International Symposium on Performance Science
ISBN 978-94-90306-02-1

Toward a multilevel model of expressive piano performance

Sebastian Flossmann[1] and Gerhard Widmer[1,2]

[1] Department of Computational Perception, Johannes Kepler University, Linz, Austria
[2] Austrian Research Institute for Artificial Intelligence, Vienna, Austria

Expressive performance modeling requires different information for each expressive dimension. Most systems, however, rely on a single approach for all dimensions. Further, tempo and timing are mostly treated as one atomic entity instead of being decomposed into elements and treated separately. We propose a performance model that discriminates expressive dimensions with regard to the modeling approach and, additionally, uses separate subsystems to model tempo and timing.

Keywords: expressive piano performance; performance model; support vector machines; probabilistic reasoning; multi-level model

Modeling expressive musical performance is a complex task with information requirements that vary from one expressive dimension to another. For example, dynamics is guided to a considerable extent by annotations in the score, whereas the overall performance tempo is more closely related to phrasing (Todd 1989). Timing and articulation, however, may depend more on local aspects of the score.

Performance modeling systems normally rely on one of three common approaches: (1) probabilistic models (e.g. Grindlay and Helmbold 2006), (2) rule-systems (e.g. Friberg *et al.* 2006), or (3) case-based reasoning (e.g. Widmer and Tubodic 2003). The system we discuss in this study differs from the bulk of performance rendering systems in two significant aspects. Firstly, common to all the systems is that they use the same approach for all performance dimensions. The system we present takes a modular approach that treats dynamics, articulation, and tempo differently. Secondly, with the noteworthy exception of Widmer and Tubodic (2003), most systems view tempo and timing as an atomic dimension of performance. We consider the tempo curve of a performance to be an aggregate of different components which we

treat separately: timing of individual notes (*note timing*), phrase-related tempo trends (*local tempo*), and global performance tempo (*tempo markings*).

Our performance model has its roots in the probabilistic rendering system YQX, which won the Rendering Contest RENCON 2008 (Hashida *et al.* 2008). A detailed description of YQX can be found in (Widmer *et al.* 2009). In the original system, a simple Bayesian Network predicted tempo, loudness, and articulation. While the prediction of articulation remains the same in the current system, tempo prediction is replaced by three subsystems, one for each of the components mentioned above. The loudness model is replaced by a model relying on a decomposition of the dynamic annotations in the score (Grachten and Widmer 2011). In this study, we discuss how we handle two of the aspects of performance tempo: local tempo and note timing.

METHOD

Data and score representation

The system is trained using two unique corpora of performances: 13 complete Mozart piano sonatas performed by Roland Batik and the complete works for solo piano by Chopin performed on stage by Nikita Magaloff. All pieces were played and recorded on a Bösendorfer computer-controller grand piano and converted from Bösendorfer's proprietary format to MIDI. All performed notes were aligned to their counterparts in symbolic representations of the score. This resulted in a collection of performances with detailed performance and complete score information for each note.

We describe the score using a combination of local descriptors (rhythmic and melodic) and higher-level features from the Implication-Realization (I-R) model of melodic expectation by Narmour (1990):

- *Duration ratio* describes the ratio between the score duration of a note and its successor.
- *Rhythm context* is an abstract description of a note's duration in relation to its neighbors (e.g. long-short-long).
- *Metrical strength* describes the metrical importance of a note-onset.
- *Pitch interval* measures the distance to the next note in semi-tones.
- *IR-label* is the name of the I-R situation applicable to a note.
- *IR-arch* measures the distance to the next point of strong closure according to an I-R analysis of the score.

Modeling "tempo" as a composite phenomenon

In musical performances, tempo usually refers to a combination of three aspects: (1) *global tempo* refers to the initial tempo prescription at the beginning of a score; (2) *local tempo* describes localized tempo trends which, for example, outline larger musical units (e.g. phrases) and realize annotations in the score; (3) *(local) note timing* refers to local (note-wise) deviations from the local tempo that emphasize single notes through delay or anticipation. In order do make the tempo as observed in a performance independent of *global tempo*, we transform it into a series of logarithmic ratios between score and performance inter-onset intervals (IOIs). We call the result *complete tempo curve* and view it as a composite of local tempo and note timing. More precisely, we associate *local tempo* with the low-frequency content of the complete tempo curve, which we extract by applying a moving average. The residual, the curve that remains after subtracting the local tempo from the complete tempo curve, is associated with *note timing*.

Predicted tempo and timing

Assuming that note timing is a local phenomenon, we model and predict it using the simple Bayesian approach of the original YQX system. The predictions depend only on the immediate score characteristic of each note. With respect to local tempo, we consider two methods: (1) the performance-context-aware Bayesian model presented in Flossmann *et al.* (2009; YQX-global) and (2) support vector machines with and without local performance context (SVM and SVM-C, respectively). The Bayesian network approach is an adaptation of the Viterbi-Algorithm for Hidden Markov Models that results in a tempo prediction that is optimal in the sense that at each point the prediction is the value with the highest probability given the current score characteristics and performance predictions. The SVM we use is a regression model with a Gaussian kernel. To incorporate performance context, we use the previously predicted tempo value as an additional input feature.

RESULTS

In this section, we first discuss the results of experiments using both the Mozart and the Chopin corpora and then inspect qualitative aspects of the different predictions. The experiments were conducted on subsets of the corpora, selected according to stylistic criteria—fast and slow movements for the Mozart sonatas, different categories (ballades, nocturnes, etc.) for the Chopin data—as they might contain different interpretational concepts that could

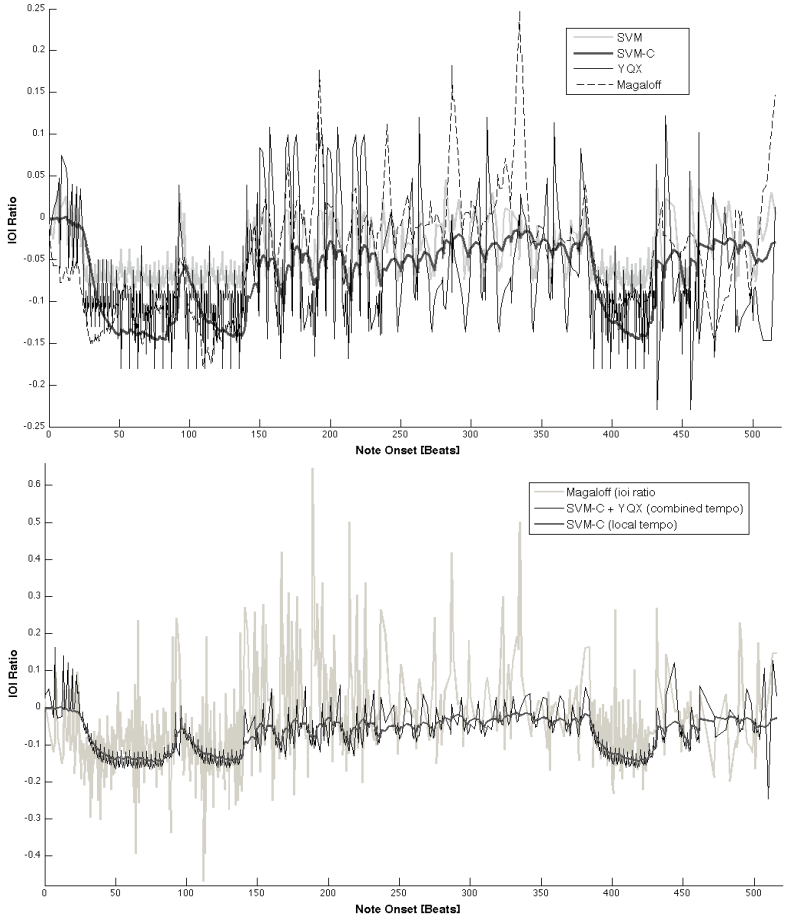

Figure 1. Local tempo predictions by three algorithms for Waltz Op. 34 No. 3 (upper panel) and tempo curves for Waltz Op. 34 No. 3 (lower panel), as observed in Magaloff's performance and combined from separate predictions for local tempo and note timing.

also be reflected in the predictions. As a numerical quality indicator, we use the correlation coefficient that measures the similarity between the predicted and the real tempo curve.

The quality of the results of the different algorithms depends heavily on the selected subset of score features and the tempo aspect they are trained to

predict. Their respective best results are based on different sets of features, which suggests that different dimensions of tempo indeed depend on different aspects of the score. Trained with suitable feature sets, the note timing predictions of the different algorithms are numerically comparable, with the context-free algorithms YQX and SVM performing slightly better. For predictions of local tempo it seems beneficial to incorporate the performance context: both context-aware algorithms, YQX-global and SVM-C, outperform the context-free algorithms by roughly 15%.

Qualitative evaluation

Although the quantitative evaluation does not indicate significant differences between the algorithms, the curves they predict exhibit discriminating characteristics. Figure 1 shows tempo curves predicted for Waltz Op. 34 No. 3 by Chopin. The upper panel illustrates a typical situation often found in predictions for local tempo: the prediction made by the context-free Bayesian approach (YQX), while reasonably similar to the original (r=0.37), exhibits sharp fluctuations, which is not a desired characteristic for local tempo. The curve predicted by the context-free SVM is more similar to the original (r=0.49) but comparably unsteady. Both algorithms have only information about the local score context, which explains the similar behavior. Integrating the performance context seems to have the desired effect: the curve predicted by the context-aware SVM (r=0.66) is much smoother with distinctive trends. The lower panel shows a complete tempo curve assembled from a local tempo prediction with a context-aware SVM, and a note timing prediction with the context-free YQX. The resulting curve retains the coherent tempo trends from the local tempo prediction and is enriched by the local variations from the note timing prediction. Compared with the result of a context-aware SVM trained to predict the complete tempo curve directly, the correlation of the combined prediction is slightly lower (r=0.38 and r=0.41, respectively). However, the combined curve is steadier than the directly predicted curve and displays much clearer trends.

DISCUSSION

We have presented a performance model that takes a multi-level approach to tempo prediction: instead of searching for one model for all aspects of tempo, components relating to different levels of locality are modeled by specialized subsystems and afterwards combined to form the tempo. The resulting tempo curves seem to reproduce better the musicality of the performances. This is not always reflected in the numerical similarity, but rather than suggesting an

aesthetically inferior result, this casts doubt on the suitability of correlation as a quality indicator. Further research will explore evaluation criteria that are musically more meaningful. Another fundamental problem is the score model: simple score descriptors cannot capture abstract musical concepts such as phrase boundaries, cadences, and harmonic suspensions. We believe that the most promising line of investigation lies in creating a more musically meaningful score model that describes the score at several different levels.

Acknowledgments

This research is supported by the Austrian Science Fund (FWF) under project numbers TRP 109-N23 and Z159 (Wittgenstein Award).

Address for correspondence

Sebastian Flossmann, Department of Computational Perception, Johannes Kepler University, Altenbergerstr. 69, Linz 4040, Austria; *Email:* sebastian.flossmann@jku.at

References

Flossmann S., Grachten M., and Widmer G. (2009). Expressive performance rendering: Introducing performance context. *Proceedings of the SMC 2009: The 6th Sound and Music Computing Conference*. Porto, Portugal.

Friberg A., Bresin R., and Sundberg J. (2006). Overview of the KTH rule system for musical performance. *Advances in Cognitive Psychology, 2,* pp. 145-161.

Grachten M. and Widmer G. (2011). Explaining musical expression as a mixture of basis functions. *Proceedings of the SMC 2011: The 8th Sound and Music Computing Conference* (submitted). Padova, Italy.

Grindlay G. and Helmbold D. (2006). Modeling, analyzing, and synthesizing expressive piano performance with graphical models. *Machine Learning, 65,* pp. 361-387.

Hashida M., Nakra T. M., Katayose, H., *et al.* (2008). Rencon: Performance rendering contest for automated music systems. *Proceedings of the 10th International Conference on Music Perception and Cognition (ICMPC 10)*. Sapporo, Japan.

Narmour E. (1990). *The Analysis and Cognition of Basic Melodic Structures*. Chicago: University of Chicago Press.

Todd N. P. (1989). A computational model of rubato. *Contemporary Music Review, 3,* pp. 69-88.

Widmer G., Flossmann S., and Grachten M. (2009). YQX Plays Chopin. *AI Magazine, 30,* pp. 35-48.

Widmer G. and Tubodic A. (2003). Playing Mozart by analogy: Learning multi-level timing and dynamics strategies. *Journal of New Music Research, 32, pp.* 259-268.