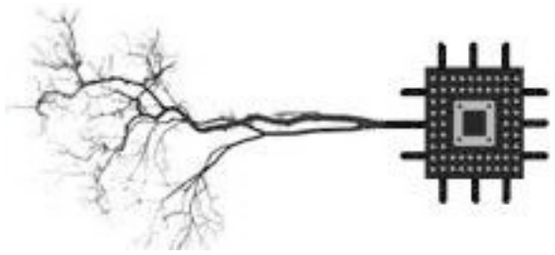
Measuring & Modeling Musical Expression

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

NIPS 2007 Music, Brain and Cognition Workshop



Université de Montréal

Brams

International Laboratory for
Brain, Music and Sound Research

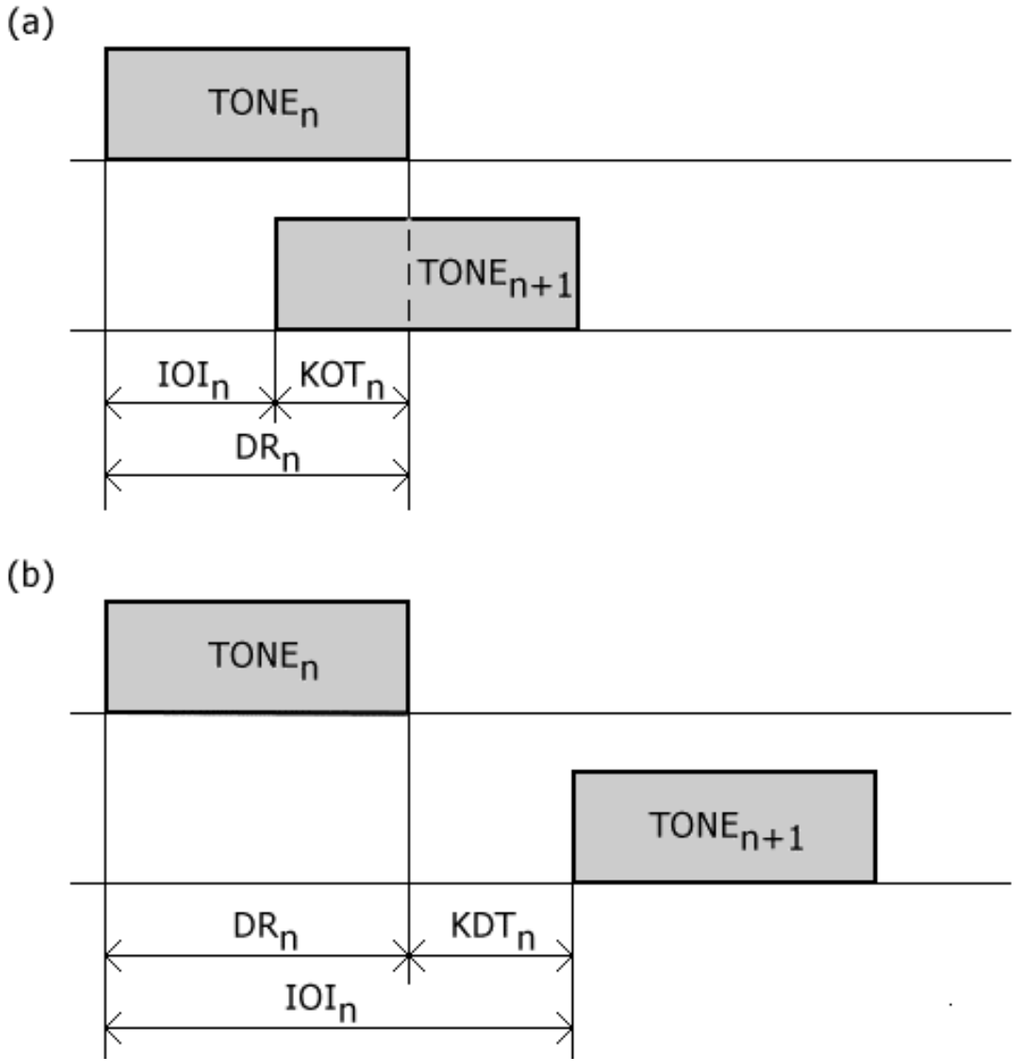
Overview

- Why care about timing and dynamics in music?
- Previous approaches to measuring timing and dynamics
- Models which predict something about expression
- Working without musical scores
- A correlation-based approach for constructing metrical trees

Note-level measures (MIDI)

- Pitch
- Velocity
- Duration

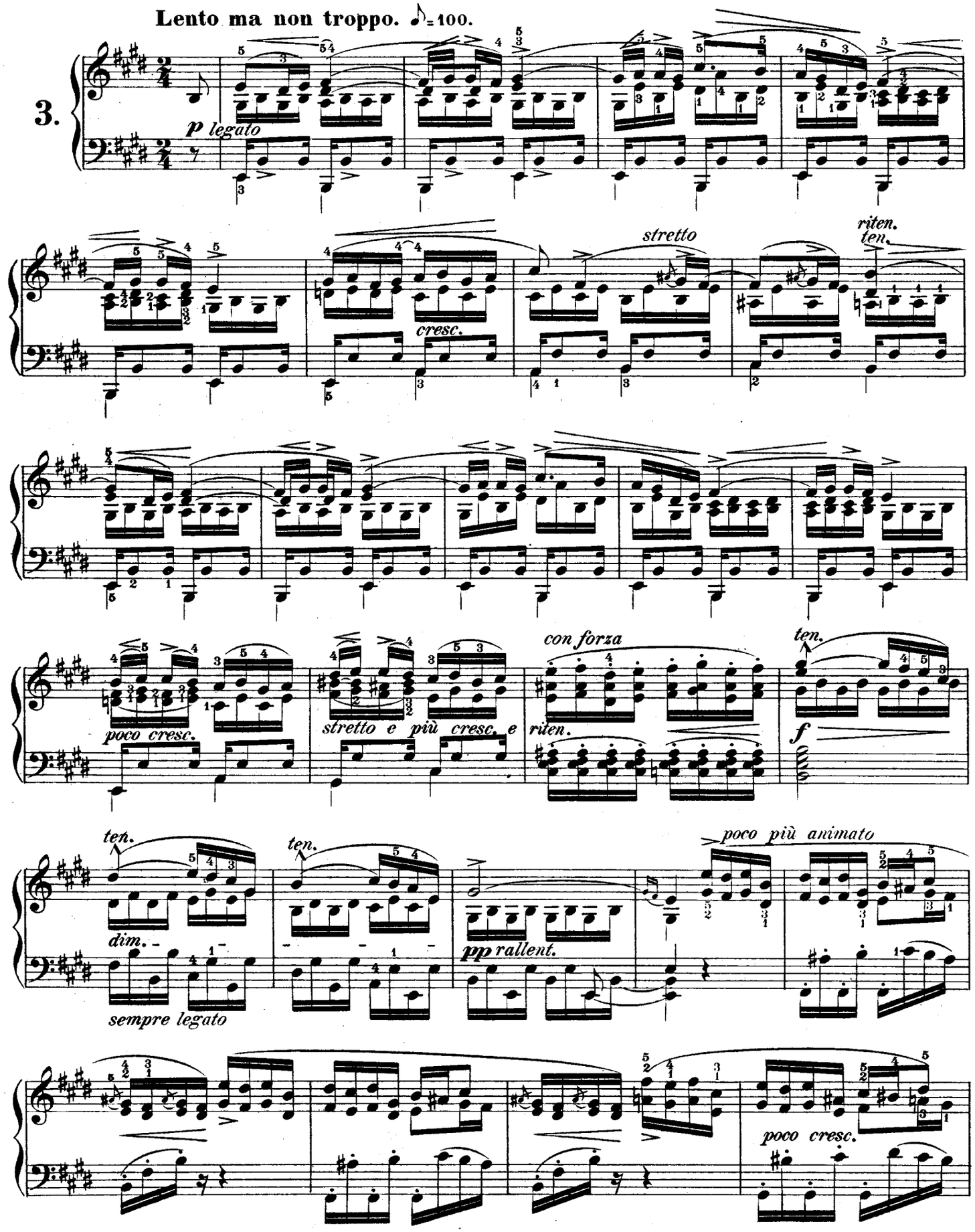
- IOI (inter-onset interval)
- KOT (key overlap time)
- Pedaling (piano)



From R. Bresin "Articulation Rules for Automatic Music Performance"

Figure 1. (a) Definition of inter-onset interval (IOI_n), duration (DR_n) and key overlap time (KOT_n) for $TONE_n$ followed by an overlapping $TONE_{n+1}$. (b) Definition of inter-onset interval (IOI_n), duration (DR_n) and key detached time (KDT_n) for $TONE_n$ followed by a non-overlapping $TONE_{n+1}$.

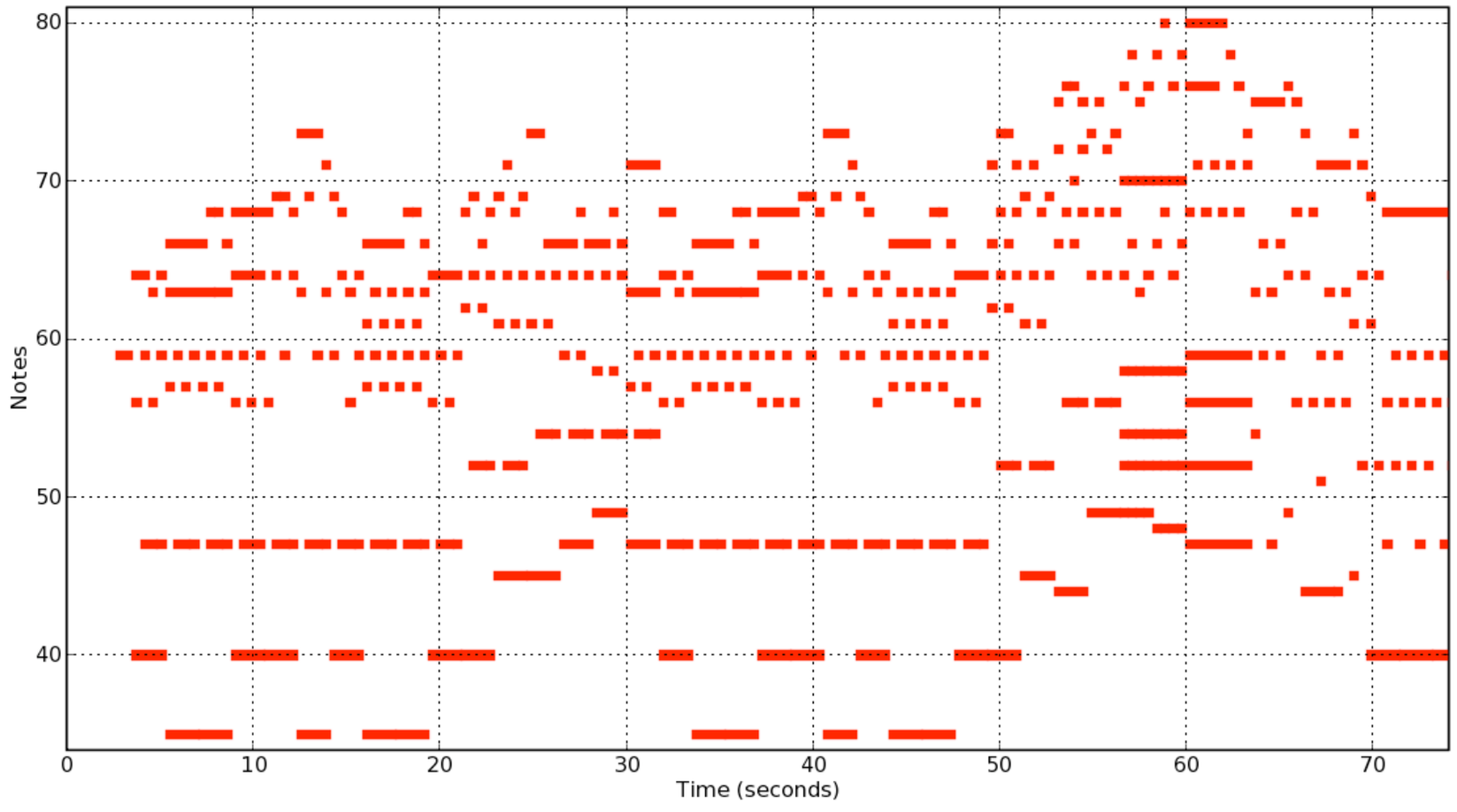
Example:
Chopin Etude
Opus 10 No 3



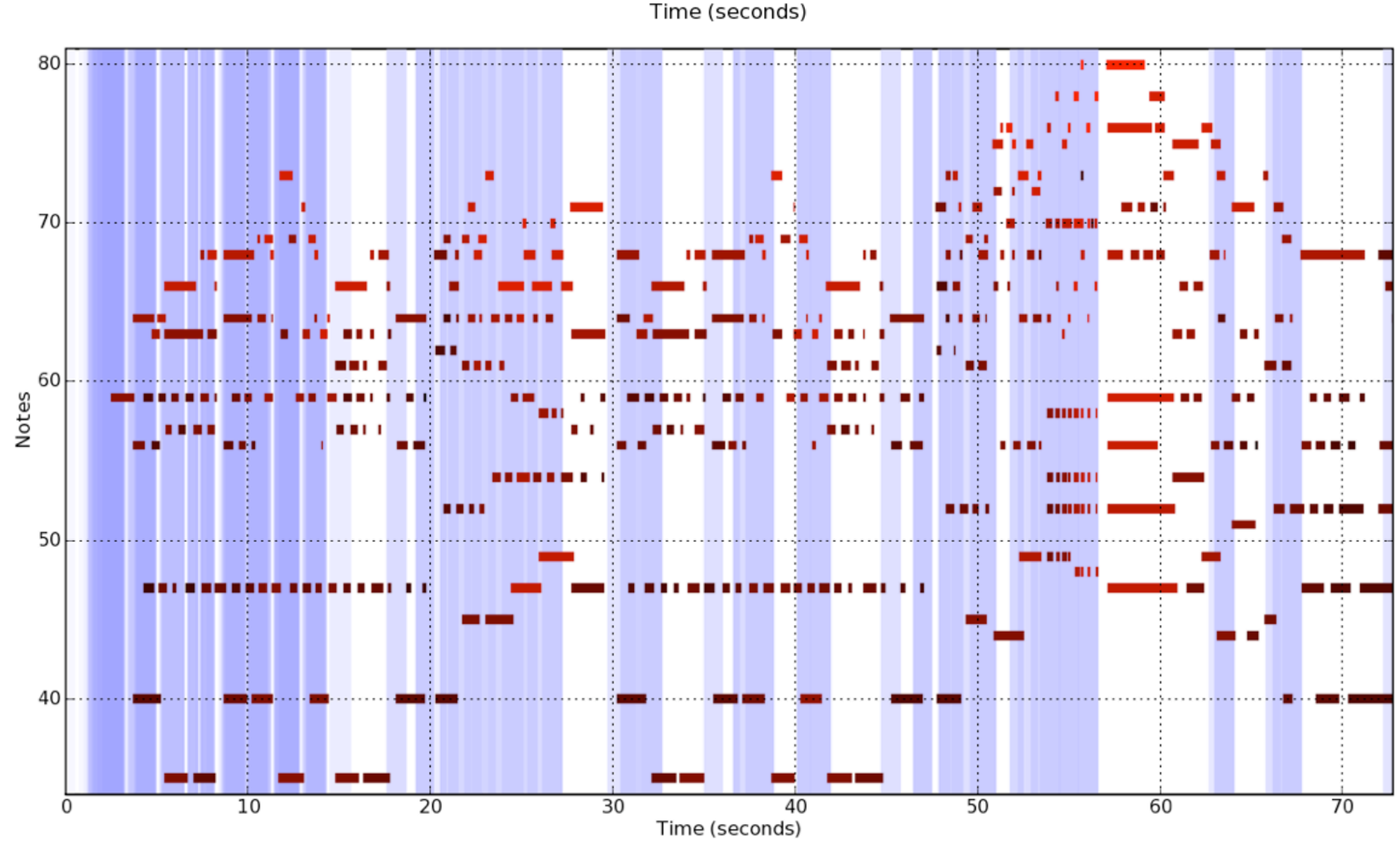
Bösendorfer

Example: Chopin Etude Opus 10 No 3

Deadpan
(no expressive timing or dynamics)



Human performance
(Recorded on Boesendorfer ZEUS)

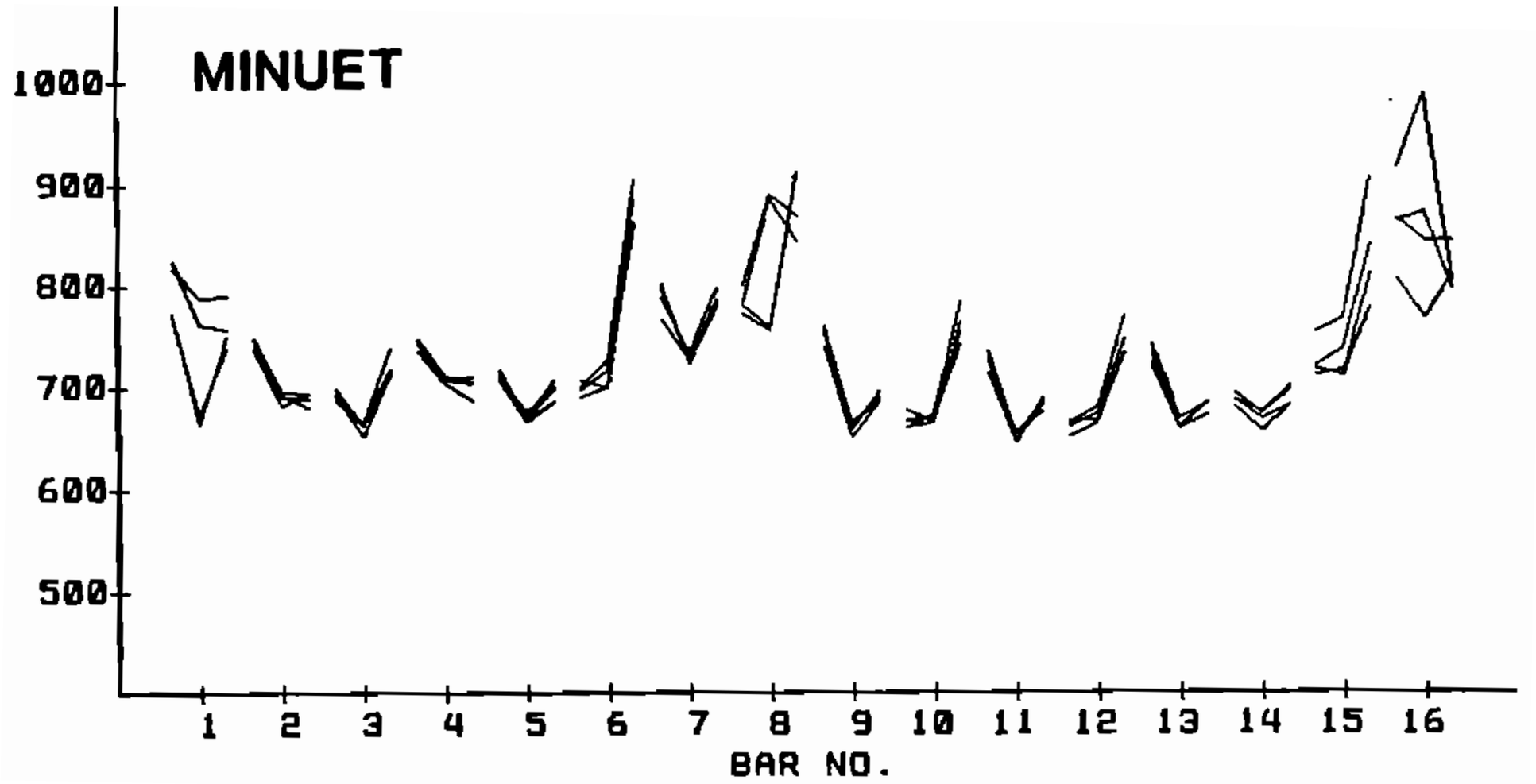


Differences limited to:

- timing (onset, length)
- velocity (seen as red)
- pedaling (blue shading)

What can we measure?

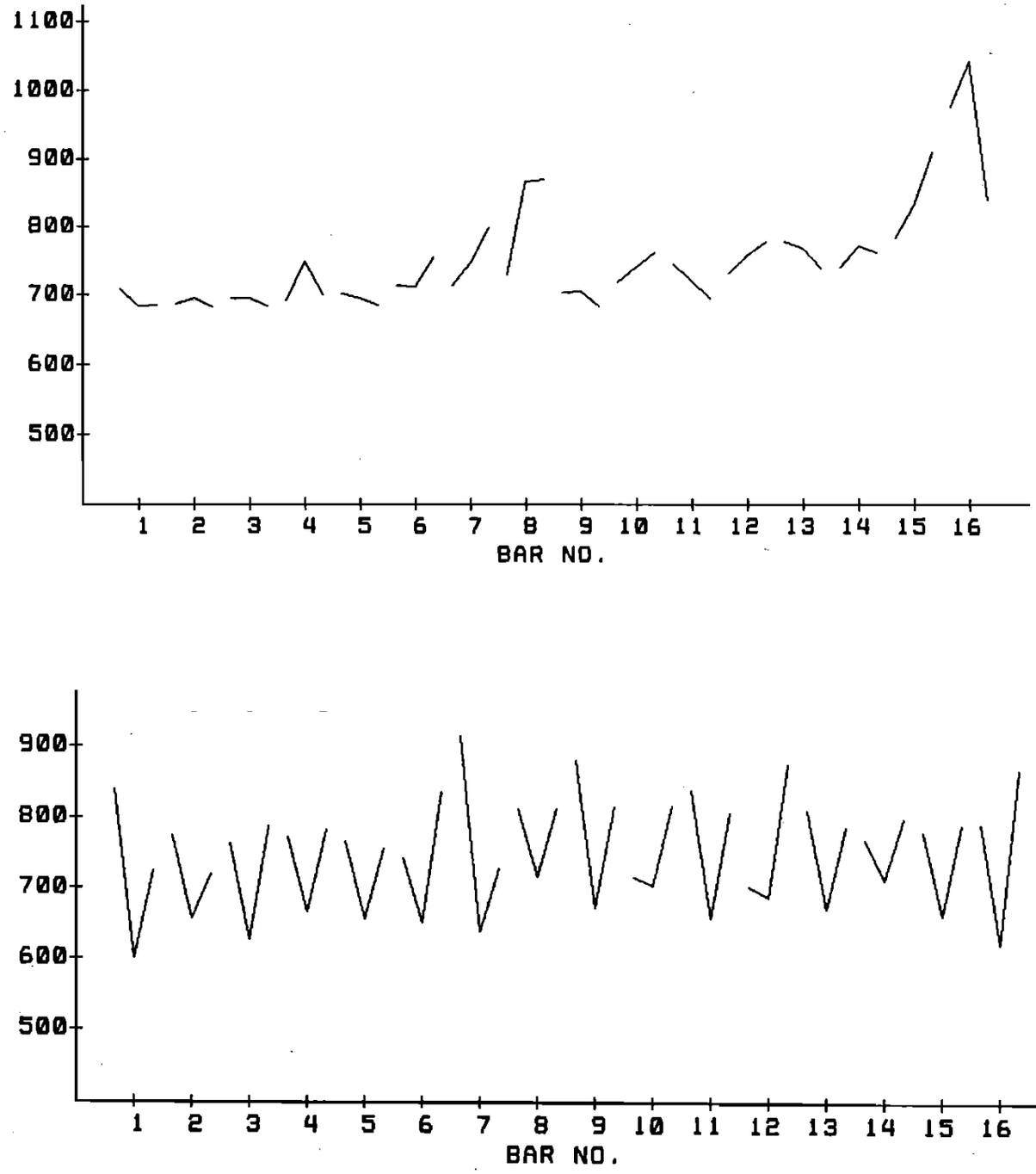
- Repp (1989) measured note IOIs in 19 famous recordings of a Beethoven minuet (Sonata op 31 no 3)



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.
  - Timing and expressive dynamics may be inter-dependent
  - However, interesting that EA, SA sound good at all

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.

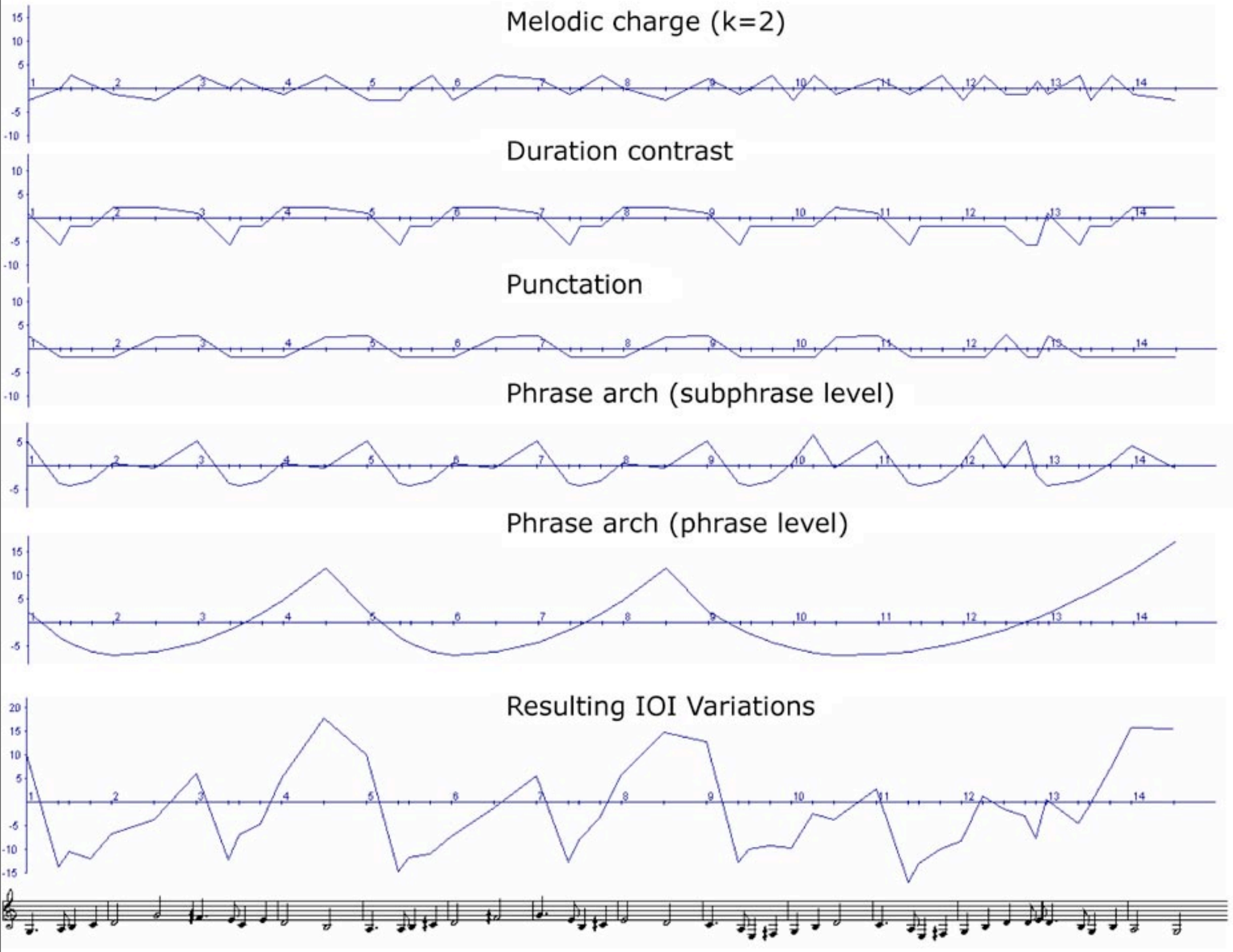


Figure 2.

The resulting IOI deviations by applying Phrase arch, Duration contrast, Melodic charge, and Punctuation 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 (Widmer

RULE TL2:
`abstract_duration_context = equal-longer`
`& metr_strength` $\leqslant$ `1`
$\Rightarrow$ `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' $\leqslant 1$)."

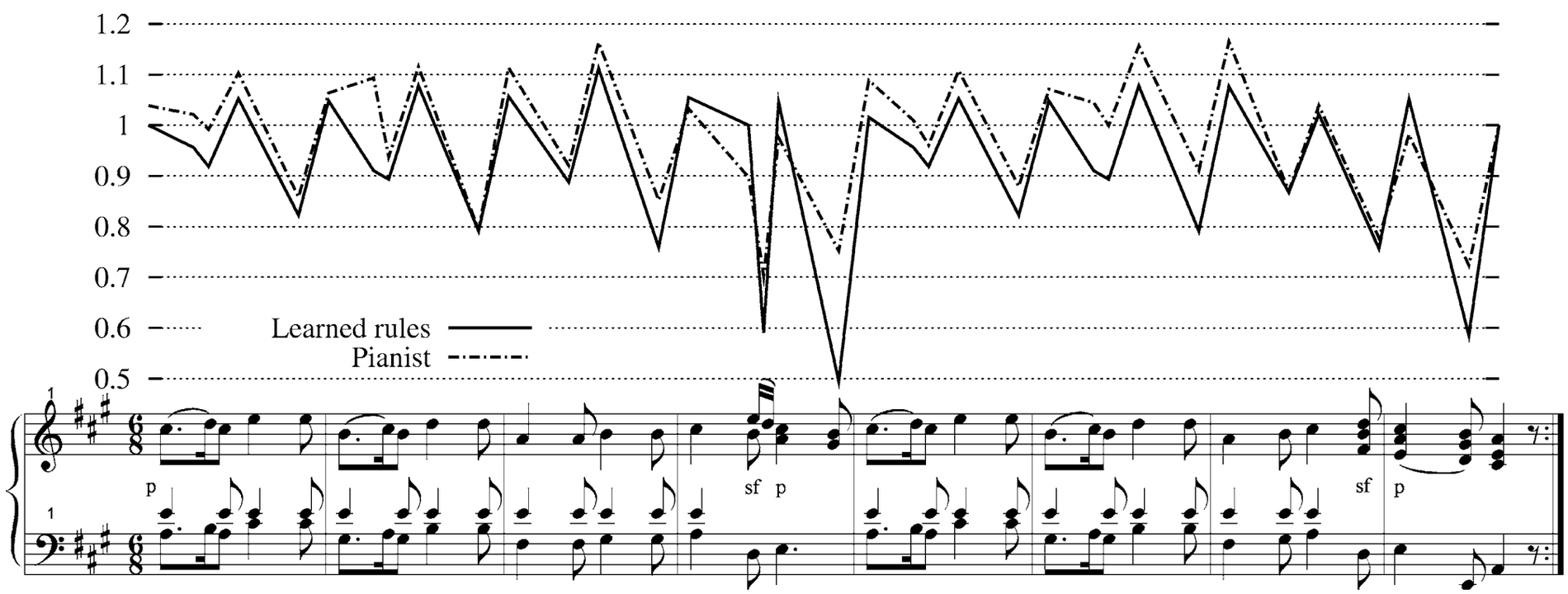


Fig. 5. Mozart Sonata K.331, 1st movement, 1st part, as played by pianist and learner. The curve plots the relative tempo at each note—notes above the 1.0 line are shortened relative to the tempo of the piece, notes below 1.0 are lengthened. A perfectly regular performance with no timing deviations would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

Music Plus One (C. Raphael)

Task 1 : Listen

Inputs:

- sampled acoustic signal
- musical score

Output:

- Time at which notes occur

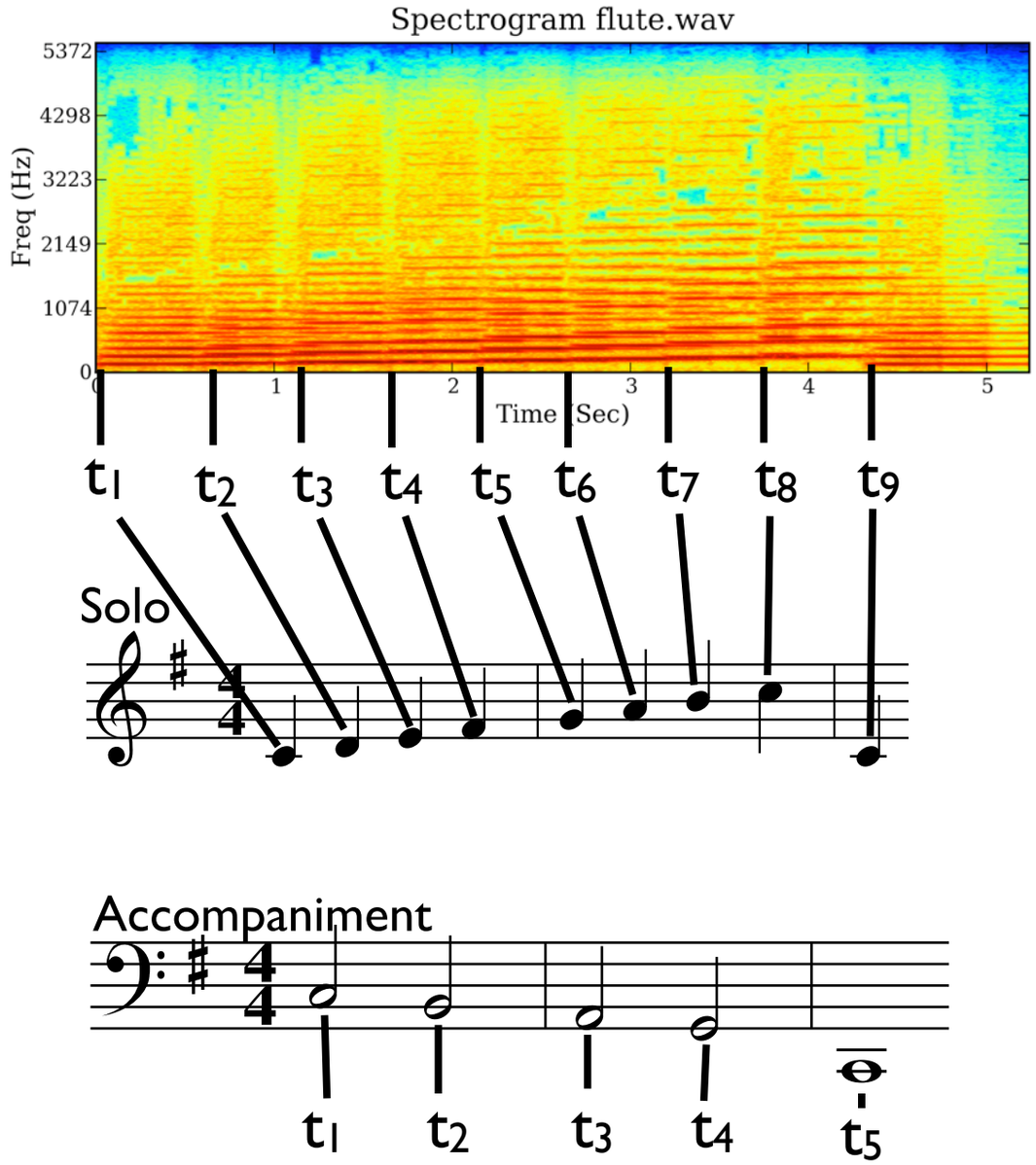
Task 2 : Play

Inputs:

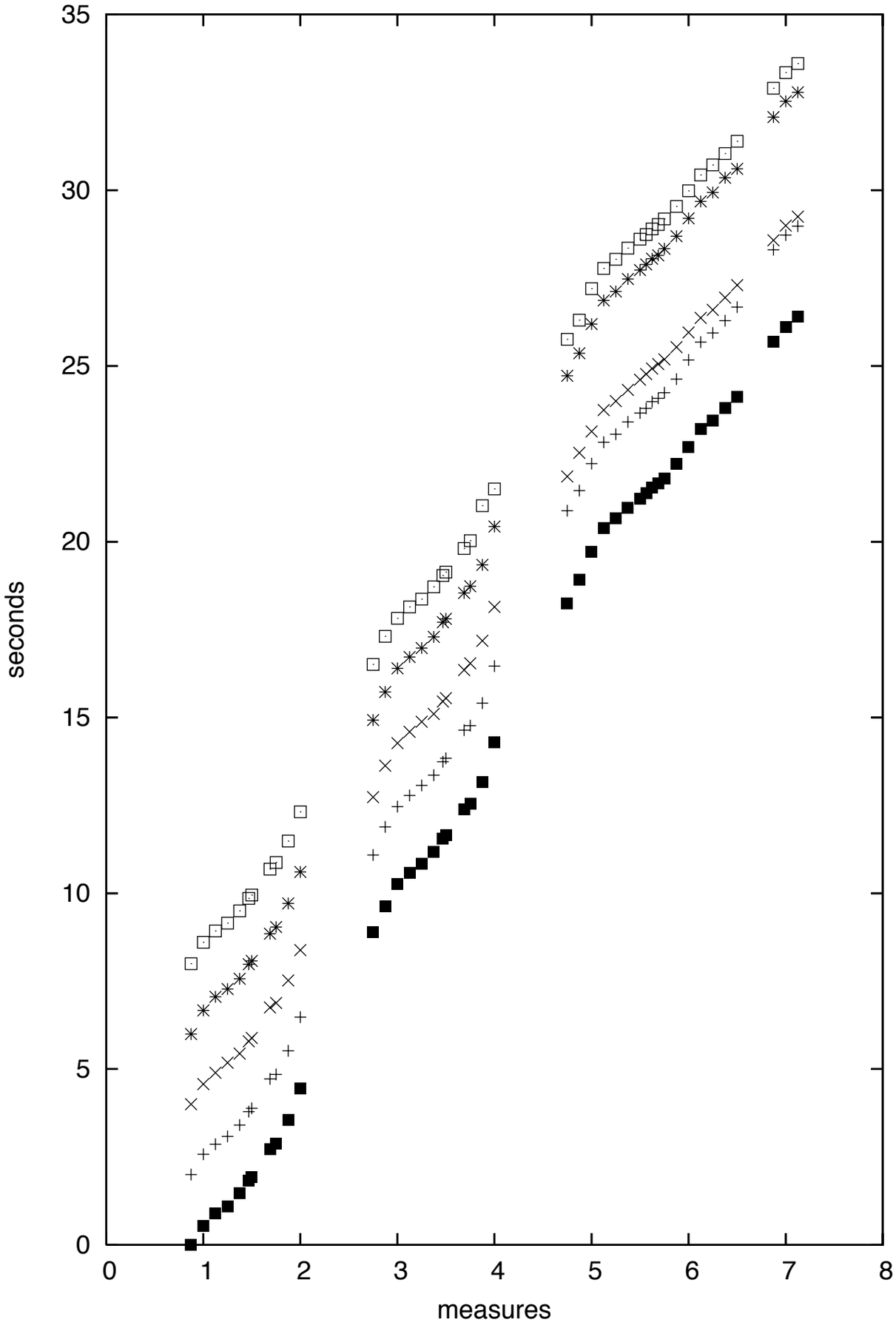
- output from Listen module
- musical score
- rehearsal data from musician
- performances of accompaniment

Output:

- Music accompaniment in real time



Text and graphics on following pages from slide presentation by Chris Raphael. Thanks Chris!



Five performances of same musical phrase

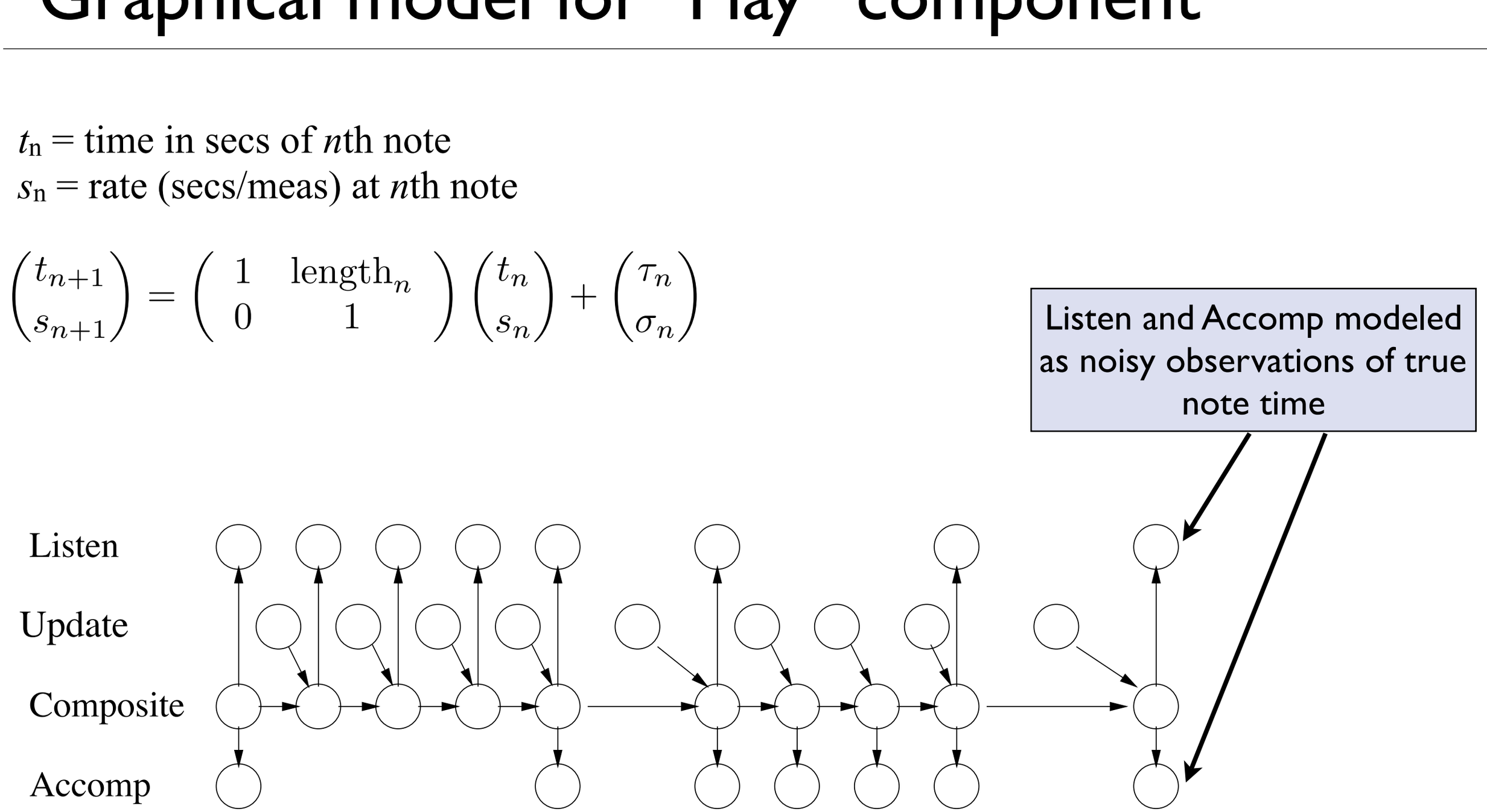
Intuition: there are regularities to be learned

Graphical model for "Play" component

t_n = time in secs of nth note
s_n = rate (secs/meas) at nth note

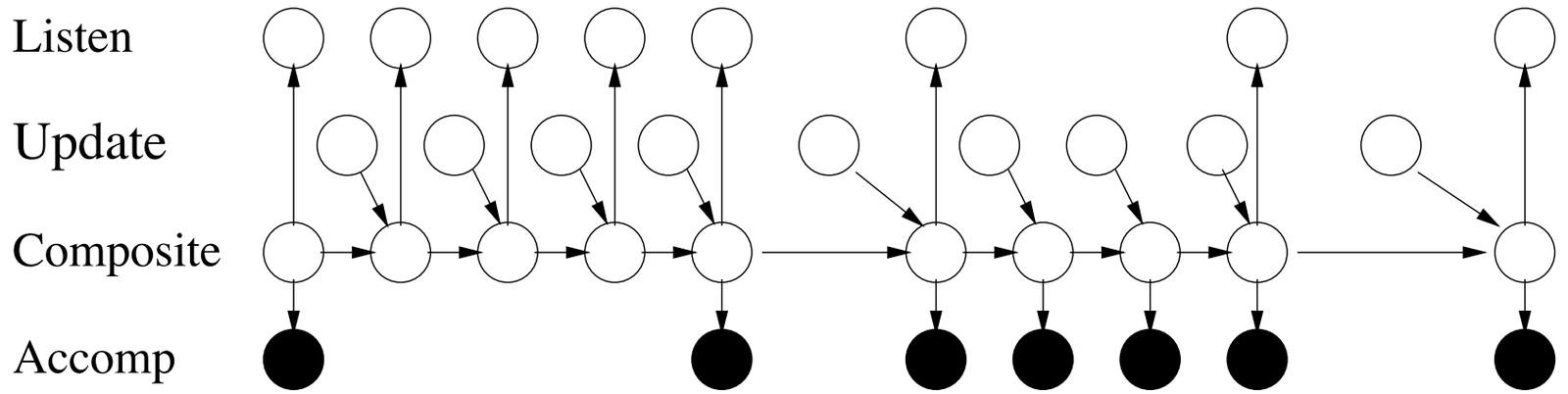
$$\begin{pmatrix} t_{n+1} \\ s_{n+1} \end{pmatrix} = \begin{pmatrix} 1 & \text{length}_n \\ 0 & 1 \end{pmatrix} \begin{pmatrix} t_n \\ s_n \end{pmatrix} + \begin{pmatrix} \tau_n \\ \sigma_n \end{pmatrix}$$

Listen and Accomp modeled as noisy observations of true note time

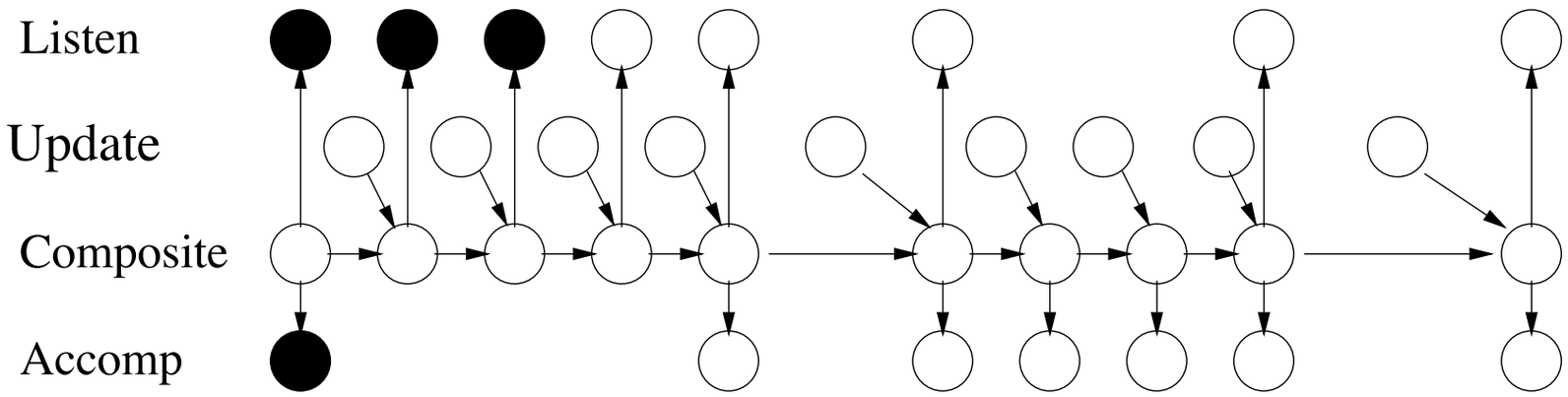


Inference and generation in “Play” component

Inference: Model trained using EM, first on accompaniment data then solo data.



Real time accompaniment: Each time new info observed recompute marginal for next accomp. note and schedule.



Yesterday's talk so I'll keep it short...

KCCA (Dorard, Hardoon & Shawe-Taylor)

- Want to fit specific performer style (unlike, e.g., Widmer et.al.)
- Correlate musical score to performance
- Score features: melody and chords projected into vector using Paiement et.al.

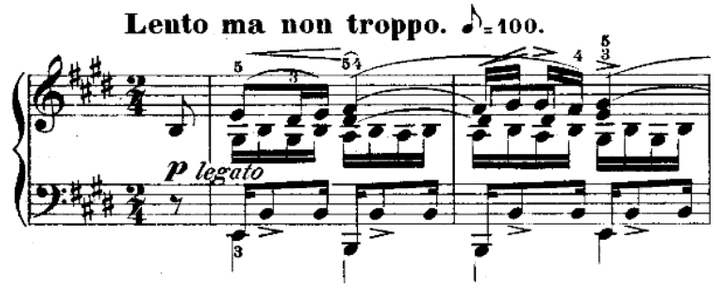
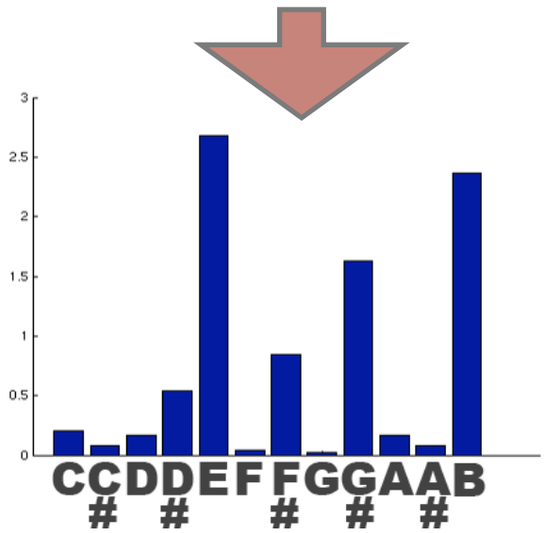


Figure 3: *First two bars of Etude 3 Opus 10 by Chopin*

Beat	Melody	Chord
1	$B3$	$B3$
2	$E3$	$[E2\ B2\ G\#3\ B3\ E4]$
3	$D\#3$	$[E2\ B2\ G\#3\ D\#3]$
...	...	...

Figure 4: *Feature representation of the score in Figure 3*



Yesterday's talk so I'll keep it short...

KCCA (Dorard, Hardoon & Shawe-Taylor)

- Audio performance features: instantaneous tempo and loudness of onsets ("worm"of Dixon et al)
- Use KCCA (a kernel version of Canonical Correlation Analysis) to correlate these two views.
- Required kernel for score features and kernel for audio (worm) features
- Currently only preliminary results.

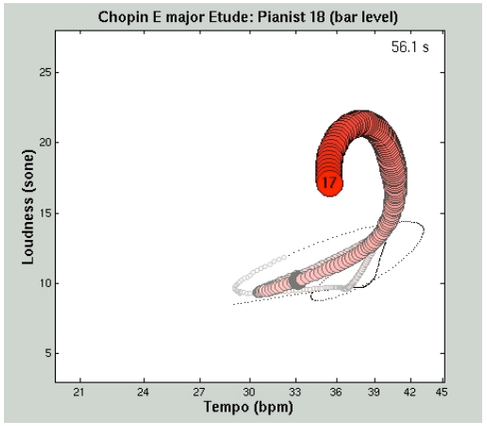


Figure 1: *Smoothed graphical view of a worm*

Beat	Tempo (bpm)	Loudness (sone)
1	22.3881	3.2264
2	22.3881	2.3668
3	21.4286	6.7167
4	19.0597	4.2105
5	28.1426	8.3444
6	30.0000	10.2206
7	26.7857	14.1084
8	25.8621	14.0037
9	35.7143	7.8521
...	...	...

Figure 2: *Machine representation of a worm*

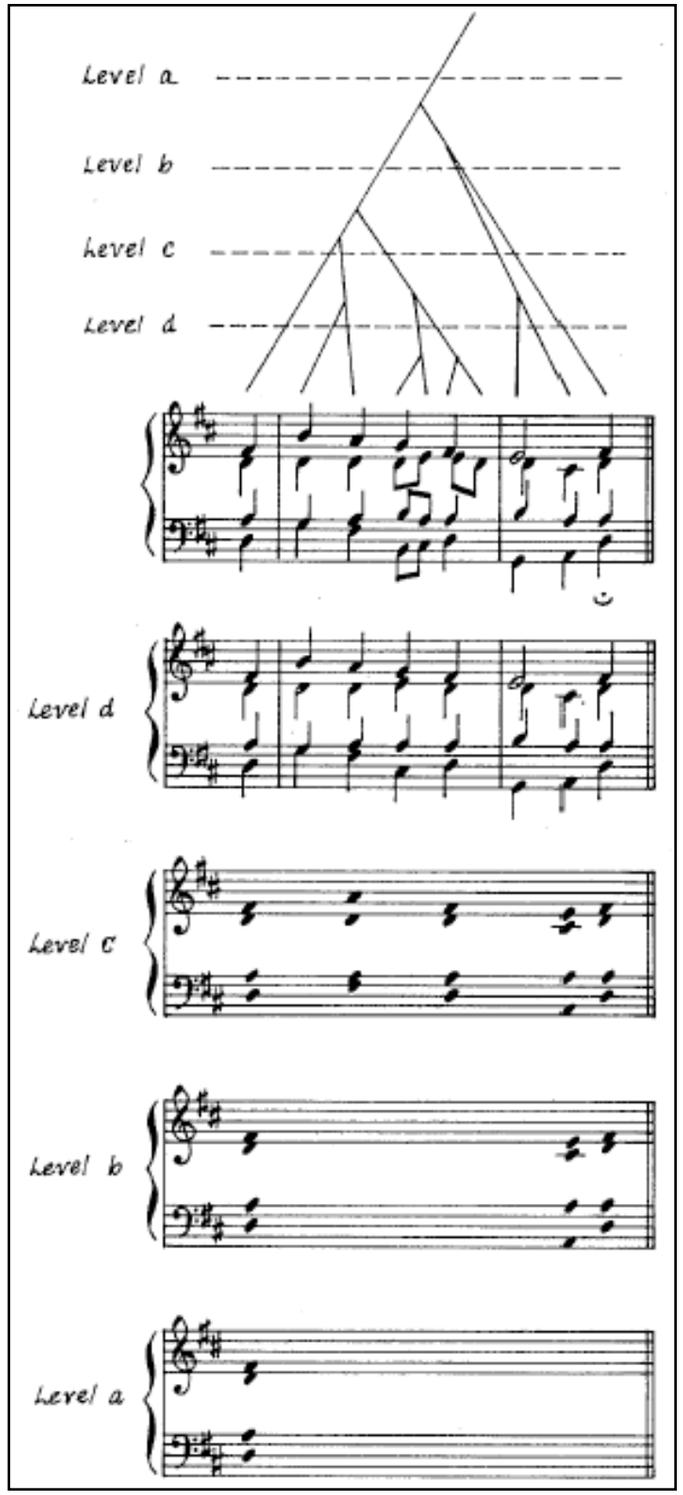
Summary

- Important information in timing and dynamics.
- Artificial expressive performances can be pleasing
- We saw four approaches to automatic performance:
  - "classic AI" rules-based system (KTH)
  - rules induction (Widmer)
  - generative model (Raphael)
  - kernel approach (Dorard et. al.)

But: these all make use of a musical score.
(Some less than others....)
Can we get away from that?

Challenges in score-free expressive performance

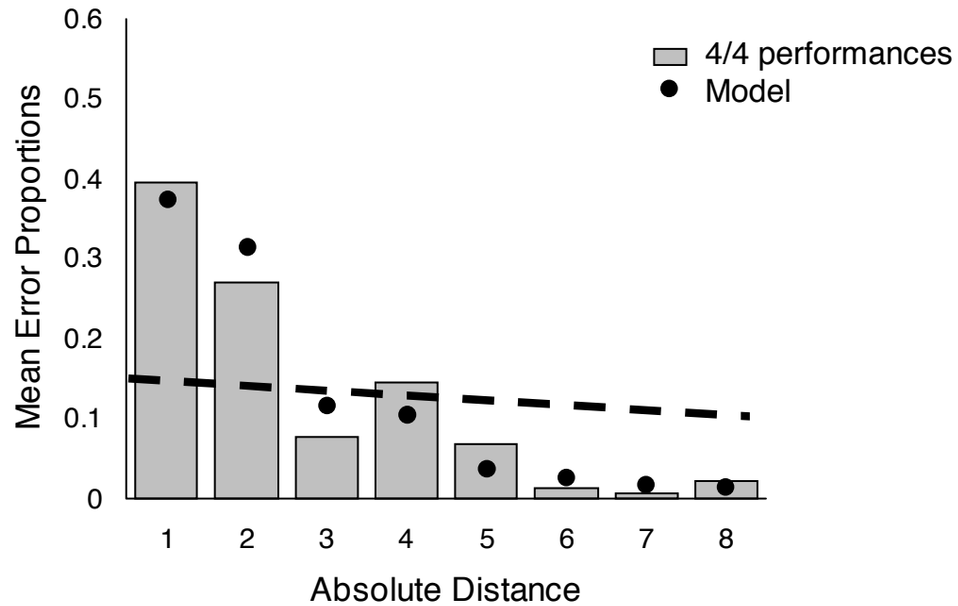
- Local information is not sufficient for modeling music expression
- Score contains long-timescale information about phrasing and metrical organization
- Automatic methods exist for estimating deep hierarchical structure in music from a score
- Without score, this task is more difficult

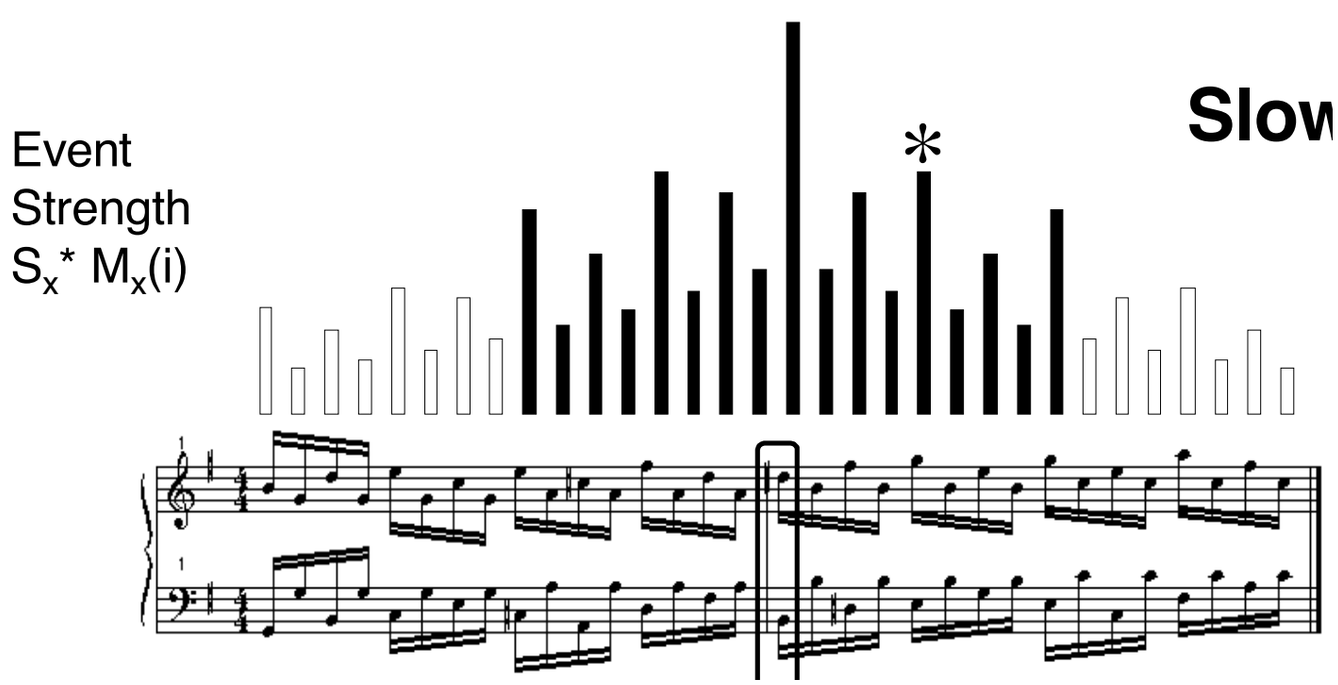


Graphic from AITEC Department of Future Technologies (ftp.icot.or.jp)

Focus: musical meter

- Meter provides long-timescale framework for music
- Meter and performance are closely related
- Example: performance errors correlate with meter.
  Palmer & Pfordresher (2003)
- Rest of the talk: use meter as proxy for musical score to gain access to nonlocal information



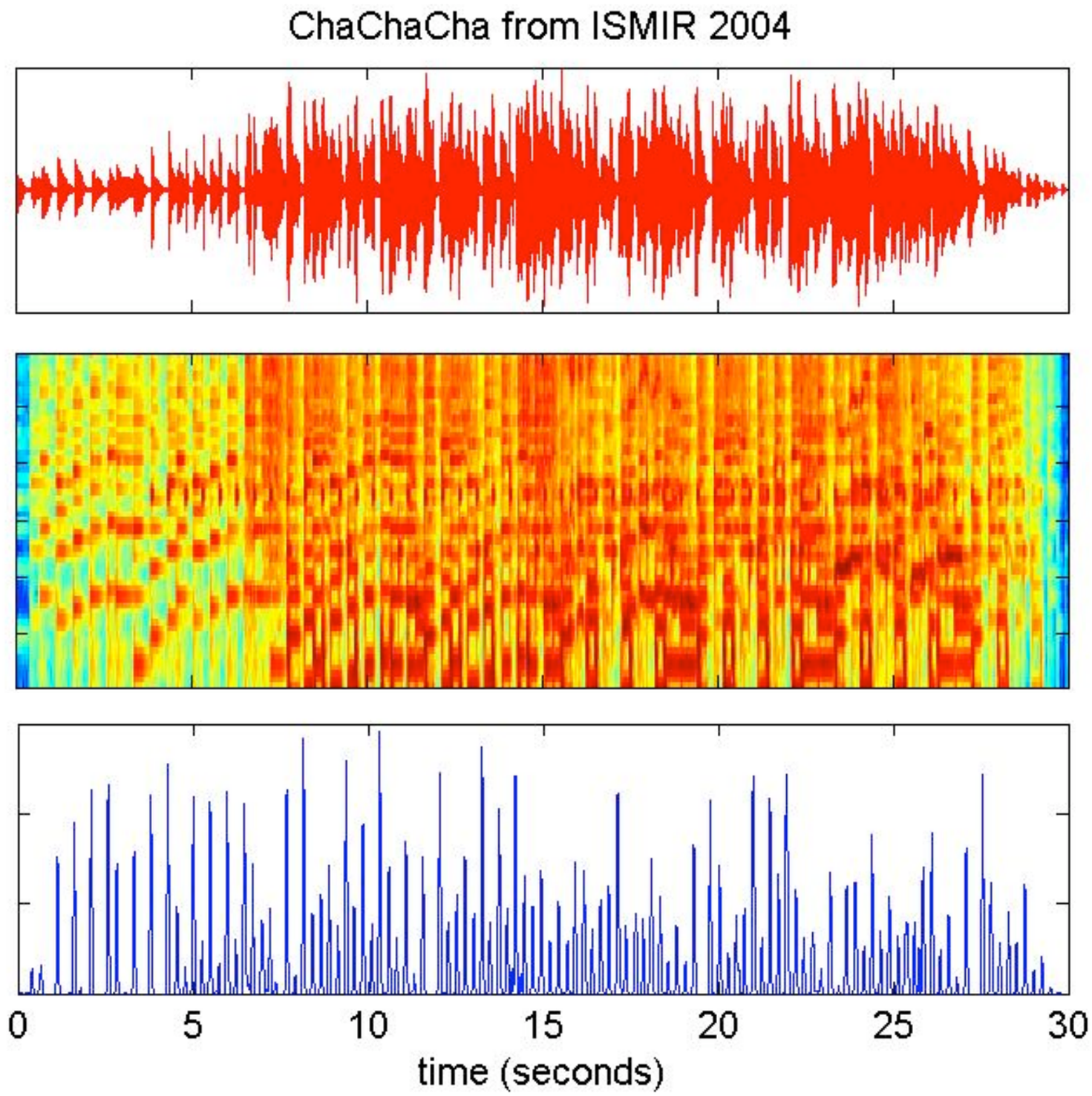


Audio pre-processing (not necessary for MIDI)

Waveform at original sampling rate

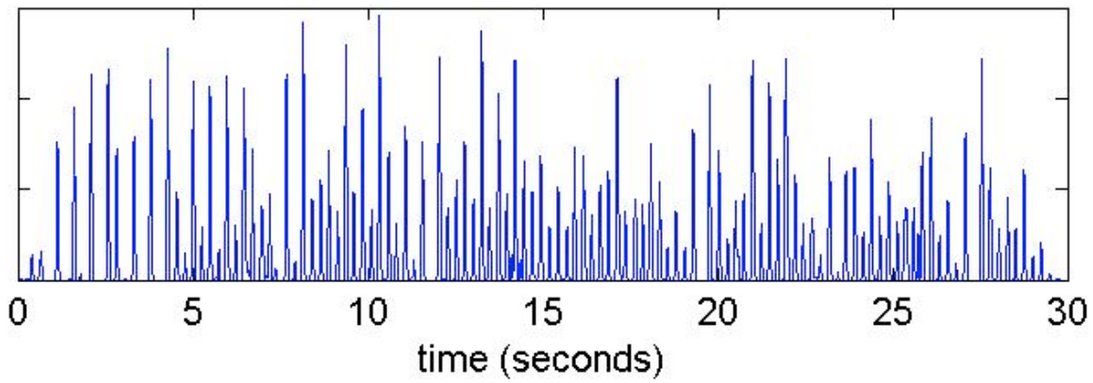
Log spectrogram with ~10ms frames

Sum of gradient yields ~100hz signal

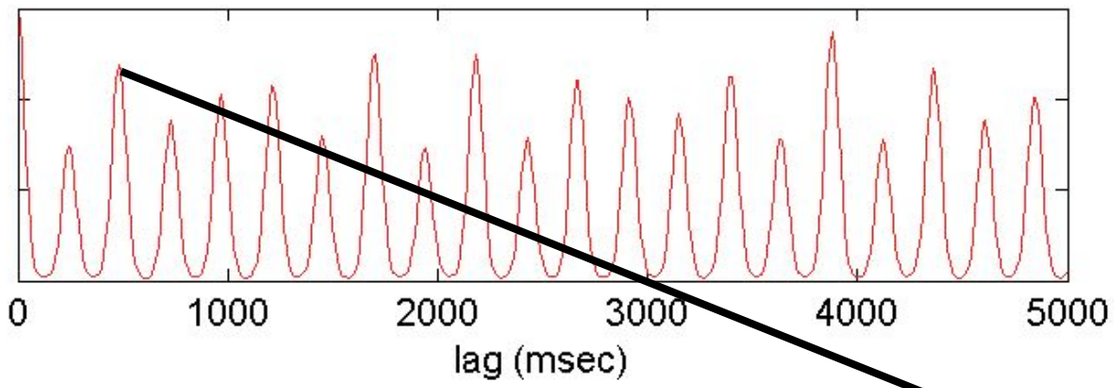


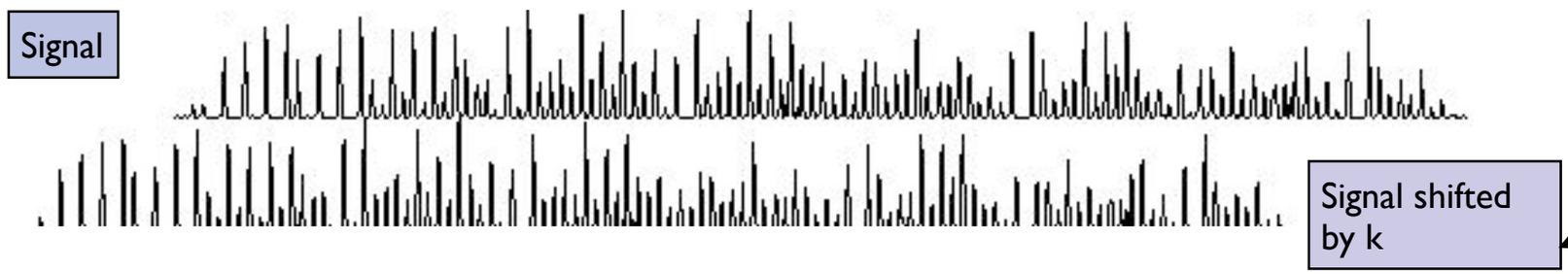
Computing Autocorrelation

100Hz signal



Autocorrelation

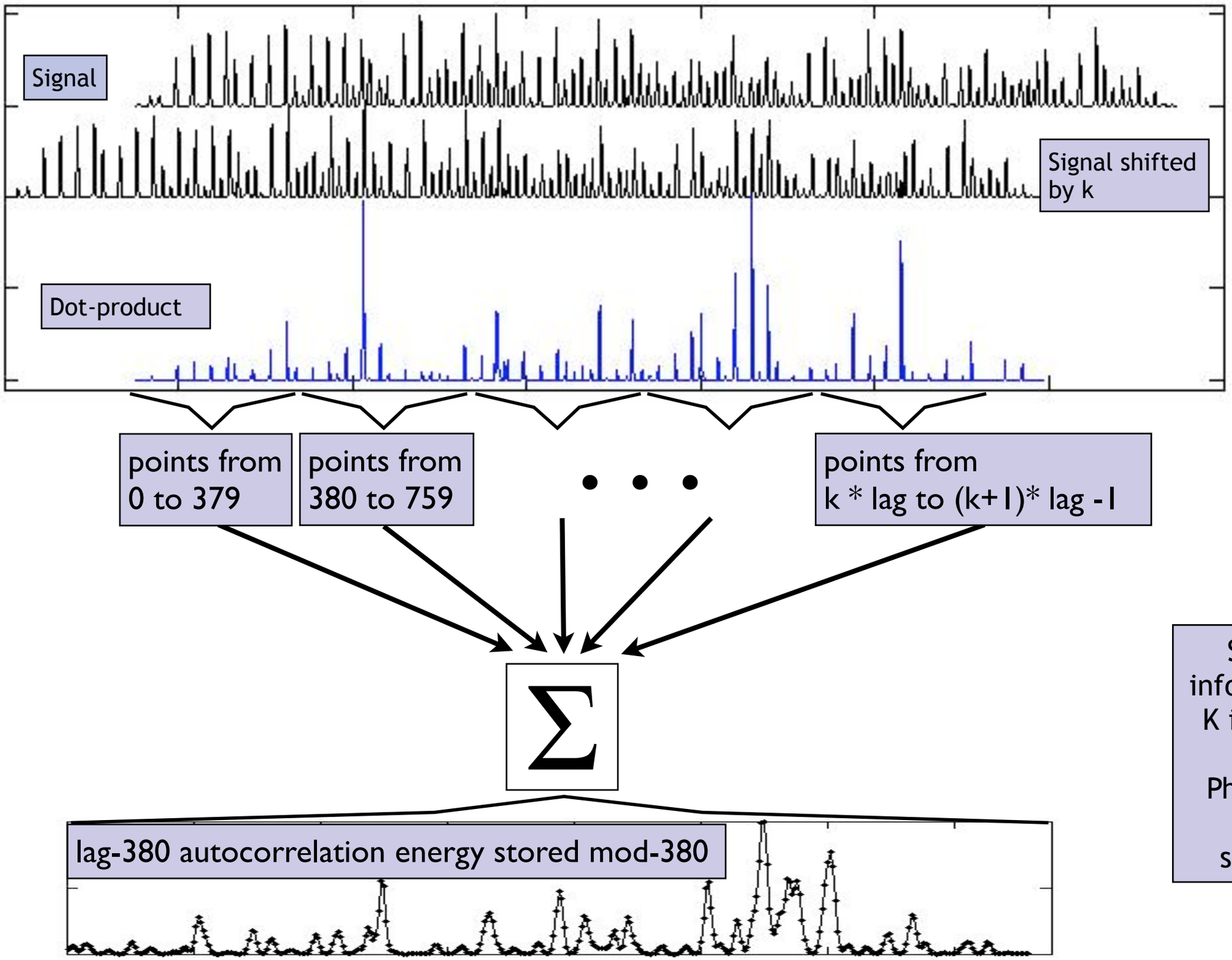




Autocorrelation value *a(k)* for a single lag k is the sum of dot-product between signal and signal shifted k points.

$$a(k) = \sum_{i=k}^{N-1} x(i)x(i-k)$$

Preserving phase (example: lag 380)

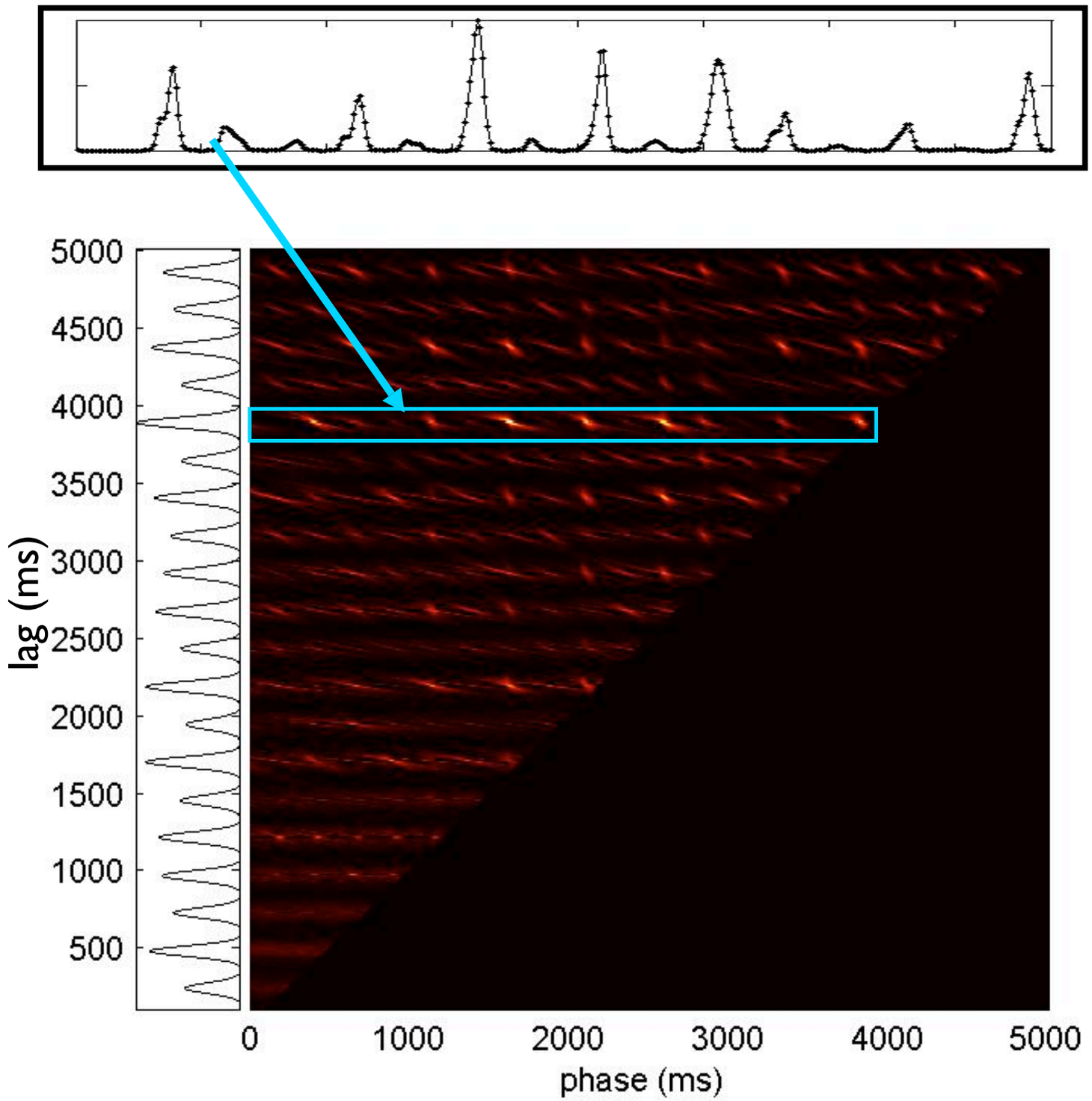


Store autocorrelation information for a single lag K in a vector of length K.

Phase of autocorrelation energy is preserved spatially in the vector.

The Autocorrelation Phase Matrix (APM)

- The autocorrelation phase matrix (APM) has a row for each lag.
- Rows are ordered by lag.
- Phase is stored in milliseconds. Thus the matrix is triangular (long lags take more time before they cycle around).

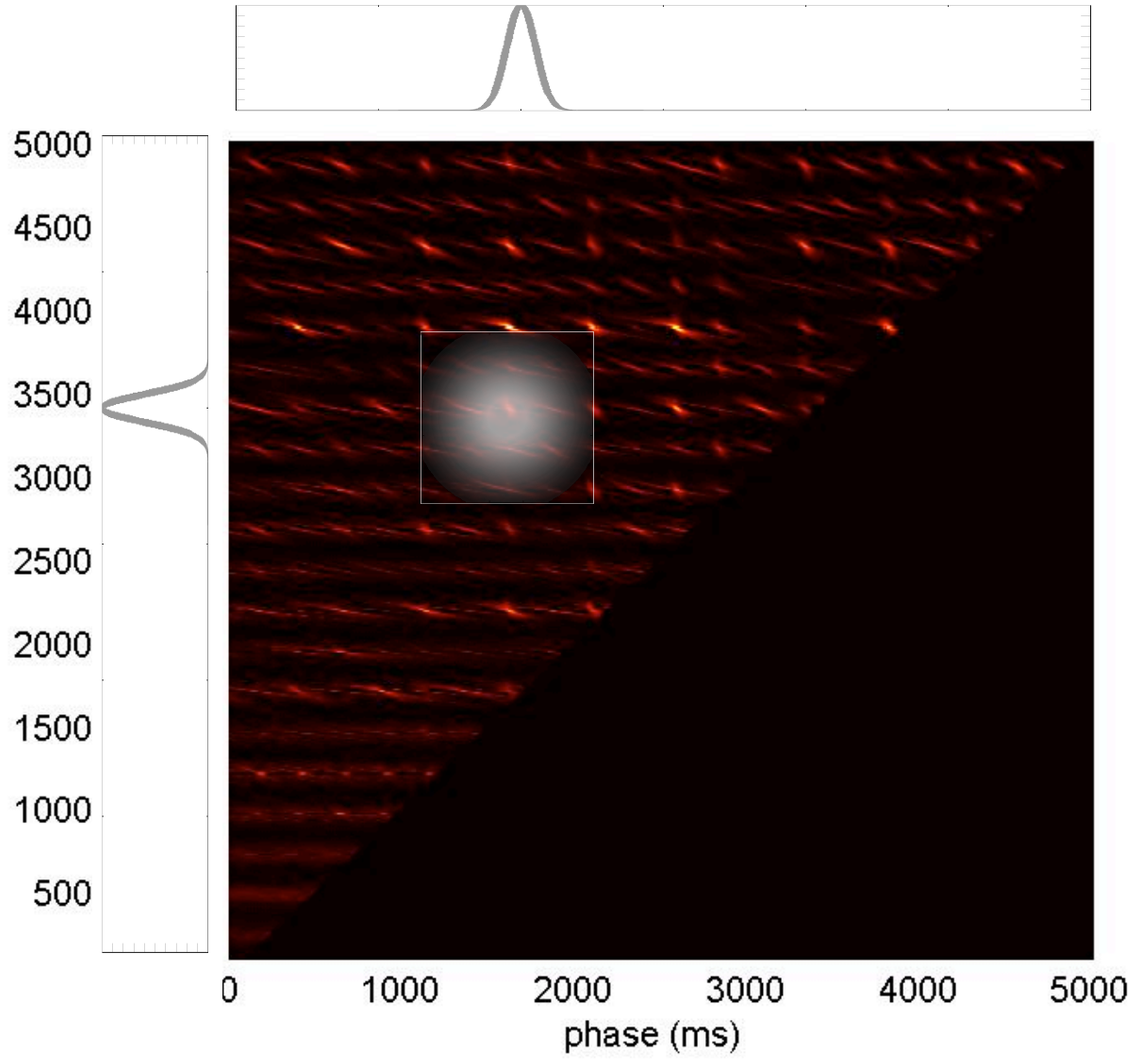


The Autocorrelation Phase Matrix (APM)

- The APM provides a **local representation** for tempo variations and rhythmical variations

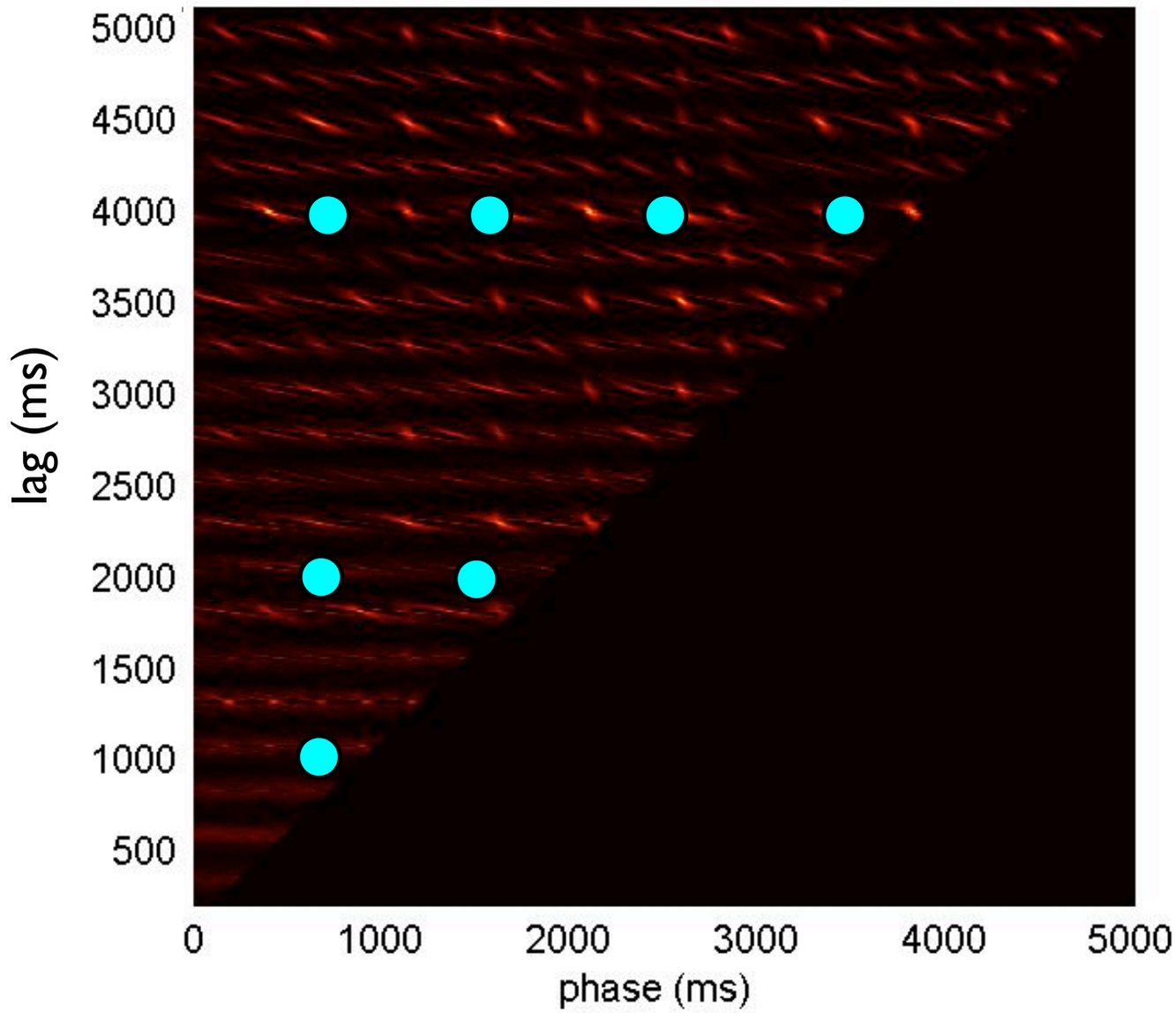
- Small horizontal changes on APM reach near-neighbors in frequency

- Small vertical changes on APM reach near-neighbors in phase

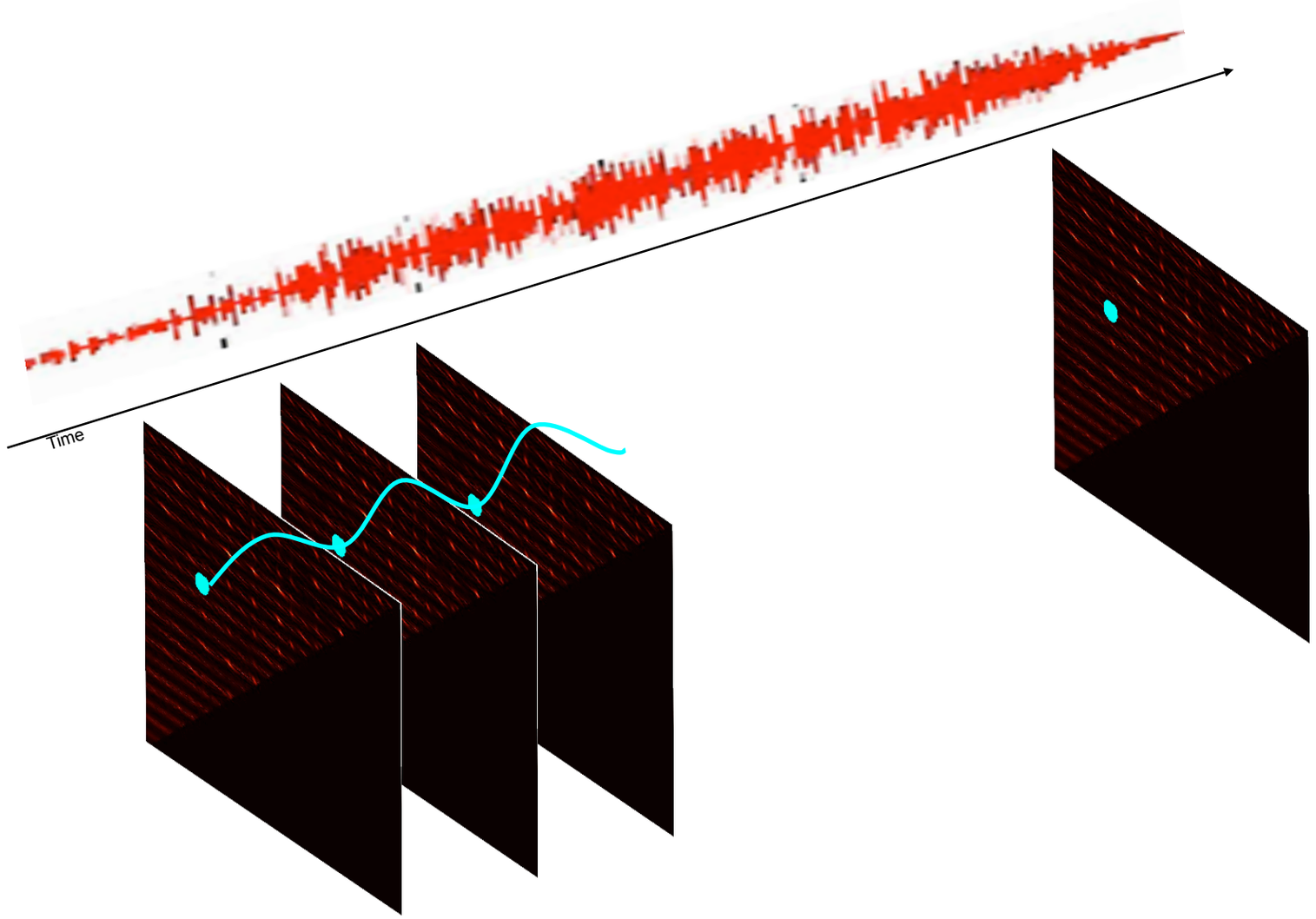


Metrical Interpretation

- A metrical tree can be specified as a set of metrically related points on the APM
- Search is thus done in space of meter and tempo



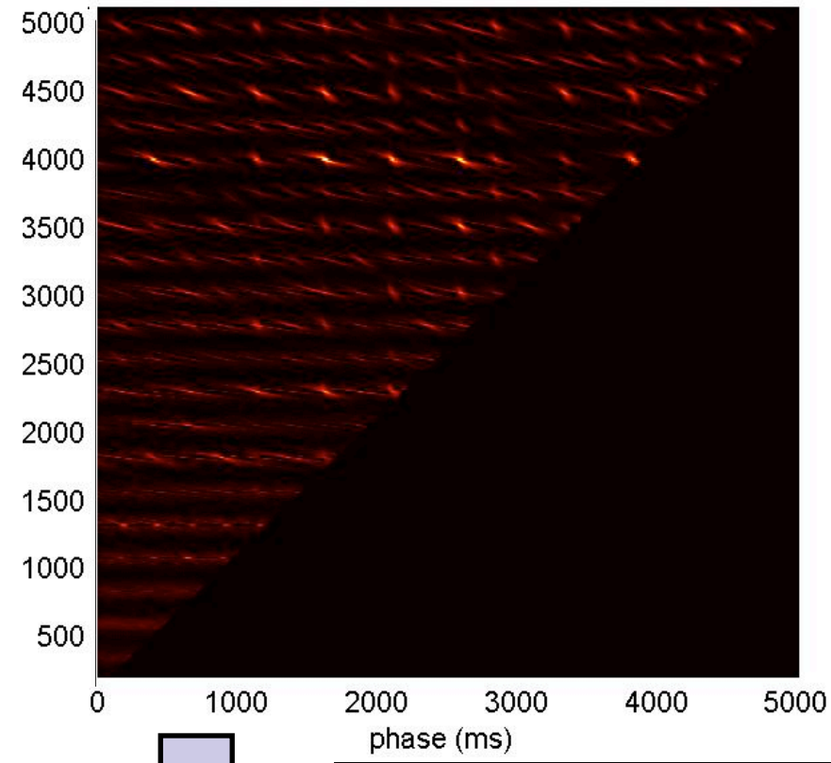
Finding beat and meter



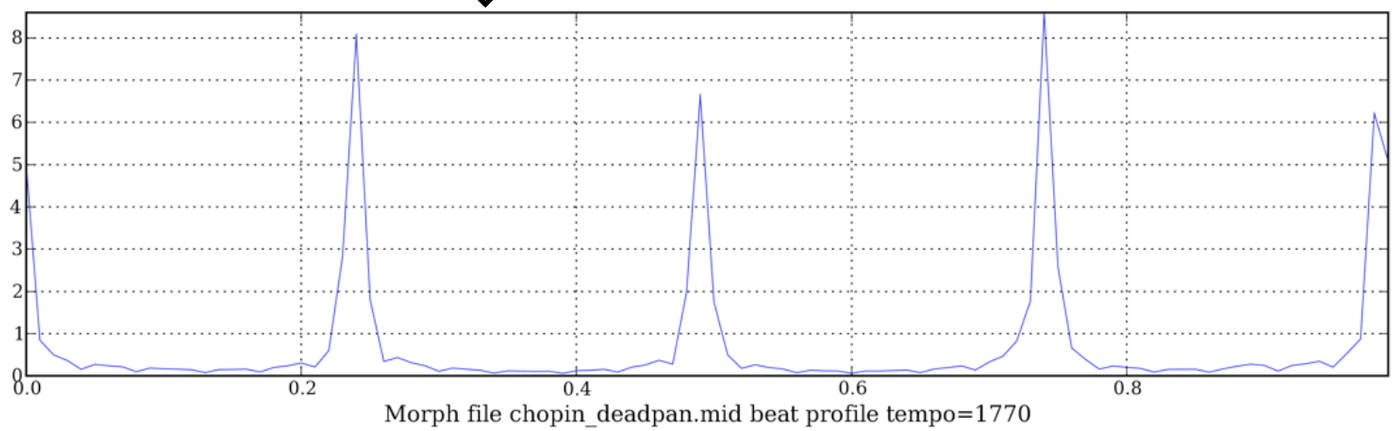
- Search is done through the space of metrical trees using Viterbi alignment.
- Resulting metrical tree "contracts" and "expands" with changing tempo.
- Details in Eck (2007).

Expressive performance dynamics

- Use the APM to identify meter as it changes in time.
- Measure expressive dynamics and timing with respect to the APM
- Measurements made in milliseconds (time) but stored in radians (phase)
- Allows us to generalize to new pieces of music with different tempi and meter

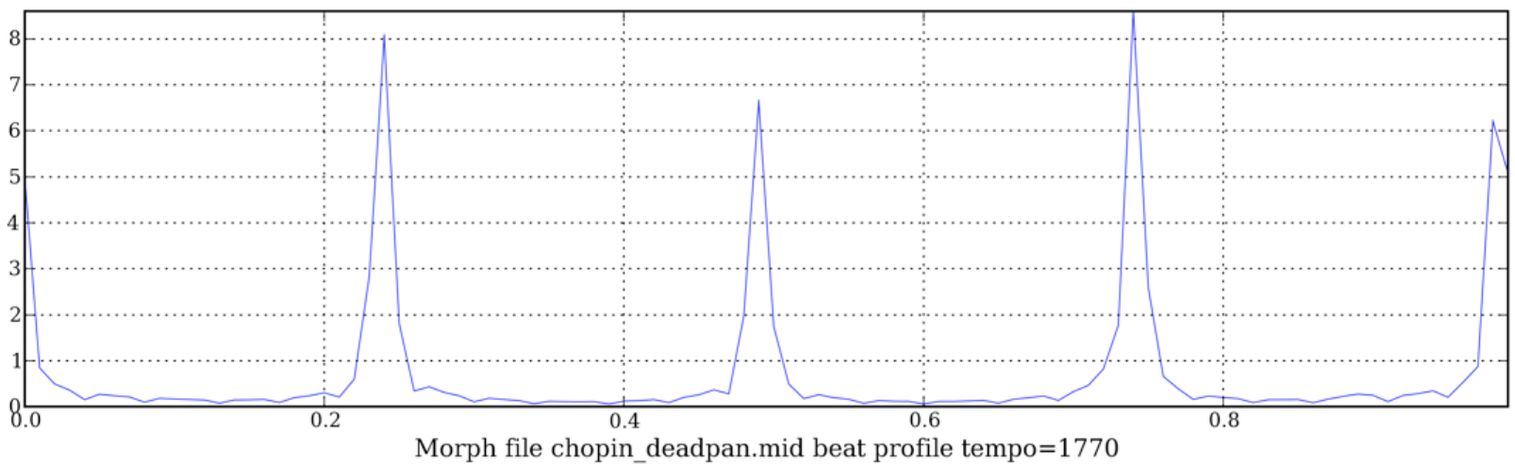


Integrate over time the winning metrical tree.



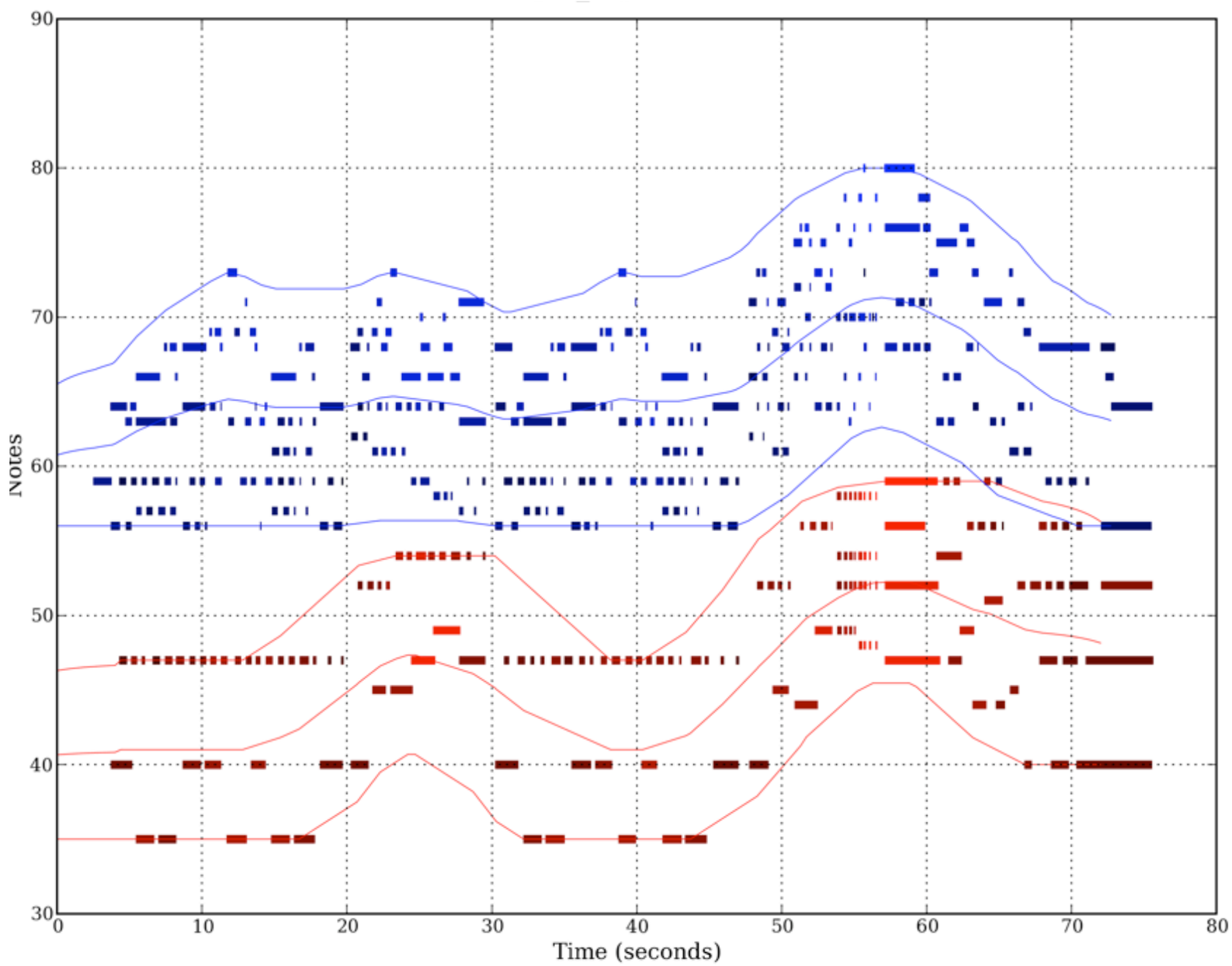
Modest example

- Morph the Chopin etude to sound a bit like me playing Heart and Soul after a couple of beers.
- Use hill climbing to find nearest maxima in target vector.
- Provides rudimentary measure-level perturbation only (preliminary and unrealistic).
- Timing, velocity, chord spread.

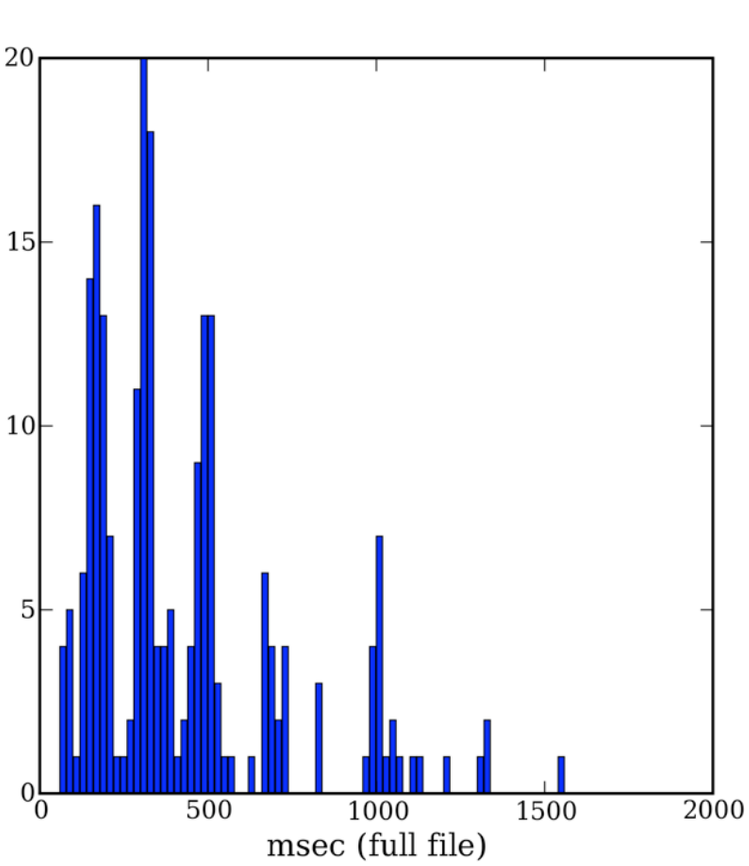


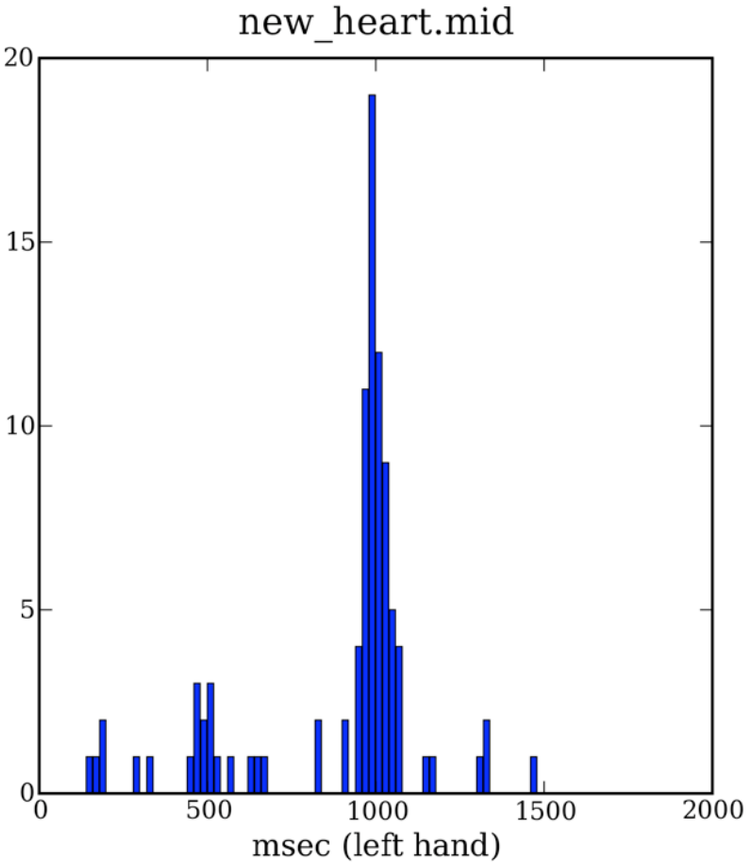
Collecting performance stats for the piano

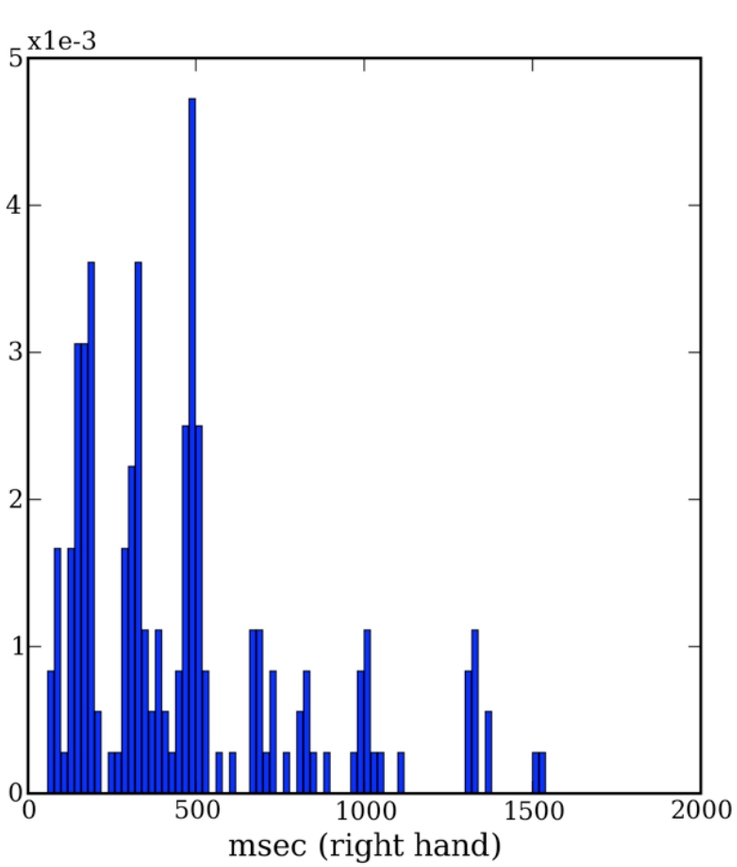
- For piano, identify hands using clustering
- Easier than finding leading melodic voice. No melodic analysis required
- Once hands are identified, identify chords
- Measure duration, velocity, legato,chord spread



Hand-specific statistics for piano







- Hands are somewhat rhythmically independent
- Measurements with respect to single hand are different than those for both hands (here: duration)

Conclusions

- Expressive timing and dynamics are important part of music
- Short overview of approaches
- Discussed task of score-free expressive performance
- Suggest using metrical structure as proxy for musical score

- Related this to APM model
- Future work:
  - There remains more future work than completed work!
  - So this list would be too long....
  - Thank you for your patience.

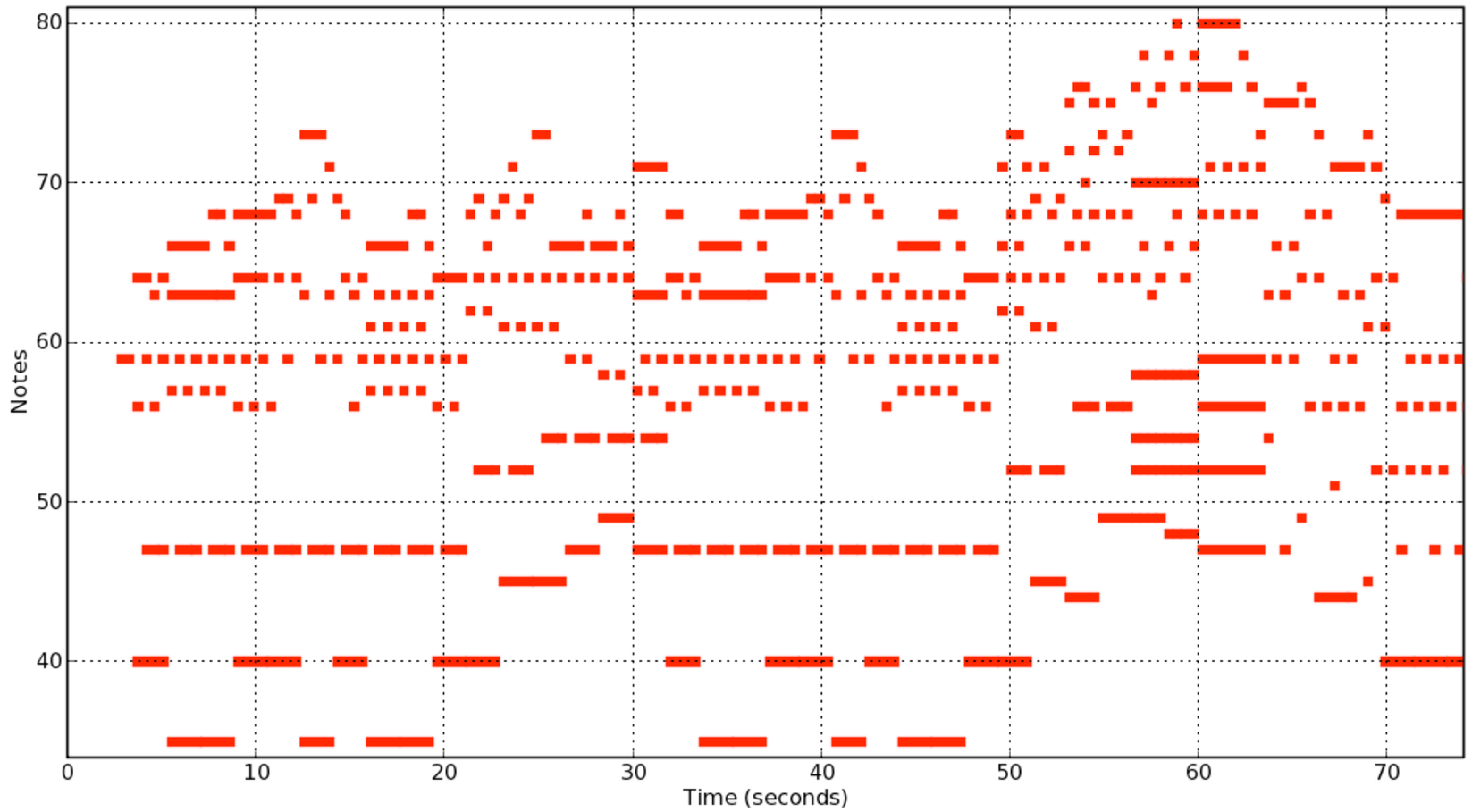
Bibliography

- 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*.
- L. Dorard, D. Hardoon & J. Shawe-Taylor (2007). Can Style be Learned? A Machine Learning Approach Towards 'Performing' as Famous Pianists. NIPS Music Brain Cognition Workshop, Whistler.
- 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)*.
- C. Palmer & P. Pfordresher (2003). Incremental planning in sequence production. *Psychological Review* 110:683-712.
- 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.

Following are deleted slides

Example: Chopin Etude Opus 10 No 3

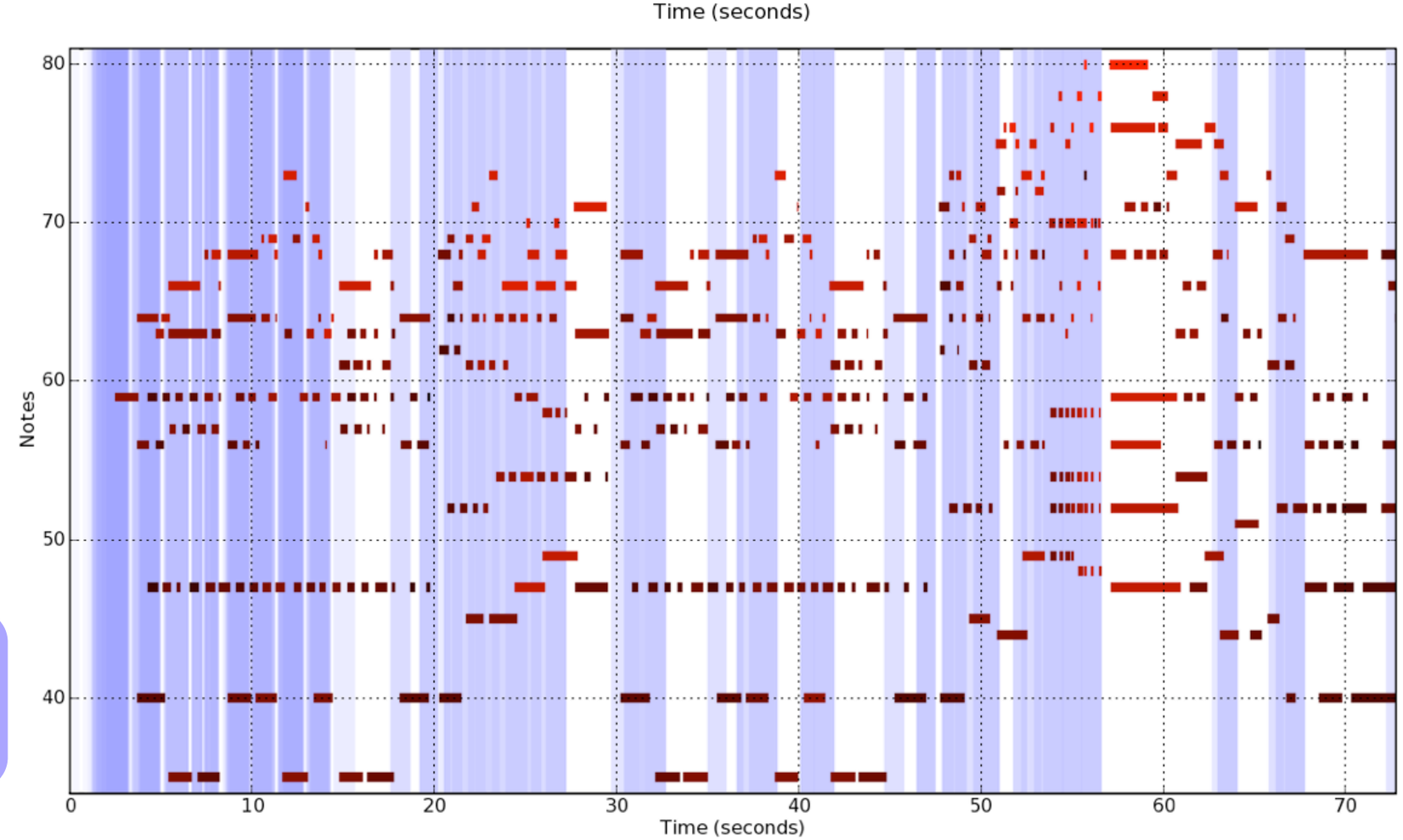
Deadpan
(no expressive timing or dynamics)



Human performance
(Recorded on Boesendorfer ZEUS)

Differences limited to:

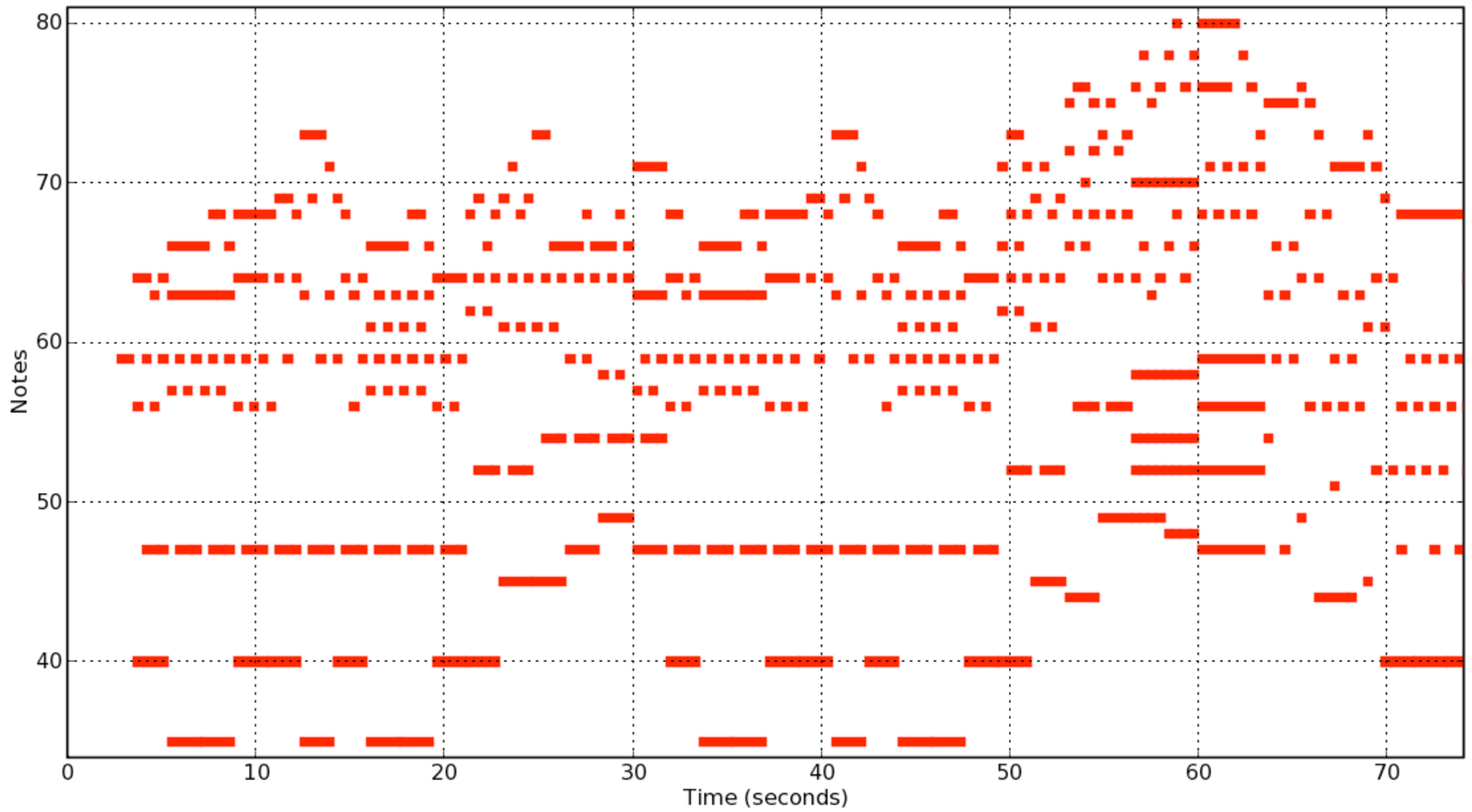
- timing (onset, length)
- velocity (seen as red)
- pedaling (blue shading)



Flat timing Flat velocity	Expressive timing Flat velocity	Expressive timing Expressive velocity

Example: Chopin Etude Opus 10 No 3

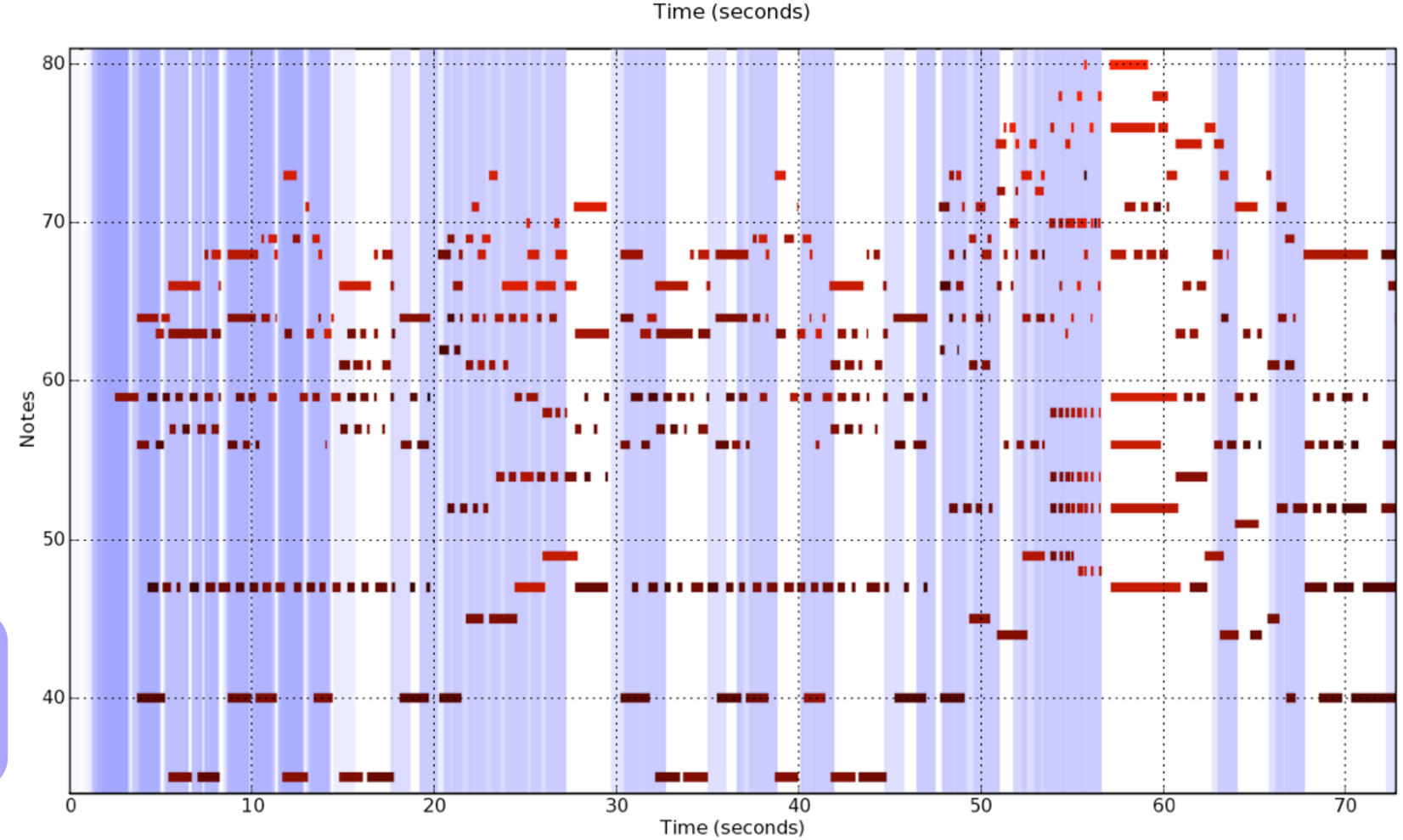
Deadpan
(no expressive timing or dynamics)



Human performance
(Recorded on Boesendorfer ZEUS)

Differences limited to:

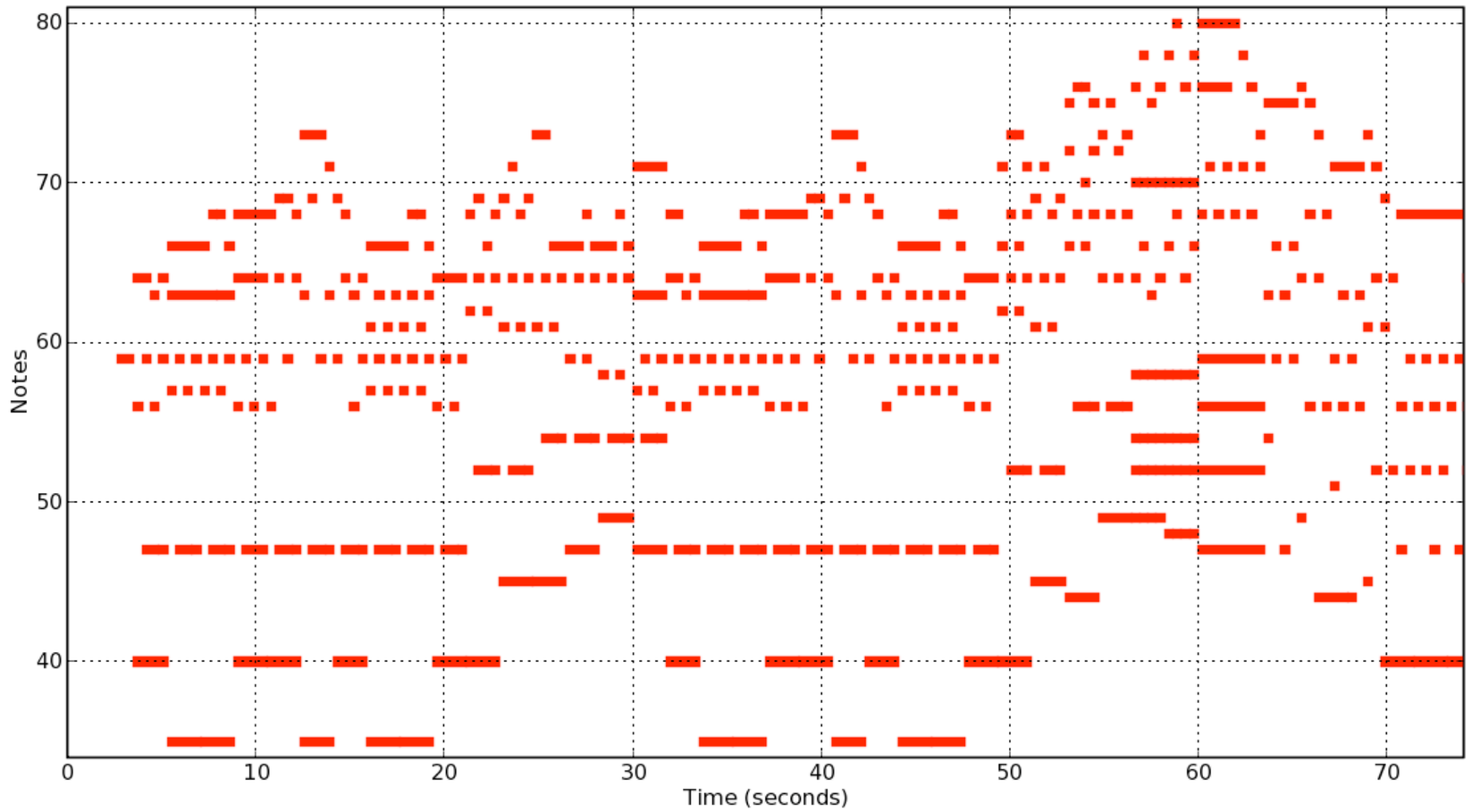
- timing (onset, length)
- velocity (seen as red)
- pedaling (blue shading)



Flat timing Flat velocity	Expressive timing Flat velocity	Expressive timing Expressive velocity

Example: Chopin Etude Opus 10 No 3

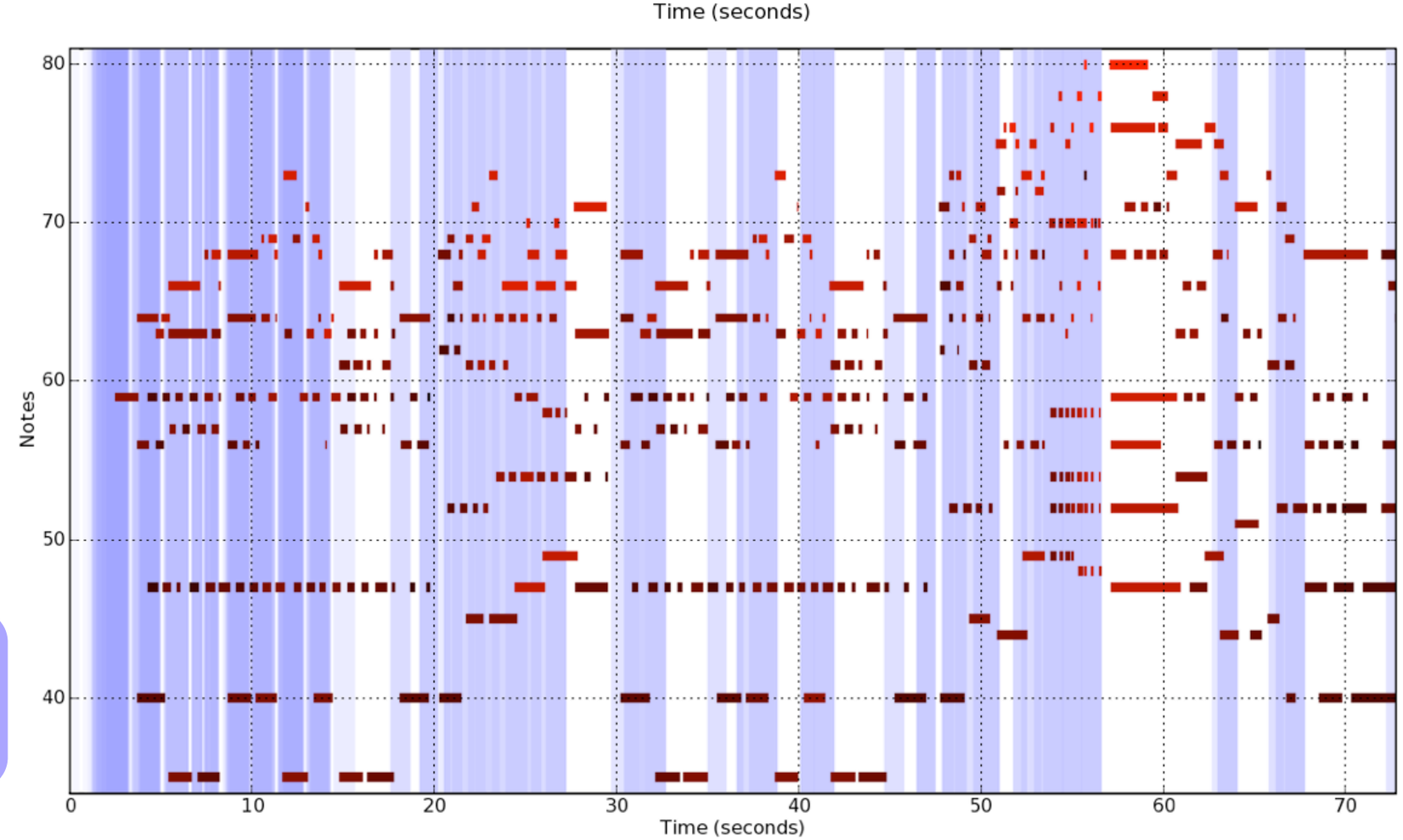
Deadpan
(no expressive timing or dynamics)



Human performance
(Recorded on Boesendorfer ZEUS)

Differences limited to:

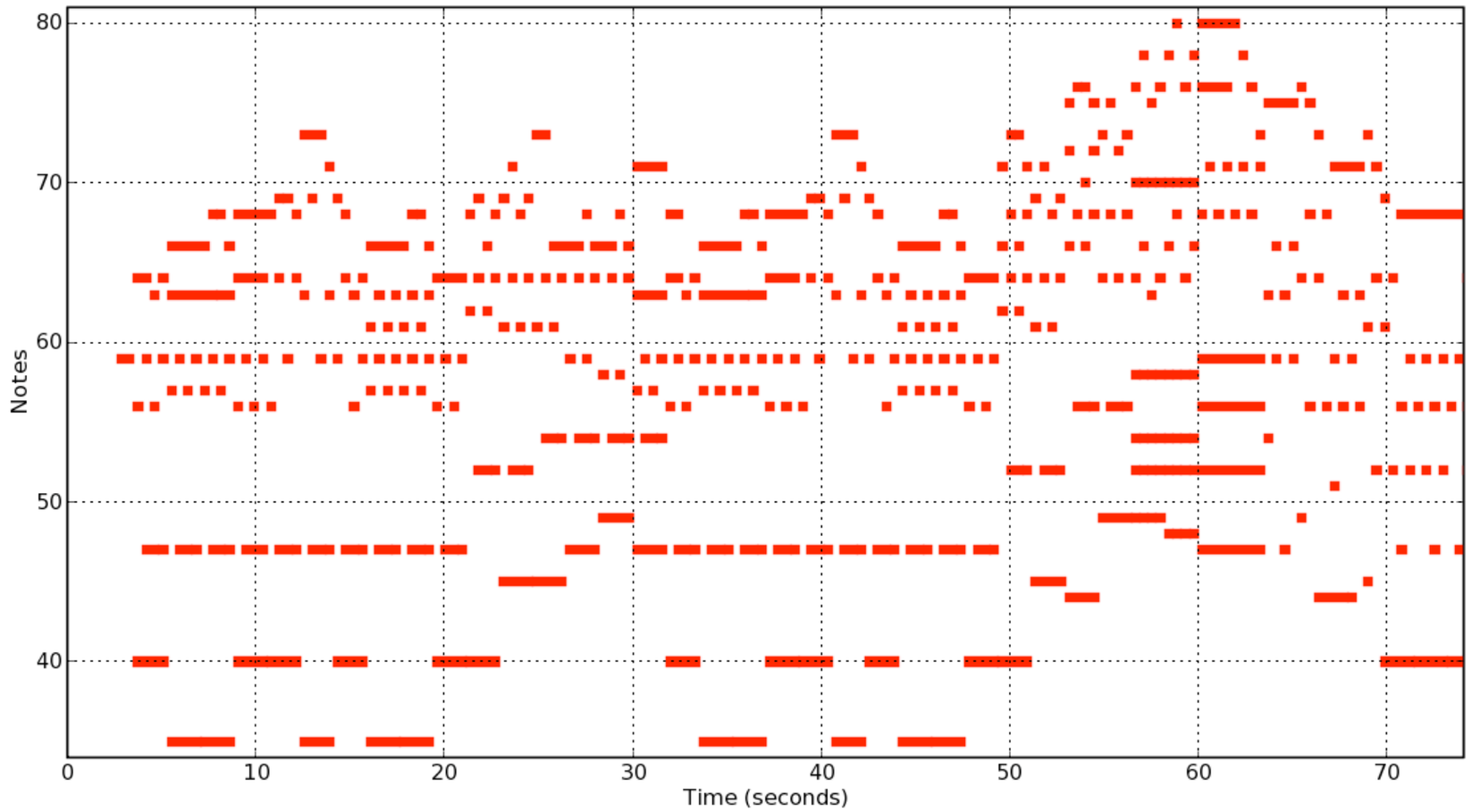
- timing (onset, length)
- velocity (seen as red)
- pedaling (blue shading)



Flat timing Flat velocity	Expressive timing Flat velocity	Expressive timing Expressive velocity

Example: Chopin Etude Opus 10 No 3

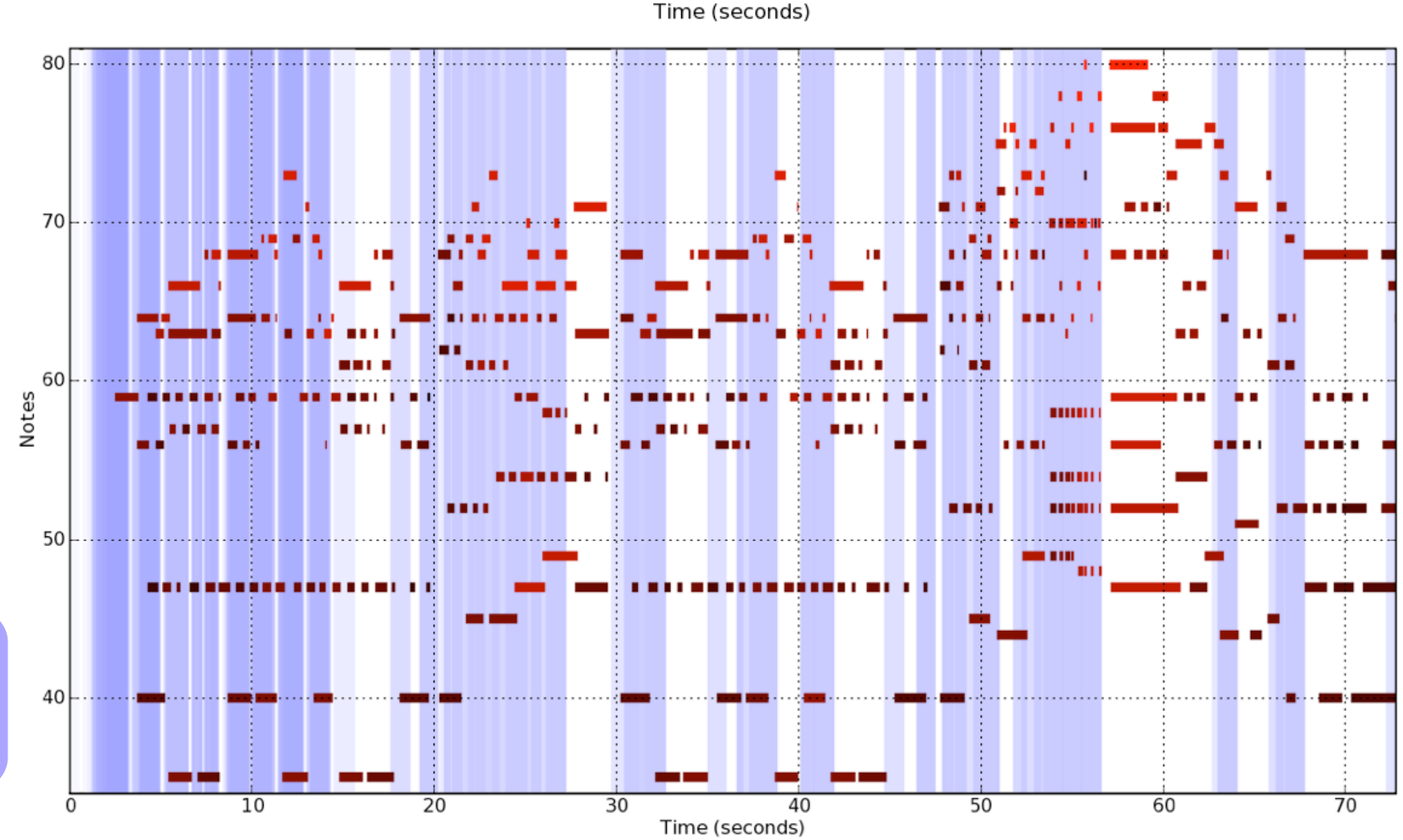
Deadpan
(no expressive timing or dynamics)



Human performance
(Recorded on Boesendorfer ZEUS)

Differences limited to:

- timing (onset, length)
- velocity (seen as red)
- pedaling (blue shading)



Flat timing Flat velocity	Expressive timing Flat velocity	Expressive timing Expressive velocity

Focus: musical meter

- Meter is the measurement of a musical line into measures of stressed and unstressed "beats", indicated in Western music notation by the time signature.

- Many methods for (imperfectly) estimating metrical structure in audio and MIDI

-