 154 trials. An additional 26 trials had to be removed due to missing raw physiological data or the inability to extract all the corresponding physiological features[5].

We evaluated all four resulting models in terms of precision, which is the ratio of correct predictions to all predictions[6]. To assess the importance of each feature in predicting the induced emotions, we also examined the respective importances of the features as determined by the trees computed by the algorithm. In scikit-learn, this can be calculated based on the reduction in Gini impurity that the feature brings within each tree.

In the context of machine learning, the selection of input features generally impacts the performance of the model. Correlated features can cause issues such as multicollinearity, which can lead to less reliable predictions. Conversely, focusing on a subset of features reduces the risk of overfitting, especially when the training data size is limited. Finally, selecting a subset of less correlated features allows to better understand which features are driving the predictions, which can enhance the interpretability of the model. In order to select our input features, we therefore relied on their correlations. Figure 2 shows the Spearman's correlations of the main physiological features we extracted and of the music features from Spotify. The process for selecting our features involved trying different combinations while avoiding multicollinearity, as indicated by the correlation matrix.

3. Results

The precision results obtained through a 5-fold cross-validation procedure with 80/20 splits are shown in Figure 3. A first interesting observation is that using only musical features resulted in better precision compared to using only physiological features (64.2% vs 60.1%). This result is expected as induced emotions often align with perceived emotions [3], and the collected physiological data may only partially correspond to the emotional response, given that physiology can be influenced by various factors. Combining both types of features further increased the precision, which reached 65.7%. This result confirms the importance of making the distinction between perceived emotions and induced emotions, with the former being the most prevalent in the domain of music emotion recognition. It also confirms that integrating both types of data can be instrumental in predicting actual emotional responses to music, even when those emotions fall into discrete music-specific categories. Finally, also including the familiarity ratings as input of the classifier further improved the accuracy, which reached 68.1%. This result further confirms the importance of taking into account this factor for the detection of induced emotions.

Next, we looked at the respective importance of the input features of the four versions of the model. As explained in the previous section, we selected a subset of the physiological features and of the music features based on the correlation matrix to avoid redundancy. The values of the resulting best combinations we found are shown in Figure 4. Regarding the version with only physiological features, we can notice that there was no clear winner between EDA and HRV features, as both types of features

---

[4]https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

[5]Our final dataset and our code can be accessed here: https://homepages.loria.fr/gbonnin/music-mouv/.

[6]Note that for each prediction, since each trial corresponds to a single induced emotion, the number of false negatives is equal to the number of false positives, making recall equivalent to precision. In this case, the metric is also often referred to as hit ratio.
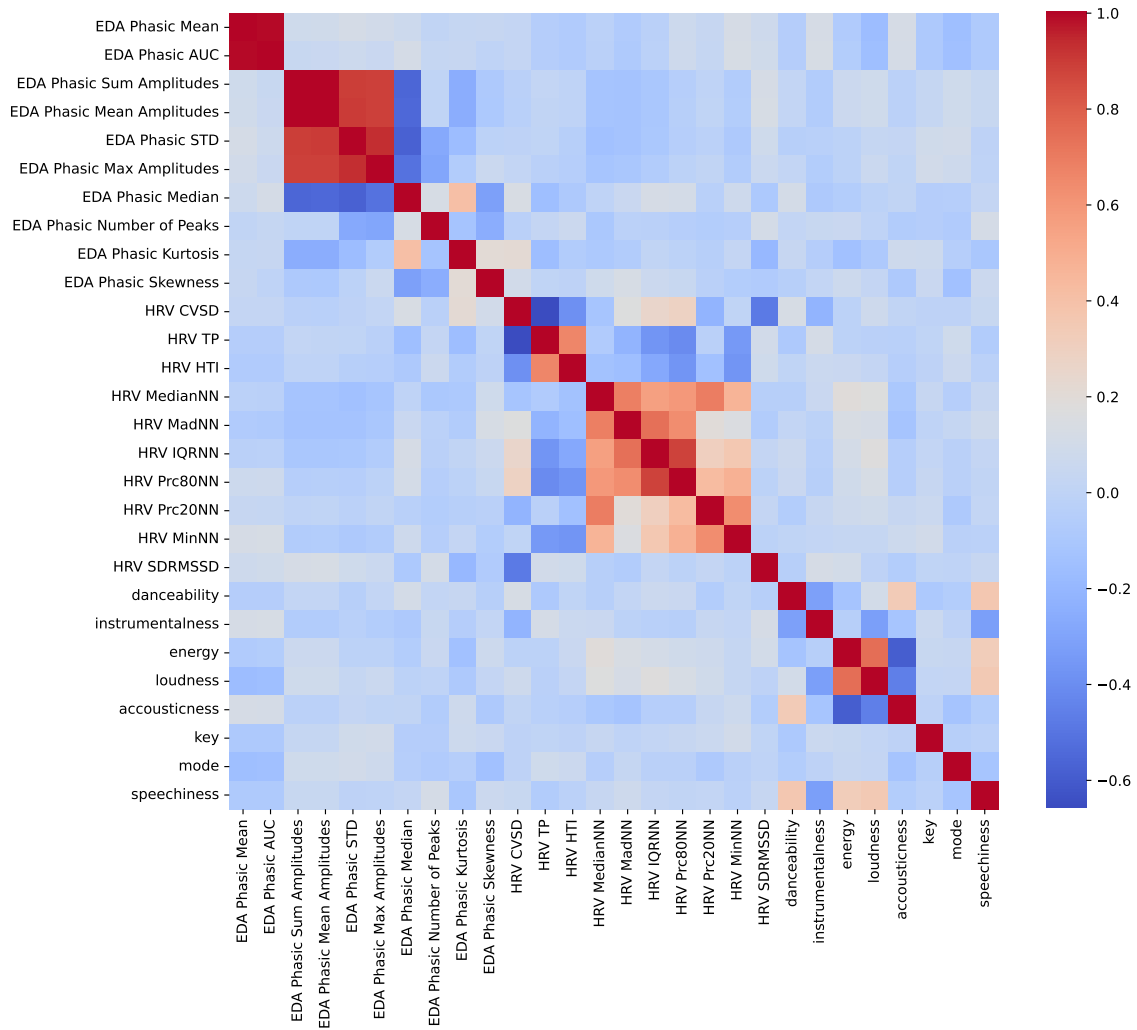
Figure 2: Correlation Heat Map for the main features we extraced and of the music features from Spotify.

were quite balanced and as no strong difference appeared between successive features. The version with only music features led to a very different outcome, with a strong difference of importance between the features, except for danceability and speechiness, which were the two most important features.

Combining both types of features further validated the significance of danceability and speechiness, with importance values far higher than the rest. This strong difference aligns with the better results obtained from using music features alone compared to using physiological features alone: since induced emotions often correspond with perceived emotions, music features are the strongest predictors. Nevertheless, the importance of physiological features remains substantial, with their importance values being comparable to, or even exceeding, those of other music features. This tends to further support the importance of personalising the recognition of induced emotions with physiological features. One interesting outcome was the reduced importance of acousticness when physiological features were incorporated. This may suggest a relationship between acousticness and certain physiological features, which was not evident in the correlation matrix. Finally, the importance of the familiarity feature when also including it was unexpectedly low, especially given the previously observed importance of this feature in terms of correlation with induced emotions. A relationship may also exist between familiarity and certain physiological features.

Overall, these results confirm that, while musical features are established strong predictors of
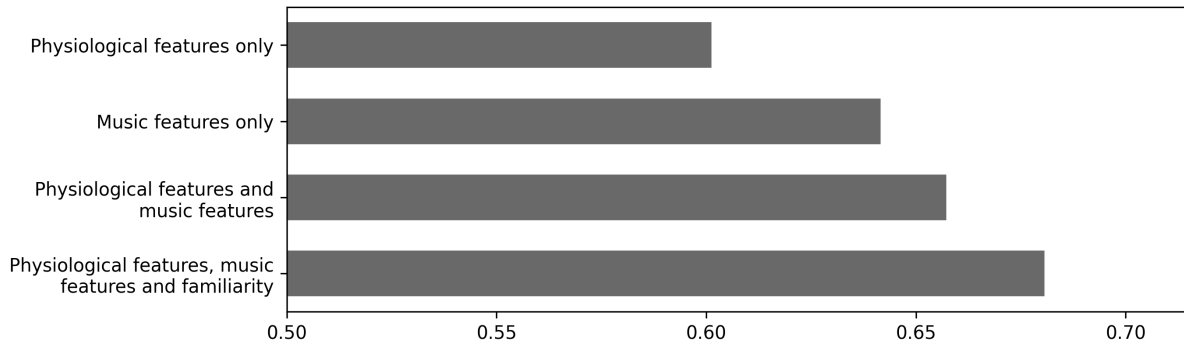
Figure 3: Precision of the model depending on the type of input features.

perceived emotions, they also are, to a certain extent, good predictors of induced emotions. However, these factors alone are insufficient, as also incorporating physiological data enables more accurate predictions tailored to each participant. Finally, the results further validate the importance of music familiarity in predicting induced emotions, as including this feature in our model further enhanced the results.

4. Conclusion

This paper investigated the use of physiological data and music features to predict discrete emotional responses induced by music listening, as defined by the music-specific discrete GEMS model. While the use of music features alone led to a precision of 64.2%, incorporating physiological data improved the results to 65.7%. Additionally, including music familiarity further enhanced the precision to 68.1%, confirming the importance of this feature. Feature importance analysis revealed danceability and speechiness as key music features, with physiological features showing balanced but substantial contributions. In our future work, we will explore additional physiological and contextual factors to refine our understanding of music-induced emotions. By doing so, we seek to uncover how emotion indicators can be effectively utilised to enhance music recommendations.

5. Acknowledgments

References

[1] M. Doumbia, M. Renard, L. Coudrat, G. Bonnin, Characterizing the Emotional Context Induced by Music Listening and its Effects on Gait Initiation: Exploiting Physiological and Biomechanical Data, in: Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization, Association for Computing Machinery, New York, NY, USA, 2023. URL: https://doi.org/10.1145/3563359.3596982.

[2] D. Han, Y. Kong, J. Han, G. Wang, A survey of music emotion recognition, Frontiers of Computer Science 16 (2022).

[3] Y. Song, S. Dixon, M. Pearce, A. Halpern, Perceived and Induced Emotion Responses to Popular Music: Categorical and Dimensional Models, Music Perception: An Interdisciplinary Journal 33 (2016).
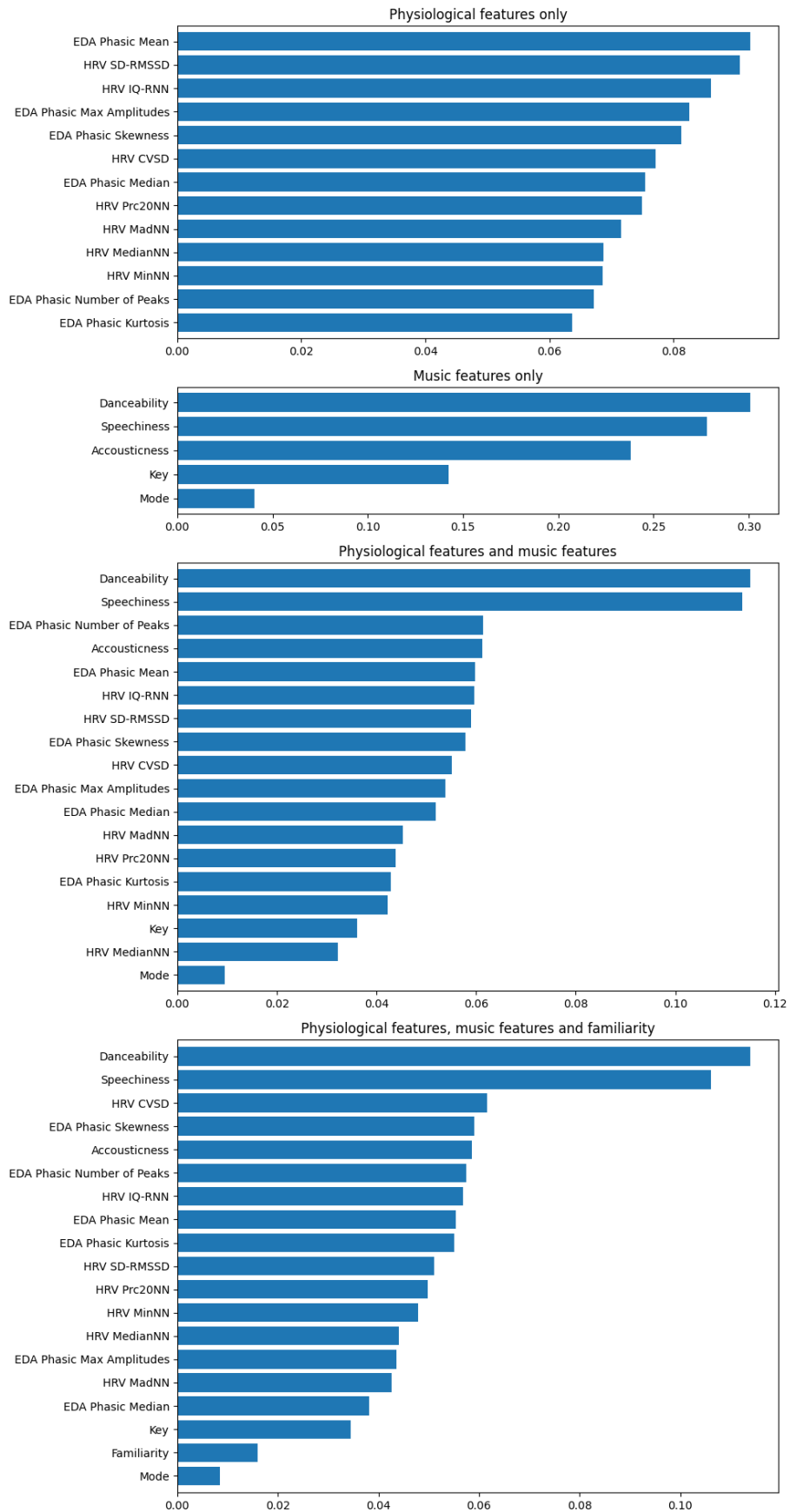
---

[7]https://www.grid5000.fr/

Figure 4: Feature importances of each variant of the model.

[4] X. Hu, F. Li, R. Liu, Detecting Music-Induced Emotion Based on Acoustic Analysis and Physiological Sensing: A Multimodal Approach, Applied Sciences 12 (2022).

[5] N. Vempala, F. Russo, Modeling Music Emotion Judgments Using Machine Learning Methods, Frontiers in psychology 8 (2018).

[6] M. Zentner, D. Grandjean, K. Scherer, Emotions evoked by the sound of music: characterization, classification, and measurement, Emotion 8 (2008).

[7] F. Shaffer, J. Ginsberg, An Overview of Heart Rate Variability Metrics and Norms, Frontiers in public health (2017).

[8] D. Makowski, T. Pham, Z. J. Lau, J. C. Brammer, F. Lespinasse, H. Pham, C. Schölzel, S. H. A. Chen, NeuroKit2: A python toolbox for neurophysiological signal processing, Behavior Research Methods 53 (2021) 1689–1696. URL: https://doi.org/10.3758%2Fs13428-020-01516-y. doi:10.3758/s13428-020-01516-y.