Technical Report RENCON 2025

Generating piano performances with three models in Director Musices using a semi-automatic approach.

Anders Friberg, KTH Royal Institute of Technology, Stockholm, Sweden, afriberg@kth.se
Gabriel Jones, University of Music and Performing Arts Graz, Austria, gabrieljimijones@gmail.com

Introduction

The Director Musices (DM) system is a collection of algorithms for shaping musical performances. The initial research started in the 1980's (Sundberg et al., 1983) and the system is still periodically supported and developed. The system uses a combination of different modelling tools to emulate aspects of human performance, such as phrasing, articulation, timing, dynamics, intonation, and groove. Many of these different tools have been verified and tested (for an overview see Friberg, et al., 2006). It is therefore not a "black box" system that generates performances according to a set of training data, though some machine learning methods have been used. The main purpose has been to get an understanding of how musicians are communicating different aspects through the performance variations, such as the musical structure, the motional quality, or the emotional quality.

We chose to use only three different models to generate our performances of the set examples. The first one was a recent model of dynamics, the second was the previously defined phrasing rule for the timing, and the third was the final ritardando model.

The dynamics model was recently developed and applied within the DM system (Jones and Friberg 2023). This model was developed through machine learning predicting dynamics from a set of features. Initially 48 features were defined for each note that captured different aspects of the musical structure. They were later reduced to 34. The features were divided in five different groups: pitch, timing, meter, and phrasing, which made it possible to estimate the influence of each group on the performed dynamics. Support Vector Regression was used to make the prediction from the features to the performed dynamics. The model was trained on a set of 30 diverse melodies selected from the baroque, romantic and post-tonal repertoire. All the melodies were performed by 20 pianists on a Disklavier piano. The dynamics for each note was averaged across the pianists resulting in a total of 1701 data points used in the training. For a detailed description, see Jones and Friberg (2023).

The phrase arch rule has been part of DM for some time (Friberg, 1995) and has been used in many studies (e.g. Juslin et al., 2001). Its basic function is to apply "arches" over the phrases that modulate the tempo and the dynamics, typically specifying a slow-fast-slow shape for the tempo and a soft-loud-soft shape for the dynamics. It was designed to be flexible to account for the variation found in expressive performances and has several parameters that can be used to shape the arches in a more detailed way. In this instance, we only used the phrase arch for timing variations in this case.

The final ritardando model was based on experimental data taken from the tempo variation of ritardandi at the end of performed pieces, as well of the deceleration of runners when slowing to a stop (Friberg and Sundberg, 1999).

Scores

We chose to render the following pieces.

1. Handel: Capriccio in G minor, HWV 483
2. Beethoven: 32 Variations in C minor, WoO 80 - Theme and the first 5 variations
3. Mozart: Pianosonata No. 16, K545, first 28 measures.

Procedure

For the first two examples, we opened the MusicXML files in Sibelius and exported the corresponding MIDI files. The MIDI files were imported into DM and then edited so that they corresponded to the initial score notation. This was necessary since the exported MIDI did not correspond exactly to the quantified notation in the score for some notes. The Mozart example was previously imported into DM. Phrasing marks were manually inserted. For the Handel example, the phrasing followed the legato marks in the score closely, with some simplifications. For the other two examples, phrasing marks were inserted at phrase boundaries, often every other bar, as well as section boundaries. The phrase structure was relatively simple and straightforward in these examples. The dynamic markings such as *pp* and *f* were inserted according to the score with some small changes. No other performance marks in the score were transferred.

The dynamics model is not yet fully implemented in DM, therefore the features were first computed in DM, exported to Matlab for the application of the trained model, and finally imported back into the score in DM. In this procedure, the right- and left-hand parts were individually processed, applying the same rules to each part as an individual melody. Although the model is fully automatic, DM still allows us to change the overall amount that is applied to the score using the dedicated quantity parameter. At this point, the score in DM was ready for the rule application and testing.

For each of the three examples we defined a "rule palette", that is a set of models with the corresponding parameters, see Figure 1. We found that a factor of two for the dynamics model was reasonable for all the examples, creating a nuanced expressive contour, without any excessive peaks. This corresponds roughly to the dynamic range used by the pianists in the model training. We used almost the same phrase arch settings for all examples with some small variations. The shape of the arch was rather flat in the beginning, slightly faster through the phrase and with a more pronounced (but still small) ritardando at the end of each phrase. This kind of phrasing is typical of "motorically" oriented pieces with a clear pulse, often in repeated patterns. The final ritardando was only applied in the very end of each example.

In summary, this can be considered a hybrid approach with a fully automatic generation of dynamics together with a phrasing selected by manual adjustments, however with rather simple and common parameter selections.

Audio rendering

The audio was rendered from MIDI using the software piano simulation Pianoteq ver. 8.4.0 with the factory patch "HB Steinway D Prelude". The only modification was that the reverb was changed to large hall. The MIDI velocity data from DM was calibrated specifically for Pianoteq so that the internal measure of dynamics in DM was in decibels (c.f. Bresin et al., 2002). Nothing was edited after the rule application.

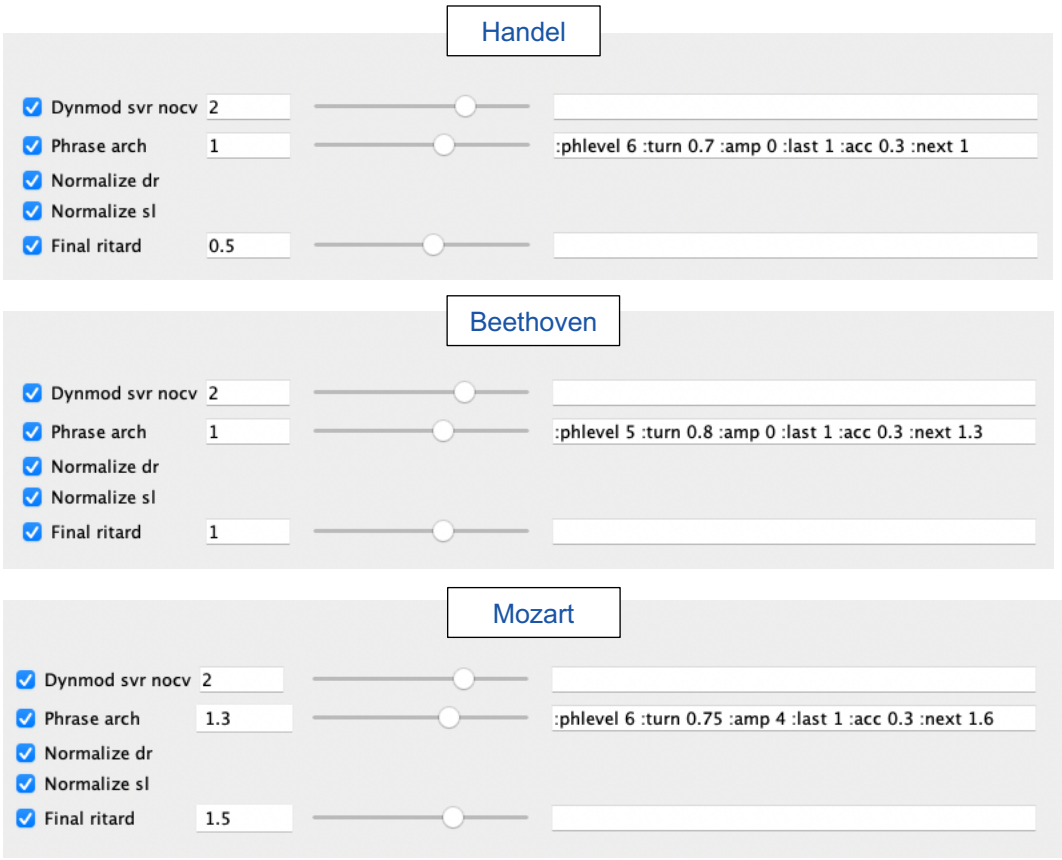

Figure 1. The specific model parameter settings for each example. "Dynmod svr nocv" corresponds to the dynamics model. The number in the box is the overall amount of each rule where a value of one corresponds roughly to an amount that is considered noticeable but not too pronounced.

References

Bresin, R., Friberg, A., & Sundberg, J. (2002). Director musices: The KTH performance rules system. In *SIGMUS-46, Kyoto, Japan* (pp. 43-48). Information Processing Society of Japan,.

Friberg, A. (1995). Matching the rule parameters of Phrase arch to performances of" Träumerei": A preliminary study. *STL-QPSR*, *36*(2-3), 063-070.

Friberg, A., Bresin, R., & Sundberg, J. (2006). Overview of the KTH rule system for musical performance. *Advances in cognitive psychology*, *2*(2), 145.

Friberg, A., & Sundberg, J. (1999). Does music performance allude to locomotion? A model of final ritardandi derived from measurements of stopping runners. *The Journal of the Acoustical Society of America*, *105*(3), 1469-1484.

Jones, G., & Friberg, A. (2023). Probing the underlying principles of dynamics in piano performances using a modelling approach. *Frontiers in Psychology*, *14*, 1269715.

Juslin, P. N., Friberg, A., & Bresin, R. (2001). Toward a computational model of expression in music performance: The GERM model. *Musicae scientiae*, *5*(1_suppl), 63-122.

Sundberg, J., Askenfelt, A., & Frydén, L. (1983). Musical performance: A synthesis-by-rule approach. *Computer Music Journal*, *7*(1), 37-43.