COMPUTATIONAL MODELS OF EXPRESSIVE MUSIC PERFORMANCE: THE STATE OF THE ART

Gerhard Widmer and Werner Goebl

1. Introduction
2. KTH model
3. Todd Model
4. Mazzola Model
5. Widmer and Goebl model
6. Quantification of individual style
7. Conclusions
8. References
INTRODUCTION

- Work dates back to Seashore and colleagues (1938)
- Expressive music performance parameters include tempo, timing, dynamics, and articulation.
- Article provides an overview of current research into computational modeling of expressive performance
- Hypothesis set out in computational models can be tested empirically with performance data
KTH MODEL

- Developed at the Royal Institute of technology in Stockholm
- Set of performance rules to predict aspects of timing, dynamics, and articulation
  - ostensibly based on the local musical context
- “analysis–by–synthesis” approach
  - pros: models one kind of performer–listener interaction
  - cons: relies heavily on the performer small number of examples presented
KTH MODEL

DURATION–CONTRAST Rule

☐ modifies the ratio between sequential notes to emphasis difference in their durations

☐ quality control parameter (k)
  ☐ 1 – full effect
  ☐ 0 – no effect
  ☐ -1 – reverse the effect

☐ problems: several rules influence the duration of the note, which makes this rule dependent on these
KTH MODEL

- To produce a predictive model the parameters of the model need to be tuned
  - analysis–by–synthesis approach presents some basic suggestions

- Empirical evaluation of recorded performances are also necessary
  - Sundberg et al. (1991) determined the perceptual threshold for the k values
  - Friberg (1995) used a greedy search method to fit parameters to the PHRASE ARCH rule based on the first nine measures a single piece
  - Sundberg et al. (2003) fitted PHRASE ARCH k values manually to a single performance of a Mozart sonata movement
KTH MODEL

- Empirical evaluations outside of KTH
  - Zanon and De Poli (2003a; 200b) tested both fixed and time varying k values
  - Gabrielsson and Juslin (1996) related model to emotional colourings
  - Juslin et al. (2002) developed a comprehensive computation model of expressive performance
    - G – generative KTH model
    - E – Juslin’s earlier work on emotional models
    - R – random variability
    - M – analogies to physical motion
Widmer and Gobel’s conclusion

KTH rule model is a viable representation language for describing expressive performance. To what extent it can account for the observed variations in large collections of performances of truly complex music is still an open issue.
TODD MODEL

- Developed by Neil Todd from late 1980s to early 1990s
- Structure-level models of expressive timing and dynamics
- “analysis–by–measurement” approach
  - empirical evidence obtained directly from measurements of human performances
  - assumptions:
    - direct link between musical structure and performance
    - relationship can be modeled with a single rule
- pros: appeals to a theoretical framework to assess musical context (Lerdahl and Jackendoff 1983)
- cons: overly simplistic (“the faster, the louder”)
TODD MODEL

THE FASTER, THE LOUDER

☐ intensity is proportional to the squared tempo

☐ used a recursive look-ahead procedure to allow the hierarchical grouping in the music to control the instantaneous tempo

☐ leads to increased dynamics and tempo at the middle of phrases and reduced dynamics/slowing down at points of stability, such as phrase boundaries

☐ this is modeled at each level of the piece’s hierarchy
TODD MODEL

☐ Empirical evaluation

☐ Todd (1992) compared the model’s output with tempo and dynamic curves of one or two performances of a small number of pieces by Haydn

☐ Windsor & Clarke (1997) used regression analysis of several performances generated by Todd’s model against two repeated human performances

☐ residuals: idiosyncrasies of human performance not explained by the model

☐ Clarke & Windsor (2000) had human listeners evaluate performances generated by Todd’s model
TODD MODEL

☐ Widmer and Gobel’s conclusion

• The authors note that “the faster, the larger” model is overly simplistic, but doesn’t provide an overall assessment of the approach
MAZZOLA MODEL

- Developed by Guerino Mazzola and colleagues in Zurich
- Mathematical music theory and performance model
- Analysis and performance components
  - Computer-aided analysis tools for musical structure
  - Each aspect implemented in a Rubbette (plugin)
- Performance is generated with the Rubettes
- Uses “Stemma/Operator” theory
MAZZOLA MODEL

METRO RUBETTE

- (inner) metrical analysis
  - result is different than Lerdahl and Jackendoff--esque (outer) metrical analysis
  - used linear mapping between metrical weight and tone intensity to generate a performance
  - not compared with real performances
MAZZOLA MODEL

- Widmer and Gobel’s conclusion

The (EspressoRUBETTE) plugin visualization the extracted performance data in several ways: alongside classical pianoroll notation it displays the extracted performance vector fields as two-dimensional colour contour plots. Lack labels, legends, or explanations, their meaning remains rather unclear to the reader.
MACHINE LEARNING MODEL

- Developed at Vienna
- Multi-level model of expressive timing and dynamics
- Uses large amounts of empirical data extracted from a performance to train a machine learning model to predict local, note-level expressive deviations and higher-level phrasing pattern
- Note that they are using INDUCTIVE MACHINE LEARNING and DATA MINING techniques (emphasis is Widmer’s)
MACHINE LEARNING MODEL

NOTE-LEVEL MODEL

- Inductive rule learning algorithm to learn note-level rules for timing, dynamics, and articulation
  - i.e., how the performer will play a particular note
  - complementary to higher-level manipulations

- Training method
  - recordings of 13 Mozart piano sonatas by one performer
  - each note melody described by 29 attributes
  - computer learned a set of 17 simple classification rules

  E.g., “given two notes of equal duration followed by a longer note, lengthen the note (i.e., play it more slowly) that precedes the final, longer one, if this note is in a metrically weak position (‘metrical strength’ ≤ 1).” (Widmer and Goebl 209)
MACHINE LEARNING MODEL

MULTI-LEVEL MODEL

- Attempts to account for the role of tempo, dynamics, and articulation in shaping abstract structures, such as motifs, groups, and phrases

- Work discussed in the article focused on having the computer predict the expressive timings and dynamics shapes that a performer would use for a phrase

- Assumptions
  - expressive timing or dynamics gestures can be reasonably approximated quadratic curves
  - a multi-level performance can be represented as a linear combination of these shapes at different hierarchical levels
  - similar phrases will be played similarly by different pianists
MACHINE LEARNING MODEL

MULTI–LEVEL MODEL

☐ Inputs to the system
  ☐ example performances by musicians
  ☐ hierarchical phrase analysis of the music
  ☐ tempo and dynamics curves

☐ System fits quadratic approximation functions to the curves associated with each phrase

☐ Predicts elementary expressive shapes for similar phrases in different pieces

☐ Can be combined with the note–level model, such that the note–level model compensates for the “residuals”
MACHINE LEARNING MODEL

- Empirical evaluation
  - Widmer (2002) gives “a very detailed picture of the relative generality and reliability of the rules” and describes quantitative results of experiments with “large numbers” of new pieces
    - found that the rules carried over to other performers
  - Widmer and Tobudic (2003a) tested the predictive performance of the multi-level model by measuring how closely the tempo and dynamics curves of the new performances matched those predicted by the model
    - results were better than chance and mechanical performances
  - Tobudic and Widmer (2003b) optimized the case-based learning algorithm and used first-order logic and structural similarity to model the phrases’ hierarchical context
    - produced some quantitative improvements
MACHINE LEARNING MODEL

- Widmer and Gobel’s conclusion
  Not really stated
  Stated results are somewhat vague
  - “better than chance”
  - “some quantitative improvements”
QUANTIFICATION OF INDIVIDUAL STYLE

- Repp (1992) demonstrated the differences in pianists’ styles
  - observed that there were characteristic phrasing behaviors that could be statistically attributed to certain pianists
  - limited to a single piece
- Widmer et al. (2003) used computation methods to acquire timing and dynamic information from a large number of recordings
QUANTIFICATION OF INDIVIDUAL STYLE

- Visualization of performance trajectories
- Beat-level tempo and dynamics curves can be integrated into a tempo-loudness trajectory

Smoothed tempo-loudness trajectory for a performance of Chopin’s Ballade Op. 47

Horizontal axis: bpm
Vertical axis: loudness (in sone)
QUANTIFICATION OF INDIVIDUAL STYLE

- Performance alphabets
- Trajectories are cut into short fixed-length segments and normalized
- Similar segments are grouped into clusters
  - The center of the clusters represent “a set of typical elementary tempo-loudness patterns”

Performance alphabet for six performances of Mozart piano sonatas

Dots indicate directionality
Shaded regions indicate variance
QUANTIFICATION OF INDIVIDUAL STYLE

- Automatic identification of performers
  - Stamatatos and Widmer (2003) showed that computers could differentiate between different pianists given a training set of only one piece per pianist
  - Zanon and Widmer (2003) showed that computers can learn to distinguish between different pianists based on high-level trajectories
  - “Results are still very preliminary and we have not yet managed to pinpoint precisely which features are the most distinguishing ones...”
CONCLUSIONS

- Much room for further research
- Some work has also been done on “expressive intentions” or emotions
- Results of such research may be useful in creating new control devices
- Ultimately though these types of predictive models will never account of all the idiosyncrasies of human performance because they can only account for what is written in the score and not for any external context
THANK YOU

Any questions?
REFERENCES

03.12.08 IFT 6080 – MUSIC AND MACHINE LEARNING JOHANNA DEVANEY