

Ching-Hua Chuan

Mar 1 2006

Synopsis and Response Report of Computational Models of Expressive Music Performance: The State of the Art by Gerhard Widmer and Werner Goebel

This paper provides an overview of state-of-the-art computational modeling of expressive music performance. Four selected computational models are reviewed in detail, including the rule-based performance model from KTH (Sundberg et al.), the structure-level models of timing and dynamics by Todd, the mathematical model of musical structure and expression by Mazzola, and the machine learning model combining note-level rules with structure-level expressive patterns by Widmer. Apart from these four models, Widmer and Goebel briefly describe their recent project on differentiating artists by performance styles at the end of the paper, with a focus on the commonalities between performances and performers,

The KTH rule-based model consists of a set of performance rules that predict or prescribe aspects of timing, dynamics, and articulation, based on local musical context. The KTH model is developed by “analysis-by-synthesis approach,” that involves a professional musician who directly evaluates the rules generated by the researcher. A feedback loop between evaluating and reformatting is built, aiming to find the best formulation and parameter settings for each rule. For every rule, a quantity control parameter (a “switch”) is designed in order to formulate the final performance the cumulative sum of all rules multiplied by the switches. In 1995, Friberg applied a simple greedy search method to fit the parameters of a single rule to the timing data of 28 performances for the first nine bars of Schumann’s *Träumerei*. The results indicate that the model is a useable description language for expressive performance. More recently, Zanon and De Poli attempt to fit the model to real-world data using several human performances with some specific emotional intentions. The result shows that KTH rules are quasi-orthogonal to each other with only few exceptions.

In the late 1980s and early 1990s, Todd proposed several structural-level models of expressive timing and dynamics. Todd’s model is categorized as the “analysis-by-measurement” approach, obtaining empirical evidence directly from measurements of human expressive performances. The model is based on two assumptions: there is a direct link between certain aspects of the musical structure and the performance, and this relation can be modeled by one simple, single rule. The well-known Todd’s rule, “the faster the louder, the slower the softer,” has been evaluated in several studies. In Clarke & Windsor’s study in 2000, a panel of human listeners evaluated both human performances and algorithmic performances created by Todd’s model. However, the results demonstrate that expressive timing and dynamics did not relate to each other in the simple manner as suggested by Todd’s model. In another empirical study in 1997, Windsor and Clarke tune the model’s parameters to human performances. The best

algorithmic performance, the so-called hybrid performance, is the one with different level weights for timing and dynamics – timing requires more emphasis on lower structure levels and dynamics on higher levels.

The Mazzola model is based on an enormous theoretical background, the “mathematical music theory,” and consists of an analysis part and a performance part. The analysis part applies computer-aided analysis tools for various aspects of music structure, including meter, melody, and harmony. Each aspect is implemented into a plugin, the RUBETTE, which assigns a particular weight to each note in a symbolic score. The performance part transforms structural features into an artificial performance. The model is implemented into a software package call EspressoRUBETTE, which has the capabilities of analyzing MIDI format data input, performing score-to-performance matching, and extracting performance vector fields for a given human performance.

The last model described in this paper is a machine learning system developed by the authors themselves. The model produces general performance rules by *inductive machine learning* and *data mining* techniques from a large amount of empirical data. Two types of models are designed: the note-level model and the multi-level model. One example of the learned note-level rules is: “*Given two notes of equal duration followed by a longer note, lengthen the note that precedes the final, longer one, if this note is in a metrically weak position.*” In the note-level model, the system learns note-level rules for timing, dynamics, and articulation, i.e., how a pianist is going to play a particular note in a piece. For the multi-level model, the goal of the system is learning how to predict the kind of elementary tempo and dynamics “shape” at a given level of the phrase hierarchy.

The last part of the paper describes a recent project, automatic identification of performers. Each performance is analyzed into trajectories in a tempo-loudness visualization space proposed by Langner and Goebel. The trajectories are cut into short segments and clustered into groups. The trajectory at the center of each group is chosen to be the *performance alphabet*. Then the performances can be distinguished from each other by a sequence of coded performance alphabet “letters.”

This paper is a comprehensive summary of state-of-the-art performance research. Widmer and Goebel introduce four major models, describing the basic ideas and assumptions of each model, and reporting any empirical results. The advantages and disadvantages of the models are also compared, which help readers get better understanding of each model. It is always interesting to read about the different approaches to the same research topic. The most intriguing part is to see how people look at the same problems and with what assumptions, not the technique details.