

Introducing the Mazurka Project

Nick Cook and Craig Stuart Sapp

Music Informatics research group seminar
School of Informatics, City University, London
22 February 2007

Introducing [∇] the Mazurka Project

the *Sensationalized* Story of

An exclusive exposé by
Nick Cook and Craig Stuart Sapp
about their role (or lack thereof) in the alleged Hatto Hoax.

Music Informatics research group seminar
School of Informatics, City University, London
22 February 2007

Source material: mazurka recordings

Mazurka in G minor Op. 24, No. 1

29 performances:

- 1,500+ recordings of 49 mazurkas
- = 30 performances/mazurka on average
- about 65 performers, 75 CDs

number of mazurka performances in each decade

248	Auldrey (1981)	Decca 443 191-2
346	Bart (1999)	Virgin 5 55619
239	Bloch (1995)	Erato 003 224507
204	Bruckner (1960)	Empire 28 28 4 3237
249	Chiu (1999)	BMG 249 1005 10
230	Chick (1994)	Polster 1552040703
250	Curtis (1951)	Crescent Arbel 910312
244	Faluy (1989)	Harmonia 3 356236
306	Furstenberg (1942)	Crescent Arbel 8204-6
301	Fleiss (1977)	MCA 66 0049
141	Francis (1996)	BMG CD2 74 741 1 2
312	Gatti (1997)	Crescent Arbel 927012
312	Gold (2001)	Decca 4551
237	Harari (1951)	RCA 09 06 4899-2
252	Lewis (1996)	DG 47 3554 2
311	Lindell (1994)	Crescent Arbel 9103
312	Magaloff (1977)	Philips 436 81329-2
245	Mass (2000)	BMG 60 2615 10
247	Polubinski (1999)	Decca 455 10112 13
308	Rosen (1989)	Decca 455 10112 13
203	Rubinstein (1939)	Star 8 110056-57
332	Rubinstein (1952)	BMG 60 2615 10
248	Rubinstein (1966)	BMG 60 2615 10
257	Shebanova (2002)	BMG 60 2615 10
345	Smith (1975)	BMG 60 2615 10
304	Tsing (1989)	Empire 28 28 4 3237
259	Tsing (2005)	BMG 60 2615 10
241	Trapp (2005)	BMG 60 2615 10
150	Trasky (1959)	Polster 1552040703

Performance-Correlation Timescapes

A multi-correlation timescape is like a photograph taken inside a forest.

The photographer is standing at the position of the performance being analyzed.

The colours are images of trees which are close by (large splotches of colour), or far away (small splotches of colour).

Black indicates that the performer tree is not particularly close to any other tree (but there could have been ones that died and have fallen to the ground).

Purpose of Timescapes

• To examine influences between performers

Perahia & Rubinstein?

- Rubinstein 1939
- Uninsky 1971
- Rubinstein 1966
- Indjic 1988
- Hatto 1997
- Shebanova 2002
- Magaloff 1977
- Horowitz 1985
- Horowitz 1971

Perahia 1994
Mazurka in A minor Op. 17, No. 4

for more: <http://mazurka.org.uk/ana/pcor>

Boring Timescape Pictures

Occasionally we get over-exposed photographs back from the store, and we usually have to throw them in the waste bin.

The same performance by Magaloff on two different CD re-releases:

Philips 456 898-2 mazurka 17/4 in A minor Philips 426 817/29-2

- Magaloff 1977
- Magaloff 1977b

• Structures at bottoms due to errors in beat extraction, measuring limits in beat extraction, and correlation graininess.

Boring Timescape Pictures?

Two difference performances from two different performers on two different record labels from two different countries.

Calliope 3321 mazurka 17/4 in A minor Concert Artist 20012

Indjic 1988 Hatto 1997

see: http://www.charm.rhul.ac.uk/content/contact/hatto_article.html

Beat-Event Timing Differences

Hatto beat location times: 0.853, 1.475, 2.049, 2.647, 3.278, etc.
 Indjic beat location times: 0.588, 1.208, 1.788, 2.408, 3.018, etc.

difference plot

Hatto / Indjic beat time deviations

deviation (seconds)

beat number

remove 0.7% timeshift

Pataphysical Analysis of Journalism

Reported:

"Conclusive? Perhaps not, as there is still a one in 1000 possibility that Hatto made her own recordings, but certainly troubling."

Actually:

Based on reasonable scientific assumptions, the probability that the Hatto performance of Mazurka in A minor Op. 17 No. 4 is not the same performance as that of Indjic is less than

$$1 \text{ in } \sqrt{\text{googol}}$$

This probability is equivalent to 1 atom out of all atoms in a typical star

[1] MFT

Slightly More Interesting

Same work; same pianist; different performance

Rubinstein 1939 Rubinstein 1952 Rubinstein 1966

24/2

Yet More Interesting

p s

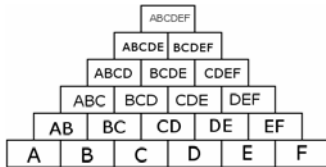
n t

The conspiracy goes deeper? or Not?

f c b

Timescape's Plotting Domain

- Examine the internal tempo/harmony structure of a performance/composition
- 1-D data sequences transformed into 2-D plot
- Example of a piece with 6 beats at tempos A, B, C, D, E, and F:

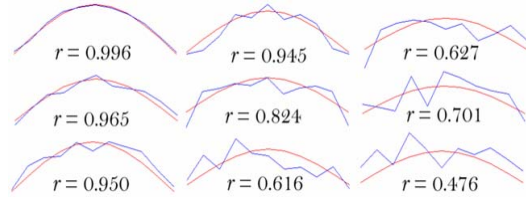


In performance correlation timescapes, correlation measurements calculated for each cell. Also tempo average timescapes can be calculated.

Under the Hood (Bonnet)

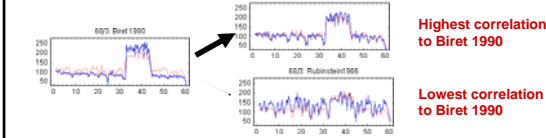
$$\text{Pearson correlation: } \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

- Measures how well two shapes match:
- $r = 1.0$ is an exact match.
- $r = 0.0$ means no relation at all.

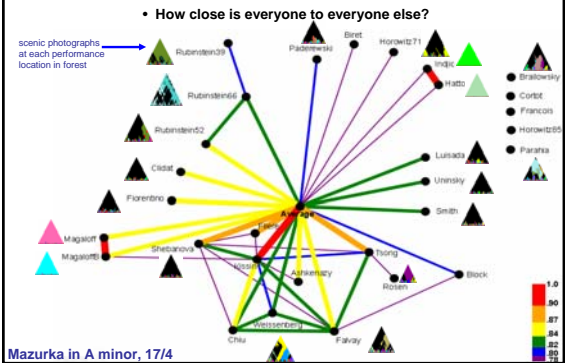


Overall performance correlations

	Bi	Br	Ch	Fl	In	Lu	R8	R6	Sm	Un
Biret	1.	0.92	0.81	0.83	0.95	0.85	0.62	0.5	0.55	0.86
Brailowsky	0.92	1.	0.81	0.86	0.91	0.84	0.66	0.55	0.65	0.85
Chiu	0.81	0.81	1.	0.86	0.86	0.81	0.76	0.74	0.67	0.89
Friere	0.83	0.86	0.86	1.	0.88	0.84	0.73	0.7	0.74	0.89
Indjic	0.95	0.91	0.86	0.88	1.	0.88	0.66	0.59	0.63	0.9
Luisada	0.85	0.84	0.81	0.84	0.88	1.	0.67	0.61	0.56	0.89
Rubinstein 1938	0.62	0.66	0.76	0.73	0.66	0.67	1.	0.77	0.62	0.75
Rubinstein 1966	0.5	0.55	0.74	0.7	0.59	0.61	0.77	1.	0.59	0.69
Smith	0.55	0.65	0.67	0.74	0.63	0.56	0.62	0.59	1.	0.64
Uninsky	0.86	0.85	0.89	0.89	0.9	0.89	0.75	0.69	0.64	1.

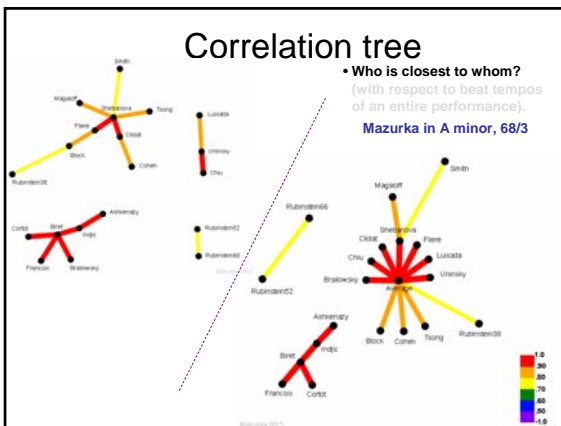


Map of the Forest: Correlation Network



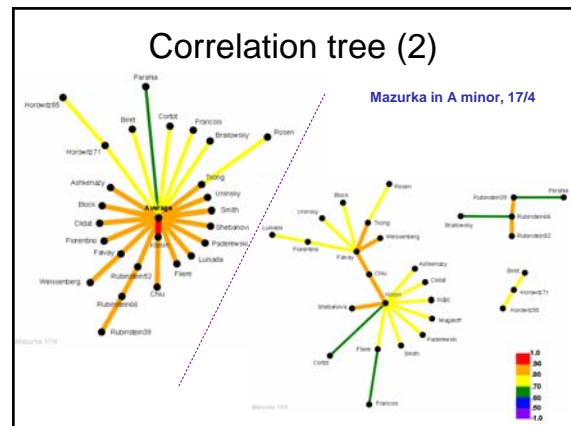
Correlation tree

- Who is closest to whom? (with respect to beat tempos of an entire performance).
- Mazurka in A minor, 68/3



Correlation tree (2)

Mazurka in A minor, 17/4



Performance data extraction

Reverse conducting

- Listen to recording and tap to beats.
- Tap times recorded in Sonic Visualiser by tapping on computer keyboard.

Align taps to beats

tempo by beat
loudness by beat

- Reverse conducting is real-time response of listener, not actions of performer.
- Adjust tap times to correct beat locations.
- A bit fuzzy when R/LH do not play in sync, or for tied notes.

Automatic feature extraction

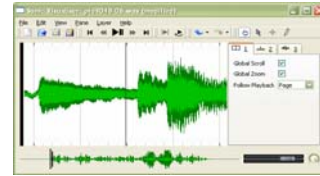
off-beat
timings

individual
note timings

individual note
loudnesses

Sonic Visualiser

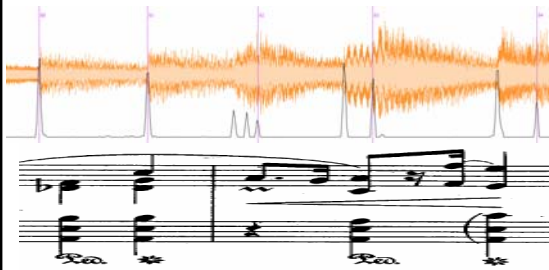
- Underlying data for timescapes extracted using Sonic Visualiser



Developed at the Center for Digital Music, Queen Mary, University of London

Main webpage for Sonic Visualiser: <http://www.sonicvisualiser.org>
Online documentation: <http://sv1.sourceforge.net/doc/reference/en>
Analysis plugin resources: <http://sv1.sourceforge.net/vamp.html>
Mazurka plugins: <http://sv.mazurka.org.uk>

Mz SpectralReflux plugin for SV



<http://sv.mazurka.org.uk/MzSpectralReflux>
<http://mazurka.org.uk/cgi-bin/tapsnap> = Move taps to nearest onset

Additional Slides

iTunes not Examining Musical Content

- iTunes does not examine musical content: only describes CD contents by track count and hash of track times.
- 1st CD identified by Brian Ventura using iTunes: Liszt's Transcendental Etudes, on or before 13 Feb 2007.
- Then 2nd CD of Rachmaninov Piano Concertos identified by Jed Distler on or before 15 Feb 2007 using iTunes.
- 4th CD identified of Brahms 2nd Piano Concerto identified by Steve de Mena on morning of 16 Feb 2007 or most likely before.
- 18 CD sources identified by morning of 21 February 2007 (methods vary)

references: <http://www.classicstoday.com/features/021807-joycehatto.asp>
<http://www.gramophone.co.uk/news/MainTemplate.asp?storyID=2759&newssectionID=1>
rec.music.classical.recording
http://en.wikipedia.org/Joyce_Hatto

iTunes / CDDB

<http://en.wikipedia.org/wiki/Cddb>

- Gramophone groups initial discovery was by using iTunes CDDB technology

[John Oswald w Gilles Arteau & Émile Morin / .Parcours Scénographique](#)

Tracks: 1
Total time: 60:10
Year: 1997
Disc-ID: classical / [020e1a01](#)

1. Parcours Scénographique

[Craig Sapp / Digital Whitenoise](#)

Tracks: 1
Total time: 60:10
Year: 1999
Disc-ID: blues / [020e1a01](#)

1. Whitenoise

Track/Time Hash is identical: **020e1a01**

first six digits are hash of track times last two digits are track count

discid problem usually resolved by placing in another category (classical v blues)

Purely Coincidental?

One of the major characters in the recent Hatto story is from this school:

Andrew Rose from Pristine Audio has a music degree from the **City University of London**, and spent 14 years as a sound engineer at BBC Radio in London.

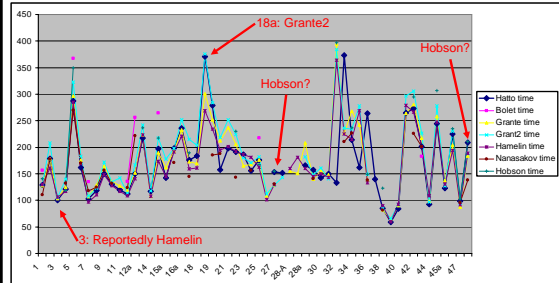
• 3rd CD identified by Gramophone group on or before 15 Feb 2007 was not found using iTunes. Andrew Rose of Pristine Audio verified first two results by expert analysis of audio content. He also discovered Godowsky/Chopin Studies to be a copy of a previous commercial release by Carlo Grante (Later it was reported by a contributor to rec.music.recording.classical that one of the tracks, #3, is by Marc Hamelin)

http://en.wikipedia.org/wiki/Leopold_Godowsky

These [Studies] are so taxing even for virtuosos that only five have ventured to record the entire set: Geoffrey Douglas Madge, Carlo Grante, Marc-André Hamelin and Joyce Hatto.

So: Godowsky is the best choice for a manual search for a match.

Godowsky/Chopin Study track times



• Study 18a is the most suspicious coincidence.