

Measuring Similarity in Performances of Chopin Mazurkas

Craig Stuart Sapp
CCRMA Colloquium; Stanford University
31 October 2007



Mazurka Project

AHRC Research Centre for the History and Analysis of Recorded Music

<http://mazurka.org.uk>

• 2732 recordings of 49 mazurkas by Frédéric Chopin (1810-1849)

= Average of 56 performances/mazurka

least: 39 performances of 41/3

most: 89 performances of 17/4

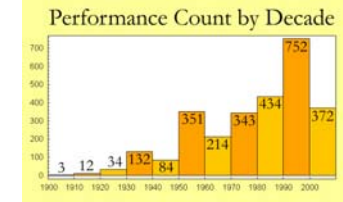
• 157 performers

• on 209 CDs/records

• 123 hours of music

• Earliest performance from 1902 by Alfred Grünfeld: mazurka 67/4

Masako Ezaki, 2006:



Many Performances of Same Composition

89 performances of mazurka 17/4

• how to compare and navigate through all the performances?



(measure 31)

Afanassiev 2001
Andnes 1990
Ashkenazy 1981
Bacha 1998
Barbosa 1983
Belavsky 2004
Ben-Or 1989

Biret 1990
Blet 2003
Block 1995
Brailowsky 1960
Brunhoff 1963
Casadeus 1930
Chiu 1999
Clidat 1994
Cohen 1997
Coop 1987
Cortot 1951
Csalog 1996
Czerny-Stefanska 1949
Czerny-Stefanska 1949
Czerny-Stefanska 1989
Ezaki 2006
Falvay 1989
Ferenczy 1958
Fiorentino 1990
Flügel 1977
Fou 1978
François 1956
Gieseking 1938
Ginzburg 1957
Goldmann 1997
Guller 1956
Hatto 1993

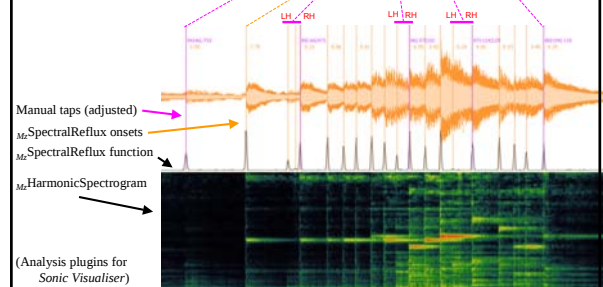
Hatto 2006
Horowitz 1971
Horowitz 1985
Indjic 1988
Kapell 1951
Kiepusa 1999
Kilenyi 1937
Kissin 1993
Kitain 1937
Kushner 1990
Lévy 1951
Lear 1994
Lefebvre 1950
Lilamand 2001
Luisada 1990
Lushtak 2004
Lympany 1968
Lympany 1990
Magaloff 1977
Magaloff 1977b
Magin 1975
Milikina 1970
Mohovich 1999
Nadelmann 1956
Ohlsson 1999
Olejniczak 1990
Olejniczak 1991

Osinska 1989
Pöntinen 2003
Paderewski 1912
Paderewski 1923
Paderewski 1924
Perahia 1994
Perlemuter 1986
Poblocka 1999
Rangell 2001
Risler 1920
Rosen 1989
Rubinstein 1939
Rubinstein 1952
Rubinstein 1966
Rummel 1943
Shebanova 2002
Simon 1991
Smith 1975
Szpliman 1948
Szczepanska 1959
Tanyel 1992
Uninsky 1971
Vardi 1988
Wasowski 1980
Weissenberg 1971
Zecchi 1942
Zecchi 1942b

Data Entry Example

(Mohovich 1999)

• pickup longer than 16th
• LH before RH
• triplets played same speed as sextuplets
• first sextuplet note longest



Difficult Data Extraction

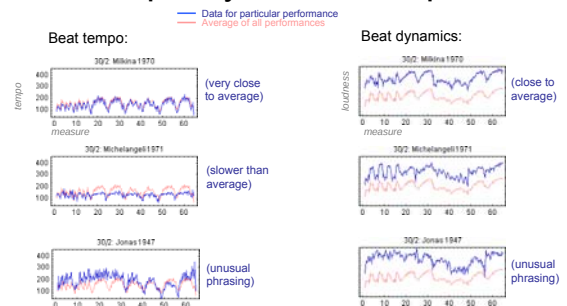
(Risler 1920)

• pickup longer than 16th
• LH before RH
• triplets played same speed as sextuplets
• first sextuplet note longest

lots of false positives
lots of noise/clicks

cc:ur|cyc

Tempo/Dynamics Graphs



• Data extracted for ~300 performances of ~6 mazurkas (~10% of collection)
• Using Sonic Visualiser / Vamp Plugins developed at C4DM, Queen Mary, U. of London
<http://www.sonicvisualiser.org> & <http://sv.mazurka.org.uk> (vamp plugins)

Similarity Measure

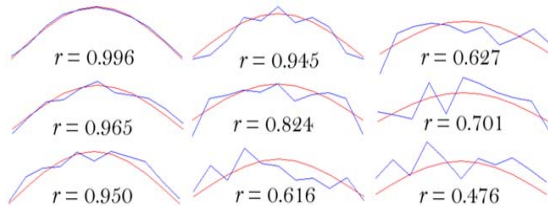
Pearson correlation:

output range: -1.0 to +1.0

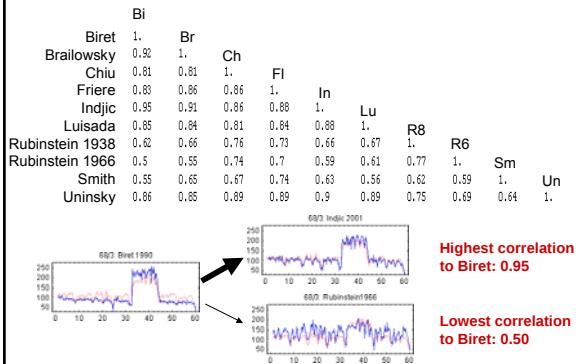
$$\frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$

• Measures how well two shapes match:

$r = +1.0$ is an exact match
 $r = 0.0$ means no relation
 $r = -1.0$ means upside-down

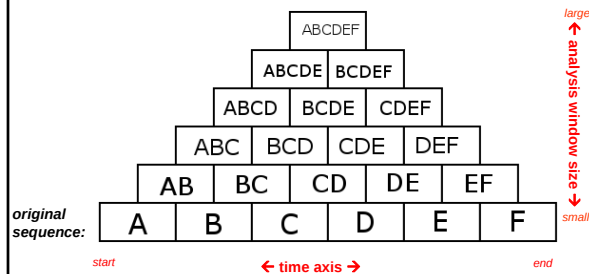


Correlation Examples



Scape Plotting Domain

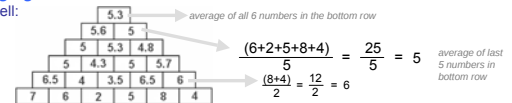
- 1-D data sequence chopped up into all possible n -grams to form a 2-D plot
- Example of a composition with 6 beats at tempos A, B, C, D, E, and F:



Scape Plotting Range

- **Any** operation can be applied to each cell in plotting domain:

- Example using **averaging** in each cell:



Comparative Timescapes

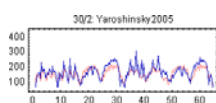
- **Cell operation:** Index of performance which has highest correlation to reference performance (excluding reference).

$$\text{cell}_{ij} = \max_{k \neq r} \text{Correlation}(\text{performance}_{ijr}, \text{performance}_{ijk})$$

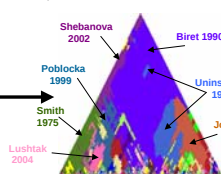
- Each performance assigned unique (arbitrary) index color:

- performer 1
- performer 2
- performer 3
- etc.

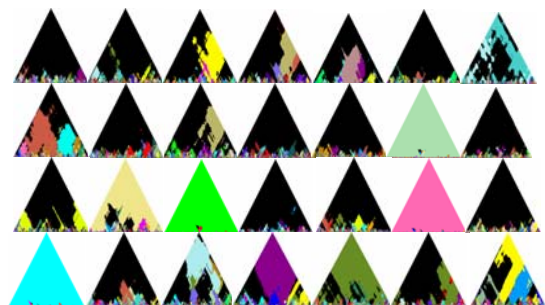
- Example reference performance:



Resulting plot shows best correlated performance at all timescales:



Performance Correlation Scares



mazurka.org.uk/anal/pcor (tempo) mazurka.org.uk/anal/pcor-gbdyn (dynamics)

<http://mazurka.org.uk/anal/pcor-joint> (tempo+dynamics)

<http://mazurka.org.uk/anal/pcor-all> (all features displayed)

Calliope 3321

Concert Artist 20012

Indjic 1988

Hatto 1997

~ 100 CDs of Joyce Hatto performances issued on the Concert Artist label in 2003-2006. (also 70 cassettes in 80's and 90's)

Indjic mazurka performances first borrowed for a Hatto cassette release in 1993.

- Origins of ~65 CDs on other commercial recordings have been identified (23 in the first week after story broke)

http://en.wikipedia.org/wiki/Joyce_Hatto
<http://www.farhanmalik.com/hatto/cdlist.html>

- Borrowed performances from at least 70 pianists

<http://www.farhanmalik.com/hatto/pianistslist.html>

LETTER FROM ENGLAND

FANTASIA FOR PIANO

Joyce Hatto's incredible talent

BY MARK SINGER

September 17, 2007

The New Yorker

Anniversary

Date: Wed, 1 Nov 2006 17:01:56 +0000
 From: "Craig Sapp" <craig@sapp@gmail.com>
 To: "Nicholas Cook" <nicholas.cook@rhul.ac.uk>,
 "Georgia Volioti" <georgia_volioti@hotmail.com>,
 "Daniel Leech-Wilkinson" <daniel.leech-wilkinson@kcl.ac.uk>
 Subject: Mazurka 66-3 notes

Hi Everyone,

I make new plots of the correlation timescapes:
<http://mazurka.org.uk/ana/pcor/mazurka66-3>

The most interesting pair are Hatto 1997 and Indjic 2001. They are both very close to each other – in fact, impossibly close. I double-checked and made sure that I did not make a mistake, and seem to not have done so. Those two recordings are exactly the same, with Hatto playing 3% slower than Indjic and with a different starting time offset in the recording.

The Hatto CD recordings are from 1997 and 2005 (with no indication of which is which in the liner notes), so you can't go by the date as to which is the original. I suspect that Hatto is a copy of Indjic's performance, particularly since it was released this past year.

Since they are from different labels (Indjic = Calliope 3321 and Hatto = Concert Artist 92712), we might be making big money testifying for the winning side in a court case. :-)

--+Craig

Fiorentino Fakes

Sergio Fiorentino

Con. Artist: CACD 9200-2 (2003)

Mazurka 6/2
 Mazurka 7/1
 Mazurka 7/2
 Mazurka 7/3
 Mazurka 7/4
 Mazurka 7/5
 Mazurka 17/2
 Mazurka 17/4
 Mazurka 24/1
 Mazurka 24/2
 Mazurka 24/3
 Mazurka 24/4
 Mazurka 30/1
 Mazurka 30/2
 Mazurka 30/3
 Mazurka 30/4
 Mazurka 33/1
 Mazurka 33/3
 Mazurka 33/4
 Mazurka 50/2
 Mazurka 50/3
 Mazurka 59/3
 Mazurka 59/1
 Mazurka 59/2
 Mazurka 59/3
 Mazurka 63/2

Janusz Olejniczak
 Naive/OPUS 111:
 OP20002 (1991)

Idil Biret
 Naxos 8.550358
 (1990)

Anna Malikova
 Classical Records
 CR-053 (1999)

Same Performer over Time

• Timescapes were developed to examine "soft plagiarism" in performances
 -- more interesting than Con. Artist's copyright infringement

Rub52
 Rub66
 Rub39
 Rub66
 Rub52
 (Rub39)

Rubinstein 1939
 57.4% Rubinstein 1952
 20.3% Rubinstein 1966
 2.4% Zak 1937
 1.9% Miliana 1970
 1.9% Lutsak 1991
 1.8% Rachmaninoff 1923
 1.2% Czerny 1949
 1.2% Athanasazy 1981

Rubinstein 1952
 37.5% Rubinstein 1966
 35.2% Rubinstein 1939
 4.3% Falvey 1989
 2.5% Poblecka 1999
 2.5% Rachmaninoff 1923
 2.2% Mokrych 1999
 1.6% Kapell 1951
 1.4% Afanasiev 2001
 1.2% Shevtsova 2002
 1.2% Miliana 1970

Rubinstein 1966
 59.3% Rubinstein 1952
 8.3% Kuchner 1989
 8.2% Rubinstein 1939
 3.2% Ozuska 1989
 2.2% Kapell 1951
 2.1% Hatto 1988
 1.8% Chou 1999
 1.5% Miliana 1970
 1.3% Zak 1937
 1.1% Falvey 1989
 1.1% Lutsak 2004

(Mazurka 63/3, 60 performances)

Strong Interpretive Influences

Mazurka in C Major 24/2 comparing 30 performances

Poblecka 1999
 Nezu 2005

• Nezu did graduate studies with Poblecka

Purely Random Matches

• Small blotches, inverted triangles & broken borders = poor matches

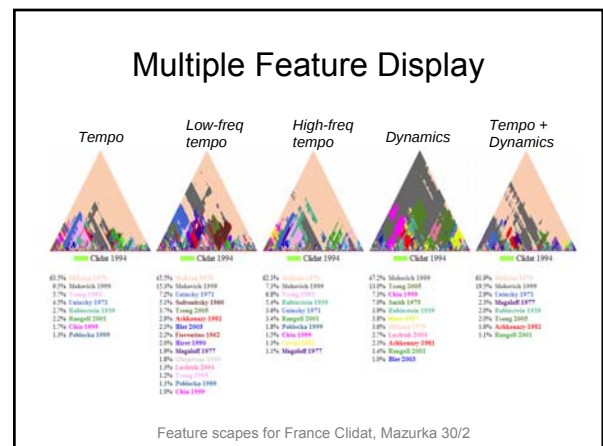
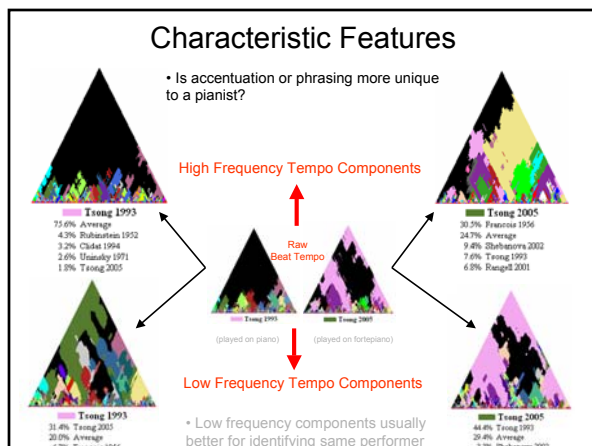
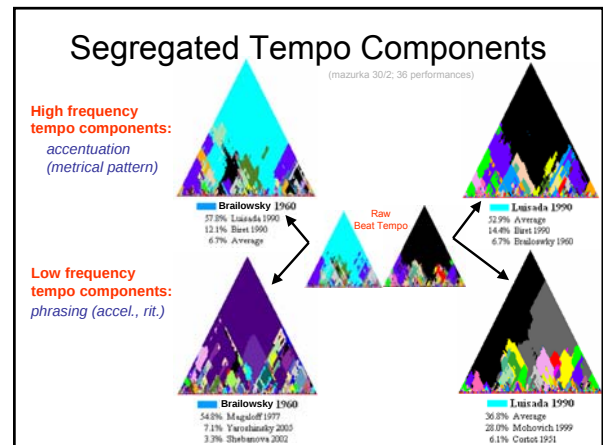
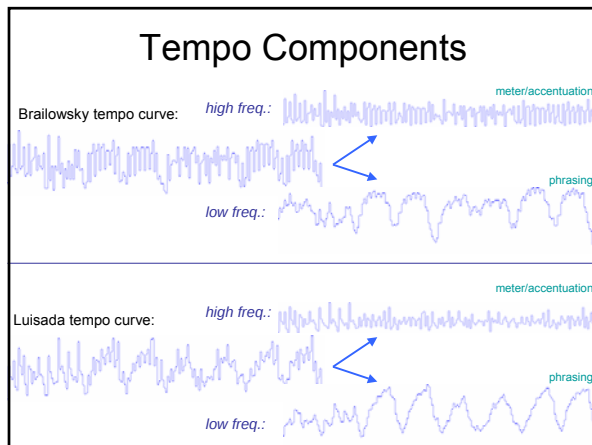
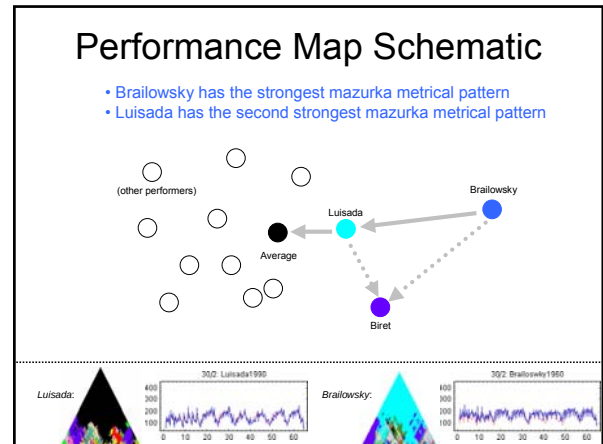
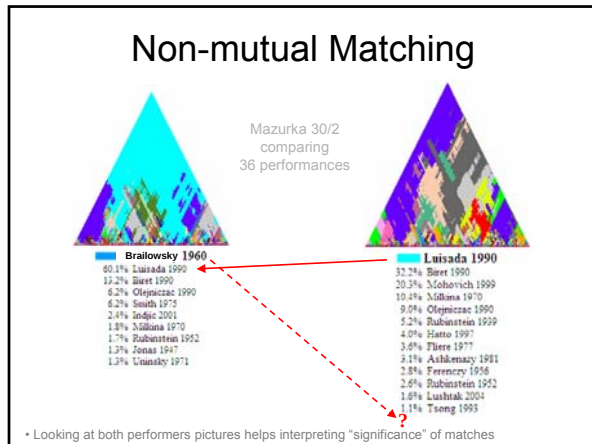
two white-noise sequences:

Uninsky 1971
 22.1% Tene 1945
 14.1% Rubinstein 2002
 13.1% Rubinstein 1966
 12.1% Rubinstein 1939
 11.1% Rubinstein 1952
 10.1% Rubinstein 1966
 9.1% Rubinstein 1939
 8.1% Rubinstein 1952
 7.1% Rubinstein 1966
 6.1% Rubinstein 1939
 5.1% Rubinstein 1952
 4.1% Rubinstein 1966
 3.1% Rubinstein 1939
 2.1% Rubinstein 1952
 1.1% Rubinstein 1966
 0.1% Rubinstein 1939

Random
 22.1% Tene 1945
 14.1% Rubinstein 2002
 13.1% Rubinstein 1966
 12.1% Rubinstein 1939
 11.1% Rubinstein 1952
 10.1% Rubinstein 1966
 9.1% Rubinstein 1939
 8.1% Rubinstein 1952
 7.1% Rubinstein 1966
 6.1% Rubinstein 1939
 5.1% Rubinstein 1952
 4.1% Rubinstein 1966
 3.1% Rubinstein 1939
 2.1% Rubinstein 1952
 1.1% Rubinstein 1966
 0.1% Rubinstein 1939

• Many performances are equidistant to Uninsky performance, none probably particularly similar to his performance.

(mazurka 30/2: 36 performances)



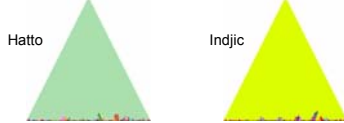
Qualitative/Quantitative

- Scape plots show the best matches, not necessarily good matches

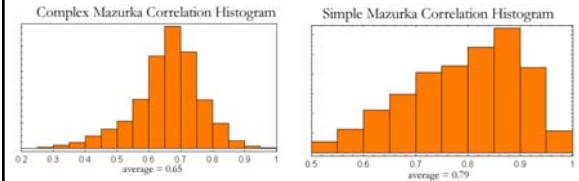
- Brailowsky not really a "good" match to Luisada in 30/2:



- Also sensitive to the "Hatto Effect" (i.e.: who is Indjic really similar to?)



Expected Correlation Values



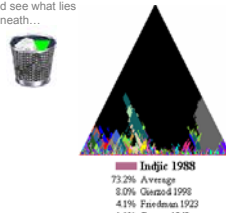
- Different mazurkas have different correlation value distributions
- Simpler/shorter mazurkas have higher average correlations
- Complex/longer mazurkas have lower average correlations

Scape Layers

- Hatto takes up nearly 100% of Indjic's scape

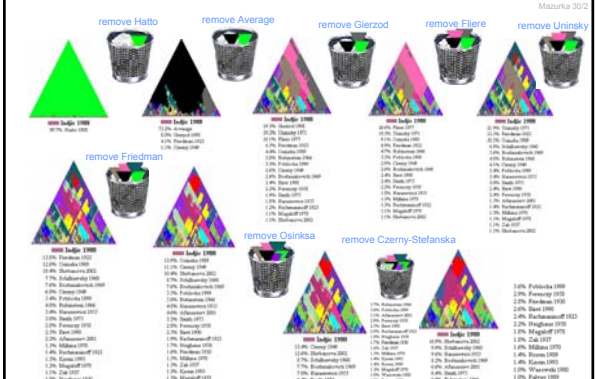


- So remove Hatto layer and see what lies beneath...



Mazurka 30/2

Peeling Layers



"Peeled" Rankings

Reference: Indjic 1988
(Mazurka 24/2, no avg.)

	Rank	Score
Hatto 1997	1	0.96
Tsujii 2005	2	0.26
Malcuzyński1951	3	0.36
Nezu 2005	4	0.33
Falvay 1989	5	0.32
Lushtak 2004	6	0.27
Poblocka 1999	7	0.21
Uninsky 1959	8	0.17
Rangell 2001	9	0.17
Florentino 1962	10	0.21
Shebanova 2002	11	0.19
Tsong 1993	12	0.21
Ashkenazy 1981	13	0.20
Biret 1990	14	0.12
Rubinstein 1939	15	0.14
Fliere 1977	16	0.11
Clidat 1994	17	0.13
Rubinstein 1966	18	0.10
Rubinstein 1952	19	0.16
Luisada 1990	20	0.14
Smith 1975	21	0.12
Rosen 1989	22	0.10
Brailowsky 1960	23	0.10
Chiu 1999	24	0.12
Francois 1956	25	0.13
Nageloff 1977	26	0.11
Perlemuter 1992	27	0.11
Cortot 1951	28	0.09
Random 1	29	0.05
Random 2	30	0.03

Area coverage in
scape plot *
(total-rank) / total

Noise Floor

Reference: Indjic 1988
(Mazurka 24/2; no avg.)

	Rank	Score
Hatto 1997	1	0.96
Tsujii 2005	2	0.26
Malcuzyński1951	3	0.36
Nezu 2005	4	0.33
Falvay 1989	5	0.32
Lushtak 2004	6	0.27
Poblocka 1999	7	0.21
Uninsky 1959	8	0.17
Rangell 2001	9	0.17
Florentino 1962	10	0.21
Shebanova 2002	11	0.19
Tsong 1993	12	0.21
Ashkenazy 1981	13	0.20
Biret 1990	14	0.12
Rubinstein 1939	15	0.14
Fliere 1977	16	0.11
Clidat 1994	17	0.13
Rubinstein 1966	18	0.10
Rubinstein 1952	19	0.16
Luisada 1990	20	0.14
Smith 1975	21	0.12
Rosen 1989	22	0.10
Brailowsky 1960	23	0.10
Chiu 1999	24	0.12
Francois 1956	25	0.13
Nageloff 1977	26	0.11
Perlemuter 1992	27	0.11
Cortot 1951	28	0.09
Random 1	29	0.05
Random 2	30	0.03

Noise Floor

Recalculation of Top Rankings

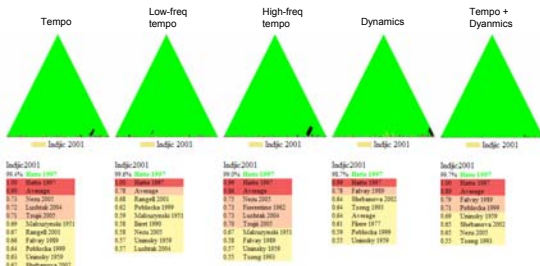
Reference:	2-Rank 2-Score		3-Rank 3-Score		
Indjic 1988					
(Mazurka 24/2; no avg.)					
Hatto 1997	1	0.96	Hatto 1997	1	1.00
Tsujii 2005	2	0.26	Tsujii 2005	2	0.76
Malczukowski1951	3	0.36	Malczukowski1951	5	0.67
Nezu 2005	4	0.33	Nezu 2005	4	0.68
Falvey 1989	5	0.32	Falvey 1989	3	0.70
Lushtak 2004	6	0.27	Lushtak 2004	6	0.64
Poblocka 1999	7	0.21	Poblocka 1999	8	0.59
Uninsky 1959	8	0.17	Uninsky 1959	7	0.61
Rangelli 2001	9	0.17	Rangelli 2001	9	0.51
Florentino 1962	10	0.21	Florentino 1962	11	0.46
Shebanova 2002	11	0.19	Shebanova 2002	10	0.50
Tsong 1993	12	0.21	Tsong 1993	12	0.43
Ashkenazy 1981	13	0.20	Ashkenazy 1981	13	0.36
Biret 1990	14	0.12	Biret 1990	14	0.22
Rubinstein 1939	15	0.14	Rubinstein 1939	17	0.14
Noise Floor					
	Fliere 1977	16	0.11	23	
	Clidat 1994	17	0.13	18	
	Rubinstein 1966	18	0.10	26	
	Rubinstein 1952	19	0.16	15	
	Luisada 1990	20	0.14	16	
	Smith 1975	21	0.12	21	
	Rosen 1989	22	0.10	25	
	Brailowsky 1960	23	0.10	27	
	Chiu 1999	24	0.12	20	
	Francis 1956	25	0.13	19	
	Magaloff 1977	26	0.11	22	
	Perlemuter 1992	27	0.11	24	
	Cortot 1951	28	0.09	28	
	Random 1	29	0.05	29	
	Random 2	30	0.03	30	

Final Similarity Metric

- 3-Scores are not symmetric, so take geometric mean of forward and reverse scores gives better global similarity measure:

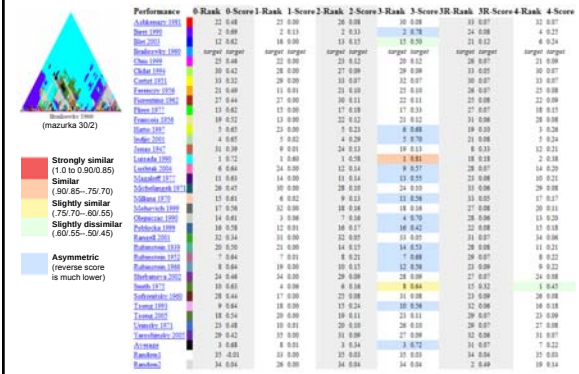
Reference:	3-Rank	3-Score	3R-Rank	3R-Score	4-Rank	4-Score
Indjic 1988						
(Mazurka 24/2; no avg.)						
Performance						
Ashkenazy 1981	13	0.36	7	0.5	13	0.42
Biret 1990	14	0.22	2	0.74	14	0.4
Brailowsky 1960	27	0.1	17	0.14	26	0.12
Chiu 1999	20	0.12	8	0.48	17	0.24
Clidat 1994	18	0.13	10	0.35	22	0.21
Cortot 1951	28	0.09	6	0.63	18	0.24
Falvey 1989	3	0.7	4	0.62	5	0.66
Florentino 1962	11	0.46	4	0.66	11	0.55
Filieri 1977	23	0.11	11	0.32	23	0.19
Francis 1956	19	0.13	10	0.15	25	0.14
Hatto 1997	1	1	1	1	1	1
Luisada 1990	16	0.14	6	0.74	15	0.32
Lushtak 2004	6	0.64	2	0.82	3	0.72
Malczukowski 1951	5	0.67	3	0.71	5	0.69
Magaloff 1977	22	0.11	24	0.1	28	0.1
Nezu 2005	4	0.68	4	0.79	2	0.73
Perlemuter 1992	24	0.11	4	0.49	19	0.23
Poblocka 1999	8	0.59	4	0.7	8	0.64
Rangelli 2001	9	0.51	2	0.87	6	0.67
Rosen 1989	25	0.1	20	0.12	27	0.11
Rubinstein 1939	17	0.14	10	0.37	20	0.23
Rubinstein 1952	15	0.16	8	0.31	21	0.22
Rubinstein 1966	26	0.1	10	0.3	24	0.17
Shebanova 2002	10	0.5	5	0.76	10	0.62
Smith 1975	21	0.12	6	0.52	16	0.25
Tsong 1993	12	0.43	5	0.57	12	0.45
Tsujii 2005	2	0.76	4	0.67	4	0.71
Uninsky 1959	7	0.63	2	0.64	9	0.63
Random 1	29	0.05	29	0.06	29	0.06
Random 2	30	0.03	28	0.04	30	0.03

Indjic Similarities by Feature

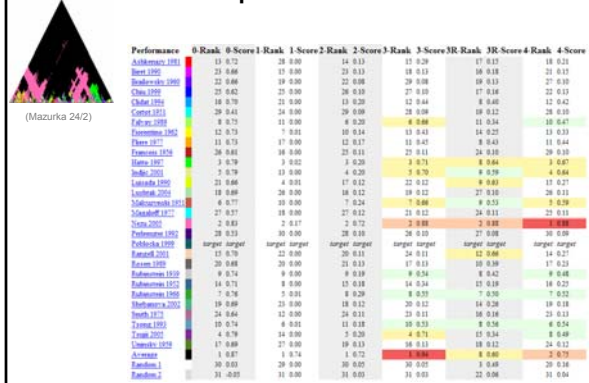


Mazurka 24/2

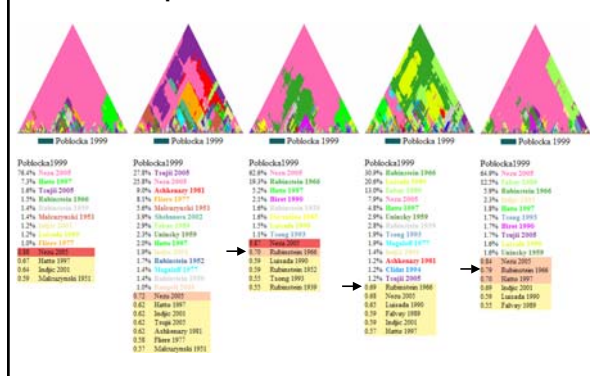
Example: Brailowsky



Example: Poblocka



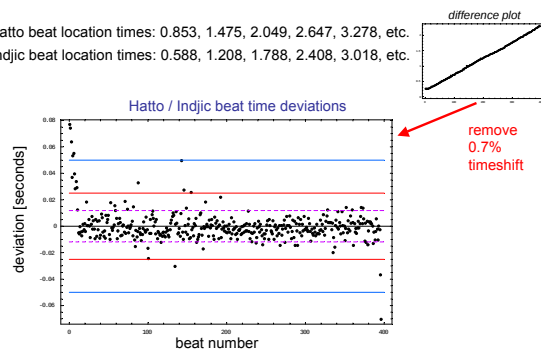
Example: Poblocka Features



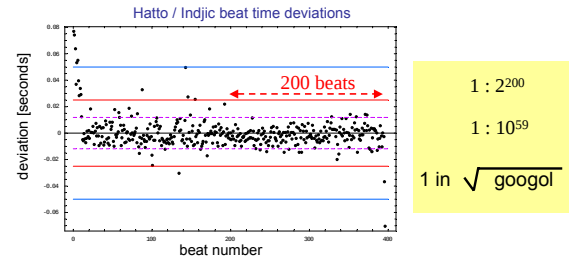
Extra Slides

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.

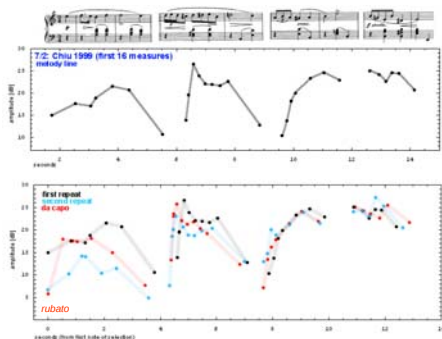


Timing Difference Probability

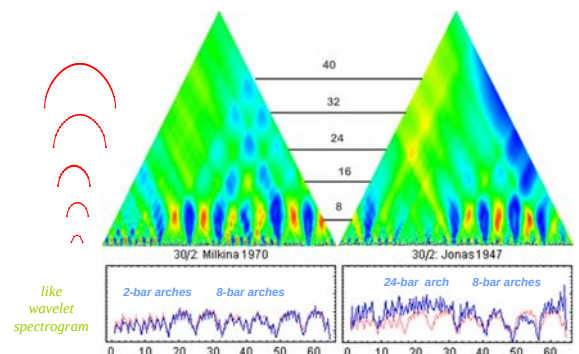


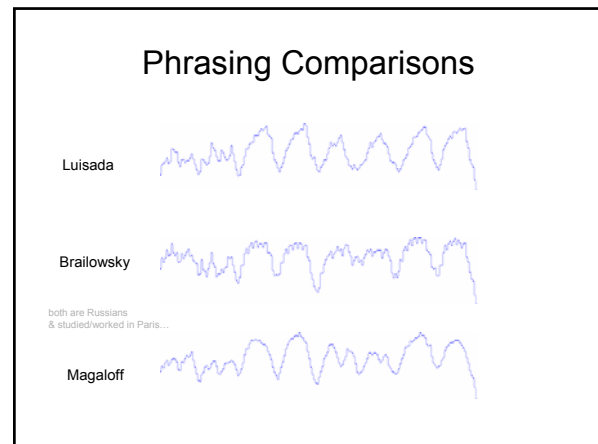
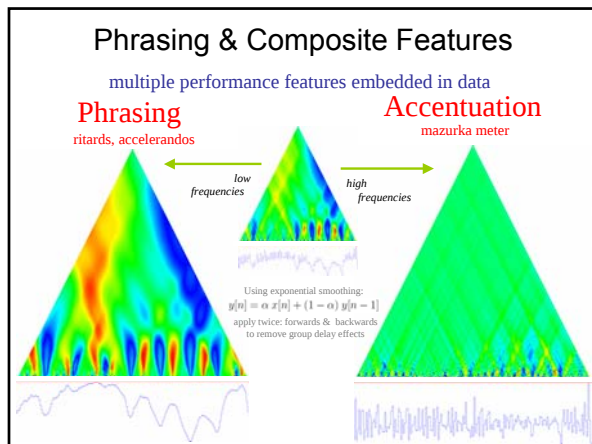
probability that same performer can produce a second performance so closely is equivalent to one atom out of an entire star.

Dynamics & Phrasing



Arch Correlation Scapes

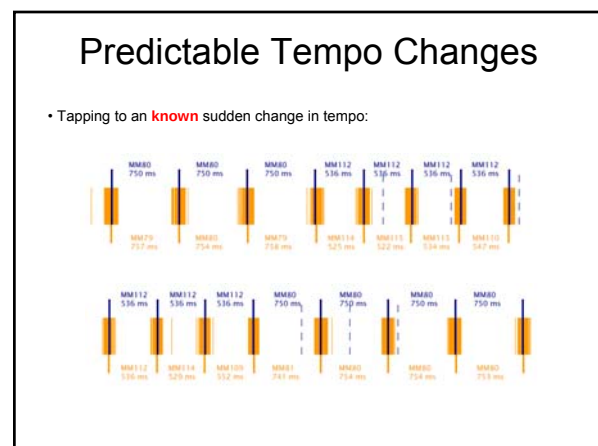
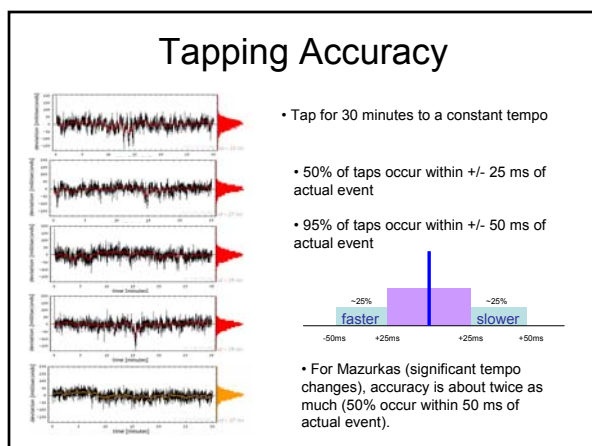
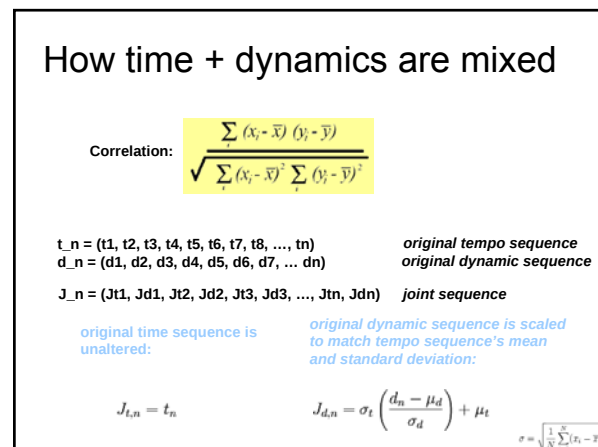




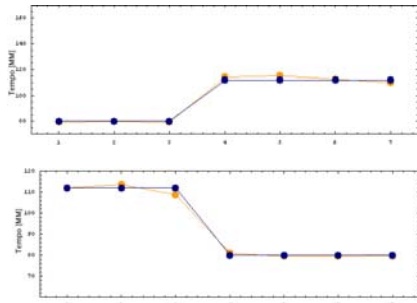
Hatto Ghost Performers

Ashkenazy, Vladimir	Duchable, François-René	Marshev, Oleg	Szokolay, Balázs
Aspaas, Tor Espen	Frith, Benjamin	Matsuzawa, Yuki	Tateno, Izumi
Babayan, Sergei	Gindin, Alexander	Moreira-Lima, Arthur	Thiollier, François-Joel
Banowetz, Joseph	Grante, Carlo	Muraro, Roger	Tipo, Maria
Baselga, Miguel	Gutierrez, Horacio	Nagy, Peter	Tomsic, Dubravka
Bellucci, Giovanni	Haebler, Ingrid	Nicolosi, Francesco	Trzeciak, Joanna
Benoit, Prisca	Hamelin, Marc-André	Nojima, Minoru	Wodnicki, Adam
Biret, Idil	Hegedus, Endre	O'Connor, John	Zarafiants, Evgeny
Bloch, Boris	Heisser, Jean-François	Ogawa, Noriko	Zilberstein, Liliya
Bronfman, Yefim	Hiseki, Hisako	Ohlsson, Garrick	
Browning, John	Hobson, Ian	Okashiro, Chitose	
Brownridge, Angela	Indjic, Eugene	Pagny, Patricia	
Budiardjo, Esther	Jandó, Jenő	Raeakallio, Matti	
Campanella, Michele	Kim, Paul	Rahkonen, Margit	
Chen, Pi-hsien	Kissin, Evgeny	Ránki, Dezső	
Collard, Jean-Philippe	Kramreiter, Tomás	Reyes, Alberto	
Dalberto, Michel	Kuzmin, Leonid	Scherbakov, Konstantin	
Didenko, Yuri	Long, Beatrice	Simon, Lazlo	
Du Plessis, Herbert	Malikova, Anna	Sterczynski, Jerzy	

<http://www.farhanmalik.com/hatto/pianistslist.html>

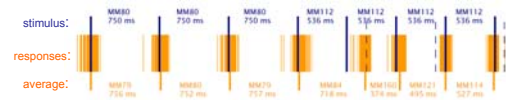


Predictable Tempo Changes

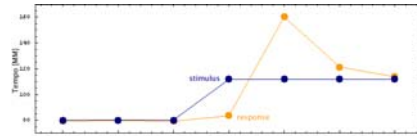


Unpredictable Tempo Changes

- Tapping to an unknown sudden change in tempo
- Suddenly faster:

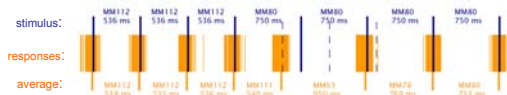


Same data as a tempo plot:



Unpredictable Tempo Changes (2)

- Tapping to an unknown sudden change in tempo
- Suddenly slower:



Same data as a tempo plot:

