

## Data-rich approaches to authorship and ascription

Peter van Kranenburg  
Meertens Institute (Amsterdam)

5 July 2012



## Outline

- Authorship
- Recent Research
- Case
- Future perspectives

5 July 2012

2

## Authorship



5 July 2012

3

## Authorship

Who is the author of this piece?

5 July 2012

4

## Authorship

What does 'authorship' mean?

5 July 2012

5

## Authorship

Four functions (Harold Love, 2002):

### Precursory Authorship

Incorporation of earlier writing

### Executive Authorship

Creation of the actual text

### Declarative Authorship

'Owning' the words

### Revisionary Authorship

Editing the text

5 July 2012

6

## Authorship

---

Thomas Merriam (2002):

**Who wrote Shakespeare?**

**What did Shakespeare write?**

5 July 2012 7

## Authorship

---

Authorship studies

- external evidence
- internal evidence (often proof by example)



5 July 2012 8

## Authorship

---

Can we measure personal style more precisely?

5 July 2012 9

## Authorship

---

What is personal style?

5 July 2012 10

## Personal Style

---

Definition L.B. Meyer:

"Style is a **replication of patterning**, whether in **human behavior** or in the **artifacts** produced by human behavior, that results from a series of **choices** made within some set of **constraints**."

Style and Music, Chicago, 1989.

Allows for quantification.

5 July 2012 11

## Personal Style

---

Three Levels for Constraints:

Laws

Rules

Strategies ← Authorship

5 July 2012 12

## Recent Research

---

Trowbridge 1986  
 Pollastri 2001  
 Hippé 2002  
 Cilibrić 2004  
 Hardoon 2004  
 Koenig 2004  
 Catalepe 2004  
 Wolkowicz 2007  
 Geertzen 2008  
 Lebar 2008  
 Bellaachia 2009  
 Hillewaere 2009  
 Kraneburg 2009  
 Pérez-Sancho 2009  
 Hillewaere 2010  
 Kalaiakatsos-Papakostas 2010  
 Dor 2011  
 Mearns 2010  
 Honingh 2011  
 Kalaiakatsos-Papakostas 2011a  
 Kalaiakatsos-Papakostas 2011b  
 Taminau 2011  
 Kraneburg 2012

5 July 2012 13

## Recent Research: Composers

---

J.S. Bach  
 G.F. Handel  
 W.A. Mozart String Quartets  
 F.J. Haydn  
 L.v. Beethoven  
 F. Chopin

5 July 2012 14

## Recent Research: Methods

---

**Classification of global feature vectors**  
 Trowbridge 1986; Hippé 2002; Van Kraneburg 2004, 2009; Hillewaere 2009, 2010; Mearns 2010; Kalaiakatsos-Papakostas 2010; Kalaiakatsos-Papakostas 2011; Bellaachia 2009; Honingh 2011; Dor 2011; Taminau 2011

**N-gram approaches**  
 Pérez-Sancho 2009; Hillewaere 2009, 2010; Wolkowicz 2007

**Compression-based approaches**  
 Cilibrić 2004; Catalepe 2005

**Neural networks**  
 Kalaiakatsos-Papakostas 2010

**Markov chains**  
 Pollastri 2001; Hardoon 2004; Kalaiakatsos-Papakostas 2011

**Alignment-based learning**  
 Geertzen 2008

5 July 2012 15

## Recent Research: Summary

---

Main approach in recent computational authorship attribution:  
 Employing Machine Learning techniques  
 Evaluation in terms of classification accuracy.

5 July 2012 16

## Recent Research: Summary

---

1. Digitize music
2. Quantize compositions
3. Derive statistical model for known compositions
4. Test the model
5. (Apply model to anonymous compositions)

E.g. Mozart vs. Haydn String quartets	Kraneburg 2004 Hillewaere 2009 Hillewaere 2010 Kalaiakatsos 2011a Kalaiakatsos 2011b Dor 2011 Taminau 2011	79% 80% 75% 72% 67% 74% 73%
--	--	---

5 July 2012 17

## Case(s)

---

5 July 2012 18

## Organ Fugues in Bach Catalog



5 July 2012

19

## BWV 534

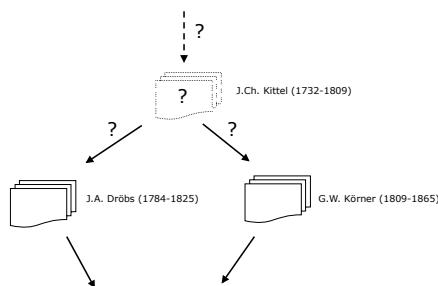
*R. Variation of Toccata in F major written by J.S. Bach.*

BWV 914  
BWV 548

5 July 2012

20

## BWV 534 Sources



5 July 2012

21

## BWV 534 in literature

- J.N. Forkel (1802): not mentioned at all.  
 F.K. Griepenkerl (1844): in edition.  
 A. Schweitzer (1908): "gewaltiger Tragik".  
 Ph. Spitta (1916): "ein etwas ungeordnetes Wesen".  
 H. Keller (1950): no doubt  
 A. Dickinson (1956): "forced counterpoint".  
 P. Williams (1980): bad copy.  
 D. Humphreys (1985): not Bach. Johann Christian Kittel?  
 P. Dirksen (2000): Wilhelm Friedemann Bach?

5 July 2012

22

## Steps

1. Digitize music
2. Quantize compositions
3. Derive statistical model for known compositions
4. Test model
5. Apply model to anonymous compositions

5 July 2012

23

## Digitize music

5 July 2012

24

## Quantization

Features for authorship attribution:

- Independent of composition
- Independent of date
- Abundance
- Measurable

Oeuvre  
Genre  
Composition  
Section  
Phrase  
Motif  
Note

5 July 2012 25

## Quantization

Intervals

5 July 2012 26

## Quantization

Parallels

5 July 2012 27

## Quantization

Dissonance

5 July 2012 28

## Quantization

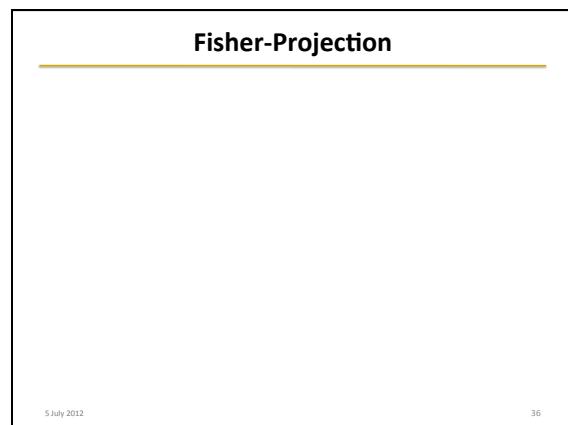
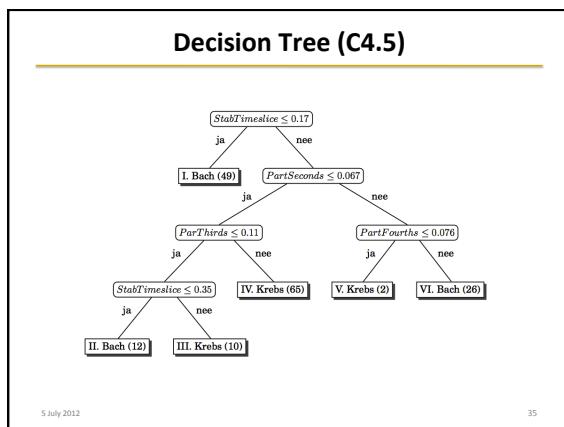
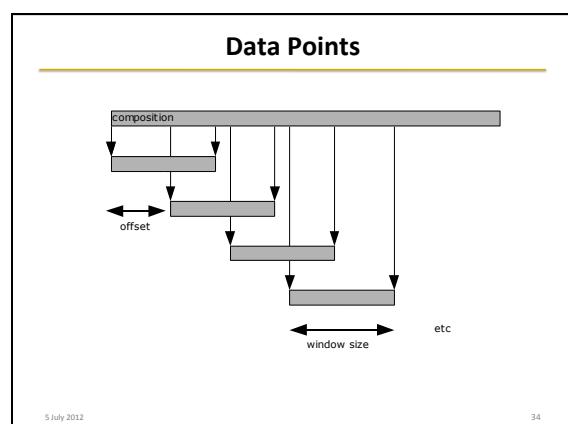
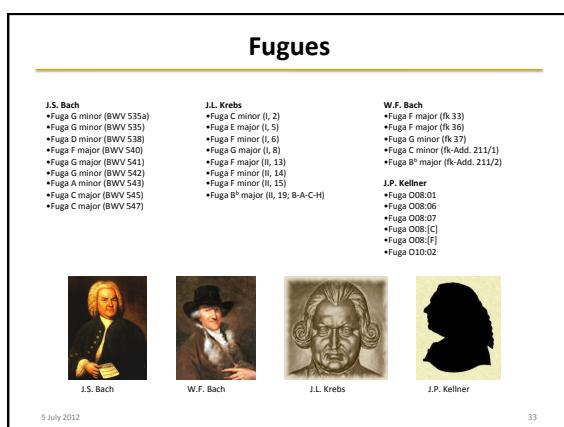
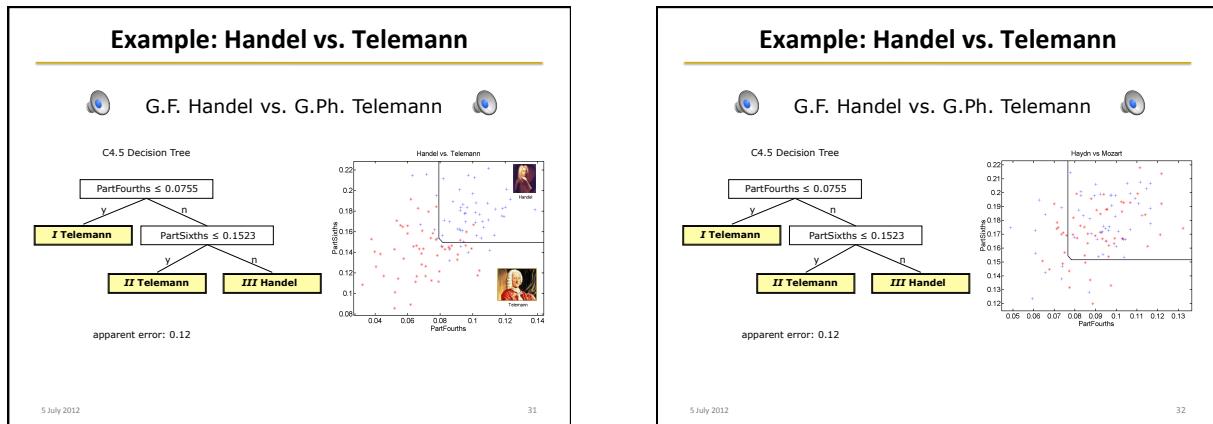
Time slices

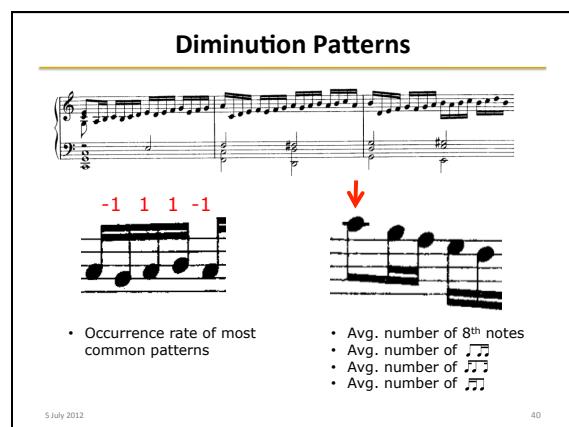
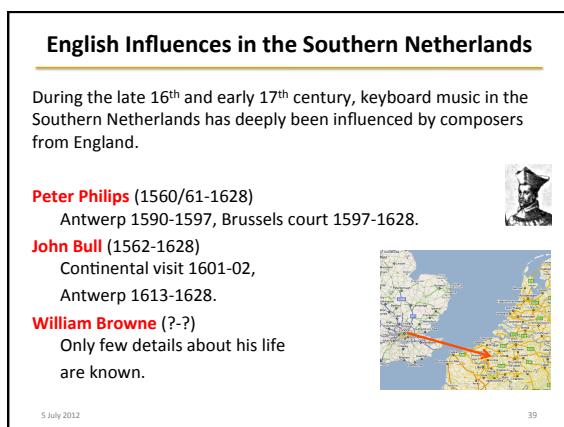
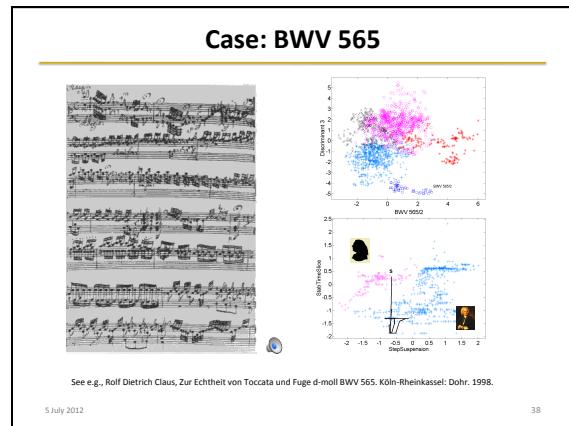
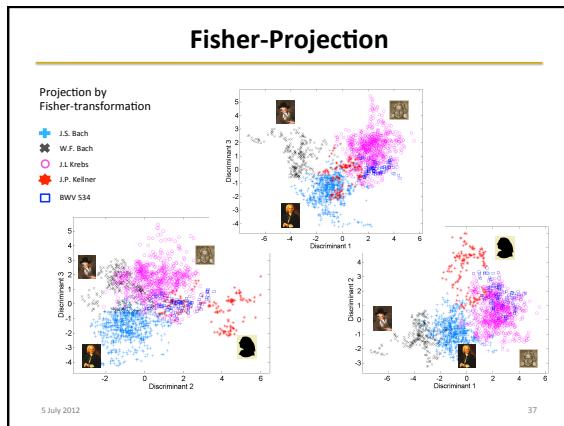
5 July 2012 29

## Features

Feature	BWV 545	Krebs I_5
PartSeconds	0.063800	0.0445205
PartThirds	0.257315	0.291065
PartFourths	0.0180956	0.0235613
PartAugFourth	0.0226244	0.0277086
PartDimFifths	0.1277553	0.131071
PartFifths	0.131976	0.139447
PartSixths	0.0794872	0.0867956
PartEighths	0.1277553	0.131071
PartOctaves	0.131976	0.139447
PartSixths	0.0214178	0.0364259
PartThirds	0.092006	0.155355
PartFourths	0.0214178	0.0364259
PartEighths	0.0766212	0.08587741
StepSuspension	0.1152124	0.1152124
VoiceDensity	0.8125	0.856076
DisPart	0.467136	0.435637
BeginBarDiss	0.616823	0.408602
HarmonyEntropy	5.86014	5.03714
PitchEntropy	4.03455	4.08664
SonorityEntropy	3.3399	4.26365
StabTimeslice	0.312442	0.383532

5 July 2012 30





### Compositions

<b>Browne</b> (ed. Vendome/Good)	<b>Bull</b> (M.Britt. XIX)	<b>Philips</b> (Fitzwilliam)
<ul style="list-style-type: none"> <li>Fantasia 8vi ton</li> <li>Allemande I</li> <li>Allemande II</li> <li>Allemande III</li> <li>Mal Sims</li> </ul>	<ul style="list-style-type: none"> <li>Revenant</li> <li>Prelude</li> <li>Go from my window</li> <li>Bull's goodnight</li> </ul>	<ul style="list-style-type: none"> <li>Freno. Seconda Parte</li> <li>Cosi moriro. 3a Parte</li> <li>Pavana Pagget</li> <li>Gallarda</li> <li>Passamezzo Pavana</li> <li>Gallarda Passamezzo</li> <li>Chi fara fede al Cielo</li> <li>Bon jour mon cuer</li> <li>Pauna Doloroso</li> <li>Gallarda Dolorosa</li> <li>Amarilli di Julio Romano</li> <li>Margott Laborez</li> <li>Fantasia</li> <li>Pavana</li> <li>Le Rossignol</li> <li>Fantasia</li> </ul>
<b>Browne</b> (Ghent Ms. 96)	<b>Bull</b> (Noëls Flamands)	
<ul style="list-style-type: none"> <li>O lieftie lief</li> <li>Lanturli</li> <li>[w.o. title]</li> <li>Daphne</li> </ul>	<ul style="list-style-type: none"> <li>Den lustelijken meij</li> <li>Een kindeken is ons... (A)</li> <li>Een kindeken is ons... (B)</li> <li>Laat ons met herten rejne</li> </ul>	
<b>Bull</b> (M.Britt. XIV)	<b>Philips</b> (ed. Harley)	
<ul style="list-style-type: none"> <li>Prelude and Fantasia</li> <li>Pr. &amp; Fantasia 8<sup>th</sup> mode</li> <li>Fantasia</li> <li>Fantasia 5<sup>th</sup> mode</li> <li>Fantasia 8<sup>th</sup> mode</li> </ul>	<ul style="list-style-type: none"> <li>Almande</li> <li>Pavana Anglica</li> <li>Fantasia</li> <li>Chi fara fed'al ciel</li> </ul>	

5 July 2012 41

### Diminutions: Inventory

In total: 2444 patterns. 237 unique patterns.

Occurrence rates of 11 most common patterns (in %) in all compositions (RH):

Pattern	Phillips	Bull	Browne
-1 -1 -1 -1	7.87	9.64	2.62
-1 -1 -1 1	6.64	3.38	2.33
-1 1 1 1	5.09	2.71	4.65
-2 1 1 -1	1.23	1.35	8.72
-2 1 1 1	2.47	2.03	4.94
1 -1 -1 -1	9.26	10.66	3.20
1 -1 -1 1	6.94	5.75	2.33
1 1 -1 -1	5.09	3.38	2.62
1 1 -2 1	9.10	5.08	4.36
1 1 1 1	13.43	13.54	11.34
2 1 1 1	2.62	2.54	4.94

5 July 2012 42

## Diminutions: Inventory

In total: 2444 patterns, 237 unique patterns.

Occurrence rates of 11 most common patterns (in %) in all compositions (LH):

Pattern	Philips	Bull	Browne
-1 -1 -1 -1	15.63	7.47	7.94
-1 -1 -1 1	1.18	2.07	1.59
-1 1 1 1	1.18	1.24	3.17
-2 1 1 -1	0.59	0.00	1.59
-2 1 1 1	1.77	4.15	9.52
1 -1 -1 -1	16.52	7.47	7.94
1 -1 -1 1	1.18	2.07	1.98
1 1 -1 -1	4.13	4.15	4.37
1 1 -2 1	1.47	6.64	4.37
1 1 1 1	13.86	9.54	11.51
2 1 1 1	1.77	4.15	9.52

5 July 2012 43

## Models: Philips vs. Bull and Browne

**Right Hand**

Selected features:

- 2 1 1 1
- 1 -1 -1 1
- 1 1 -2 1
- Avg. Number of 8<sup>th</sup> notes
- Position of 8<sup>th</sup> notes
- Direction of Motion
- Most Common Pitch
- Repeated Notes

**Left Hand**

Selected features:

- Duration of Melodic Arcs
- Most Common Melodic Interval
- 2 1 1 1
- Dominant Spread

a b <-- classified as  
27 6 | a = PP  
5 33 | b = Other

13 4 | a = PP  
1 22 | b = Other

5 July 2012 44

## Next

Is classification accuracy the end goal?

Next steps:

- Understanding of differences between personal styles
- Develop a theoretical foundation for choosing features
- Integration in Musicological discourse
- Standardized test set for composer attribution

5 July 2012 45

5 July 2012 46