# Real Time Modeling of Emotions by Linear Regression

Sergio Giraldo and Rafael Ramirez

Pompeu Fabra University, Music Technology Group,
Roc Boronat, 138 - 08018 Barcelona, Spain
{sergio.giraldo,rafael.ramirez}@upf.edu
http://www.mtg.upf.edu

**Abstract.** Expression in music performance is a well studied research topic that deals with the analysis of the deviations from the score introduced by musicians in order to achieve certain style or emotion. In the past some systems have been proposed for rendering expressive performances with a particular emotion content based on musicological knowledge. In this study we present a machine learning based expressive music performance system. We train expressive performance models using linear regression for four different moods: happy, sad, angry and tender. We then interpolate these models in order to to obtain intermediate expressive models for other emotions. Finally, we apply the resulting models to generate performances with emotional content. Our implementation allows the generation of performances with mood transitions in real-time.

**Keywords:** Expressive music performance, machine learning, linear regression, mood modeling, real-time

## 1 Introduction

Expression plays a critical role in music performance. It can be defined as the deviations from the score a musician do when performing a musical piece, in order to convey certain style or emotion. The analysis of how these deviations and its relation with emotional aspects of the performance is a topic which has been a widely studied by researchers.

Several approaches to expressive music performance have been proposed aiming at creating systems able to automatically render expressive music performances, and some can render performances with different emotional content (e.g [1], [2]). The core of these systems are expressive performance models which are used to produce new performances. These models can be obtained manually using an Analysis by Synthesis approach (e.g. [3]) or by applying machine learning techniques (e.g. [4]). However most of these models base the modeling strategies in musicological aspects of performance rather than on a systematic selection of note descriptors.

In this paper we present an approach to model four emotions: happy, sad, tender and angry. We obtain recordings of the same piece by a professional musician, in each of these four different moods. We extract descriptors of the notes

**In MML 2013: International Workshop on Machine Learning and Music, ECML/PKDD, Prague, 2013.**

and apply machine learning techniques together with standard feature selection methods (e.g. wrapper feature selection), to train models for each emotion and three performance actions: duration, articulation and loudness. We then interpolate the obtained models in order to obtain models for intermediate emotions.

In the next section we present an overview of the system, in Section 3 we present initial results and finally Sections 4 presents some conclusions.

## 2 System Overview

The general framework of the system is depicted in Figure 1. Each mood corresponds to a quadrant in the arousal-valence plane. The models are trained using recordings of a musical piece in all four moods. The input of the system is a control coordinate on the arousal valence plane. The coefficients of the models are interpolated to obtain the prediction of the performance actions, which serve as input for synthesis.

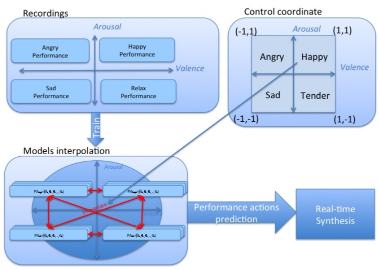


**Fig. 1.** General framework of the system for emotion music modeling.

**Data acquisition.** The first step is to obtain a MIDI representation of the score of a set of pieces. We then obtain recordings of performances of the pieces in four different emotions (happy, sad, angry and tender) by 4 professional guitarists. The recordings were obtained without metronome, so each performer could freely choose tempo and introduce tempo variations that better express each mood.

**Expressive performance actions.** Expressive performance actions are defined as the strategies and changes introduced in a performance, which are not specified by the score. In this study we will focus on three specific performance actions: duration ratio (ratio between the score duration and the performed duration ), the energy ratio (ratio between loudness of a note and average loudness), and articulation ratio (level of staccato - legato).

**Feature extraction.** Each note in the training data is annotated with a number of attributes representing both properties of the note itself (Intra-note features) and some aspects of the context in which the note appears (Inter-note features). Information about the note includes note duration, energy, pitch, while information about its context includes relative pitch and duration of neighbour notes, as well as, melodic and harmonic analysis. A complete list of the features and its exact description can be found on Giraldo 2012 Master thesis [5].

**Modeling and Feature selection.** Feature selection was performed using wrapper method with "best first" as a search method and with forward and backward elimination. Ten cross-fold validation was used for testing and evaluating the models. For each model we obtained a list of features that were most frequently selected by the feature selection method over the ten folds. For interpolation purposes we selected the same features for the four emotion models for each performance action, selecting the average of the most relevant features for the four models.

**Interpolation.** A coordinate $c(x, y)$ define 4 areas weight areas on the plane, so each corner of the arousal valence space has a weight proportional to the area of the square which is formed in the opposite corner.

**Synthesis.** The linear regression models (formulas) were implemented in pure data and MIDI note information and its descriptors were saved on a text file. Each time a note is read form the text file, the performance actions are calculated using the model and the interpolated coefficients based on a control position on a arousal valence plane. Finally, duration, energy (velocity midi control) and articulation (duration + time delay to the next note) were calculated.

## 3 Results

In table 1 the correlation coefficients for each model are shown with an average of 0.3. Variability in accuracy is due to the fact that we selected the same feature set for the four models of each of the performance actions, so that combination may work well for some models but not for all four. However in Figure 2 it can be noticed, how the actual vs. predicted performance actions (duration ratio) curves follow a similar tendency. This indicates that the models are able to capture well the deviations, but its accuracy is penalized when particular cases with high error occur.

## 4 Conclusions

In this paper we have presented an approach to model four emotions (happy, sad, angry and tender), designing an expressive music performance system to

| | Angry | Tender | Happy | Sad |
|---|---|---|---|---|
| **DR** | 0.7509 | 0.0524 | 0.5484 | 0.4497 |
| **AR** | 0.3439 | 0.3796 | 0.4216 | -0.4334 |
| **ER** | 0.2711 | 0.2752 | 0.3709 | 0.2398 |

**Table 1.** Correlation Coeficients for Duration Ratio (DR), Articulation Ratio (AR) and Energy Ratio (ER), with linear regression.

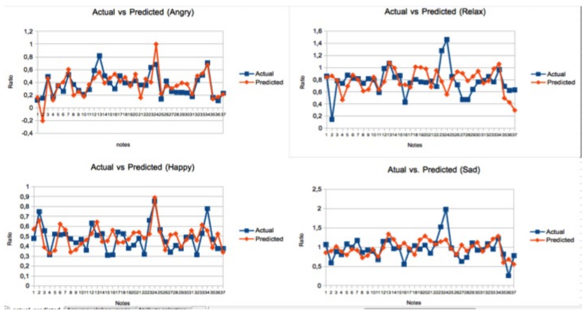


**Fig. 2.** Actual vrs predicted values for duration ratio, for each of the four moods. The models follow the tendency of the original performance actions.

render expressive performances in any mood across the arousal valence plane by interpolating the expressive parameters of the four original models. We have obtained recordings of the same musical piece in all four moods, and train linear regression models to predict performance actions in duration, loudness, and articulation. We interpolate the coefficients across the arousal valence plane to generate intermediate expressive models. The accuracy shows that models can learn the expressive parameters for each emotion.

## References

1. S. Canazza, G. De Poli, A. Rod, and A. Vidolin. An abstract control space for communication of sensory expressive intentions in music performance. Journal of New Music Research, 32(3):281294 (2003)
2. SR. Livingstone, RM., AR. Brown, and WF. Thompson. Changing musical emotion : a computational rule system for modifying score and performance. Computer Music Journal, 34(1):4164 (2010)
3. A. Friberg, R. Bresin, and J. Sundberg. Overview of the KTH rule system for musical performance. Advances in Cognitive Psychology, 2(2):145161 (2006).
4. A. Camuri. An experiment on Analisis and Synthesis of Musical Expressivity. In Proc. of 13th colloquium of music informatics, L'Aquila, Italy (2000)
5. S. Giraldo. Modeling Embellishment, Duration and Energy Expressive Transformations in Jazz Guitar. Masters thesis, Pompeu Fabra University, Barcelona, Spain (2012).