

## **Computational Models of Expressive Music Performance: The State of the Art**

Gerhard Widmer and Werner Goebel

This paper focuses on expressive music performances. The goal is to “give an overview of the current state of the art of quantitative or computational modeling of expressive performance, and to give the reader an idea of some of the current research directions in this area”. Computational models are based on mathematical models which postulate quantitative relationships between various variables that are used to define the state of some system of interest. “Computational modeling involves embodying mathematical models in computer programs that can immediately be applied to given data for testing and prediction purposes”. More specifically, the purpose of computational models of expressive music performance is to determine the physical parameters of a performance and to quantify the relationships between the score, performance context, and actual performances. This paper focuses on presenting four such computational models of expressive music performance.

The first model presented is the KTH model. This model consists of a set of rules that predict timing, dynamics, and articulation based on local musical context. Each of these rules has a quantity control parameter  $k$  which is set by the researcher. The results are good when  $k$  is set close to 1. The rule may be disabled by setting  $k$  to 0 and the rule may be inverted by setting  $k$  to a negative number. This model is developed using the “analysis-by-synthesis” approach. With this approach, musicians and researchers are in constant feedback loop trying to determine the best parameter settings for each rule. Several evaluations of the KTH model have been conducted. This paper presents seven such tests. The first study was conducted by the Sundberg group. They evaluated the perceptual responsiveness of musicians and non-musicians to expressive variations in two listening tests. In the first experiment, they found the perceptual thresholds for the settings of parameter  $k$ . In the second experiment, musicians adjusted the  $k$  values for six rules. Another experiment was conducted by Friberg which used a greedy search method to fit the parameters of a single rule to the timing data of 28 performances of one piece. Parameter settings were found that produced a reasonable fit. Another study by the Sundberg group examined how well the KTH model can be fitted to a particular Mozart piece. Kroiss used genetic algorithms to fit a set of  $k$  parameters to a large collection of pieces. He could not find a single set of parameter settings to produce a better fit than the baseline. Zanon and De Poli attempted to fit the model to real world data. It was found that the rules of the model were quasi-orthogonal to each other. In another study, they attempted to change the model so that the  $k$  parameter to vary throughout a piece. The overall results were confusing even though they were positive for particular pieces. Another study used the model to determine emotional colorings.

The second model presented is the Todd model. The assumption in this model is that “there is a direct link between certain aspects of the musical structure and the performance and, secondly, that this relation can be modeled by one single, simple rule”. This model relies on the relation between timing and dynamics of “the faster the louder, the slower the softer”. Unlike the KTH model, this one is developed using an “analysis-by-measurement” approach. In this approach, empirical evidence is obtained from measurements of human expressive performances. In an empirical study by Clarke and Windsor, listeners evaluated both human performances and algorithmic performances generated using the Todd model. It was found that timing and dynamics did not relate to each other in the manner suggested by the model. In another study, Clarke and Windsor created different algorithmic performances using different level weightings for the Todd model and evaluated them using human performances. The best

fit for the algorithmic performances was one that had different level weights for timing and dynamics.

The third model presented is the Mazzola model. This model is based on mathematical considerations. It builds on a theoretical background that covers music theory, philosophy, semiotics, and aesthetics. It consists of an analysis part and a performance part. The authors could not find empirical evaluations of this model. They glossed over this model. I was left with the impression that they probably did not fully understand this model and therefore failed to present it in any useful way.

The last model presented is the machine learning model developed by the authors. This model takes as input large amounts of data and has the computer discover significant regularities in the data using inductive machine learning and data mining. The first step in developing this model was to create a new inductive rule learning algorithm and apply it to learning note-level rules for timing, dynamics, and articulation. This note-level model can only account for a small part of the expressive patterns in performances. Therefore, more recent work as focused on inductively learning multi-level models of expressive timing and dynamics from recordings. The empirical tests described in the paper come from the authors themselves (or others in their group). In one study, coverage and precision on the training set are listed in detail for each discovered rule. This gives a “detailed picture of the relative generality and reliability of the rules”. Also, the predicative performance of the multi-level model is quantified “by measuring how well it manages to predict the details of a pianist’s performances”.

This paper also tries to “address some ongoing research that also tries to quantify and characterize, if not yet model in a predictive way, the differences between artists, that is, aspects of personal artistic performance style”. For the sake of space, I will not go further into this topic than to say that the works presented in this section come from the group the authors belong to. I found this paper to be completely biased and skewed in its view. The majority of the paper presented work done by the authors or their group. It also did a poor job of presenting the Mazzola model.