Leila Vaziri
ISE575-report#13

Computational Models of Expressive Music Performance: The State of the Art
Authors: Gerhard Widmer, Werner Goebl

Like any human intellectual activity, music performance is a complex social and cognitive phenomenon with a rich context. This paper has tried to give a comprehensive overview of the state of the art of computational modeling of expressive music performance. Four models and their approaches were presented, practical evaluations of the models on real performance data were showed. Music performance as the act of structuring and physically realizing a piece of music is a complex human activity with different aspects. This study will focus on one specific aspect which is expressive music performance.

Computational modeling is an attempt at formulating hypotheses concerning expressive performance in such a way that they can be verified on real measured performance data. Four specific models discussed in this study as: the rule- based performance model developed at KTH, the structure –level models of timing and dynamics, the mathematical model of musical structure and expression by Guerino Mazzola, Machine learning model. These models focus on commonalities between performances and performers.

Scientific model defines as a "familiar structure used as an analogy to interpret a natural phenomenon". Computational modeling involves embodying mathematical models in computer programs that can immediately be applied to given data for testing and prediction purposes. A mathematical or computational model is predictive in the sense that, assuming a specific fixed setting of all parameters, the model predicts the values of a specific set of variables from the values of other variables. The purpose of computational models of expressive music performance is to specify the physical parameters defining a performance, and to quantify systematic relationships between certain properties of the musical score, and an actual performance of a given piece. Models in human domain can not be expected to be correct as their predictions will always correspond to the behavior observed in humans. There are some implicit computational models that base their predictions on case-based reasoning.

The KTH model consists of a set of performance rules that predict aspects of timing, dynamics, and articulation, based on local musical context. The rule refers to a limited class of musical situations. Most of the rules operate at a rather low level, looking only at very local contexts and affecting individual notes, but there are rules that refer to entire phrases. The rules are parameterized with a varying number of parameters. The KTH model involves a professional musician directly evaluating any tentative rule brought by researcher. Musician and researcher are in a constant feed-back loop trying to find the best formulation and parameter settings for each rule. The important aspect of the model is that it is additive. This additivity is a particular problem when trying to fit the parameters to collections of real recordings. The model is a useable description language for expressive performance. In the experiment, the different emotional intentions as represented in the performances were differentiated by particular rules. The KTH model has been used to model certain emotional colorings that might not be immediately seen in the music structure. The extension to emotionality has led to a more comprehensive computational model of expressive performances (the GERM model). There is evidence that the KTH rule model is a viable representation language for describing expressive performance.

The Todd model in contrast to the KTH model may be summarized under the notion of "analysis by measurement ", because they obtain their empirical evidence directly from measurements of human expressive performances. The essence of these models is the assumption that there is a direct link between certain aspects of the musical structure and the performance. Also this relation can be modeled by one single, simple rule. The simplistic nature of Todd's model has the potential advantage that its theoretical assumptions can be tested relatively easily. Todd compared the model's output with the tempo and dynamics curves from one or two performances of selected pieces by Haydn, Mozart, and Chopin. Todd model was used as an analysis tool to access the idiosyncrasies of human performance.

A different model based mainly on mathematical consideration is the “Mazzola model” .The Mazzola model builds on “mathematical music theory” that not only covers various aspects of music theory and analysis through a highly complex mathematical approach, but also involves all sorts of philosophical, semiotic, and aesthetic considerations. The Mazzola model consists of an analysis part and a performance part. The analysis part involves computer-aided analysis tools for various aspects of the music structure, as meter, melody, or harmony. Each of these is implemented in RUBBETTEs that assign particular weights to each note in a symbolic score. The performance part transforms structural features into an artificial performance is theoretically anchored in “Stemma Theory” and “Operator Theory”. There is a linear mapping between metrical weight and tone intensity to generate artificial performances.The metrical, harmonic, and melodic weights as provided by the RUBATO software served as independent variables. The overall model could explain 84% of the average tempo curve of the 28 performances, each of the three analytical components contributing equally to the model.

An alternative way of building computational models of expressive performance is to start from large amounts of empirical data and to have the computer autonomously discover significant regularities in the data by inductive machine learning and data mining techniques. These learning algorithms produce general performance rules that can be interpreted and used directly as predictive computational models. The machine learning can predict local, note-level expressive deviations and higher –level phrasing patterns. These two types of models can be combined to yield an integrated multi-level model of expressive timing and dynamics. Musicians understand the music in terms of a multitude of more abstract structures, and they use tempo, dynamics and articulation to shape these structures. Music performance is a multi-level phenomenon with musical structure and performance patterns at various levels embedded within each other. There is a natural way of combining the phrase-level prediction model wit the rule-based learning model. After fitting quadratic approximation polynomials to given tempo or dynamics curve and subtracting from the original curve, what is left is residuals. Residuals are those low-level, local timing and dynamics deviations that can not be explained by phrases. The learning algorithm can be used to learn a rule based model of these local effects. Both the note-level rule model and the multi-level model have been tested on real performances. The machine processes the performance data and it produce reasonable performance pattern.

The study showed the quantification of individual styles. Predictive models like those described above generally focused on fundamental, common performance principles. The study shows the differences between artists that is aspects of personal artistic performance style. A statistical analysis revealed a number of characteristics and distinctive phrasing behaviors, some of which could be associated with certain pianists. The resulting performance data can be represented in an integrated way as trajectories in a tempo-loudness space that show the joint development of tempo and dynamics over time. Various ways toward the characterization of individual performance style are performance alphabet, and automatic identification of performers. The performance trajectories must first be converted into a form that is accessible to the automated date analysis machinery provided by data mining. The trajectories are cut into short segments. The resulting segments can be grouped into classes of similar patterns via clustering. They can see as a simple alphabet of performance, restricted to tempo and dynamics. Such performance alphabets support a variety of quantitative analysis. Another way to trying to quantify individual performance style is to develop computer programs that attempt to identify artists on the basis of their performance characteristics.

The result of this study shows that there is ample room for further research, and the field of computational performance modeling continues to be active. The new research could be finding the new control spaces and devices to control emotional aspects of music performance. The person and personality of the artist should be considered as one of the main factors in the models described in the study. There are some others aspects to be considered in predictive models, like listener expectations, performance context, artistic intentions, and personal experience.