Measuring & Modeling Musical Expression

Douglas Eck
University of Montreal
Department of Computer Science
BRAMS Brain Music and Sound

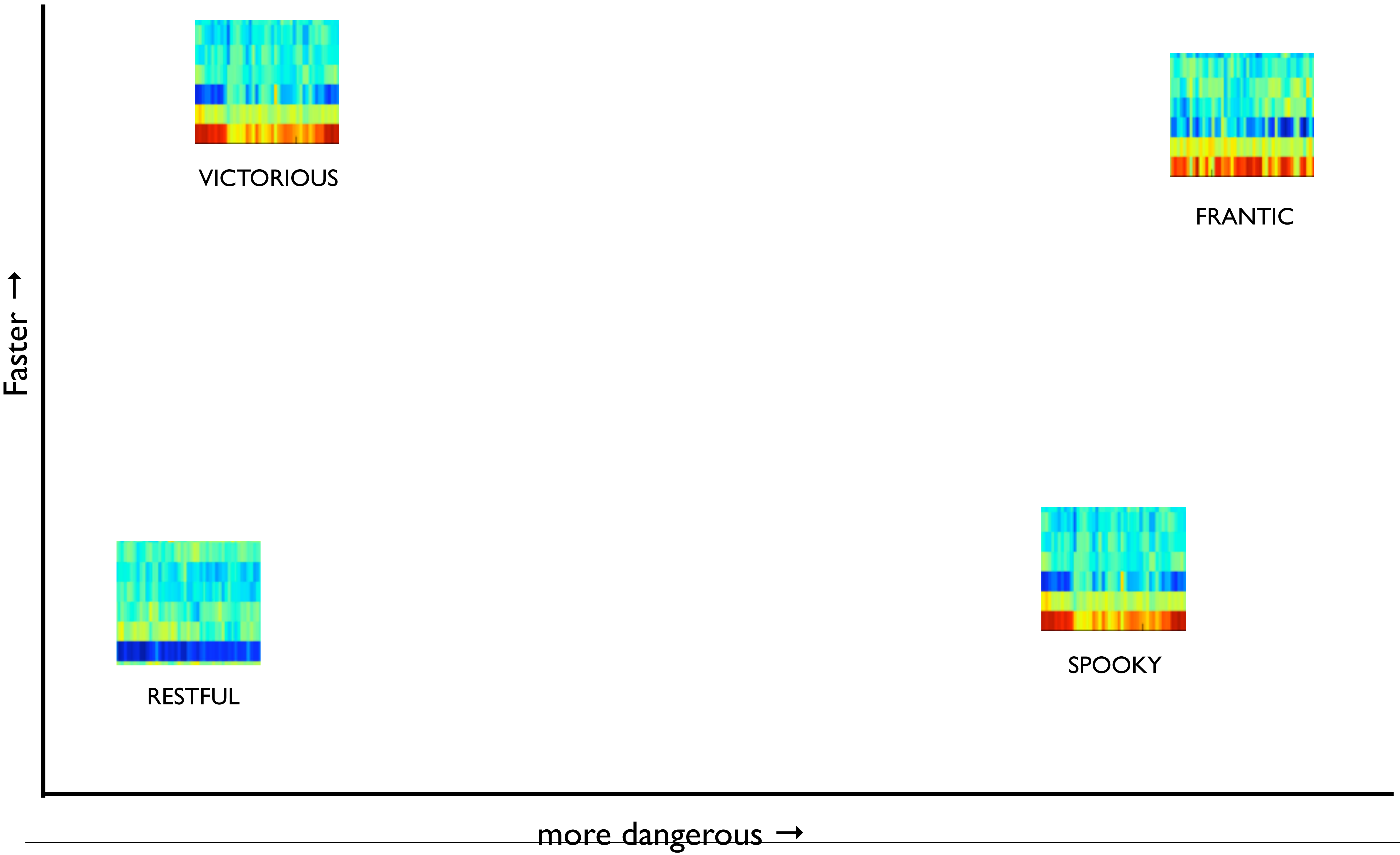
Université de Montréal

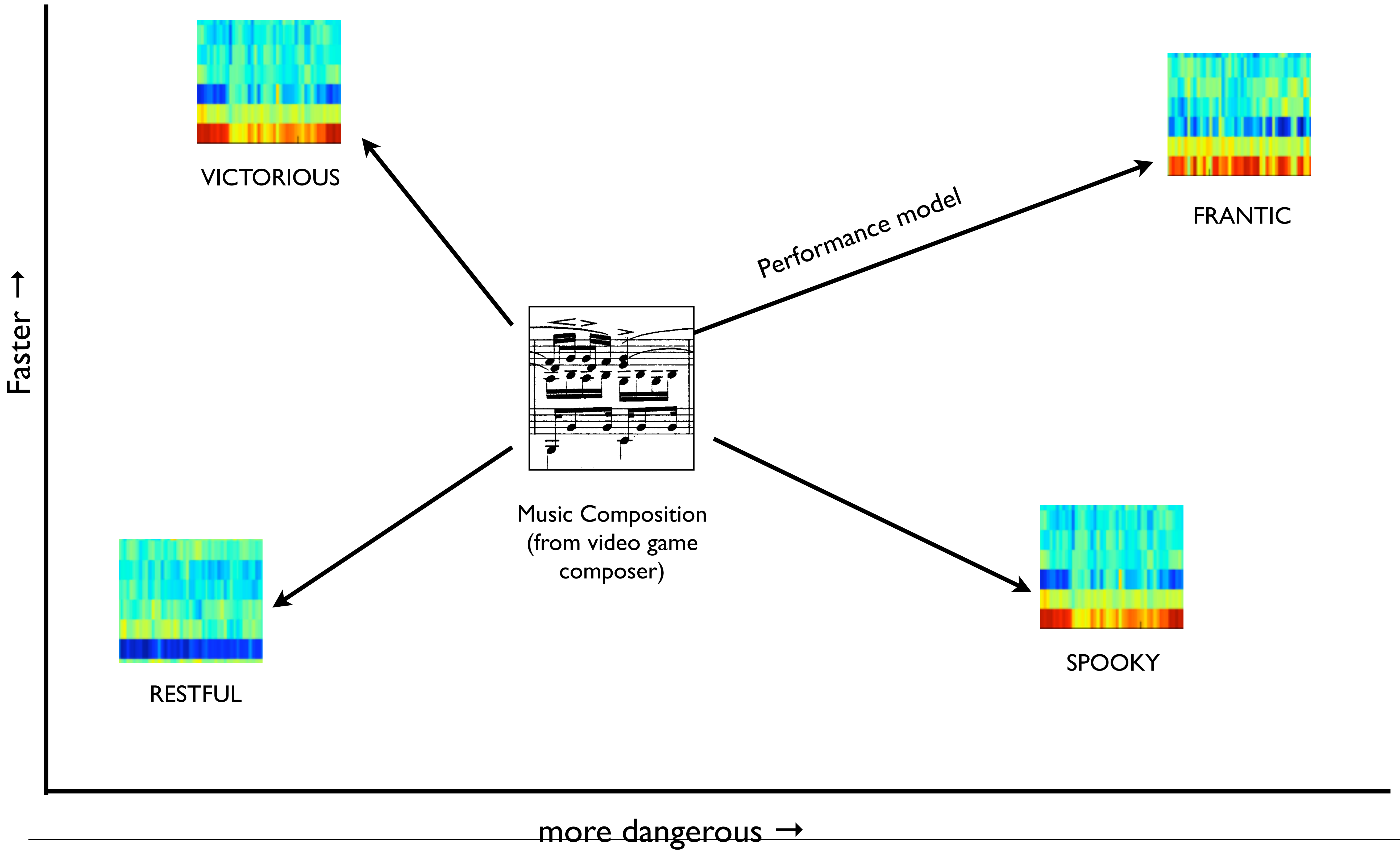
Brams
International Laboratory for
Brain, Music and Sound Research

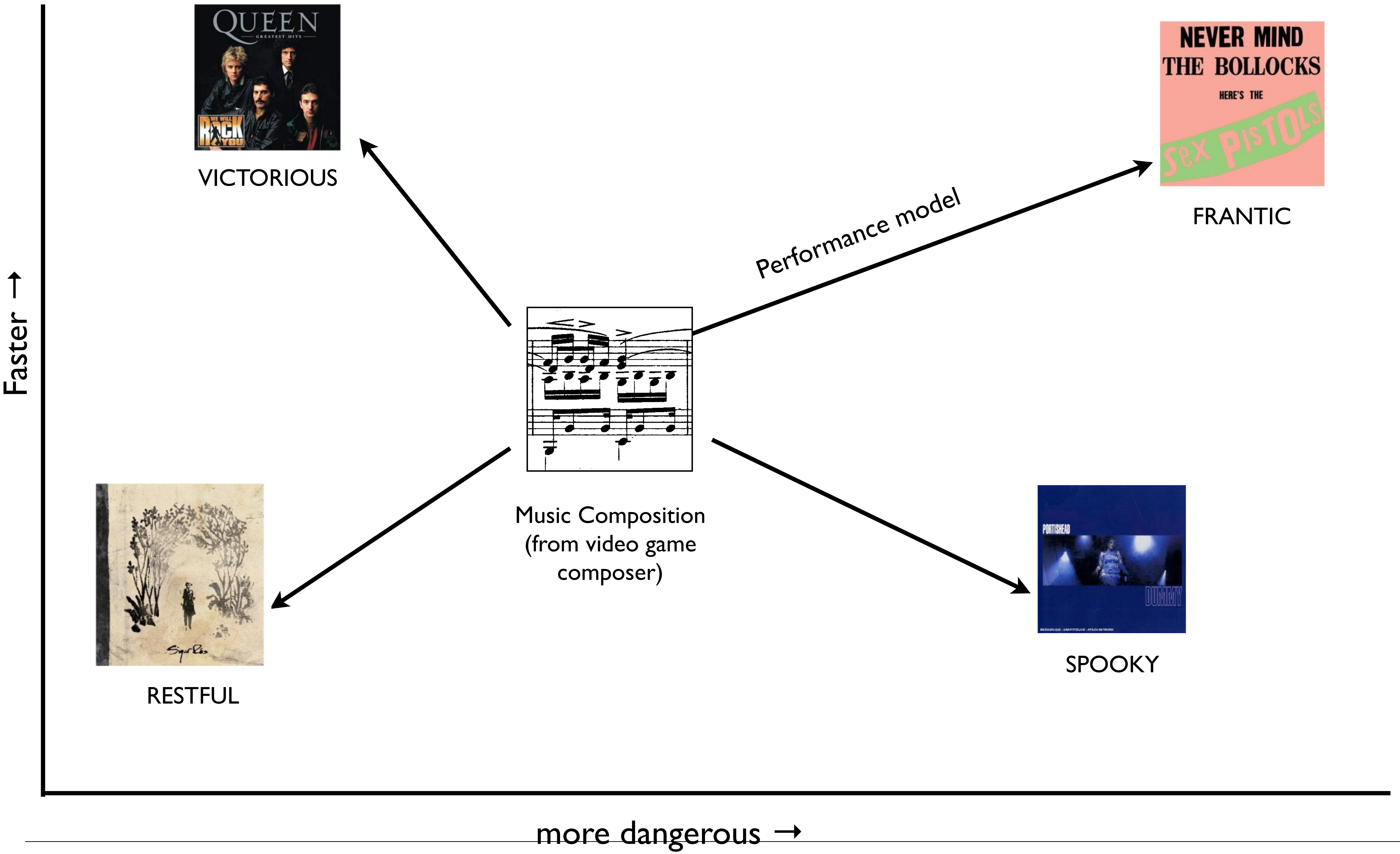
Overview

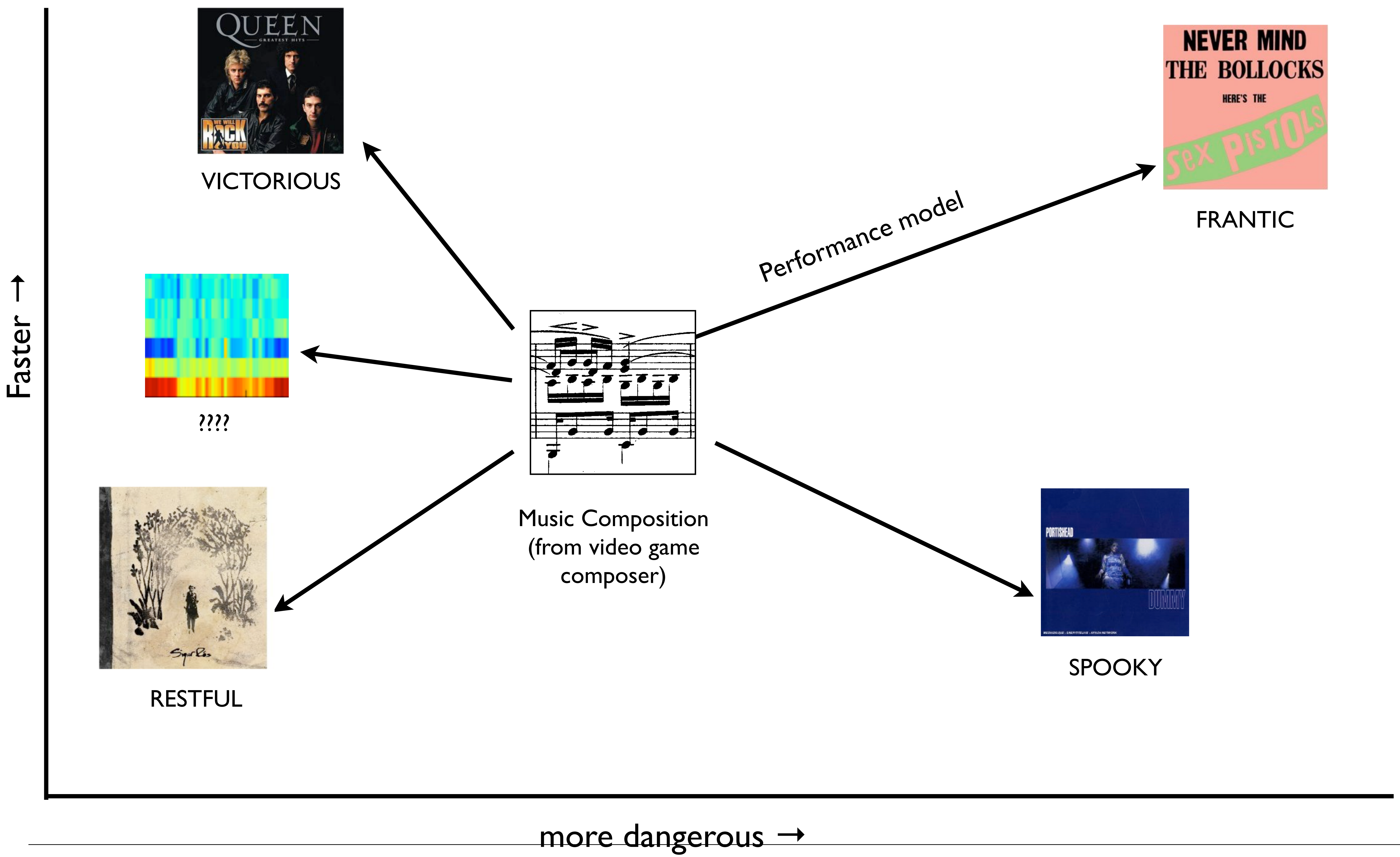
- Task: human-realistic music performance
- Challenges:
  - expressive timing and dynamics
  - generating musical variations
  - choosing appropriate timbres (instruments)
- Today: Learning expressive timing and dynamics for the piano
- Applications: music generation for film and video games
- Work done in collaboration with Stanislas Lauly

An interesting task... context-aware music generation




VICTORIOUS
FRANTIC
Performance model
Music Composition (from video game composer)
RESTFUL
SPOOKY
Faster →
more dangerous →

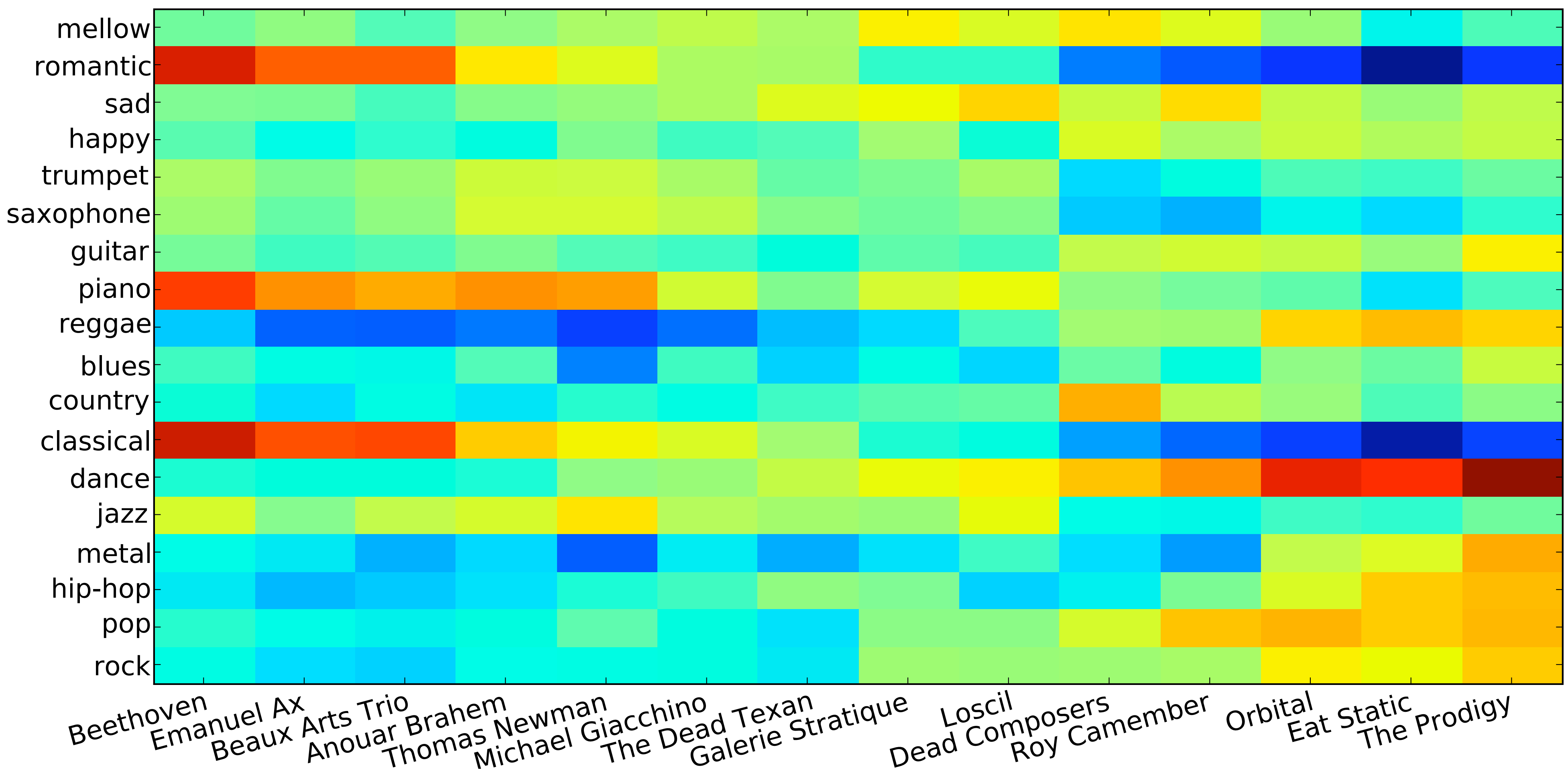

QUEEN
VICTORIOUS
NEVER MIND THE BOLLOCKS HERE'S THE Sex Pistols
FRANTIC
Performance model
Music Composition
(from video game
composer)
RESTFUL
PORTISHEAD
DUMMY
SPOOKY
Faster →
more dangerous →


QUEEN
VICTORIOUS
NEVER MIND
THE BOLLOCKS
HERE'S THE
Sex Pistols
FRANTIC
Performance model
????
Music Composition
(from video game
composer)
RESTFUL
PORTISHEAD
DUMMY
SPOOKY
Faster →
more dangerous →

Audio similarity + morphing

- We can predict words like "sad" and "jazzy" from audio. Resulting wordset useful for music recommendation (Eck et al. NIPS 07)

- We can also morph between artists based on word vector similarity



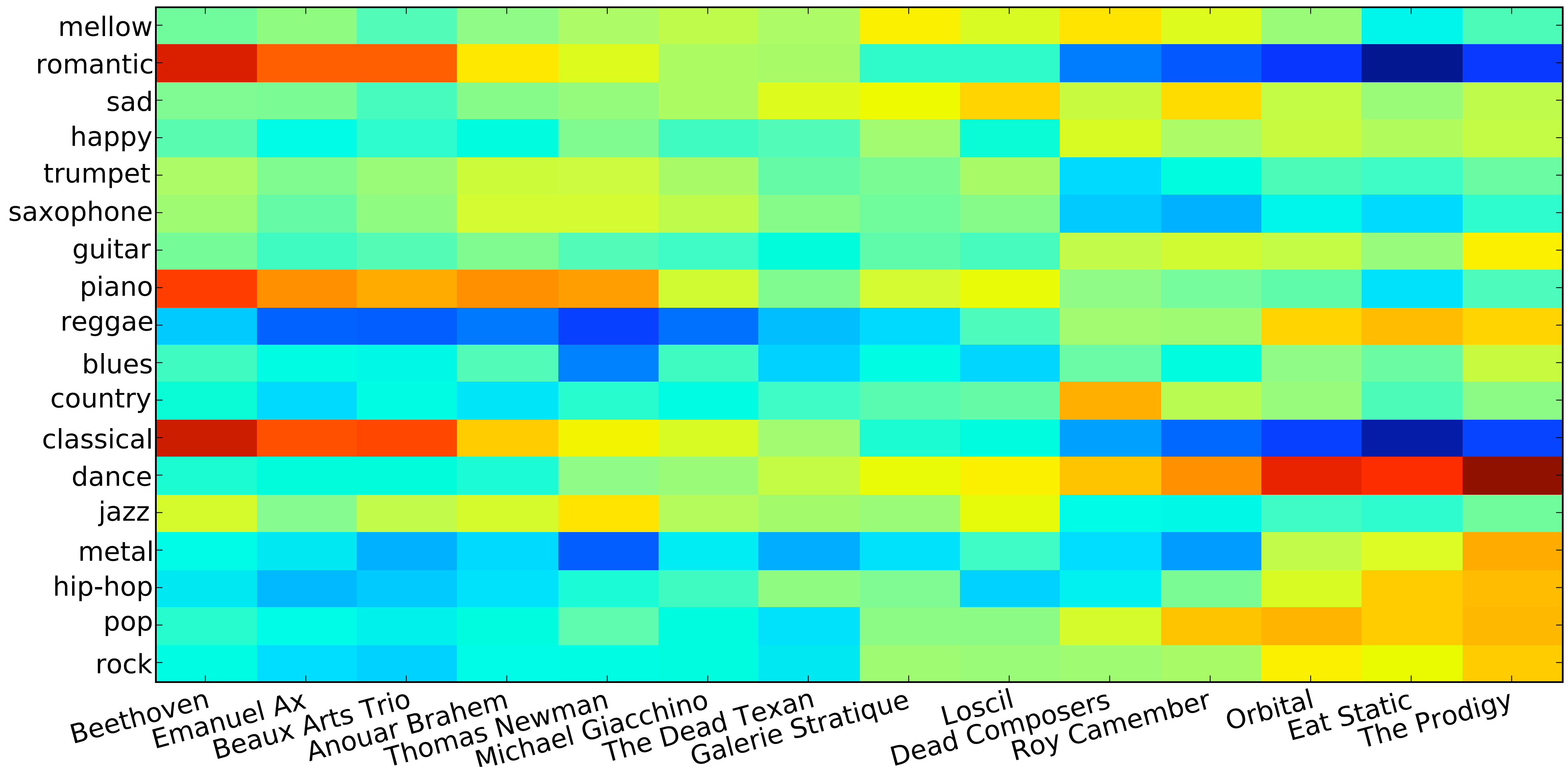
- Similar technique may allow us to generate a "dangerous" sound based on analysis of songs people think sound dangerous.

This work is a part of Sun Labs "Project Aura" recommendation framework.

Sun microsystems

Audio similarity + morphing

- We can predict words like "sad" and "jazzy" from audio. Resulting wordset useful for music recommendation (Eck et al. NIPS 07)
- We can also morph between artists based on word vector similarity



- Similar technique may allow us to generate a "dangerous" sound based on analysis of songs people think sound dangerous.

This work is a part of Sun Labs "Project Aura" recommendation framework.

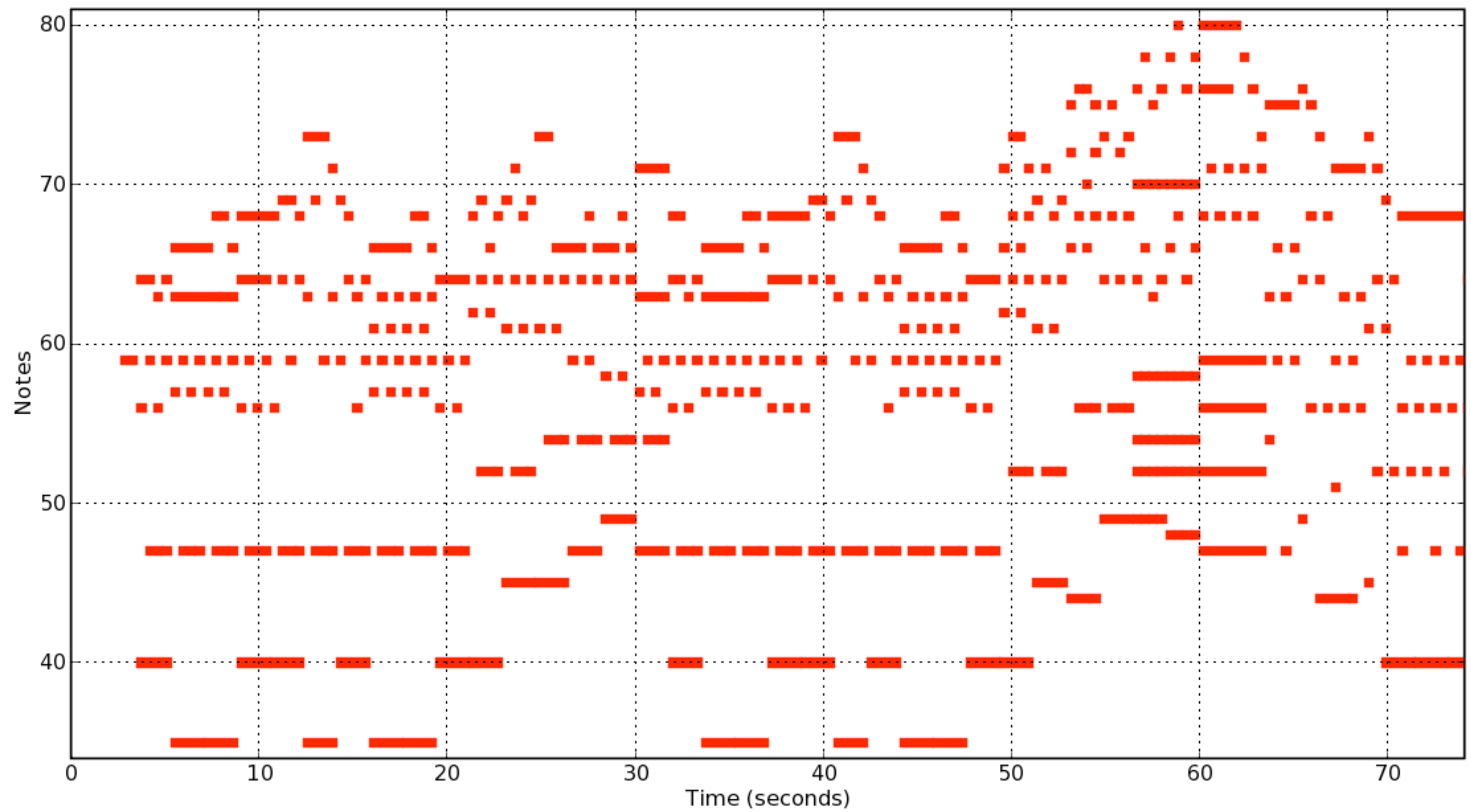
Sun microsystems

Chopin Etude
Opus 10 No 3

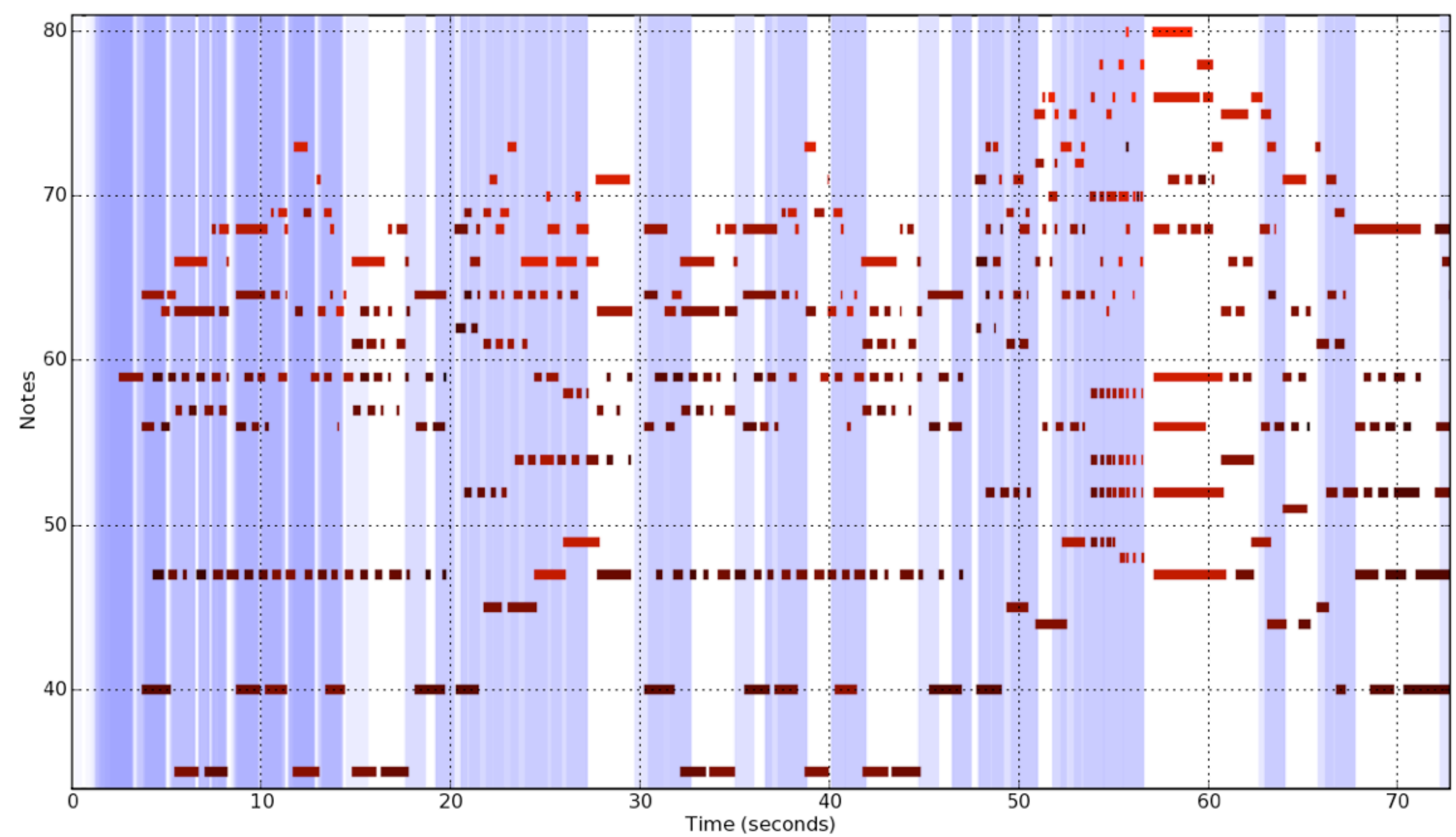
- Many challenges:
  - expressive timing and dynamics
  - generating musical variations
  - choosing appropriate timbres (instruments)
- Today: Learning expressive timing and dynamics for the piano

Example: Chopin Etude Opus 10 No 3

Deadpan
(no expressive timing or dynamics)



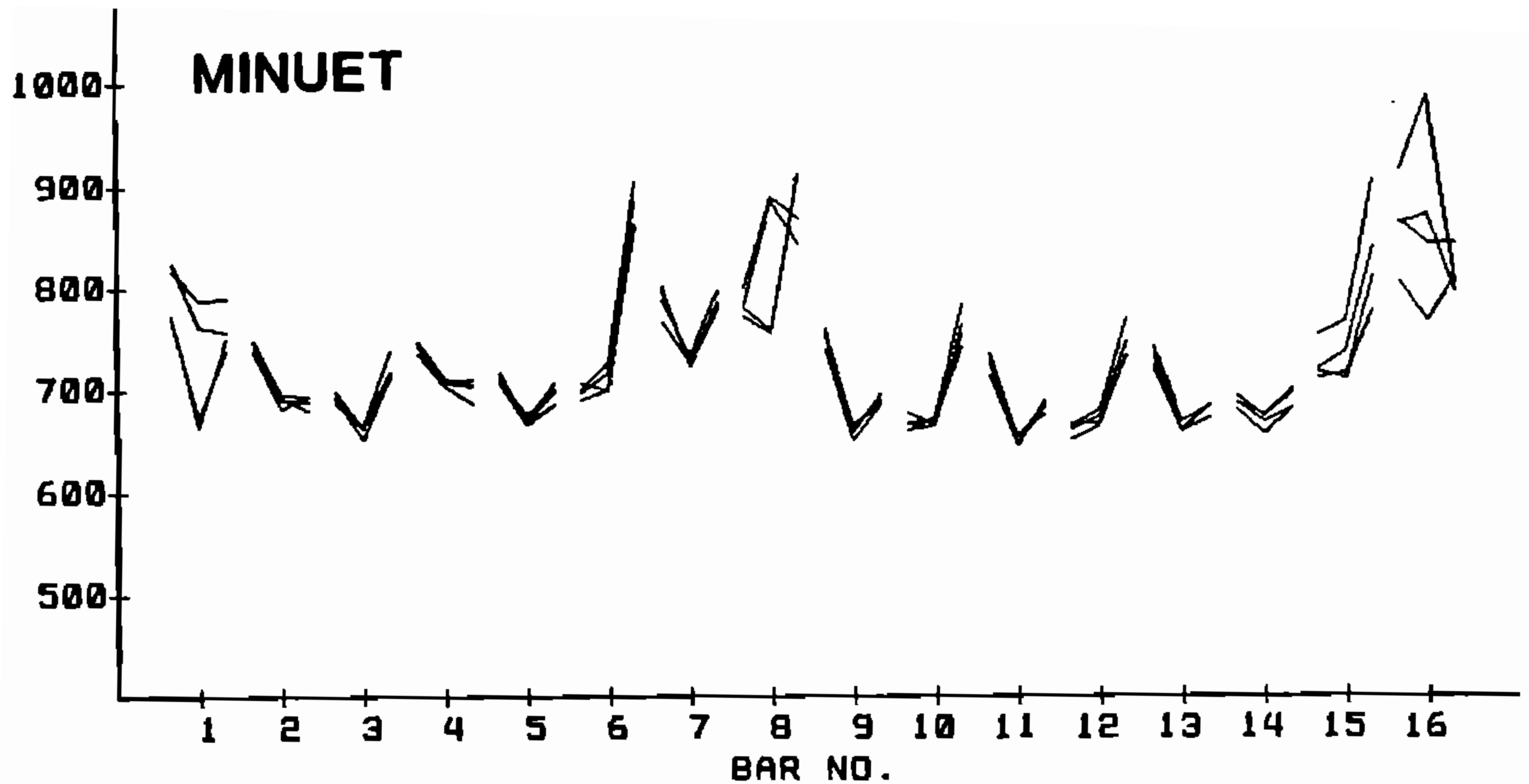
Human performance
(Recorded on Boesendorfer ZEUS)



Differences limited to:
- timing (onset, length)
- velocity (seen as red)
- pedaling (blue shading)

What can we measure?

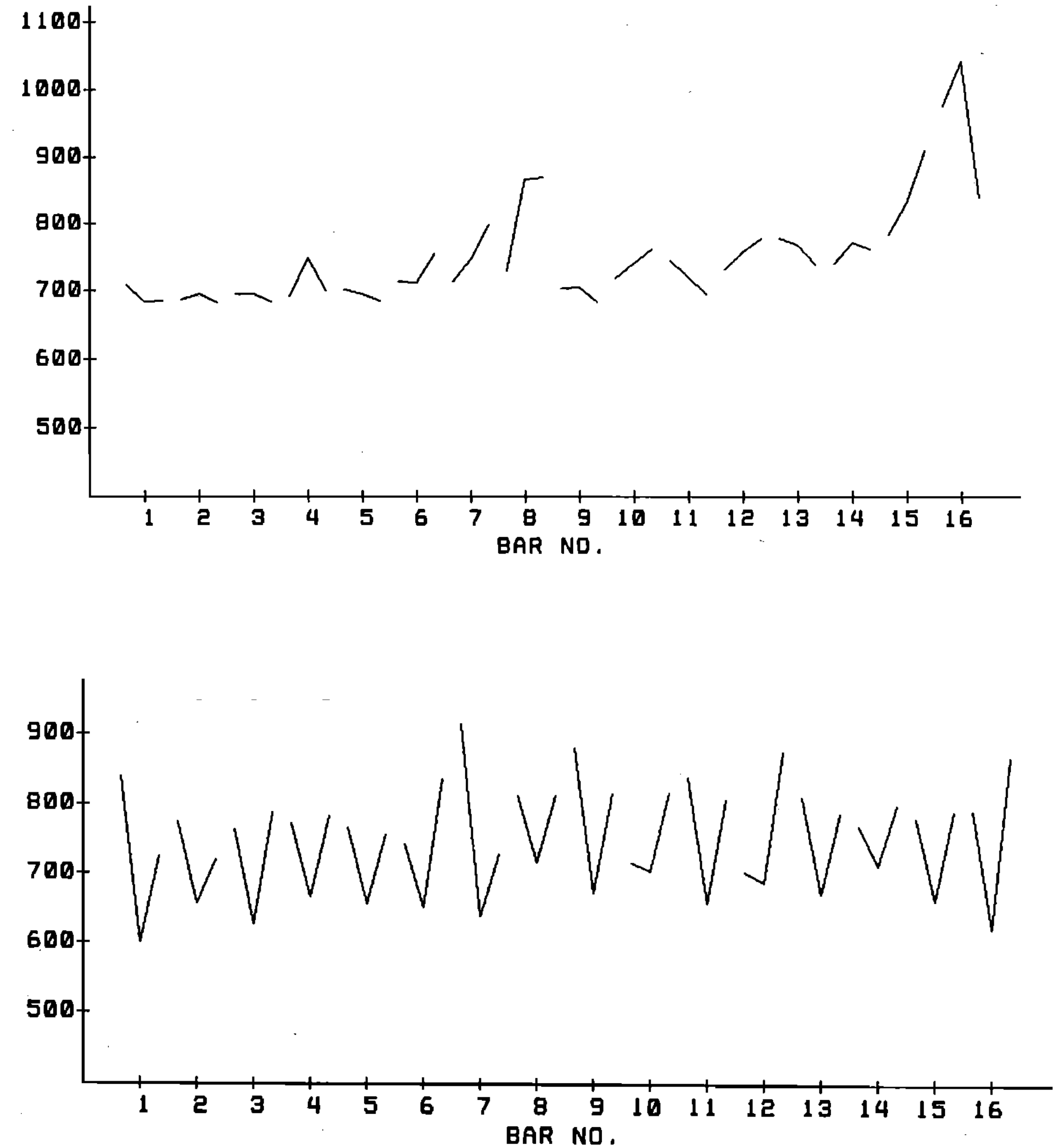
- Repp (1989) measured note IOIs in 19 famous recordings of a Beethoven minuet (Sonata op 31 no 3)



Grand average timing patterns of performances with repeats plotted separately. (From B. Repp "Patterns of expressive timing in performances of a Beethoven minuet by nineteen famous pianists",1990)

What can we measure?

- PCA analysis yields 2 major components
  - Phrase final lengthening
  - Phrase internal variation
- Simply taking mean IOIs yields can yield pleasing performance
- Reconstructing using principal component(s) can yield pleasing performance
- Concluded that timing underlies musical structure



Adapted from Repp (1990)

Timing versus expressive dynamics

- Repp (1997; experiment 2): generated MIDI from audio for 15 famous performances of Chopin's op. 10 No 3; Added 9 graduate student performances
- Retained only timing (no expressive dynamics)
- Judges ranked the average timing profile of the expert pianists (EA) highest, followed by E11, S1, S3, S9, S2, and SA.
- Conclusions:
  - EA, SA sound better than average but "lack individuality" (Repp)
  - Something is lost in discarding non-temporal expressive dynamics.
  - **Crucial point: EA and SA sound good**

KTH Model

- Johan Sundberg, Anders Friberg, many others
- Models performance of Western music
- Rule-based system built using
  - *analysis-by-synthesis*: assess impact of individual rules by listening
  - *analysis-by-measurement*: fit rules to performance data
- Incorporates wide range of music perception research (e.g. meter perception, pitch perception, motor control constraints)

Table 1.

An overview of the rule system

Phrasing	
Phrase arch	Create arch-like tempo and sound level changes over phrases
Final ritardando	Apply a ritardando in the end of the piece
High loud	Increase sound level in proportion to pitch height
Micro-level timing	
Duration contrast	Shorten relatively short notes and lengthen relatively long notes
Faster uphill	Increase tempo in rising pitch sequences
Metrical patterns and grooves	
Double duration	Decrease duration ratio for two notes with a nominal value of 2:1
Inégales	Introduce long-short patterns for equal note values (swing)
Articulation	
Punctuation	Find short melodic fragments and mark them with a final micropause
Score legato/staccato	Articulate legato/staccato when marked in the score
Repetition articulation	Add articulation for repeated notes.
Overall articulation	Add articulation for all notes except very short ones
Tonal tension	
Melodic charge	Emphasize the melodic tension of notes relatively the current chord
Harmonic charge	Emphasize the harmonic tension of chords relatively the key
Chromatic charge	Emphasize regions of small pitch changes
Intonation	
High sharp	Stretch all intervals in proportion to size
Melodic intonation	Intonate according to melodic context
Harmonic intonation	Intonate according to harmonic context
Mixed intonation	Intonate using a combination of melodic and harmonic intonation
Ensemble timing	
Melodic sync	Synchronize using a new voice containing all relevant onsets
Ensemble swing	Introduce metrical timing patterns for the instruments in a jazz ensemble
Performance noise	
Noise control	Simulate inaccuracies in motor

From: A. Friberg, R. Bresin & J. Sundberg (2006). Overview of the KTH rule system for musical performance. *Advances in Cognitive Psychology*, 2(2-3):145-161.

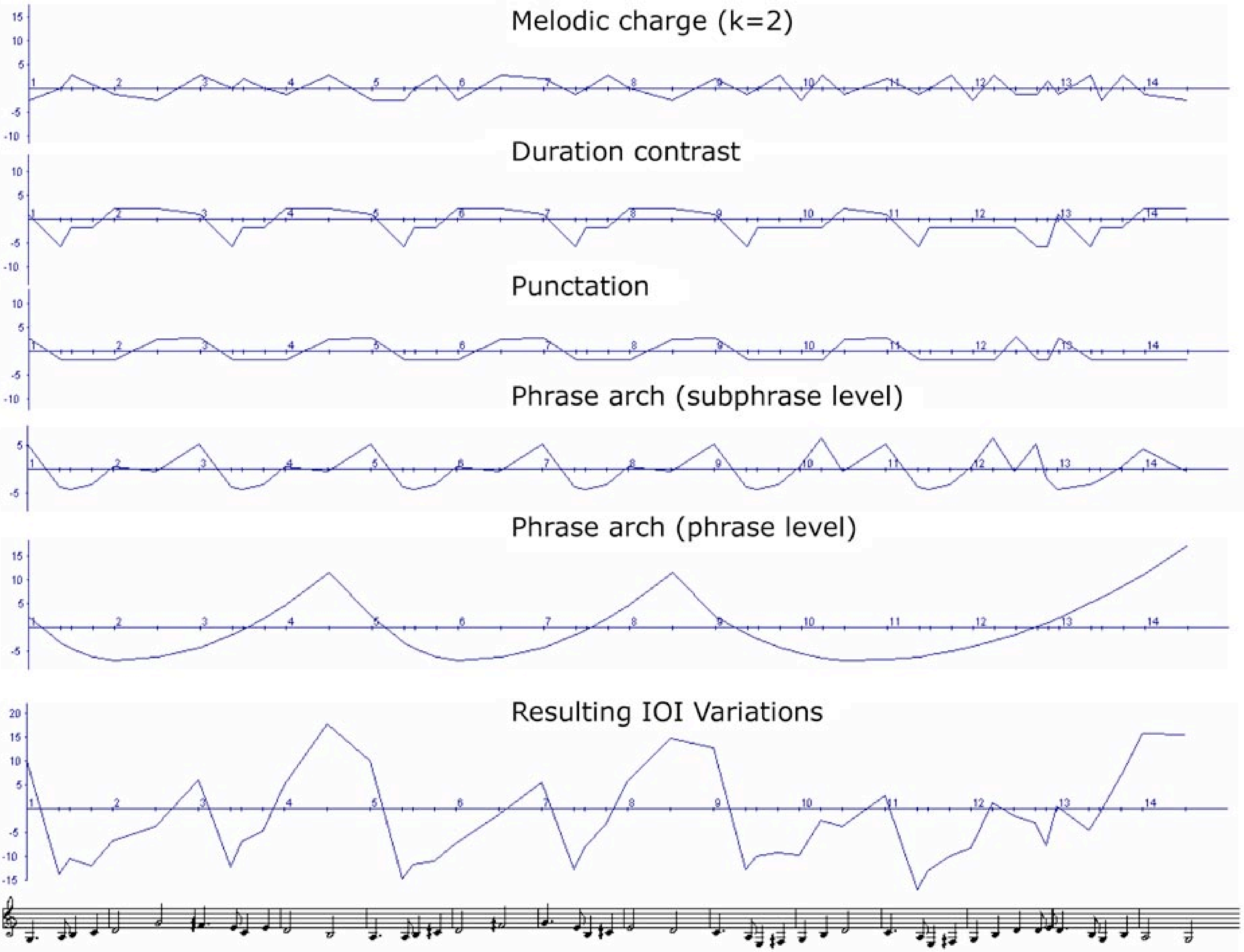


Figure 2.

The resulting IOI deviations by applying Phrase arch, Duration contrast, Melodic charge, and Punctation to the Swedish nursery tune "Ekorr'n satt i granen". All rules were applied with the rule quantity k=1 except the Melodic charge rule that was applied with k=2.

From: A. Friberg, R. Bresin & J. Sundberg (2006). Overview of the KTH rule system for musical performance. *Advances in Cognitive Psychology*, 2(2-3): 145-161.

Widmer et al. performance model

- Automatic deduction of rules for music performance
- Rich feature set (29 attributes including local melodic contour, scale degree, duration, etc)
- Performance is matched to score (metrical position).
- PLCG: Partition Learn Cluster Generalize (Widmer, 2003)
  - Discovery of simple partial rules-based models
  - Inspired by ensemble learning
- PLCG compares favorably to rule learning algorithm RIPPER
- Rules learned by PLCG similar to some KTH rules

RULE TL2:

```
abstract_duration_context = equal-longer
& metr_strength ⩽ 1
⇒ ritardando
```

"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)."

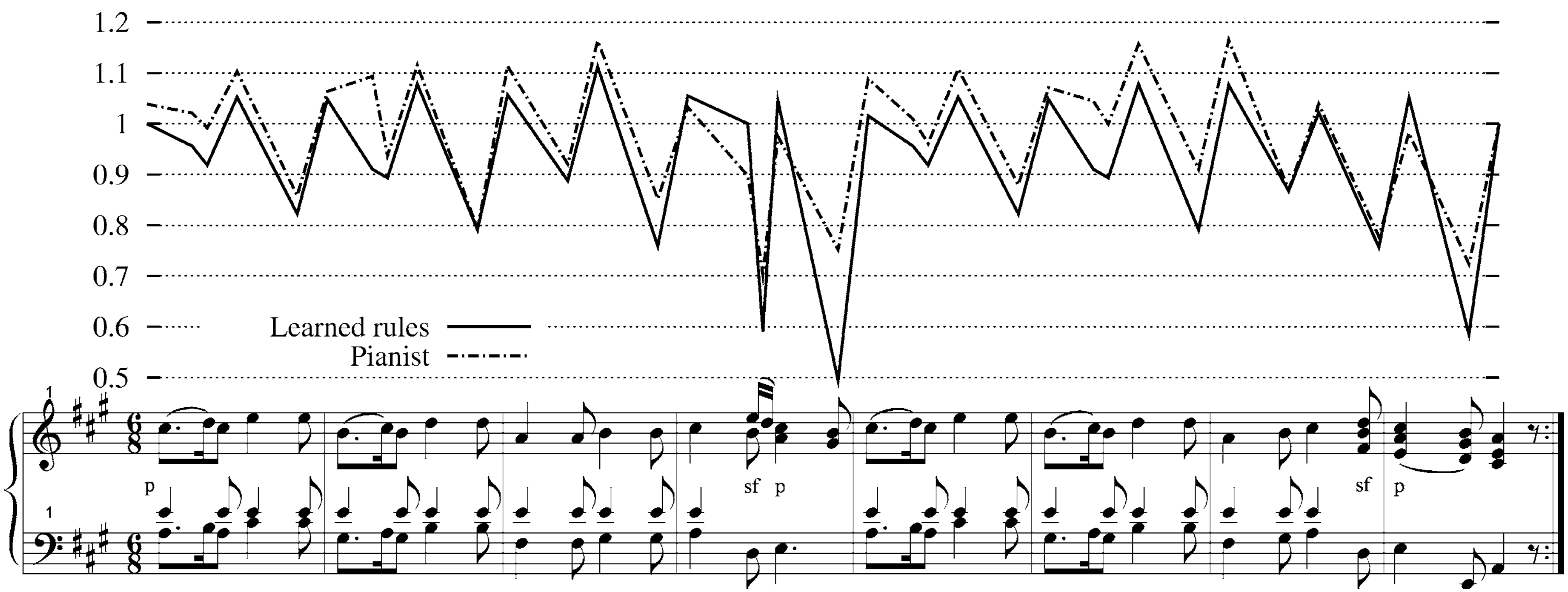


Fig. 5. Mozart Sonata K.331, 1st movement, 1st part, as played by pianist and learner. The curve plots the relative tempo at each note—notes above the 1.0 line are shortened relative to the tempo of the piece, notes below 1.0 are lengthened. A perfectly regular performance with no timing deviations would correspond to a straight line at $y = 1.0$.

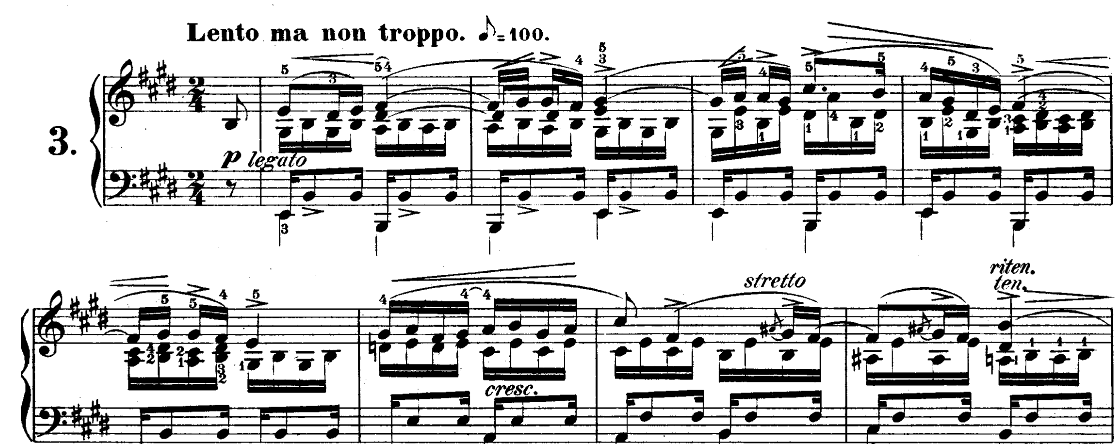
From: G. Widmer (2003). Discovering simple rules in complex data: A meta-learning algorithm and some surprising musical discoveries. *Artificial Intelligence* 146:129-148.

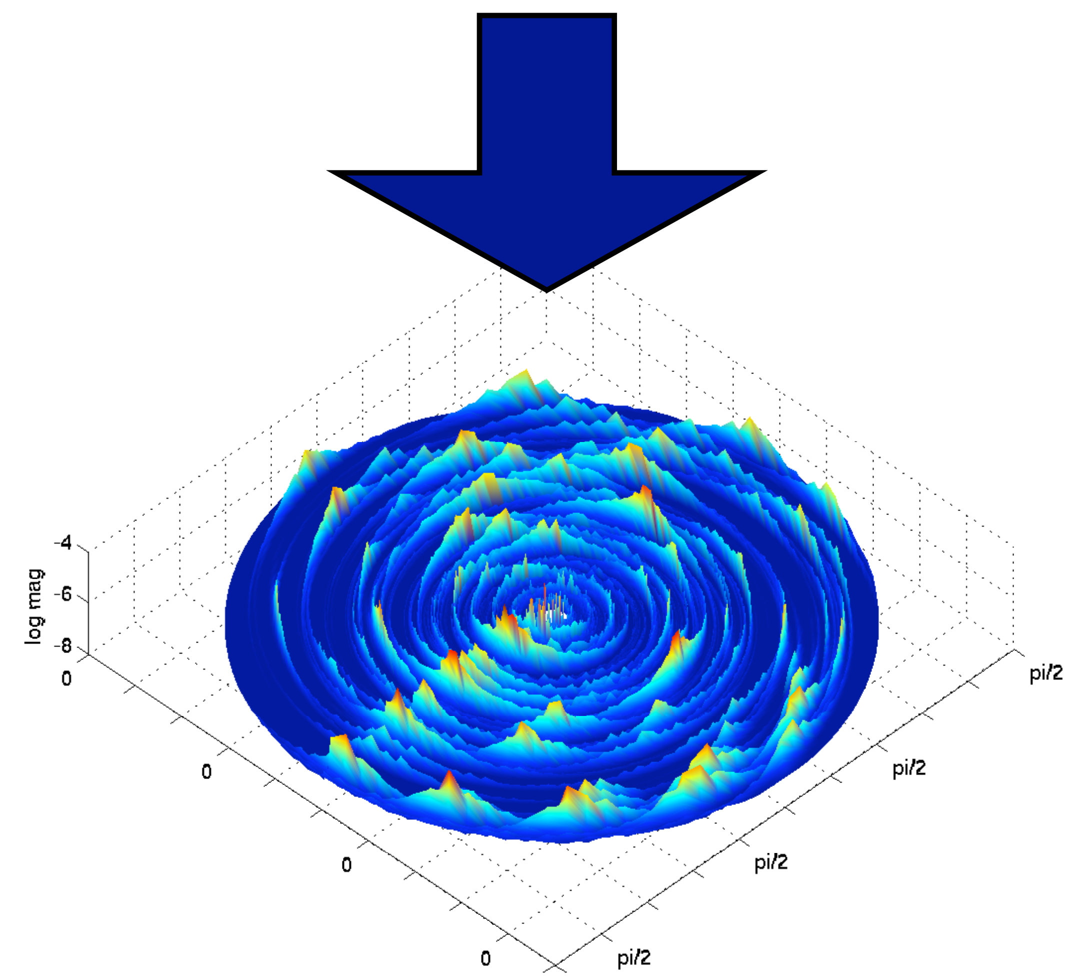
Another approach...

- KTH model has many rule-weighting parameters to set by hand
- Widmer improves this by using optimization to set rules
- Both models rely heavily on score for feature extraction
- Our goals:
  - Rely less on scores in order to work with non-scored music
  - Treat as a regression task in order to use standard machine learning techniques

Relying less on scores...

- Score provides crucial info about phrasing and meter
- But... musical score not always available
  - Jazz, pop, blues use simple scores or none
  - Millions of audio examples available, but audio-to-score is hard
- Solution: estimate phrasing and meter from audio or MIDI (Eck, 2007)
- In current study we use scores but rely on features (mostly) obtainable using estimation.





D. Eck. (2007). Beat tracking using an autocorrelation phase matrix. In *Proceedings of the 2007 International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1313-1316.

Treat as regression task ...

y = f(x)

x is a note or set of notes (chord) described by:

Durations (quarter note, half note, ...)
Amplitudes (piano, forte, ...)
Accelerations (crescendo, decrescendo, ...)
Position in measure
Position in phrase

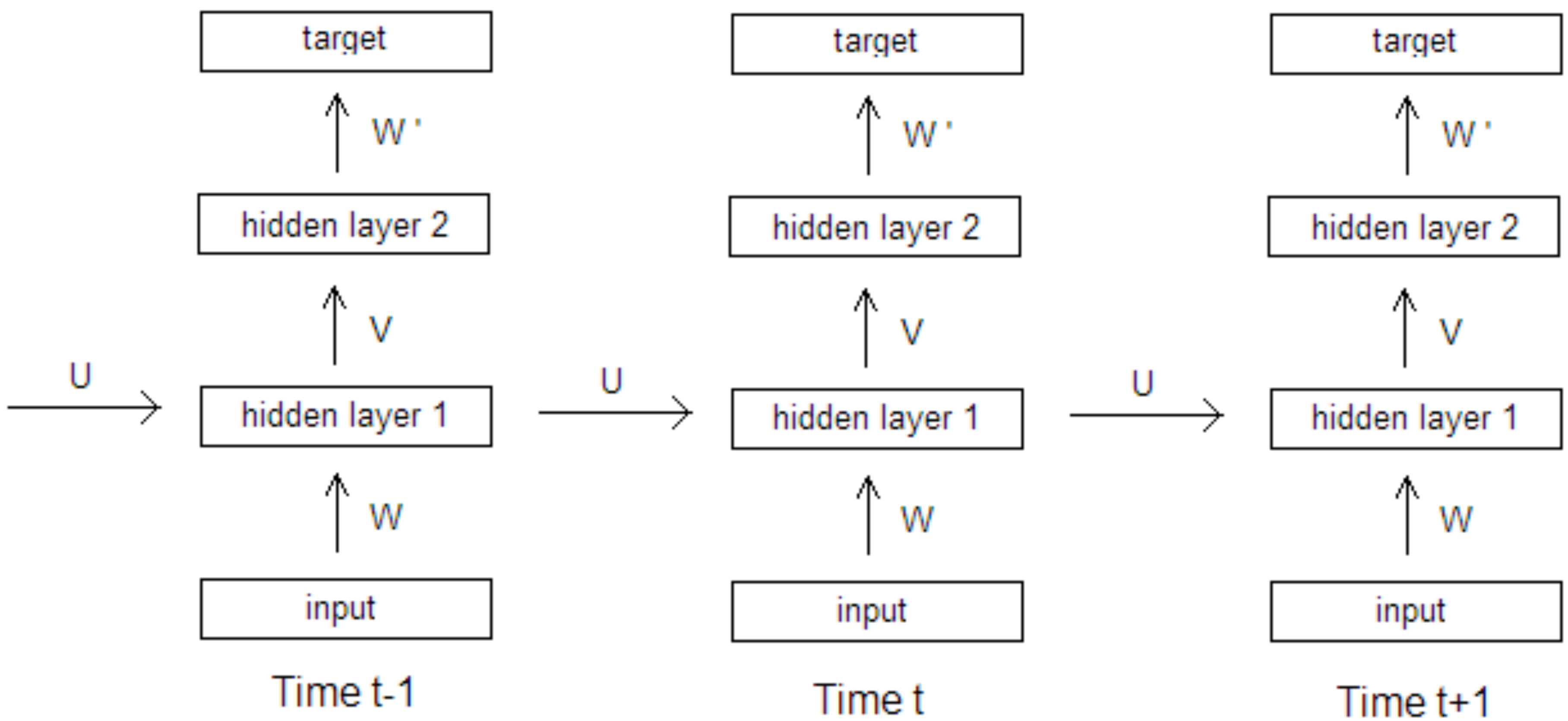
When exact note durations are known (i.e. when a score is available) we use a binary input encoding.

y is expressive deviation described by:

Note velocities
Local time deviations (chord spread...)
Overall tempo deviation

Regression algorithms

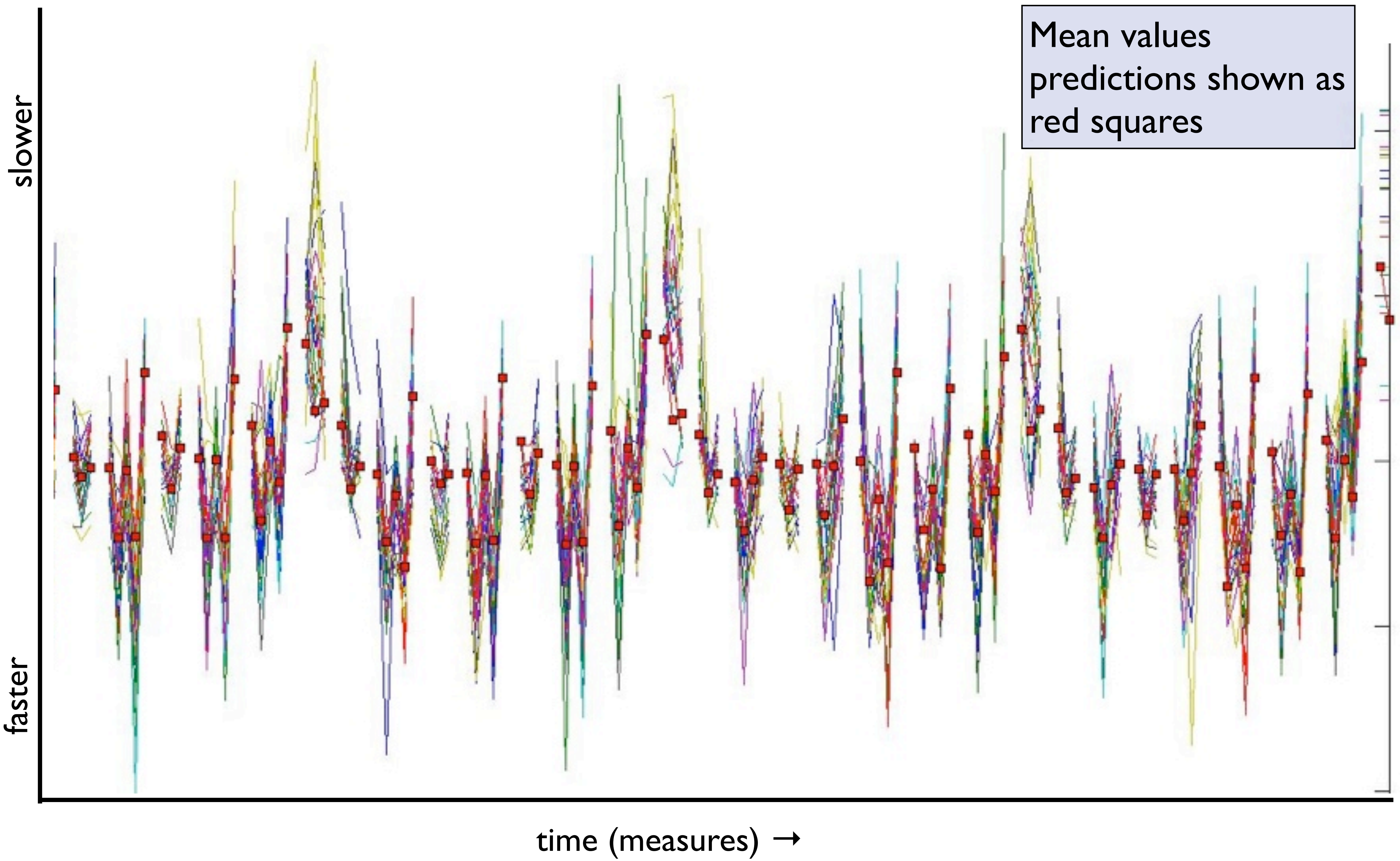
- Local variations likely learnable by any good regression method
- We also want to learn long timescale structure not encoded locally
- Baseline: recurrent neural network trained using BackProp Through Time
- Alternative: Deep Belief Network (Hinton et.al.) trained using contrastive divergence



Experiment: Learn to Perform Schubert Waltzes

- 10 highly trained pianists (performance PhD, University of Montreal Faculty of Music)
- 5 similar waltzes by Schubert
- Recorded multiple performances for each pianist on Bösendorfer ZEUS reproducing imperial grand piano
- Store as MIDI (note times and durations; pedaling)
- At least 2 performances per piece per pianist; for each performance, the piece was repeated
- 115 total performances; 38284 notes in all

Timing deviations for all 20 performances of a single waltz.



Training and generation

Training:

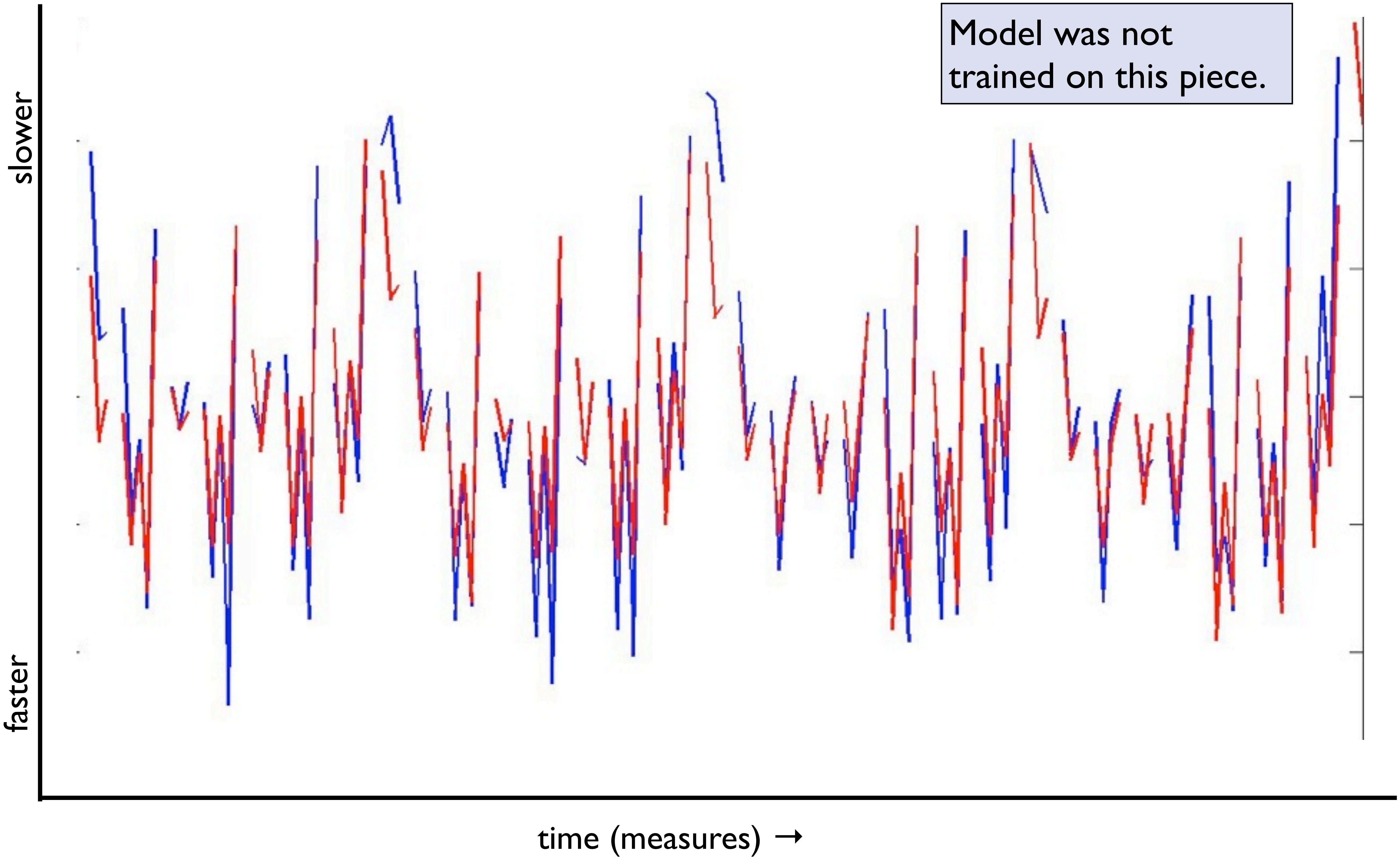
- Train algorithms on 4 pieces using MIDI performances captured from Bösendorfer ZEUS.
- Ensure generalization using out-of-sample data

Generation:

- Predict note velocities, local time deviations and overall tempo deviation for 5th piece
- Generate machine performance as MIDI from predictions
- Record performance from MIDI on Bösendorfer ZEUS

Pianist pedaling was ignored. We generated pedaling from note timing profile. (Future work)

Mean timing deviations (blue) versus predicted deviations (red)



Discussion

- Model learned:
  - phrase final lengthening
  - basic waltz feel (“lilt”)
  - voice leading
- Model did not learn:
  - more complex melodic phrasing
  - good pedaling
  - ability to make “radical” performance (regression to the mean)
- Baseline algorithm (standard BPTT recurrent network) performed as well as more complex algorithm (Deep Belief network)

Conclusions

- Expressive timing and dynamics can be learned using straight-forward machine learning approach
- Score-related information is relatively easy to obtain from MIDI performance or audio
- Can form core "performace module" for online music generation software

Bibliography

- B. Repp. (1990). Patterns of expressive timing in performances of a Beethoven minuet by nineteen famous pianists. *Journal of the Acoustical Society of America*, 88: 622-641.
- B. Repp. (1997). The aesthetic quality of a quantitatively average music performance: Two preliminary experiments. *Music Perception,* 14: 419-444.
- A. Friberg, R. Bresin & J. Sundberg (2006). Overview of the KTH rule system for musical performance. *Advances in Cognitive Psychology*, 2(2-3):145-161.
- G. Widmer (2003). Discovering simple rules in complex data: A meta-learning algorithm and some surprising musical discoveries. *Artificial Intelligence* 146:129-148.
- C. Raphael (2004). A Bayesian Network for Real-Time Musical Accompaniment, *Neural Information Processing Systems (NIPS) 14.*
- J.F. Paiement, D. Eck, S. Bengio & D. Barber (2005). A graphical model for chord progressions embedded in a psychoacoustic space. *In Proceedings of the 22nd International Conference on Machine Learning (ICML)*, Bonn, Germany.
- S. Dixon, W. Goebl & G. Widmer (2002). The performance worm: Real time visualisation of expression based on Langner's tempo-loudness animation. In *Proceedings of the International Computer Music Conference (ICMC).*
- D. Eck. (2007). Beat tracking using an autocorrelation phase matrix. In *Proceedings of the 2007 International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1313-1316.