


# Comparative Analysis of Multiple Musical Performances

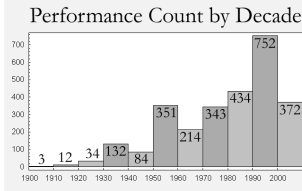
Craig Stuart Sapp  
ISMIR, Vienna, Austria  
27 September 2007



## 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:




**Performance Count by Decade**

Decade	Count
1900-1910	3
1910-1920	12
1920-1930	34
1930-1940	132
1940-1950	84
1950-1960	351
1960-1970	214
1970-1980	343
1980-1990	434
1990-2000	752
2000-2010	372

## Many Performances of Same Composition

89 performances of mazurka 17/4

- how to compare and navigate through all the performances?



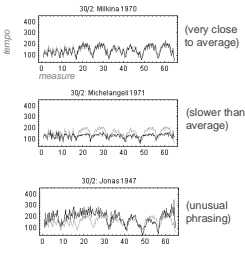
(measure 31)

Biret 1990	Hatto 2006	Osinaka 1989
Blet 2003	Horowitz 1971	Pöntinen 2003
Block 1995	Horowitz 1985	Paderewski 1912
Brailowsky 1960	Indjic 1988	Paderewski 1923
Brunhoff 1963	Kapell 1951	Paderewski 1924
Casadesus 1930	Kiepura 1999	Perahia 1994
Chiu 1999	Kilényi 1937	Perlemuter 1986
Cidat 1994	Kissin 1993	Poblocka 1999
Cohen 1997	Kitain 1937	Rangell 2001
Coop 1987	Kushner 1990	Risler 1920
Cortot 1951	Lévy 1951	Rosen 1989
Csalog 1996	Leár 1994	Rubinstein 1939
Czerny-Stefanska 1949	Lefebure 1950	Rubinstein 1952
Czerny-Stefanska 1949	Lilamand 2001	Rubinstein 1966
Czerny-Stefanska 1989	Luisada 1990	Rummel 1943
Ezaki 2006	Lushtak 2004	Shebanova 2002
Falvay 1989	Lympary 1968	Simon 1991
Ferenczy 1958	Lympary 1990	Smith 1975
Fiorentino 1990	Magaloff 1977	Szpliman 1948
Flière 1977	Magaloff 1977b	Sztopka 1959
Fou 1978	Magin 1975	Tanyel 1992
François 1956	Milkina 1970	Uninsky 1971
Gieseking 1938	Mohovich 1999	Vardi 1988
Ginzburg 1957	Nadelmann 1956	Wasowski 1980
Goldmann 1997	Ohlsson 1999	Weissenberg 1971
Guller 1956	Olejniczak 1990	Zecchi 1942
Ben-Or 1989	Hatto 1993	Olejniczak 1991
		Zecchi 1942b

## Tempo/Dynamics Graphs

— Data for particular performance  
— Average of all performances

**Beat tempo:**

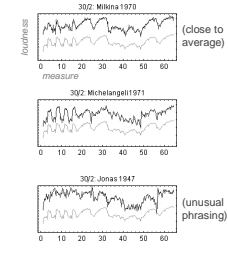


302: Milkina 1970 (very close to average)

302: Michelangeli 1971 (slower than average)

302: Jonas 1947 (unusual phrasing)

**Beat dynamics:**



302: Milkina 1970 (close to average)

302: Michelangeli 1971

302: Jonas 1947 (unusual phrasing)

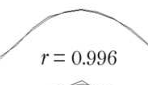
Data extracted using Sonic Visualiser / Vamp Plugins developed at C4DM, Queen Mary, U. of London  
<http://www.sonicvisualiser.org> & <http://sv.mazurka.org.uk>

## Correlation

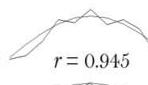
**Pearson correlation:**  
output range: -1.0 to +1.0

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

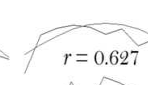
• Measures how well two shapes match:  
 r = +1.0 is an exact match  
 r = 0.0 means no relation  
 r = -1.0 means upside-down



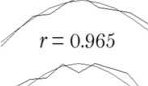
r = 0.996



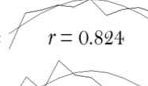
r = 0.945



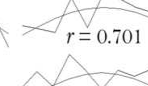
r = 0.627



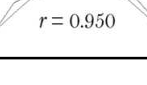
r = 0.965



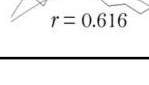
r = 0.824



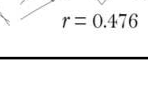
r = 0.701



r = 0.950



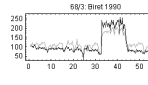
r = 0.616



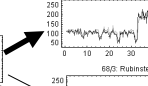
r = 0.476

## Performance Correlations


Bi		Br		Ch		Fi		In		Lu		R8		R6		Sm		Un	
Biret	1.																		
Brailowsky	0.92	1.																	
Chiu	0.81	0.81	1.																
Friere	0.83	0.86	0.86	1.															
Indjic	0.95	0.91	0.86	0.88	1.														
Luisada	0.85	0.84	0.81	0.84	0.88	1.													
Rubinstein 1938	0.82	0.66	0.76	0.73	0.66	0.67	1.												
Rubinstein 1966	0.5	0.55	0.74	0.7	0.59	0.61	0.77	1.											
Smith	0.55	0.65	0.67	0.74	0.63	0.55	0.62	0.59	1.										
Uninsky	0.86	0.85	0.89	0.89	0.9	0.89	0.75	0.69	0.64	1.									



683: Biret 1990

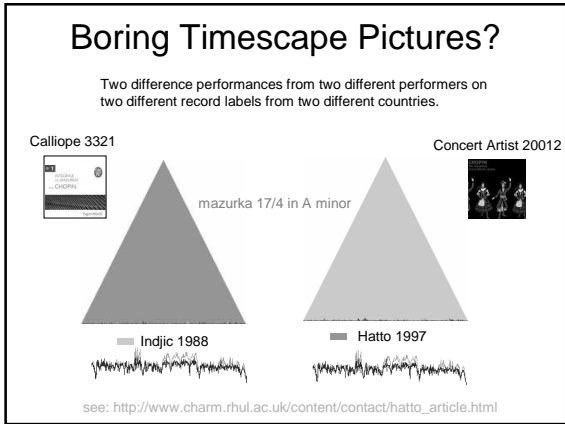
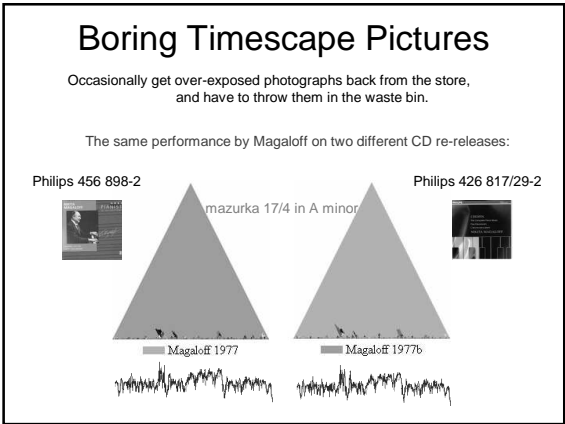
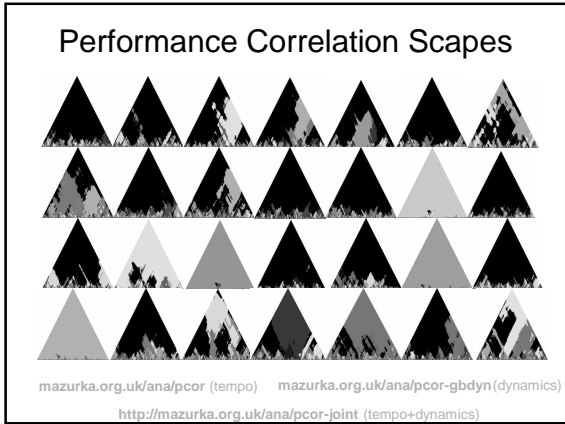
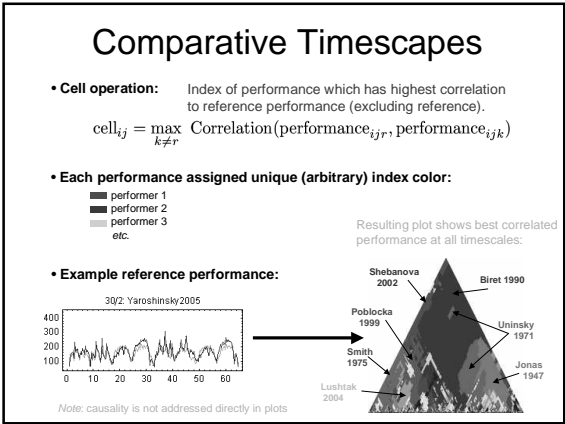
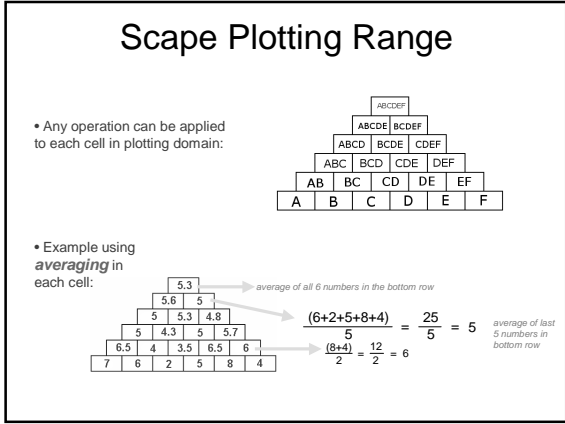
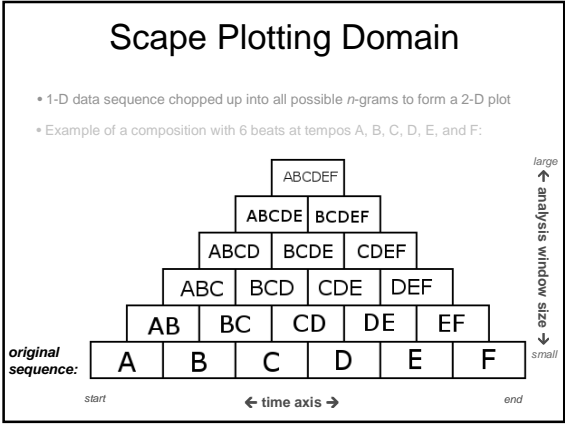


683: Indjic 2001



683: Rubinstein 1966

Highest correlation to Biret: 0.95  
Lowest correlation to Biret: 0.50



## Hatto Hoax

~ 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://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 career  
 BY MARK SINGER



September 17, 2007  
 The New Yorker

## 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 56/3
- Mazurka 59/1
- Mazurka 59/2
- Mazurka 59/3
- Mazurka 63/2
- Mazurka 7/1
- Mazurka 7/2
- Mazurka 7/3
- Mazurka 7/4
- Mazurka 7/5
- Mazurka 68/1
- Mazurka 68/2
- Mazurka 68/3
- Mazurka 68/4
- Mazurka 24/1
- Mazurka 24/2
- Mazurka 24/3
- Mazurka 24/4
- Mazurka 30/1
- Mazurka 17/1
- Mazurka 17/2
- Mazurka 17/3
- Mazurka 17/4
- Mazurka 30/1
- Mazurka 30/2
- Mazurka 30/3
- Mazurka 30/4
- Mazurka 41/2
- Mazurka 67/4

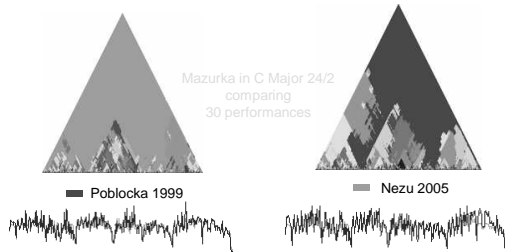
Janusz Olejniczak



Naive/OPUS 111:  
 OP2002 (1991)

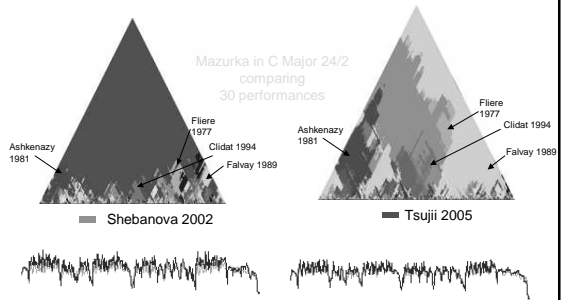
## Strong Interpretive Influences

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



• Nezu did graduate studies with Poblocka

## Strong Interpretive Influences (2)

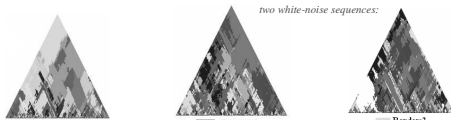


• note parallel colors: same neighboring pianists (clustering in performance space)

## Purely Random Matches

• Plots have to show some match at all points  
 -- not necessarily a good one

• Small color regions, inverted triangles & broken borders = poor matches



Uninsky 1971  
 20.9% Lisinski 2001  
 12.9% Solima 1972  
 10.5% Jones 1987  
 8.2% Chua 1999  
 7.9% Paderewski 1969  
 6.5% Jones 1987  
 4.5% Paderewski 1969  
 4.1% Ashkenazy 1981  
 3.9% Czerny 1949  
 3.8% Tsang 1993  
 3.7% Czerny 1949  
 3.6% Jones 1987  
 3.4% Lisinski 2001  
 3.2% Tsang 1993  
 3.1% Paderewski 1969  
 3.0% Ashkenazy 1981  
 2.9% Paderewski 1969

Random  
 20.7% Jones 1987  
 14.1% Shebanova 2002  
 12.9% Blot 2003  
 8.0% Chua 1999  
 7.9% Paderewski  
 6.0% Tsang 2005  
 4.9% Mohornik 1999  
 4.8% Paderewski 1969  
 3.3% Cidat 1994  
 3.2% Jones 1987  
 2.8% Lisinski 2001  
 2.7% Michalowski 1969  
 2.1% Smith 1973  
 2.1% Schillhawy 1960  
 1.7% Tsang 1993  
 1.6% Paderewski 1969  
 1.4% Ashkenazy 1981

Random  
 17.7% Shalinsky 1960  
 13.0% Paderewski 1969  
 10.1% Shalinsky  
 7.0% Blot 2003  
 7.1% Tsang 1993  
 6.0% Paderewski 2005  
 4.0% Paderewski 1969  
 3.0% Schillhawy 1960  
 3.0% Cidat 1994  
 3.1% Paderewski 1956  
 3.0% Jones 1987  
 2.0% Tsang 2005  
 2.0% Lisinski 2001  
 2.0% Lisinski 2001  
 1.9% Paderewski 1969  
 1.9% Paderewski 1969  
 1.9% Paderewski 1969  
 1.9% Paderewski 1969

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

(mazurka 30/2: 36 performances)

## Same Performer over Time

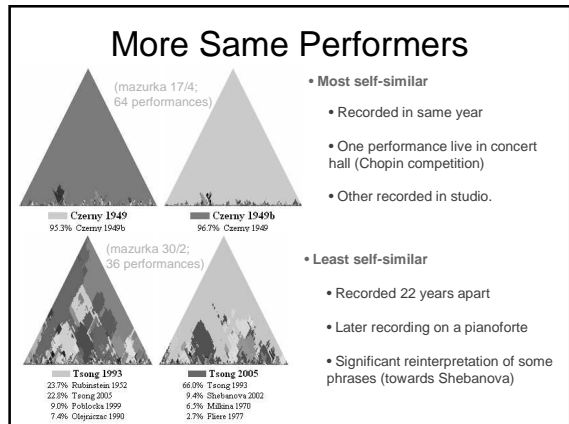
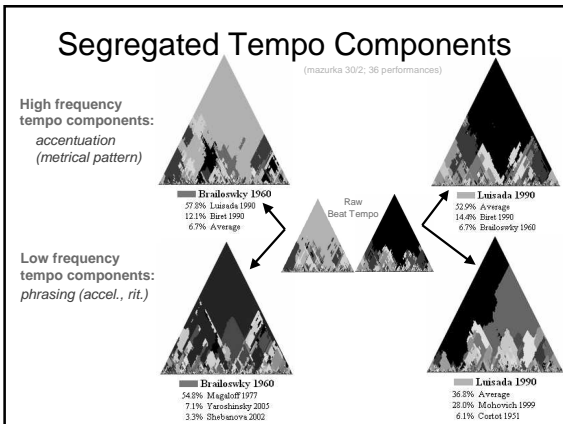
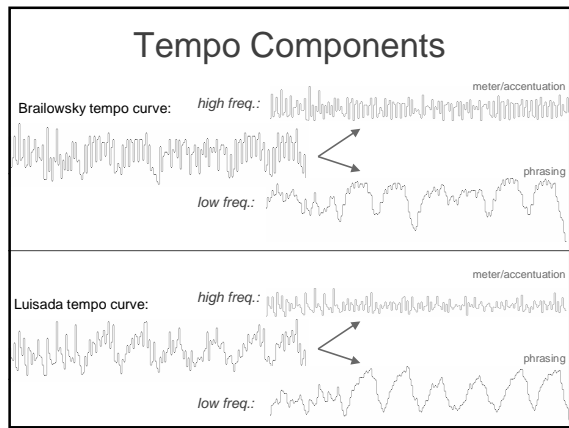
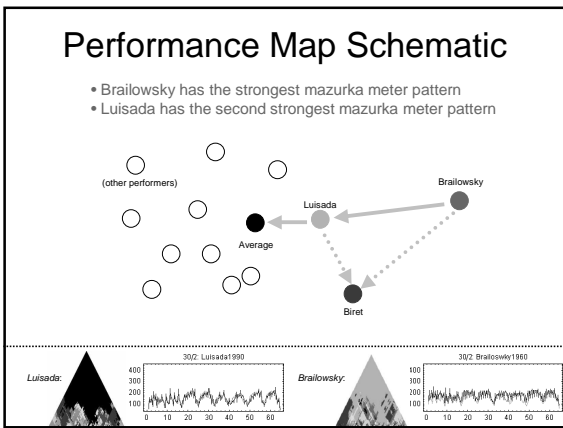
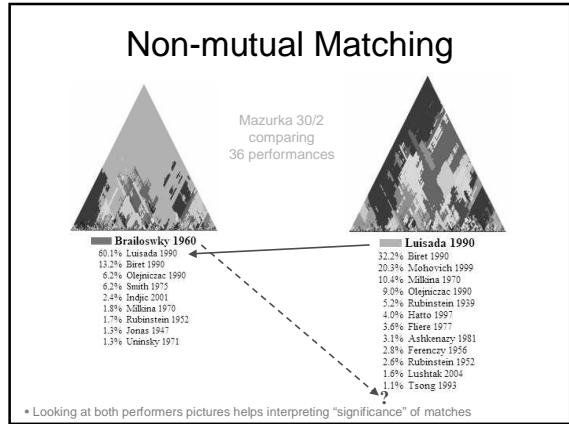
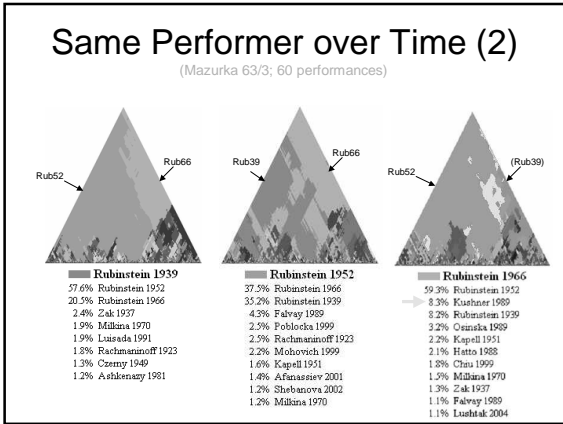
(mazurka 63/3; 60 performances)

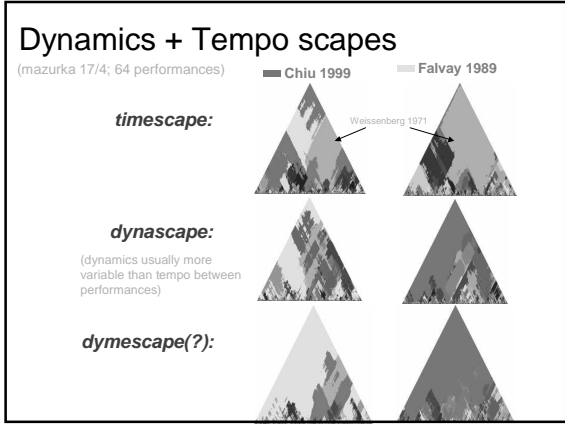
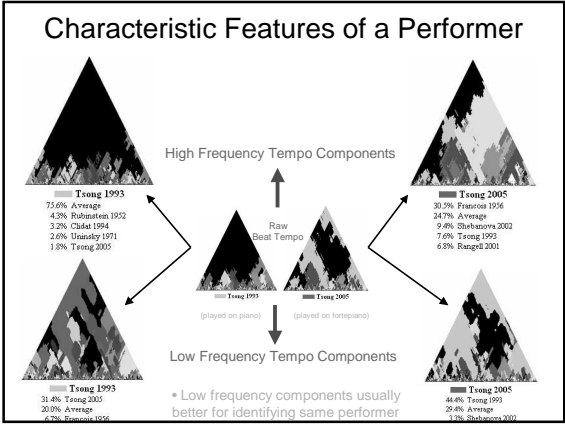


Uninsky 1932  
 74.1% Uninsky 1971  
 4.9% Moravec 1969  
 2.8% Czerny 1955  
 1.7% Schillhawy 1960  
 1.2% Czerny 1949  
 1.1% Rabecwiczowa 1932

Uninsky 1971  
 66.1% Uninsky 1932  
 6.0% Indjic 1988  
 5.2% Bortnikovich 1969  
 2.6% Czerny 1949  
 2.6% Czerny 1955  
 2.1% Schillhawy 1960  
 1.9% Midans 1970  
 1.7% Huzarszewicz 1955

• 40 years between recordings  
 • 78 rpm recording / 33.3 rpm recording  
 • France in 1932 / Texas in 1971

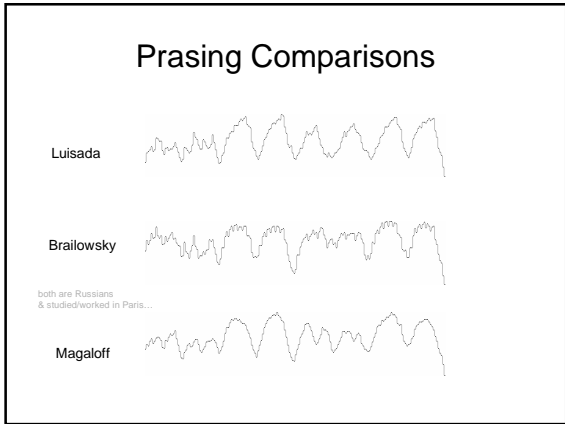




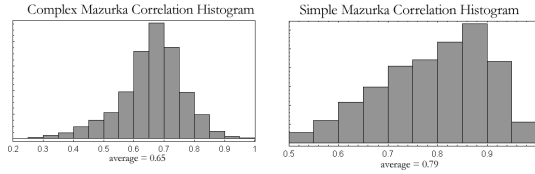
- ### Future Work
- More detailed performance features  
LH/RH synchrony; off-beat rhythms; ornaments, etc.
  - More composite/segregated feature analysis  
separating features: Brailowsky/Luisada example for 24/2  
joining features: Chiu/Falvy example for 17/4
  - Quantitative measurements of performance similarity  
timescapes are primarily qualitative, but not exclusively
  - Authenticity analysis of Con. Artist's Cortot mazurkas  
and more on Fiorentino...
- 
- <http://mazurka.org.uk/info/present/ismir-20070927>



### Extra Slides



## Expected Correlation Values



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

## 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	Hegedüs, Endre	O'Conor, John	Zarafiants, Evgeny
Bloch, Boris	Heisser, Jean-François	Ogawa, Noriko	Zilberstein, Liya
Bronfman, Yefim	Hiseki, Hisako	Ohlsson, Garrick	
Browning, John	Hobson, Ian	Okashiro, Chitose	
Brownridge, Angela	Indic, Eugene	Pagny, Patricia	
Budiarjo, Esther	Jandó, Jenő	Raekallio, Matti	
Campanella, Michele	Kim, Paul	Rahkonen, Margit	
Chen, Pi-hsien	Kissin, Evgeny	Ránki, Dezős	
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>

## How time + dynamics are mixed

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

$t_n = (t1, t2, t3, t4, t5, t6, t7, t8, \dots, tn)$   
 $d_n = (d1, d2, d3, d4, d5, d6, d7, \dots, dn)$

*original tempo sequence*  
*original dynamic sequence*

$J_n = (Jt1, Jd1, Jt2, Jd2, Jt3, Jd3, \dots, Jtn, Jdn)$  *joint sequence*

original time sequence is unaltered:

$$J_{t,n} = t_n$$

original dynamic sequence is scaled to match tempo sequence's mean and standard deviation:

$$J_{d,n} = \sigma_t \left( \frac{d_n - \mu_d}{\sigma_d} \right) + \mu_t$$

$$\sigma = \sqrt{\sum (x_i - \bar{x})^2}$$

## Peeling Back the Layers

