Computational Models of Expressive Music Performance: The State of the Art

Gerhard Widmer and Werner Goebl
Austrian Research Institute for Artificial Intelligence
Vienna, Austria

About the Authors

Gerhard Widmer
- Professor and Head of the Department of Computational Perception at the Johannes Kepler University at Linz, Austria.
- Head of the Machine Learning, Data Mining, and Intelligent Music Processing Group at the Austrian Research Institute for Artificial Intelligence (ÖFAI), Vienna.
- Research Interests
  - Machine Learning and Pattern Recognition
  - Knowledge Discovery in Databases / Data Mining / Text Mining
  - Intelligent Music and Audio Processing
  - Computational Models of Expressive Music Performance
  - Music Information Retrieval (MIR)

Werner Goebl
- Postdoctoral fellow of the Sequence Production Lab at the McGill University Montreal.
- Receiving his PhD at Graz in 2003, recently received his concert diploma for piano chamber music.
- Research Interests
  - Expressive piano performance and artificial intelligence methods of visualization and machine learning of music

What We Should Know about Computational Modeling

- A scientific model: definition in Encyclopedia Britannica
  - "A familiar structure or mechanism used as an analogy to interpret a natural phenomenon"
- The purpose of computational models of expressive music performance is to specify precisely the physical parameters defining a performance, e.g., onset timing, inter-onset intervals, loudness levels, note durations, etc.
- An important characteristic of computational model is PREDICTABLE.
- Four Comprehensive Models
  - the KTH model
  - the Todd model
  - the Mazzola model
  - the Machine Learning model
### The KTH model – Developer

- It was developed by Royal Institute of Technology ("Kungliga Tekniska Högskolan" in Swedish, KTH) in Sweden more than 20 years.
- Johan Sundberg
  - Ph.D. in musicology Uppsala University 1966
  - 1976-1981 President of the Music Acoustics Committee of the Royal Swedish Academy of Music
  - a fellow of the Acoustical Society of America

### the KTH model – Basic concepts

- It is a rule-based performance model.
- 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.
- For example, Friberg (1995b) used 18 parameters to fit the PHRASE ARCH rule in the analysis of performance of the first nine bars of Schumann’s Träumerei (Dreaming, from “Scenes from Childhood”).
- The quality control parameter ‘k’ is given to have a fairly good result when its value is around 1. With a ‘k’ value of 0, a particular rule is switched off; when ‘k’ is set to a negative value, the rules are inverted.

### the KTH model – Evaluations

- It is developed by an “analysis-by-synthesis” approach, which involves a professional musician directly evaluating the tentative rule brought forward by the researcher. Musician and researcher are in a constant feedback loop trying to find the best formulation and parameter setting for each rule.
- The model 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 judgements.
- An special aspect of the model is it is additive, that is, the effects of the individual rules are added cumulatively to give the combined effect. This aspect also results in a problem when trying to fit the parameters to collections of real recordings.

### the KTH model – the Perceptual Evaluations

- Sundberg group evaluated the perceptual responsiveness of musicians and non-musicians to expressive variations.
- In the first experiment, they determined the perceptual threshold for the individual rule quantities (‘k’ values) and found that musicians were more sensitive than non-musicians. The quantity thresholds depended strongly on the type of rule. In the second experiment, skilled musicians adjusted the ‘k’ values for six rules.
- More, Sundberg group examined how well the KTH model can be fitted to a particular performance of the slow movement of Mozart’s piano sonata K332, Adagio, manually by trial and error.

---

Widmer & Goebel: Computational Models of Expressive Performance

Presented by T.H. Chiang
the KTH model – the Perceptual Evaluations

- They focused exclusively on expressive timing of the melody and determined similarity between the performance and the model with the correlation coefficient. They tested 'k' values for each phrase individually, and the PHRASE ARCH rule, the shaping phrases on different structural levels, got the highest correlations, and it seems dominate the performance.

- Recently, Zanon and De Poli attempted to fit the model to real-world data with fixed parameter. They found the most of KTH rules were quasi-orthogonal to each other with only few exceptions. The different emotional intentions in the performance could be differentiated by particular rule, i.e., the PUNCTUATION rule separated fearful or sad form happy or neutral.

the KTH model – Evaluations

- with time-varying ones, they tried to infer 'k' parameters for windows of 1-2 bars, shifting the windows notewise along with the score. They restricted the number of included rules in first estimation phrase, and then added the window size and the number of included rules from second phrase. This 'k' parameter estimation was very sensitive to artefacts, such as timing of grace notes.

- Some specifically selected subsets of rules and 'k' parameter settings called 'emotional rule palettes' were derived from measured performances in order to model emotional colouring. For example, the DURATION CONTRACT rule distinguish between a sad and a happy performance.

the Revised KTH Model & Conclusions

- As a result, The GERM model is derived with the concept of emotionality by Juslin in 2002. The model contains:
  - Generative KTH model
  - Emotional colouring mentioned on last page
  - Random variability
  - physical Motion
  - recently included factor, Stylistic unexpectedness (2003)

- The KTH rule model is a viable representation language to describing expressive performance, but to what extent it can account for the observed variations in large collections of performance of complex music is still an issue.

The Todd model – Developer

- Neil Todd
  - a BSc in Theoretical Physics in 1982
  - working with Henry Shaffer from 1983 to 1989
  - PhD in the Psychology of Music
the Todd Model – Basic Concepts

- It is developed by the method “analysis-by-measurement”, which means it obtains empirical evidence directly from measurements of human expressive performances.
- The preliminary assumptions of the model are that there is a direct link between certain aspects of the musical structure and the performance, and this relation can be modeled by one single, simple rule.
  - For example, a performer slows down at that point in order to let the listener perceive the hierarchical structure of music.
  - The interaction of timing and dynamics was modeled by a simple relation “the faster, the louder”, in mathematical view, the intensity is proportional to the squared tempo.
- The hierarchical structure dominates the instantaneous tempo, that is, the more closure, and thus, the more important the phrase boundary, the slower tempo.

the Todd Model – Evaluations

- An evaluation was conducted by Clarke and Windsor in 1997. They tuned the Todd model’s parameters using different model level weightings to Schubert’s Impromptu D990 No3 and evaluated them against two human repeated performances. This way is called hybrid performance.
  - They found that the timing required more emphasis on lower structural levels, whereas dynamics on higher ones.
- Again, in 2000, they chose some participants to evaluate both human performance and model performance. The first four bars of Mozart’s piano sonata K331 were performed by the model and two pianists.
  - The result revealed that the timing and dynamics didn’t relate to such a simple manner suggested by the model.
- In 2003, their further experiment introduce the concept “residuals” and concluded the Todd model is used as an analysis tool to assess the idiosyncrasies of human performances.

the Mazzola Model – Developer

- Guerino Mazzola
  - PhD in Mathematics in 1971
  - An internationally recognized jazz pianist
  - His discography comprises 15LPs and CDs.
  - His scientific publications comprise over 80 papers and 11 books.

the Mazzola Model – Basic Concepts

- This model was built upon not only various aspects of music theory and a highly complex mathematical approaches, but some philosophical, semiotic and aesthetic considerations.
  - In one word, it is hard to understand.
- The model consists of an analysis part and a performance part.
  - 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 weights 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.
the Mazzola Model – Basic Concepts

- The software package of the Mazzola model is called RUBATO, which is freely available at http://www.rubato.org.
- The model has a software called EspressoRUBETTE, which analyses MIDI-like data input, performs score-to-performance matching, and extracts performance vector fields for a given human performance. This operation procedure is also called “inverse performance theory”.

the Mazzola Model – Evaluations

- Jan Beran compared the analysis of 28 performances of “Träumerei” produced by RUBATO software with the multiple regression analysis of the same pieces previously done by Bruno Repp in 1992, and Beran found that the model could explain 84% of the average tempo curve of 28 performances, each of the three analytical components contributing about equally to the model.
- Unfortunately, there is no other empirical investigations or quantitative evaluations of the model except by Mazzola himself.

the Machine Learning model – Developer

- Gerhard Widmer
- Austrian Research Institute for Artificial Intelligence (ÖFAI), Vienna

the Machine Learning model – Basic Concepts

- The model collects large amounts of empirical data measured from the performances of professional musicians. And by using inductive machine learning and data mining techniques, the computer finds the regulations in the data.
- The model predicts both local, note-level expressive deviations and higher-level phrasing patterns.
  - Note-level part means how a pianist is going to play a note in a piece, e.g., slow or fast, louder or softer.
  - Higher-level part means the expressive strategy like the shaping of an entire musical phrase, e.g., gradual ritardando towards to the end.
the Machine Learning model – Basic Concepts

- As a result, the model has a set of rules. The rules are the observed regulations from the data analyzed by computer. The training details are:
  - A pianist performed 14 Mozart's complete piano sonatas, and 106000 performed notes, almost 4 hours of music, data were collected.
  - Only 41116 notes that performed on the melodies (usually the soprano part) were kept.
  - Each note was described by 29 attributes that represent both intrinsic properties (e.g., scale degree, duration, metrical position) and local context (e.g., local melodic contour).
  - After analyzing, the computer learned 17 classification rules that could predict most note-level choices of the pianist.

- For example
  - RULE TL2:
    \[ \text{context} = \text{equal} \rightarrow \text{longer} \]
    \[ \text{metrical strength} \leq 1 \]
    - (Given two notes of equal duration followed by a longer note, lengthen the note (i.e., play it more slowly) if the note is in a metrically weak position (metrical strength \( \leq 1 \)).

the Machine Learning model – Evaluations

- In (Widmer, 2002), 41116 played notes are listed in detail for each discovered rule, thus making the verification easy.
- Some rules tested on Chopin's pieces could significantly accurately predict the Mozart's pieces.
- Experiment revealed that the model is better for dynamics than for timing or tempo in multi-level aspects. Moreover, these poor predictive performance is due to the approximation by the quadratic functions. In other words, quadratic functions may not be a good tool to describe the timing.
- Refining the definition and representation of musical context is the currently ongoing research subjects.

Current Research: quantification of individual style

- Individual style means the differences between artists. Bruno Repp (1992) indicated all the pianists have the same performance at a global level, but there were differences between how the pianists dealt at lower levels of the structural hierarchy, such as distinctive phrasing behaviors.

Finally, it combines above-mentioned note-level and rule-based learning model to yield a multi-level model of expressive timing and dynamics “shapes”.

- That is, what kind of tempo and dynamics pattern a pianist would apply to a given musical context at a given level of the phrase hierarchy. Here are three necessary assumptions:
  - The expressive timing or dynamics gestures applied to a phrase by a performer can be approximated by a family of quadratic curves.
  - A multi-level performance can be represented as a linear combination of expressive shapes at different hierarchical levels.
  - If all other things are equal, similar phrases will be played similarly by pianists.

- The model predicts expressive shapes (curves) for new phrases in new pieces, based on the analogies between known phrases. What the new curves are subtracted from the original curves are called “residuals”, that is, the low-level, local timing and dynamics deviations.

As a result, this model is a predictive computational model of expressive timing and dynamics that takes into account of both the hierarchical music structure and local music context.

Presented by T.H. Chiang
One way to present the performer’s style is by visualization –
the beat-level tempo and dynamics curve represented in a tempo-loudness space, (see figure 1)

Fig. 1. Smoothed tempo-loudness trajectory representing a performance of Frédéric Chopin’s Ballade op.47 in Ab major by Artur Rubinstein, horizontal axis: tempo in beats per minute (bpm); vertical axis: loudness in sone. (performed by Artur Rubinstein, truncated edition)

Dixon in 2002 had developed another way is called the Performance Worm, which computes and visualizes performance trajectories via computer animation.

Characterization: performance alphabets

The performance trajectories are cut into short segments of a fixed length in order to be analyzed by computer. All segments are grouped into some similar patterns via clustering. The cluster prototypes represent a set of patterns that can be used to reconstruct a full trajectory. These prototypes can be seen as a set of performance alphabets. (see figure 2)

Fig. 2. A “Mozart performance alphabet” (cluster prototypes) computed by segmentation, mean and variance normalization, and clustering from performances of Mozart piano sonatas by six pianists (Daniel Barenboim, Roland Batik, Vladimir Horowitz, Maria João Pires, André Schiff, Mitsuko Uchida). For the ‘letters’, see the graph, Mitsuko Uchida’s performance pattern is different from all the others. Also for the overview performance shape, see the figures on the right. We can observe the her performance is special.

For studying the individual performance style, we could analyze the pianist’s performance alphabet “letters” and performance shapes.

For the ‘letters’, see the graph, Mitsuko Uchida’s performance pattern is different from all the others.
Classification: automatic identification of performers

- The ongoing researches are using computer program to identify artists on the basis of their performance characteristics, such as performance trajectories.
- These researches are still very preliminary, so there is no systematic and comprehensive model or algorithm of them.

Conclusion

- The person and personality of the artist as a mediator between music and listener is totally neglected among all the models.
- There are some limits that maybe no model can describe. For example, the intentionally interpretation.
- Even if the limits are impossible to be solved, these researches advance our understanding and appreciation of the complexity of artistic behavior.

Analysis-by-synthesis

- Steps
  - 1 - Selection of performances
  - 2 - Measurement of the physical properties of every note
  - 3 - Reliability control and classification of performances
  - 4 - Selection and analysis of the most relevant variables
  - 5 - Statistical analysis and development of mathematical interpretation models of the data
  - 6 - Synthesis of performance with systematic variations
  - 7 - Judgment of synthesized versions, paying particular attention to the different experimental aspects selected
  - 8 - Study of relation between performance and experimental variables
  - 9 - Repetition of the procedure (steps 3-9) until the results converge

Analysis-by-measure

- Steps
  - 1 - Selection of performances
  - 2 - Measurement of the physical properties of every note
  - 3 - Reliability control and classification of performances
  - 4 - Selection and analysis of the most relevant variables
  - 5 - Statistical analysis and development of mathematical interpretation models of the data