# Low-level Feature Selection for Emotional Music Preferences Based on Subjective Audience Rating

Jonghyung Lee, Min-Uk Kim, Kyoungro Yoon*, *Senior Member, IEEE*
*School of Computer Science and Engineering, Konkuk University, Korea*

***Abstract*— Recently, many researches of modeling or measuring human feeling have been conducted to understand human emotions. However, researches on music-related human emotions have much difficulty due to the subjective perception of emotions. We have selected low-level musical features which may trigger human emotions, based on subjective audience ratings. This experiment is based on the subjective audience ratings of five hundred participants in a very popular Korean TV music program. In this program, audience is requested to rate music of the contestants and to select their preferred music based on their emotional feelings. The most relevant low-level features with respect to human emotions are selected by backward elimination method and experimental results show that selected low-level features touch human emotions quite positively.**

## I. INTRODUCTION

Music has been widely used as religious, political, and social tools since the appearance of humanity. Traditionally, most of the music recommendation or searching is performed based on the alphanumeric data or metadata such as the title, composer information, singer/artist information, and publication date. However, with the advances in the media-related researches and web-related technologies, new types of services are unleashed and emotion-related researches are getting more and more attention in the various fields.

It is said that the most significant and fundamental function of the music is a triggering of emotions. Most of the people empirically know that music triggers various emotions such as happiness, sadness, and joyfulness, to name a few[1]. However it is difficult to select low-level features which may trigger human emotions due to the subjective perception of emotion and the hardness of psychological explanation as well as the difficulty of constructing objective database.

The purpose of this work is to analyze the rating information contained in the weekly or biweekly TV music program's audience rating results, in the hope of finding some model consisting of several low-level musical features which are essential to trigger the human emotions.

In this paper, we analyze the audience ratings of five hundred people to select low-level features, which trigger human emotion, based on rating information and the corresponding music. Experimental results show that the prediction of music's rating based on selected low-level features is quite reliable estimation of original rating. Due to the small size of database, k-fold cross validation is used to evaluate our selected features.

This work was supported by Seoul R&BD Program (SS100006).
* Author for correspondence.

## II. ALGORITHM DESCRIPTION

### A. Database & Evaluation

In this paper, we use the music which is broadcasted on a TV music competition program. At each round of competition which is held twice in every three weeks, each of seven singers sings a song of his/her choice. Until now there were nine rounds of competition, and there were sixty three songs sung. The audience panel consisting of five age groups of teens, twenties, thirties, forties, and fifties, each of which has fifty females and fifty males, are carefully selected not to have any bias for specific genre of song or for specific singer, at each round of competition. At the end of each competition, the audience panel is requested to rate the performance by selecting the best three out of the seven performances, which impressed him/her the most.

Every audio clip is normalized in volume and re-sampled at 16,000Hz, 16 bit depth, and mono channel.

Selected features are used to predict the ratings of the audience panel and the mean and standard deviation of the prediction errors are calculated as a measure of the usefulness of the selected features.

### B. Feature extraction

Feature extraction from songs is made by *MIRtoolbox* [2]. Extracted features are organized by six high-level features such as rhythm, tonal, timbre, dynamics, fluctuation, and spectral, twenty eight mid-level features such as RMS, peak, centroid, tempo, zero-cross, etc, and eight hundred eighty four low-level feature values.

### C. Feature selection

Feature selection is a task to select the minimum number of features needed to represent the data accurately among the candidate features. In this paper, we focus on the audience rating information rather than data itself during feature selection, as the audience rating is believed to be a measure of the emotional impression.

There are many feature selection methods such as: *All possible regression* which is a procedure considering all the possible independent variables; *Forward selection* which is a procedure adding most correlated variables rather than least; *Backward elimination* which is a procedure subtracting least correlated variables; *Stepwise selection* which is combination procedure of *forward selection* and *backward elimination*.

In this paper, we used the backward elimination method, which is implemented as the *mirmap* [4] function in the *MIR toolbox*. *mirmap* is backward elimination method where the

rating information is set as the dependent variable and feature values are set as the independent variables. *mirmap* routine detects all features whose statistics show discrepancy from the normal distribution, based on a Lilliefors test at the 5% significance level. These features with non-normal distribution are then normalized using an optimization algorithm, so that their distributions become sufficiently Gaussian to apply correlation estimation. If the normalized features still do not have normal distribution within 1% significance level, they are excluded from further analysis. By applying correlation estimation, only those features whose correlation with the ratings are statistically significant (with a p-value lower than 0.05) are selected [3].

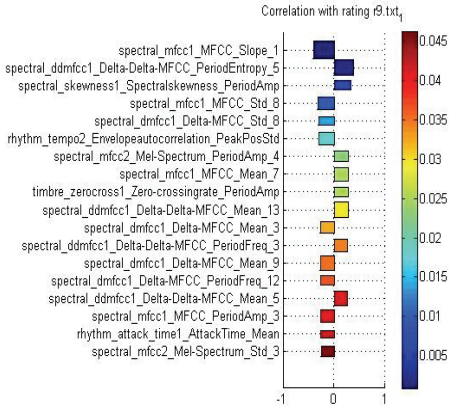


Fig. 1. Selected low-level features using backward elimination method from 1st to 9th rounds of competition

Fig. 1 shows eighteen selected low-level features using backward elimination method. In this experiment, total number of extracted low-level features is 884. The correlated features are selected from 1st rounds of competition to 9th with the highest correlation score at the top and the lowest correlation score at the bottom of the Fig.1. The p-value is indicated by colors from red to blue. Warm color, i.e., red, mean for statistical correlations of low significance and cold colors, on the contrary, mean for significant correlations [3].

## III. EXPERIMENTAL RESULT

We select features from varying number of competitions, for example, from 1st to 5th round of competition, or 1st to 6th round of competition. These experiments show very interesting result. There are features remaining in the list of highly correlated features, even though the number of competitions increases. These features are candidates of the low-level features which trigger human emotions. Once features are selected, then we use regression analysis to model these features and predict the music's rating of impressiveness. Due to the lack of database size, we use k-fold cross validation. Each round of competition becomes a query and we calculate mean error between original rating and predicted rating. Standard deviation is also calculated to evaluate how consistently the selected features represent human emotions.

Fig. 2 shows the mean error of the rating prediction. It is easy to see that there is a tendency for selected features from larger databases, for example using all nine rounds of competition, to have lower mean errors.

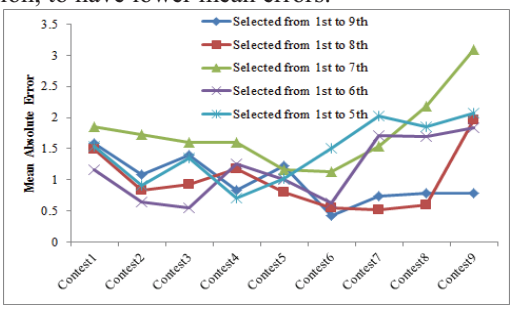


Fig. 2. Mean error results from varying feature selected

Fig. 3 shows the standard deviation of the rating prediction. Due to the lack of small database, selected low-level features from this experiment may not fully describe the factors of human emotion. However its predicted rating is closer to the original when the selected features' range is growing, and shows the possibility of correctly predicting the rating of the impressiveness if the database size grows.

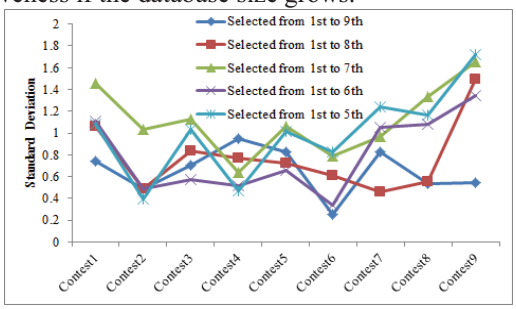


Fig. 3. Standard deviation results from varying selected features

## IV. CONCLUSION AND FURTURE WORKS

In this paper, we analyze the audience ratings of four thousand five hundred people (ratings for nine rounds of competition of five hundred people for each round of competition) to select low-level features which trigger human emotions. Experimental results show that selected low-level features are closely related to the human emotion with respect to the prediction of music rating of impressiveness.

In the future work, we will continue to experiment until this TV program is over, to figure out the most reliable low-level features which trigger human emotions. We also plan to map these selected low-level features to individual emotion. We hope this work to contribute to the researches for modeling and understanding human emotions.

## REFERENCE


[1] M. Zentner, D. Grandjean, and K. Scherer. Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion*, 8(4):494–521, 2008.

[2] O. Lartillot, P. Toiviainen, and T. Eerola. A matlab toolbox for music information retrieval. *Data Analysis, Machine Learning and Applications*, 261–268, 2008.

[3] O. Lartillot, MIRtoolbox User's Manual, http://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox

[4] Olivier Lartillot, Tuomas Eerola, Petri Toiviainen, Jose Fornari, "Multi-feature modeling of pulse clarity: Design, validation, and optimization", *International Conference on Music Information Retrieval*, Philadelphia, 2008.