

Review of

Computational Models of Expressive Music Performance: The State of the Art
By Gerhard Widmer and Werner Goebel

The goal of this paper is to give an overview of the current state of the art of quantitative or computational modeling of expressive performance. The author gave a review of the four selected computational models in detail as well as briefly discussed quantitative models and predictive models. The currently ongoing research for characterization of differences between different performing styles is presented at the last part.

The most important characteristic of computational model is PREDICTABLE. The author went through four comprehensive models (the KTH model, the Todd model, the Mazzola model, and the Machine Learning model). The KTH model is a rule-based model which means it consist a set of performance rules to predict and prescribe the timing, dynamics, and articulation. It is noted that most rules are looking at local levels. We can choose a set of quality control parameters 'k' to fit one particular rule. The model features a lot of free parameters that govern the relative strengths of the individual notes. The KTH model is developed by an "analysis-by synthesis" approach and is a kind of performance-listener interaction, but it also falls potentially short by placing high demands on the evaluating musicians and by not reliable evaluations due to the small number of perceptual judgment. The model is additive so it results in a problem when trying to fit the parameters to collections of real recordings.

The Todd Model is developed by the method "analysis-by-measurement", which means it obtains empirical evidence directly from measurements of human expressive performances. This model is based on the assumption that there is a direct link between certain aspects of the musical structure and the performance which can be modeled by one single, simple rule. According to Clarke and Windsor, the timing required more emphasis on lower structural levels, whereas dynamics on higher ones (1997), and the timing and dynamics didn't relate to such a simple manner suggested by the model (2000). In 2003, their further experiment concluded the Todd model is used as an analysis tool to assess the idiosyncrasies of human performances.

The Mazolla Model was built upon not only various aspects of music theory and highly complex mathematical approaches, but some philosophical, semiotic and aesthetic considerations. In analysis part, each music structure such as meter, melody, or harmony are implemented in a particular plug-ins called RUBETTE of computer program. The RUBETTE assigns particular weighs of each note in a score. In performance part, the music structure is transformed in to an artificial performance by the way called "Stemma Theory" and "Operator

Theory". The model iteratively modifies the "performance vector fields", each of which controls a single expressive parameter of a synthesized performance. Jan Beran found that the model could explain 84% of the average tempo curve of 28 performances (1992), but there is no other empirical investigation or quantitative evaluations of the model except by Mazzola himself.

The Machine Learning Model collects large amounts of empirical data and by using inductive machine learning and data mining techniques, the computer could find the regulations in the data. The model predicts both local, note-level expressive deviations and higher-level phrasing patterns. It combines note-level and rule-based learning model to yield a multi-level model of expressive timing and dynamics "shapes". This model could serve as a predictive computational model of expressive timing and dynamics that takes into account of both the hierarchical music structure and local music context. According to Widmer (2002), Some rules tested on Chopin's pieces could significantly accurately predict the Mozart's pieces. The model is better for dynamics than for timing or tempo in multi-level aspects.

There are some interesting facts mentioned. The person and personality of the artist as a mediator between music and listener is totally neglected among all the models. If what Bruno Repp (1992) indicated was right, all the pianists have the same performance at a global level, but there were differences between how the pianists deal at lower levels of the structural hierarchy, such as distinctive phrasing behaviors. It means all the models we have can only detect the characteristics on the global level instead of local levels. Although some recent researches are focus on the basic emotions (expressive intentions), I wonder the possibility that the different basic emotions could be modeled since it depends on individuals. Even some of the problems are not likely to be solved, these researches advance our understanding and appreciation of the complexity of artistic behavior.