Searching musical incipits by means of sequence alignment

Jelmer van Nuss, Geert-Jan Giezeman, Frans Wiering

f.wiering@uu.nl

Dept. Information and Computing Sciences
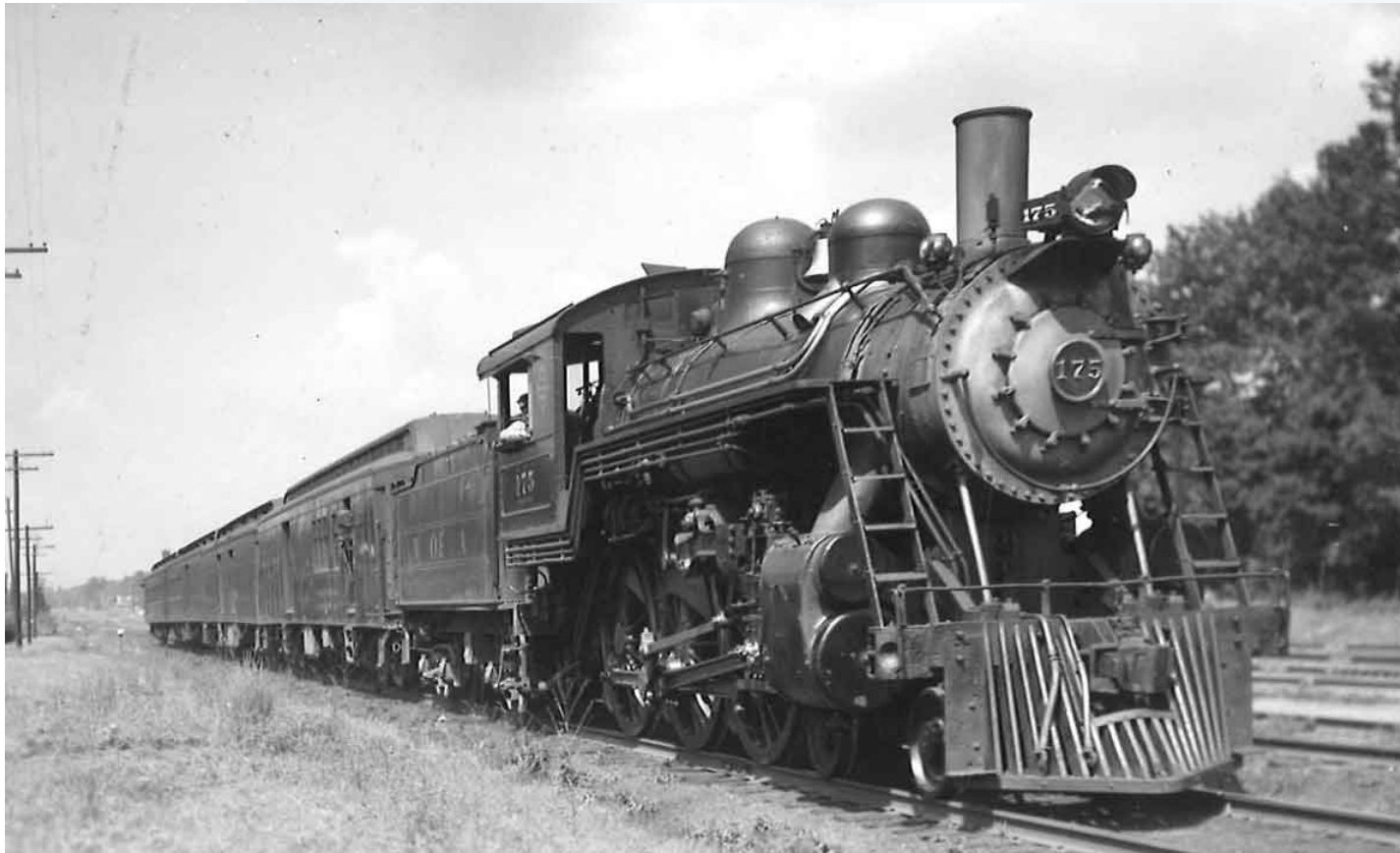
Universiteit Utrecht

Authors

- Jelmer van Nuss
  - ground truth; experiment

- Geert-Jan Giezeman
  - software development

- Frans Wiering
  - musicology and MIR context

OLD STUFF

RISM incipits

- mainly for RISM A/II (music manuscripts after 1600)
- incipits encoded in *Plaine and Easie Code* (PAEC)
- published, with other RISM products, via RISM OPAC
  - over 1 million items

RISM incipits

```
112A    1.1.1
112B    pf.
112C    G-2
112D    xFCG
112E    2/4
112F    '''2C/!{y6DCz        'C}    /2D/q3D{8Cq3D8Cq3D8Cq3D8C}/y'
        '4.B'''8Ez/
```

PAEC defies all sustainability guidelines

- mainly for RISM A/II (music manuscripts after 1600)
- incipits encoded in *Plaine and Easie Code* (PAEC)
- published, with other RISM products, via RISM OPAC
  – over 1 million items

Universiteit Utrecht

Searching RISM incipits

- OPAC interface has simple provision for melody search
  - text string
  - wildcards ? and *

900.005.286-1.1.1

GxFEBCBABBAGxF

OPAC output

- classical database search
- items either match or not
- no ranking

Geometric matching

- Rainer Typke's PhD research (2003-2007)
- Earth Mover's Distance (EMD) and variants
- idea:
  - dots in 2D space
  - dot 'weight' corresponds to duration
- the less weight has to be moved, the more similar items are
  - items are ranked
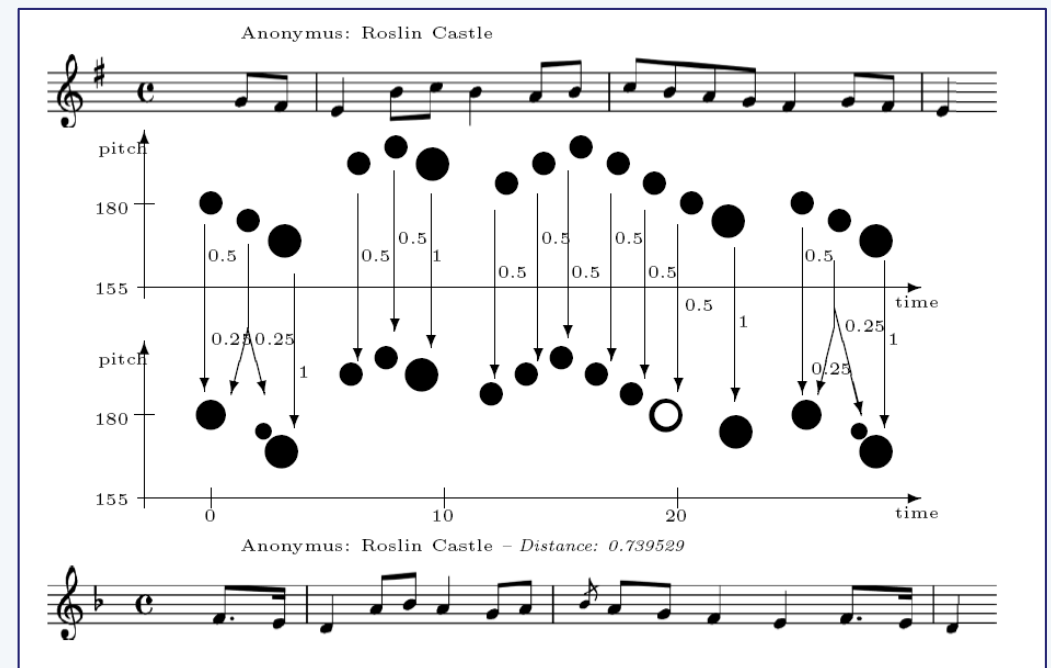
Implementation

- search engine (YahMuug!e), now defunct
  - 450.000 incipits
  - decent performance
- ground truth for 11 queries
  - used in MIREX 2005
- comparison to previous experiments
  - Roslin castle
  - deanonymisation

Universiteit Utrecht
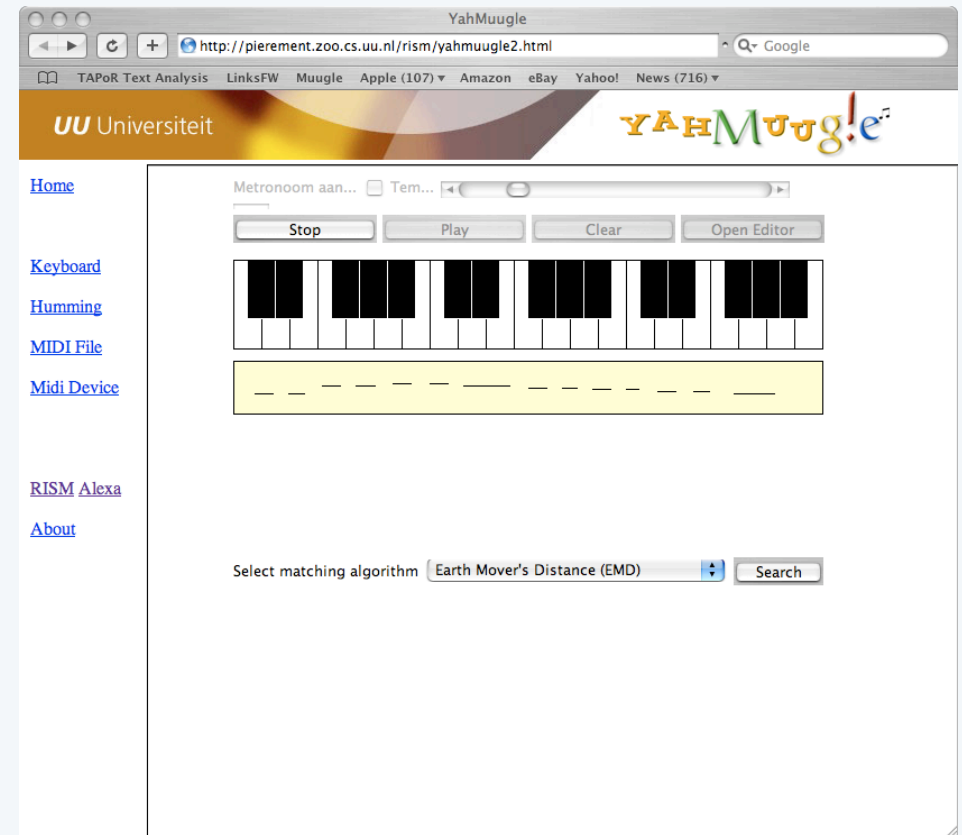
Sequence alignment

- inspiration: bioinformatics, DNA

Target Sequence
5' ACTACTAGATTACTTACGGATCAGGTACTTTAGAGGCTTGCAACCA 3'
5' ACTACTAGATT----ACGGATC--GTACTTTAGAGGCTAGCAACCA 3'
Query Sequence
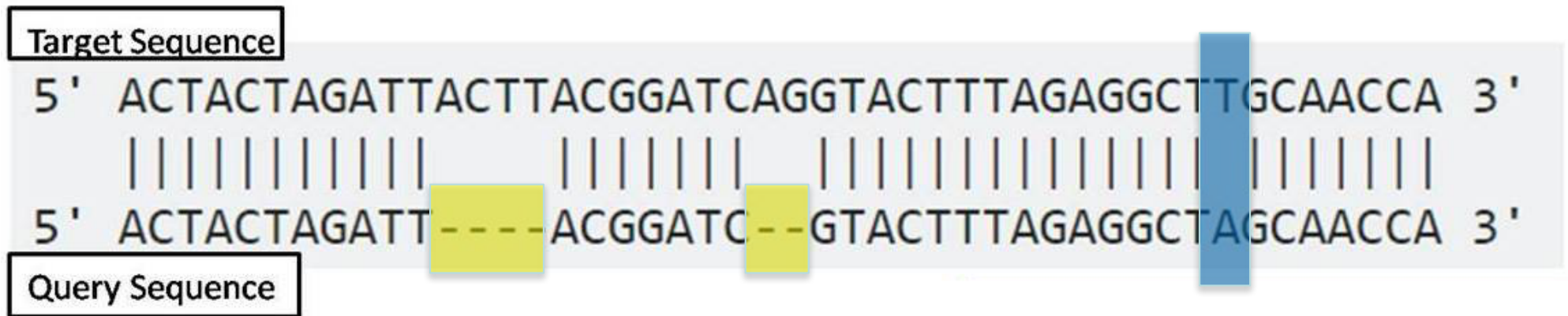
- find optimum alignment, allowing of note
  - substitutions
  - insertions, deletions → there can be gaps
- create *raters* (cost functions) for each of these
- total cost → ranking

Universiteit Utrecht

Melody sequence alignment

- Peter van Kranenburg's PhD research (2006-2010)

En daar reed eris een heer, en hij was wellegemoed En daar reed eris een heer, en hij was wellegemoed En hij droeg er d'r rozekrans on er zijn hoed En hij droeg er een ro - zekrans on erre zijn hoed.

Daar reed er een heer die was wel dideldon die was wel don da don didelon don Daar reed er een heer die was wel dergemoed

Computed optimal alignment, showing subsitution and gap scores:
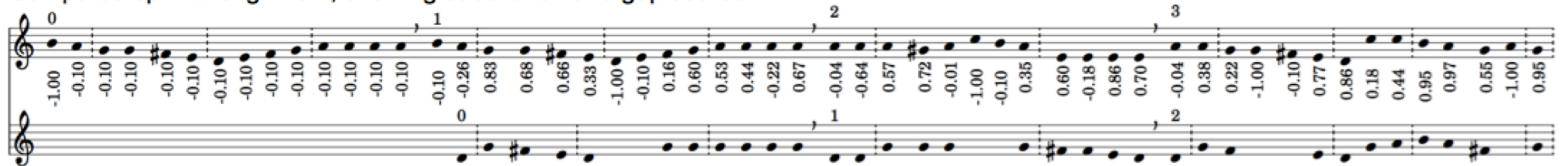
- evaluated on tune families
- oral transmission
- implemented in Dutch Song Database
  – www.liederenbank.nl

Universiteit Utrecht

NEW STUFF

Aligning RISM incipits

- investigate potential of alignment methods for different type of melodic material
  - written transmission, shorter, much larger collection

- MUSALIGN library (C++)
  - pairwise music alignment
  - combinations of raters
  - freely available

- Monochord search engine
  - RISM incipit search
  - mainly experimentation environment

https://www.projects.science.uu.nl/monochord/

RISM search by Utrecht University

create melodies | query | matching method selection | results | about

Query type: Database query with number 1982.

Matching methods

Column M1 in table is rank of matching method 1 ..

1. Exact Pitch
2. Fixed duration with ima combined Exact Pitch
3. Scaled duration with ima combined Exact Pitch

Results

Id	Score	M1	M2	M3	Name
1982	18	1	1	1	**Cristiani, Stefano** 4 Cantatas
	RISM				
778132	9.4	2	48	55	**Cordicelli, Giovanni** L' Adrasto
	RISM				
900249	9.1	3	x	x	**Schgraffer, Jakob Johann Anton** Beata es virgo Maria
	RISM				

RISM search by Utrecht University

Details of match

Method: Exact Pitch
Result id: 778132
Score: 9.4
Pitch shift: -5

Notes 1	r	r	E	C#	C#	B	C#	A	A	r	C#	B	A	G#	r	B	B	B
Pitch 1	0	0	215	204	204	198	204	192	192	0	204	198	192	187	0	198	198	198
Score	0.5	0.5	1	1	1	1	1	1	1	0.5	-0.8	-1	1	1	-0.8	1	1	-0.5
Shifted 2	0	0	215	204	204	198	204	192	192	0		192	192	187		198	198	0
Pitch 2	0	0	220	209	209	203	209	197	197	0		197	197	192		203	203	0
Notes 2	r	r	F	D	D	C	D	Bb	Bb	r		Bb	Bb	A		C	C	r

RISM search by Utrecht University

create melodies | query | matching method selection | results | about

Query type: Database query with number 1982.

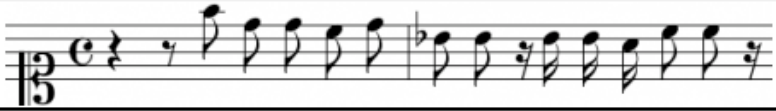## Matching methods

Column M1 in table is rank of matching method 1 ..

1. Exact Pitch
2. Fixed duration with ima combined Exact Pitch
3. Scaled duration with ima combined Exact Pitch

Results

Id	Score	M1	M2	M3	Name
1982	18	1	1	1	**Cristiani, Stefano** 4 Cantatas
	RISM				
778132	9.4	2	48	55	**Cordicelli, Giovanni** L' Adrasto
	RISM				
900249	9.1	3	x	x	**Schgraffer, Jakob Johann Anton** Beata es virgo Maria
	RISM				

Universiteit Utrecht

RISM
Répertoire International des Sources Musicales

[   ] Suchen   Erweiterte Suche

🖨 Drucken   ✉ Versenden   💾 Speichern   🔗 Permalink

‹‹ Zurück zur Trefferliste

Cristiani, Stefano ⌔

4 Cantatas

▼ **Werkinformation**

Besetzungshinweis:	S, pf
Text:	Metastasio, Pietro ⌔ [ermittelt]
Sprache:	Italienisch
Schlagwort:	Kantaten

▼ **Quellenbeschreibung**

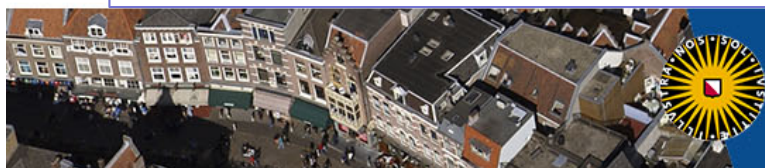|	
Originaler Titel:	*[title page:] Cantate del Metastasio ǀ Poste in Musica dal Sig.r ǀ Steffano Cristiani*
Material:	• short score (piano score): 32-39f.
	Abschrift; 20,5 x 28 cm
	Schreiber: Chelucci, Giovanni ⌔

▼ **Musikincipits**

1.1.1 S, c [f.32v-33r:] oh Dio Fileno comincia il prato.; G

– Oh Dio Fileno

Universiteit Utrecht

Alignment experiment

- evaluate raters
- relate to previous work, esp. Typke (2003, 2005)
  - case studies
  - deanonimisation not discussed here
- performance comparison difficult
  - larger dataset → revised ground truth
  - geometric search unavailable
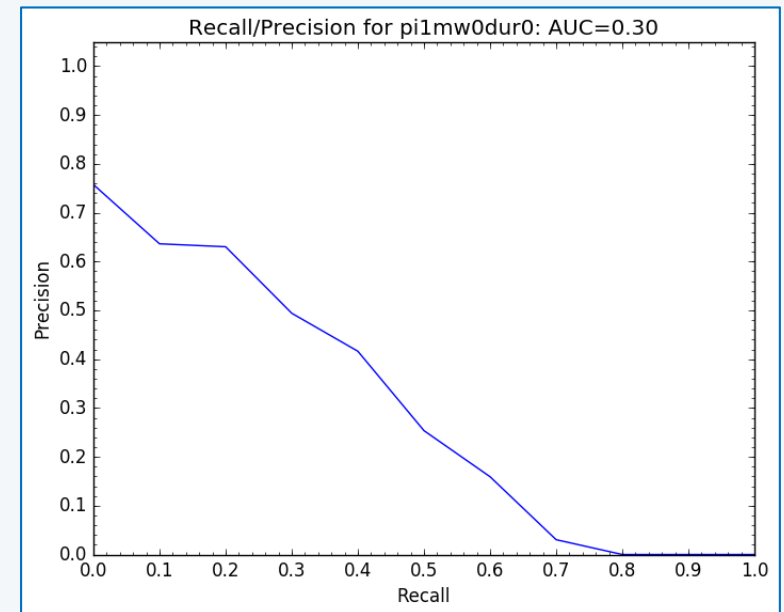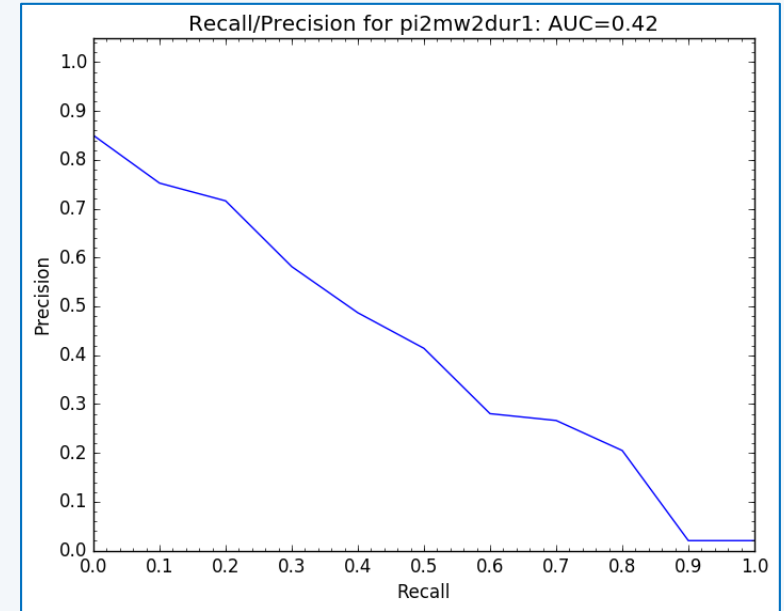
Universiteit Utrecht

Raters and methods

- 3 pitch raters
  - exact pitch
  - 2 raters using interval size (Zigzag and Kranenburg pitch)
- 3 duration raters
  - none
  - 2 raters using duration (notated and scaled)
- 3 metric raters
  - none
  - 2 raters using Inner Metric Analysis (raw and normalised)

- 3 * 3 * 3 = 27 alignment methods
- evaluation: Precision/Recall graphs
- Mean Average Precision (= Area Under Curve)

Universiteit Utrecht

typical P/R graph: black=best
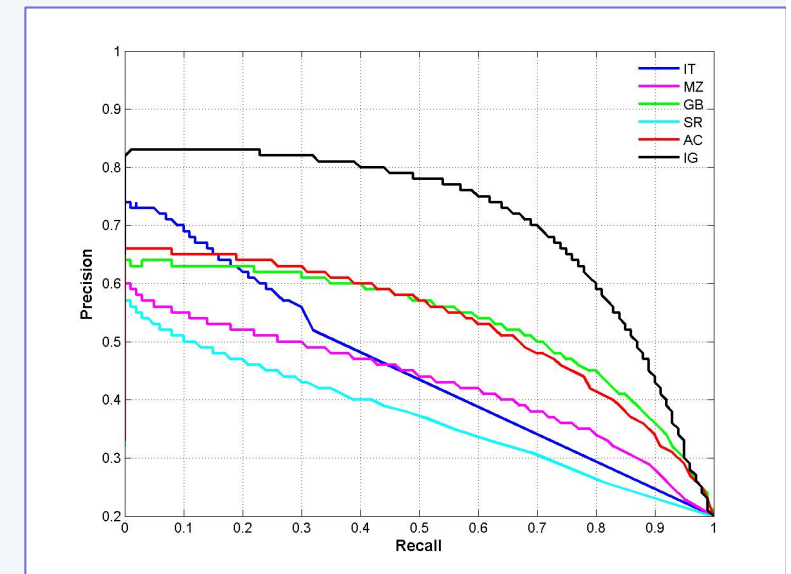
Results

- best combinations
  - exact pitch, notated duration, raw IMA (0.42)
  - exact pitch, notated duration, normalised IMA (0.41)
  - exact pitch, notated duration, no IMA (0.41)

- for comparison
  - exact pitch only (0.30)

Recall/Precision for pi2mw2dur1: AUC=0.42

Recall/Precision for pi1mw0dur0: AUC=0.30

Universiteit Utrecht

Roslin Castle

method	retrieved items	comment
Sorting DARMS code	6	Howard 1998; 13 known items
Earth Movers Distance	12	Typke et al. 2003; 17 known items
Proportional Transportation Distance	16	Typke et al. 2003, combining 3 different queries; 17 known items
RISM OPAC	17	key signature added
alignment (pi2mw0dur0)	23	exact pitch, no IMA, no duration; top-50 results only for all alignments
alignment (pi2mw2dur1)	24	exact pitch, normalised IMA, fixed duration (overall best rater combination)
alignment (pi2mw0dur1)	25	exact pitch, no IMA, fixed duration
alignment (pi2mw1dur1)	25	exact pitch, raw IMA, fixed duration
alignment (pi2mw2dur2)	26	exact pitch, normalised IMA, scaled duration
alignment (pi2mw0dur2)	27	exact pitch, no IMA, scaled duration
alignment (pi2mw1dur2)	27	exact pitch, raw IMA, scaled duration

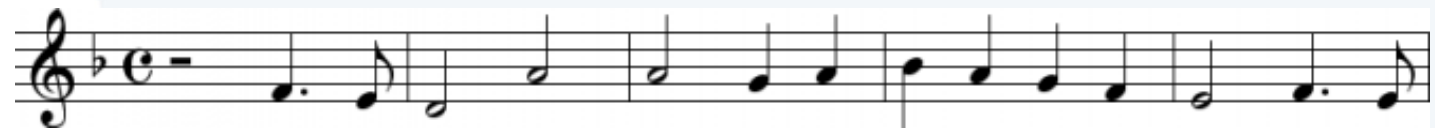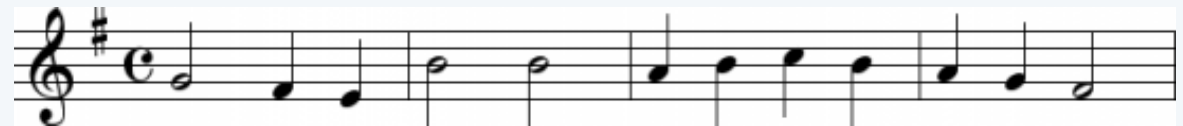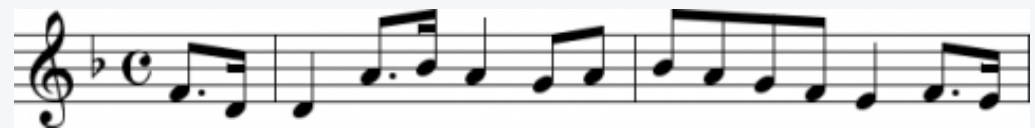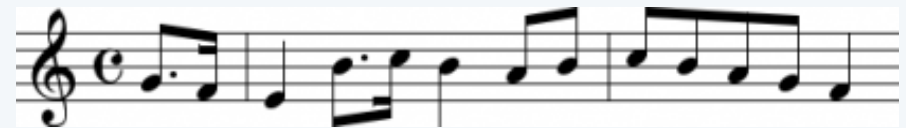29 relevant items known from RISM OPAC

Universiteit Utrecht

Additional results
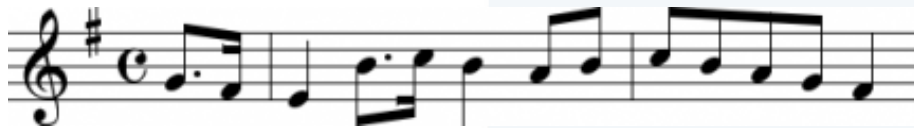
RISM OPAC query (transposition seems not to work)
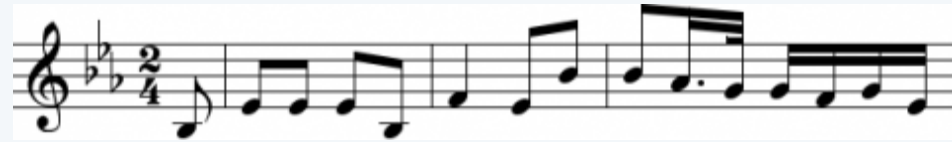
Monochord results not in RISM (transpositions excluded)

Exact pitch

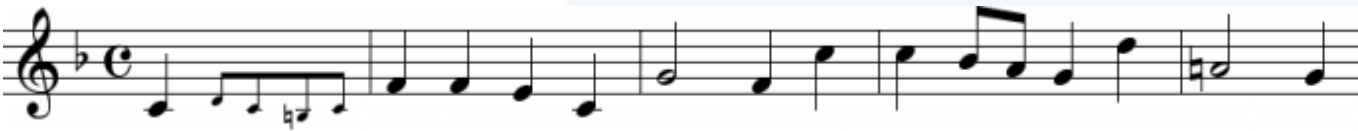Query: Anonymus. Andante Es-Dur

W.A. Mozart, Variations

A.F.J. Eberl, Variations

L. Jansa, Variations

I. Umlauff, Das Irrlicht, Excerpts

J.L. Böhner, Variations

F. Schubert. Harmonien rauschen im Hain
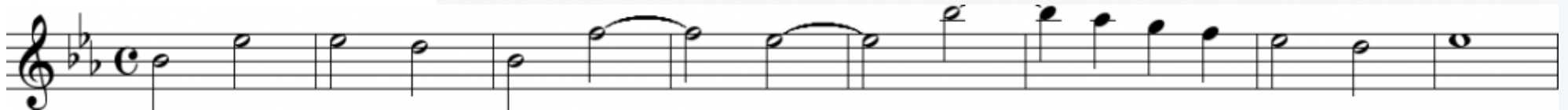
Other methods, new results

J.C.H. Rinck. Variations

J.L. Böhner, Variations

J.P.S. Schmidt. Ehre sei Gott in der Höhe   (???)

Conclusions

- alignment performs better than string matching
  - more relevant items
  - interesting non-identical matches
  - less need for query adaptation

- suggestion: integrate some combinations in RISM OPAC
  - exact pitch + fixed duration
  - exact pitch + scaled duration
  - other combinations?

- test utility in practice
  - what musicological insight do the non-identical matches provide?

- use MUSALIGN in other kinds of (MEI?) software

Universiteit Utrecht

THANK YOU

Universiteit Utrecht