

Widmer, Gerhard and Asmir Tobudic. "Playing Mozart by Analogy: Learning Multi-level Timing and Dynamics Strategies." Journal of New Music Research. Vol. 32, No. 3, pp. 259-268. 2003.

This article describes a system that learns to predict elementary tempo and dynamics "shapes" at different levels of the hierarchical musical phrase structure, and combines these predictions with local timing and dynamics effects predicted by learned note-level models. This system utilizes two types of learning algorithms: 1) a simple nearest neighbor algorithm that predicts phrase-level expressive (polynomial) shapes in new pieces by "analogy" to shapes identified in similar phrases in other pieces; and, 2) a rule learning algorithm that learns prediction rules for note-level effects from the "residuals" that cannot be attributed to the phrase structure by the expression decomposition algorithm. The inputs to the system are: 1) the scores of musical pieces; 2) tempo and dynamics curves which are represented as multiplicative factors of relevance to the average tempo and dynamics of the piece; and, 3) hierarchical phrase structure information, with four levels of phrasing, which are done by hand. For training examples of phrase-level and note-level learning, for each phrase at a given level, the second-degree polynomial is computed that best fits the part of the curve that corresponds to this phrase, and the tempo or dynamics deviations are "subtracted" resulting in a residual curve. The final curve is obtained after the fitted polynomials at all phrase levels have been "subtracted" is called the residual of the expression curve.

There are four main problems discussed about the system: 1) the limitation of the training set makes the learner's task difficult; 2) it does not take many mistakes on the part of the learner to produce a result of very poor musical quality; 3) the tempo curve cannot be well approximated by phrase-level shapes – there is too much not accounted for; and, 4) it seems that the phrase structure analysis might have been performed at too global a level, missing some relevant grouping boundaries that guide performance timing. The authors have concluded that the dependency of their approach on a "correct" or "appropriate" phrase structure analysis, where "appropriate" actually depends on the shape of the performance curves to be approximated, creates a circulatory problem. General problems associated with possible future work include: 1) the limitation of the propositional attribute-value representation that is currently used to characterize phrases does not permit the learner to refer to details of the internal structure and content of phrases; 2) the nearest neighbor learning algorithm does not produce interpretable models; and, 3) shapes applied at different levels are highly dependent on each other but are currently treated as independent because otherwise the problem is too difficult.

My attitude towards the authors who created this system is like psychologists and biologists trying to create an artificial brain. Though the authors are quite good about describing their steps and reasons, they use one pianist for their analysis as well as have a very limited learning set. The limitations of learning sets are why I dislike learning algorithms. You need a huge database in order to get any respectable results. Using only 16-pieces as the learning set and expecting a correct analysis (as the authors have done) is like making 16 connections to different parts of the artificial brain and expecting it to think. The large database and other limitations to learning algorithms are just some of the reasons why automation and artificial intelligence are disciplines in which people specialize. It is too huge of a problem to be taken lightly. I understand that this is only supposed to be viewed as a first step, but I hope the authors realize the vastness of the problem that they are attempting to solve just as in the artificial brain problem.