Goldsmiths
UNIVERSITY OF LONDON

Pragmatic Digital Musicology

Tim Crawford, David Lewis
{ t.crawford | d.lewis } @gold.ac.uk

Utopian digital musicology

> “by the year 2000 [...], we should be able to transcribe, publish, analyze, and interpret the whole surviving corpus of historical music. That would leave room for only two kinds of musicologists: one group to predict what's the composers will do next, the other group [...] to check up on the accuracy of the predictions.
>
> The rest of us could retire to Tahiti.”

Jan LaRue, ‘New directions for style analysis’, *Musicology and the Computer 1966-2000* (New York, 1970)

Pragmatic Digital Musicology

- Humanities data is messy and complicated
- Corpus building is **never** complete
- User needs, computable algorithms and music theory seldom align

- Embrace (but evaluate) imperfection
- Adopt iterative strategies robust to changing data and systems

Goldsmiths
UNIVERSITY OF LONDON

I. Corpus building

The Electronic Corpus of Lute Music (ECOLM)

Phase I

Manual data entry (c. 250 pieces)

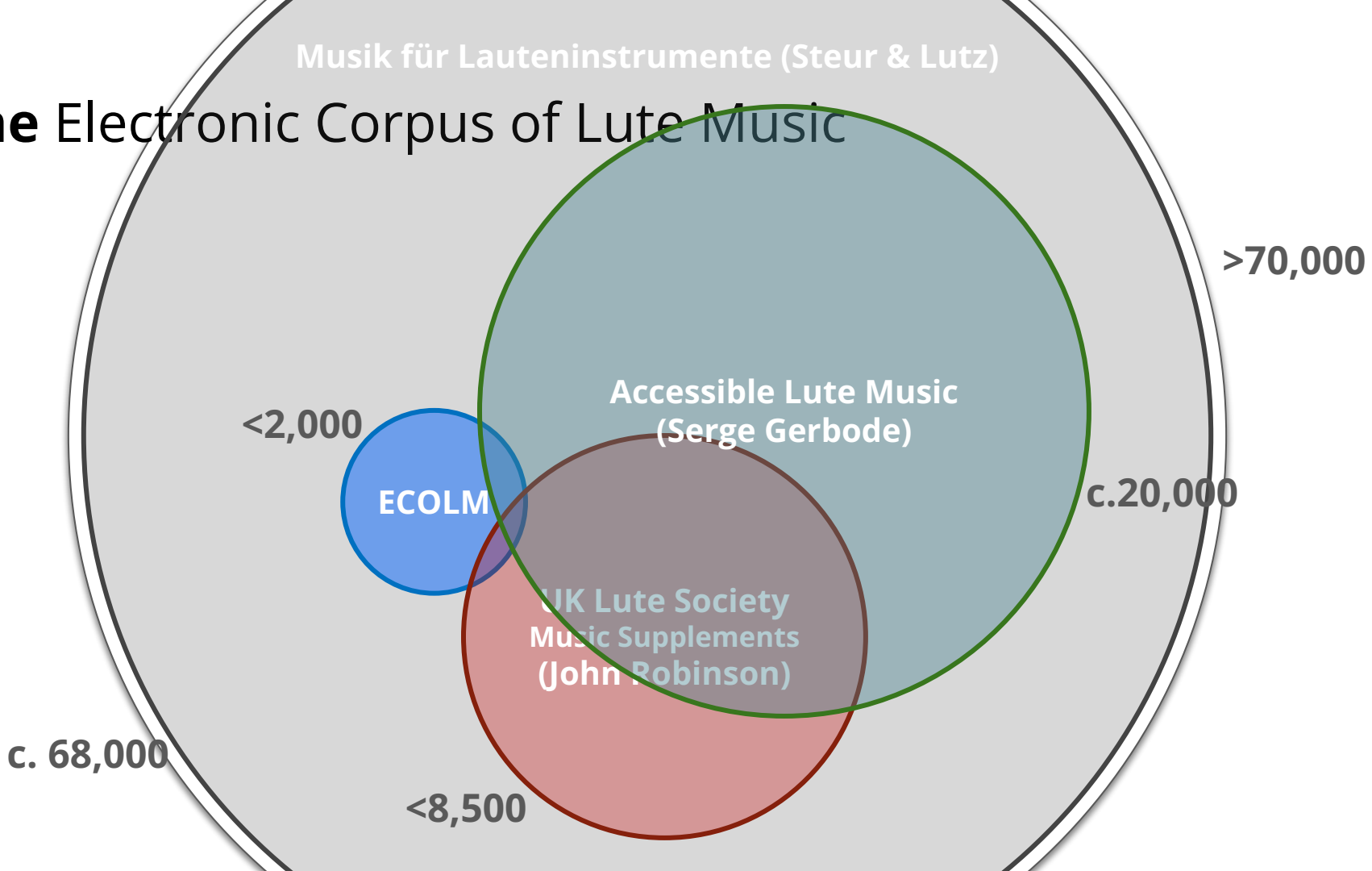
Phase II

Manual and corrected OMR (c. 850 pieces)

Phase III

Crowdsources corrected OMR (c. 1600 pieces)

The Electronic Corpus of Lute Music



Pragmatic corpus building

What sustainable steps can we take to make these useful to digital musicology?

How do we do this sustainably with the resource creators?

What happens when the resource creator stops?

Goldsmiths
UNIVERSITY OF LONDON

II. Search

Exploring renaissance music at scale with F-Tempo

Working with Golnaz Badkobeh and Alastair Porter

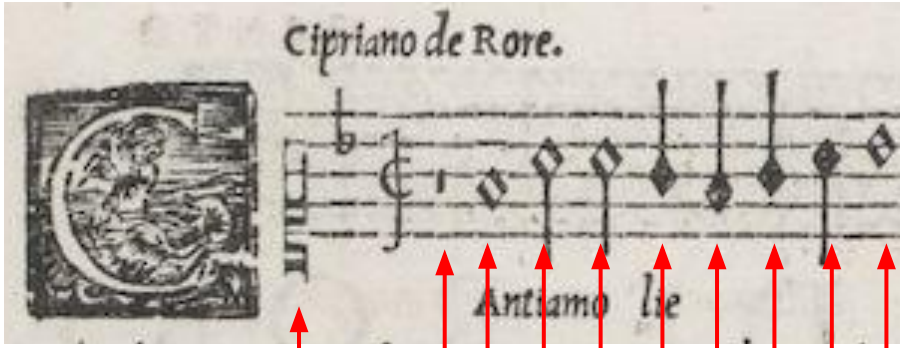
Full-text searching of musical scores

There is a shortage of large collections of machine-readable music notation...

... but there is an increasing amount of digital images of music notation

Just as Google Books uses Optical Character Recognition (OCR),

... we can use Optical Music Recognition (OMR)



Note-names: **F A A G F G A B**

Rest

Clef

Optical Music Recognition (OMR)

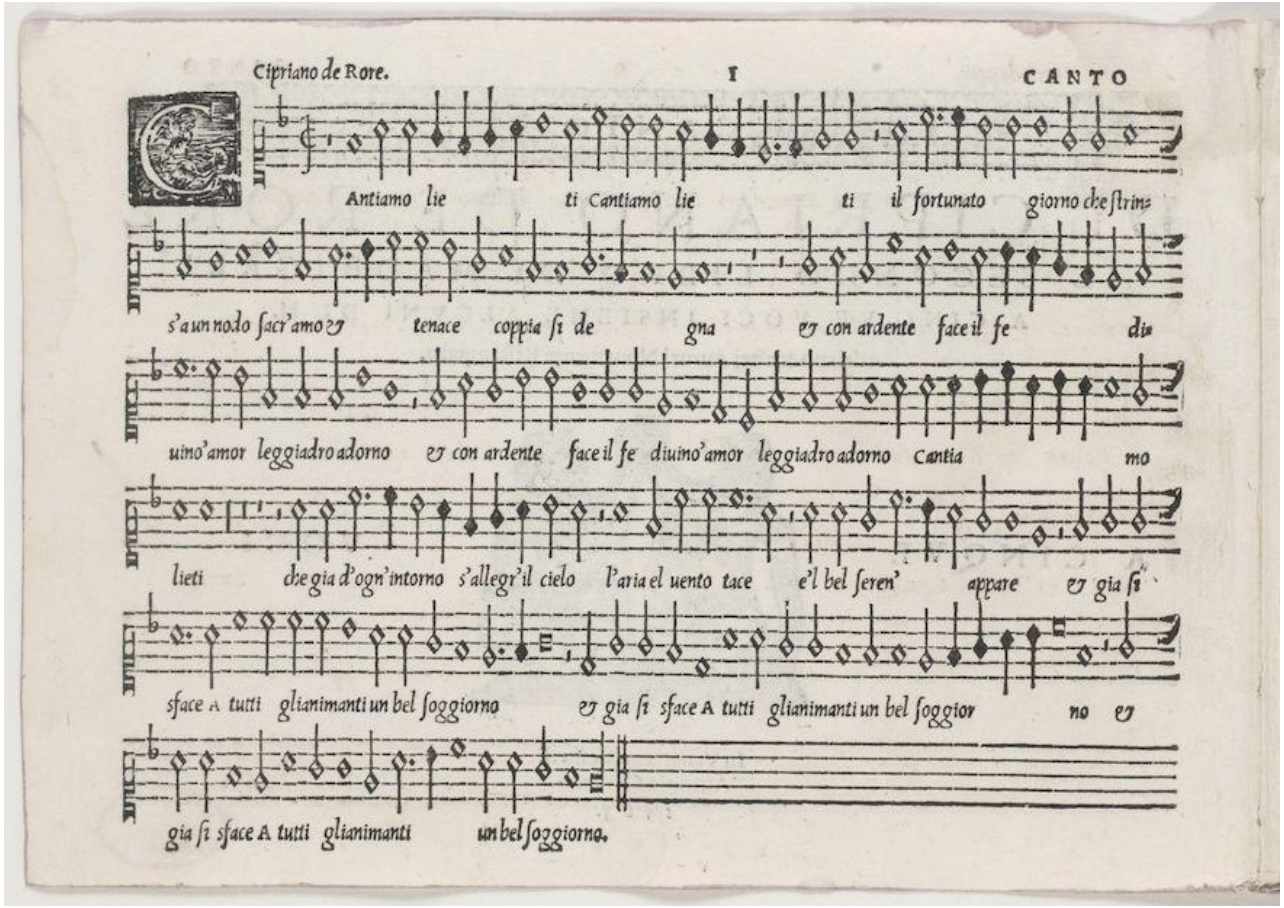
There are two types of errors that have a lasting effect:

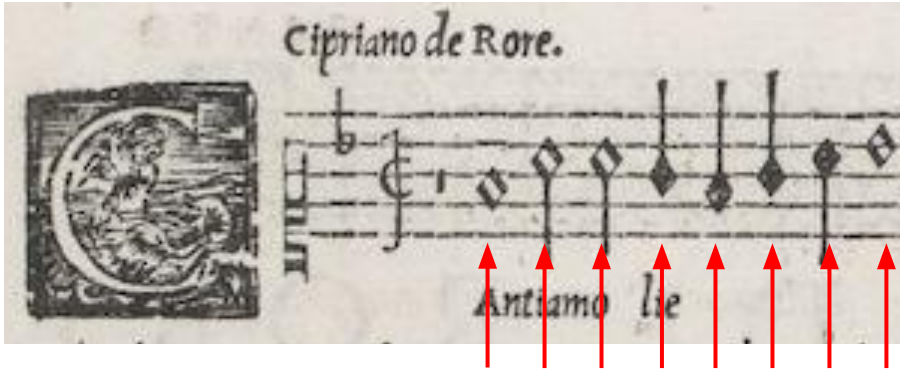
1. Rhythmic errors (everything afterwards is offset in time)
2. Clef errors (everything afterwards is offset in pitch)

Workarounds:

1. Ignore rhythm
2. Use relative pitch (intervals) – and ignore accidentals (‘diatonic pitch’)

Cipriano de Rore (1515-65), 5-voice madrigal, ‘Cantiamo lieti’, *Canto* part-book (Venice, 1563)





Note-names: F A A G F G A B

Intervals: 2 0 -1 -1 1 1 1

Encoding: B – a a A A A

Searching Music and Music Information Retrieval (MIR)

Searching music is complicated because music is a complex phenomenon

Music happens in our heads

What is music similarity? No-one agrees!

No surprise - there are probably as many kinds of musical similarity as people

Passage-Level Retrieval for Music

Given a large collection of images, from music libraries around the world, we want to be able to do two main things:

1. Find pages of music similar to a query page (***document-level*** retrieval)
2. Find occurrences of a musical pattern in the collection (***passage-level*** retrieval)

Each page has ~100 notes
We have 557k page images with 58 million notes

This demands ***indexing*** because of scale. Close analogy with Google Books, but indexing ***musical*** content, not text

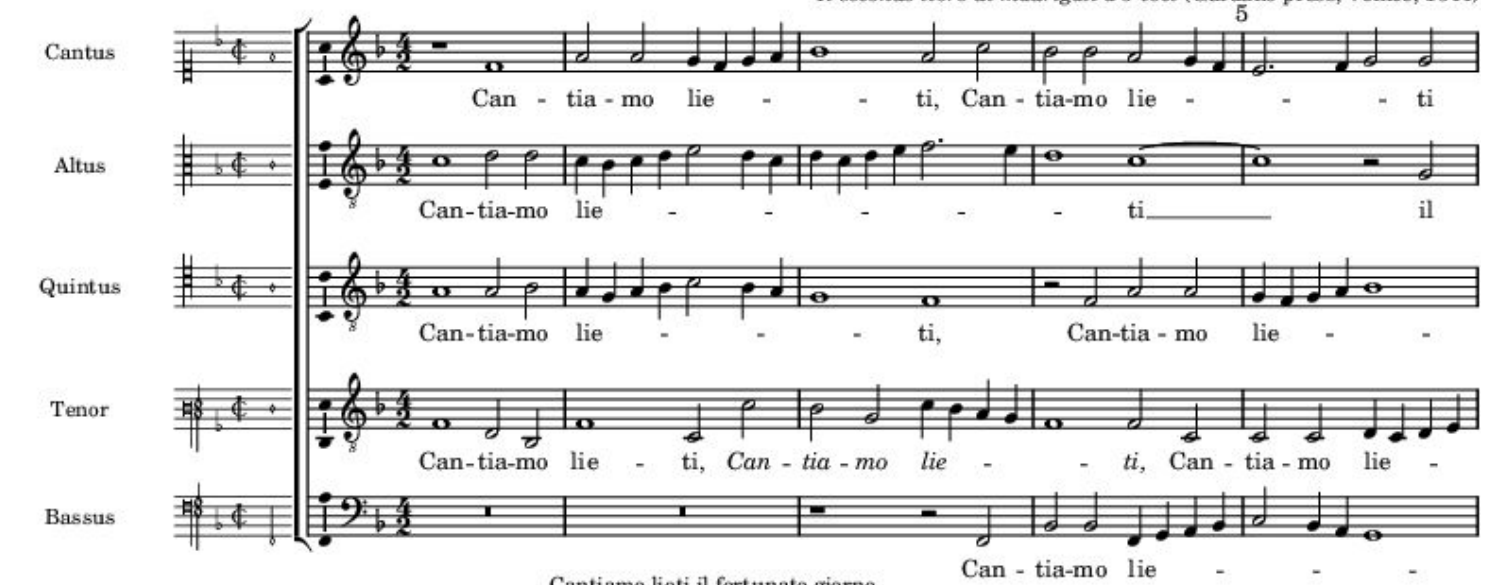
Cantiamo lieti il fortunato giorno

Prima parte

Anonymous sonnet

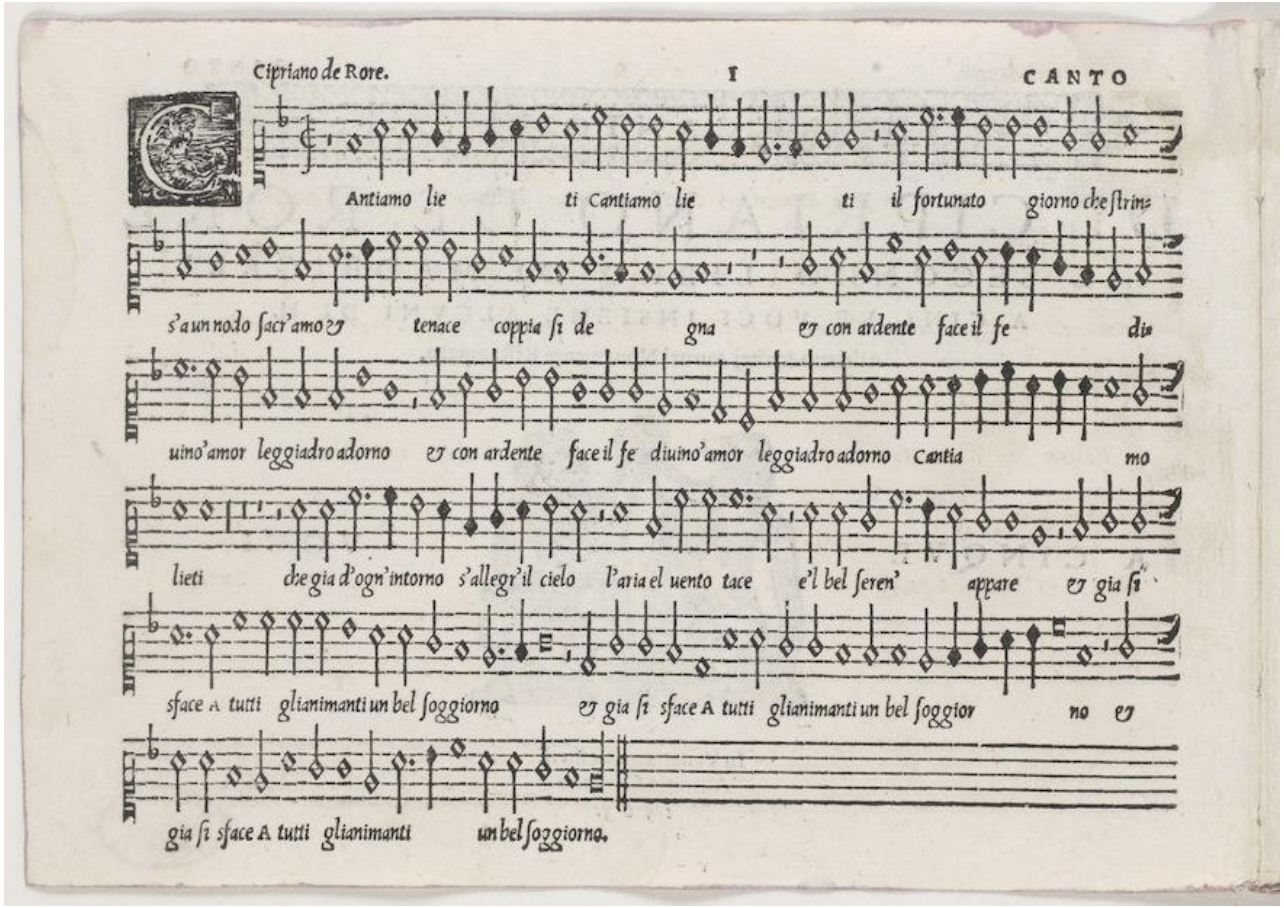
Cipriano de Rore (c.1515-1565)

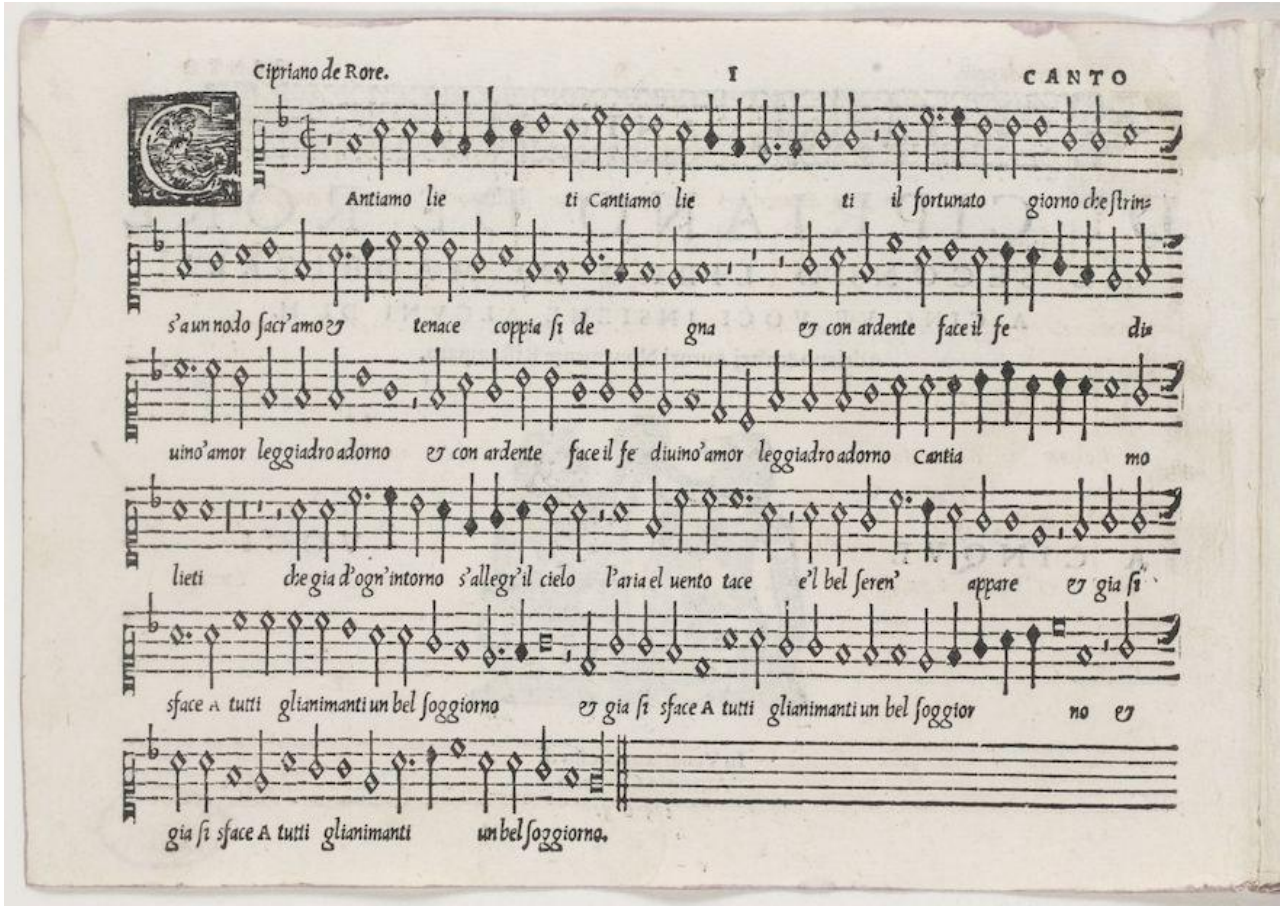
Il secondo libro di madrigali à 5 voci (Gardano press, Venice, 1544)



Cantiamo lieti il fortunato giorno
Che strins'a un nodo sacr'almo e tenace
Coppia si degna e con ardente face
Il fe divino amor leggiadro adorno.
Cantiamo lieti, che già d'ogn'intorno
S'allegr'il cielo, l'aria e 'l vento tace,
E 'l bel sereno appare, e già si sface
A tutti gli animanti un bel soggiorno.

Cipriano de Rore (1515-65), 5-voice madrigal, ‘Cantiamo lieti’, *Canto* part-book (Venice, 1563)





B-aaAAAaBa-aaaaAA-B-a--b-AbAAAcBAA-abAb-Aa-aAAbDbAaAaaaaAD-ac--CbBaBb--bAbaC--AA--AA
bAa-aA--B-aabAAAabD--b-aCaaa-bA-A-B---aa-aaaAAC-abD-a-a--aAAAAAdA-baCa-bCAAb-aaaE

Ngrams

B-aaAAAaBa-aaaaAA-B-a--b-AbAAAcBAA-abAb-Aa-aAAbDbAaAaaaaAD-ac--CbBaB b--bAbaC--AA--AAbAa-aA--B-aabAAAabD--b-aCaaa-bA-A-B---aa-aaaAAC-abD- a-a--aAAAAAdA-baCa-bCAAb-aaaE

We can split the string into smaller units for indexing
A simple distance measure is how many are shared between the query and each document

Length	ngrams				
4	**B-aa**	**-aaA**	**aaAA**	**aAAA**	**AAAa**

B-aaAAAaBa-aaaaAA-B-a--b-AbAAAcBAA-abAb-Aa-aAAbDbAaAaaaaAD-ac--CbBaBb--bAbaC--AA--AAbAa-aA--B-aabAAAabD--b-aCaaa-bA-A-B---aa-aaaAAC-abD-a-a--aAAAAAdA-baCa-bCAAb-aaaE

4 - 10grams (first 6 of each length): total is 1113 ngrams

Length ngrams

4: B-aa -aaA aaAA aAAA AAAa AAaB ...

5: B-aaA -aaAA aaAAA aAAAa AAAaB AAaBa ...

6: B-aaAA -aaAAA aaAAAa aAAAaB AAAaBa AAaBa- ...

7: B-aaAAA -aaAAAa aaAAAaB aAAAaBa AAAaBa- AAaBa-a ...

8: B-aaAAAa -aaAAAaB aaAAAaBa aAAAaBa- AAAaBa-a AAaBa-aa ...

9: B-aaAAAaB -aaAAAaBa aaAAAaBa- aAAAaBa-a AAAaBa-aa AAaBa-aaa ...

10: B-aaAAAaBa -aaAAAaBa- aaAAAaBa-a aAAAaBa-aa AAAaBa-aaa AAaBa-aaaa ...

Absent Words

```
B-aaAAAaBa-aaaaAA-B-a--b-AbAAAcBAA-abAb-Aa-aAAbDbAaAaaaaAD-a
c--CbBaBb--bAbaC--AA--AAbAa-aA--B-aabAAAabD--b-aCaaa-bA-A-B-
--aa-aaaAAC-abD-a-a--aAAAAAdA-baCa-bCAAb-aaaE
```

Absent Words (AWs): letter sequences that do not appear (v. many)

Minimal Absent Words (MAWs): AWs all of whose substrings do appear (few)

Example:

--a-

Absent Words

B-aaAAAaBa-aaaaAA-B-a--b-AbAAAcBAA-abAb-Aa-aAAbDbAaAaaaaAD-a
c--CbBaBb--bAbaC--AA--AAbAa-aA--B-aabAAAabD--b-aCaaa-bA-A-B-
--aa-aaaAAC-abD-a-a--aAAAAAdA-baCa-bCAAb-aaaE

Absent Words (AWs): letter sequences that do not appear (v. many)

Minimal Absent Words (MAWs): AWs all of whose substrings do appear (few)

Example:

--a- = **--a**

Absent Words

B-aaAAAaBa-aaaaAA-B-a--b-AbAAAcBAA-abAb-Aa-aAAbDbAaAaaaaAD-a
c--CbBaBb--bAbaC--AA--AAbAa-aA--B-aabAAAabD--b-aCaaa-bA-A-B-
--aa-aaaAAC-abD-a-a--aAAAAAdA-baCa-bCAAb-aaaE

Absent Words (AWs): letter sequences that do not appear (v. many)

Minimal Absent Words (MAWs): AWs all of whose substrings do appear (few)

Example:

--a- = --a + -a-

B-aaAAAaBa-aaaaAA-B-a--b-AbAAAcBAA-abAb-Aa-aAAbDbAaAaaaaAD-ac--C
bBaBb--bAbaC--AA--AAbAa-aA--B-aabAAAabD--b-aCaaa-bA-A-B---aa-aaa
AAC-abD-a-a--aAAAAAdA-baCa-bCAAb-aaaE

MAWs of length 4-10: total is 328 MAWs (none longer than 7)

4: ---- --a- a--- ---A --A- -A-- A--- aa-- a--A --Aa -A-a A--a A-a- Aa-- -A-A aaA- -aAa aA-a aAa- aA-A
-Aaa A-aa Aaa- -AaA A-aA AaA- -AAa A-Aa AAa- -AAA A-AA AA-A AAA- aaAa aAaA aAAa AaaA AaAA AAaa AAaA aaab aAab
aAaB Aaab AAAb aAAc AAAC aAAd AaAD AaaE a-ab aA-b aA-B -Aab Aa-b -AaB A-Ab AA-b AabA AAba AaBb aabD a-ac a-aC
A-ac A-aC -AAc -AAC -aAD -AAd -aaE --ab -a-b a-b- a--B --Ab -A-b -Ab- A--b A-b- Ab-- a-ba abA- -Aba A-bA AbA-
abAa aBaB AbAb A-bC -AbD AbD- abDb --ac --aC -aC- a--C A--C AC-- aC-a AD-- ---b -b-- b--- ---B --ba b--a b-a-
b--A b-A- bA-- B--a Ba-- B--A b-aA -bAa bA-a -bAA b-AA bAA- B-aA bAaa bAAa BAAa BAAA bAab bAaB bAAb BAAb bAAc
bAAC BAAc BAAC bAAd BAAd b-ab bA-b bA-B B-ab Ba-b bAbA BaBa bAbD b-ac B-ac B-aC b--B B--b B--B bBa- Bb-a Bb-A
--bC b--C B--C c--- ---C C--- c--a c--A C--a C-a- Ca-- C-aa Ca-a Caa- C-aA CAA- CaaA CAAa CAAA Caab CAAc CAAC
CAAd CaaE C-ac C-aC c--b c--B C--b C--B C--C D--- dA-- D--a D--A dA-a dA-A D-aa D-aA dA-B D-ab D-aC D--B DbA-
DbAA DbAb D--C

5: ---aA --aA- --aaa -a-aa -aaa- a--aa a-a-a a-aa- aa-a- --aaA -a-aA -aAA- aA--A aAA-- Aa-a- -AA-a aaa-a
aaaa- a-aaA aa-aA aAA-a Aa-aa Aaaa- aaaaa AaaaA AAAAa aaAAb aAAAc AAAAc aAAAd aaaaE AaaaE a-aab aAAb- aAAbA
-aAAC -aaAD --aab -aa-b -AAb- -AA-B AA--B AAb-A AAbAA aabAb aa-bC A-abD -AAbD a--bA -abAA -AbAa AbAaA a-bAb
Ab-Ab abAba Ab-aC AbaCa -aCa- AD-a- --B-- --bA- -b-aa b-aa- -b-Aa B-a-a B-aa- Ba-a- BAA-- b-aaA B-aaa Ba-aA
bAAAA bAAAd b-aab BAA-B -bAb- bAb-a -baC- baCaa b--b- bD-ac Caaaa CaaaA CaaaE CAAbA CAAbD C-abA Ca-bA D-a--
D--bA DbAa-

6: A--AA- -aaAA- a-aAAA -aAAAa aaaAAA aaAAAA AAAAAA aAAAab aAAAAd -aaAAC -aaaAD a-aaaE --aAAb a-a--b
AA-B-- AabD-a a--b-a -A-B-a A-B-aa AbAAAa abAAAc -abD-- --B-a- B-a--a -B-aaA b-aaaa b-aaaA bAAAaB C--AAb
D--b-A

7: -Aa-aA- -aaaAA- aa-aaaa AaaaaAA aaaaAAC -aaaaAD bAa-aAA Ba-aaaA

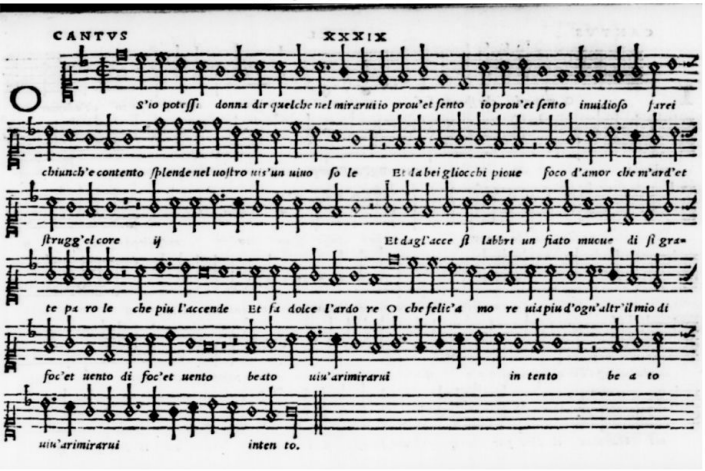
User-interface: Find same music but different page-layout

Early Music Online Search | Search within Corpus | Search with Image Upload

Show Examples | Help

Previous Book | Previous Page | Show a Random Page | Next Page | Next Book

Query page: GB-Lbl_K2h7_092_1



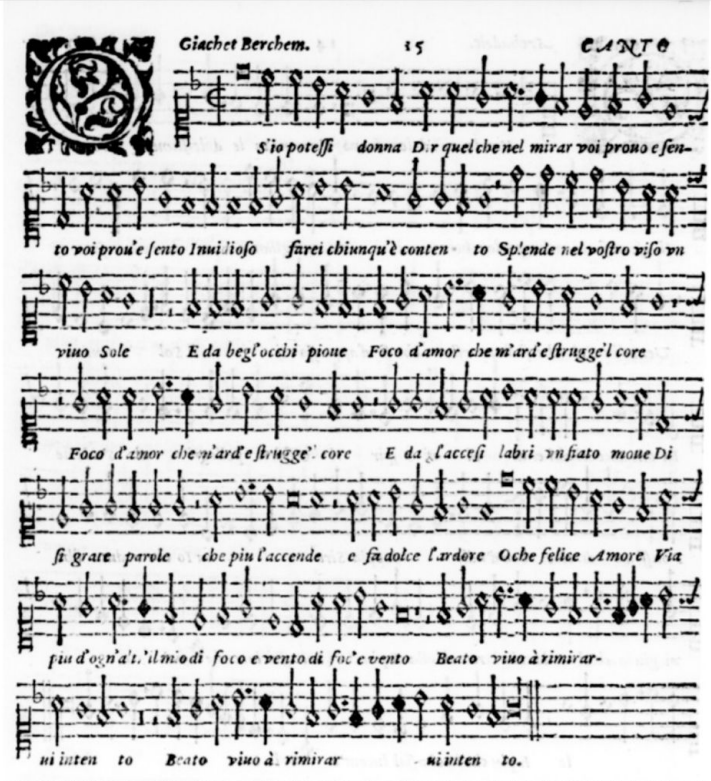
15 results - 234 words in query

Page ID	Match Score
GB-Lbl_K2h7_092_1	100.00
GB-Lbl_A220a_016_1	42.24
GB-Lbl_K2h3_010_1	41.40
GB-Lbl_K2h8_009_0	26.5
F-Pn_RES-VMC-167-0_4r_010	22.2
GB-Lbl_A324c_022_1	11.4
GB-Lbl_K2h8_016_1	11.0
PL-Wn_27517200_32144423	11.9
GB-Lbl_A220a_004_1	11.9
GB-Lbl_A277_123_1	11.4
PL-Wn_26426901_26771421	11.0
GB-Lbl_D33b_2_056_0	10.8
GB-Lbl_K4d14_015_1	10.6
GB-Lbl_K4g3_142_0	10.2
GB-Lbl_K3e19_098_1	10.6

Result Ranking: Jaccard Distance

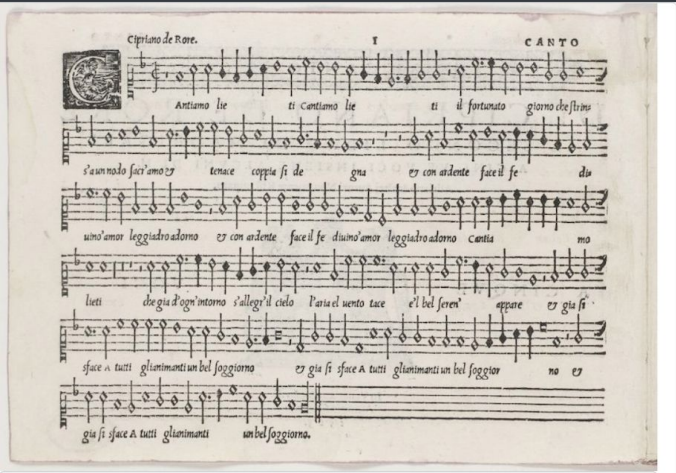
Results to display: 10

Provide judgements to help improve the system



User-interface: Unknown item search - image uploaded by user

Early Music Online Search | Search within Corpus | Search with Image Upload

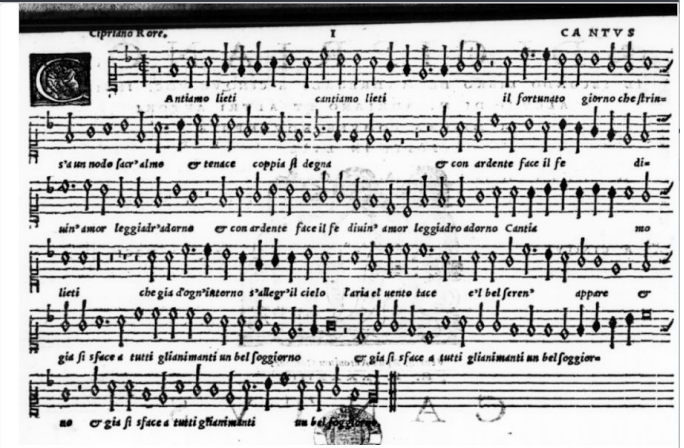


Choose File _Rore_Cantiamo.jpg

Upload and Search!

20 results - 337 words in query

Page ID	Match Score
GB-Lbl_A70b_002_0	37.68
GB-Lbl_A70g_048_0	28.60
GB-Lbl_K3g11_033_0	16.62
GB-Lbl_K8b5_066_0	15.65
GB-Lbl_A70g_071_0	15.47
GB-Lbl_C215_007_1	15.31
GB-Lbl_A337b_064_1	14.96
D-Bsb_Parangon_03_1543_060_0	14.94
GB-Lbl_D33b_2_104_0	14.93
GB-Lbl_A254_013_0	14.87
PL-Wn_32432623_33526443	14.75
GB-Lbl_K2a4_214_0	14.60
GB-Lbl_A337b_017_0	14.59
GB-Lbl_A70n_024_0	14.58
GB-Lbl_A70b_044_0	14.52
PL-Wn_27315807_33917392	14.50
GB-Lbl_A239_020_1	14.48
GB-Lbl_A324c_050_0	14.47
GB-Lbl_A254_009_0	14.45
GB-Lbl_A70g_059_1	14.39



182/337 words in page match the query

Uploaded image

Best match in database

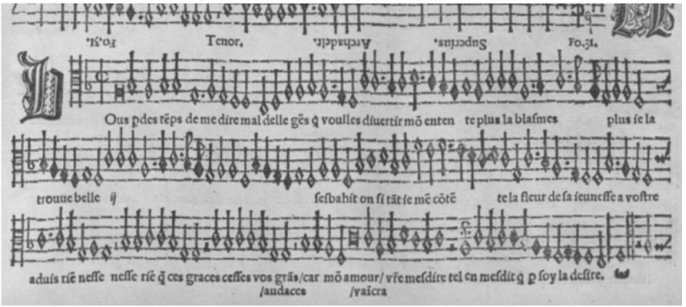
User-interface: Unexpected ‘concordance’: Same music - different work title and text

Early Music Online Search | Search within Corpus | Search with Image Upload

Show Examples | Help

Previous Book | Previous Page | Show a Random Page | Next Page | Next Book

Query page: D-Bsb_Parangon_03_1543_inv_060_0



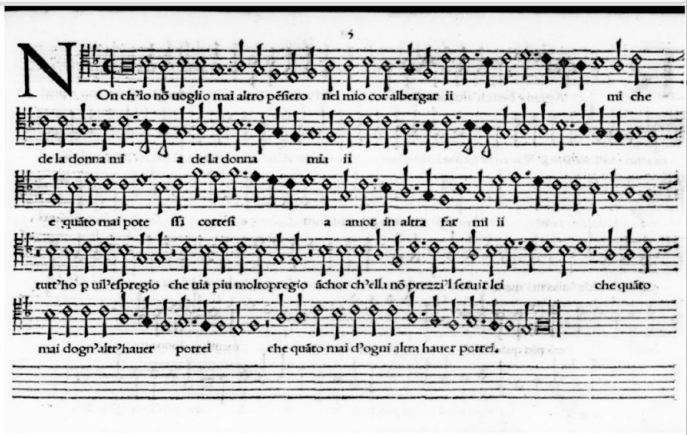
15 results - 171 words in query

Page ID	Match Score
D-Bsb_Parangon_03_1543_inv_060_0	100.00
GB-Lbl_K2h3_031_1	20.96
GB-Lbl_K2d19_006_0	14.43
PL-Wn_26426901_26771428	11.83
GB-Lbl_K2c3_065_0	10.97
GB-Lbl_K4g3_126_0	10.65
GB-Lbl_K8b7_006_0	10.58
GB-Lbl_A431_028_1	10.20
GB-Lbl_K3g4_197_0	10.04
PL-Wn_20209512_24063153	10.00
GB-Lbl_A251b_060_1	9.96
GB-Lbl_K3a15_179_0	9.88
PL-Wn_26426907_26770166	9.88
GB-Lbl_K3a14_118_0	9.83
GB-Lbl_K3f9_048_0	9.78

Result Ranking: Jaccard Distance

Results to display: 10

Provide judgements to help improve the system



70/233 words in page match the query

French chanson, ‘Vous perdez temps’ by Arcadelt (in database)

Best next match in database: Italian madrigal, ‘Non ch’io non volgio’ by Arcadelt

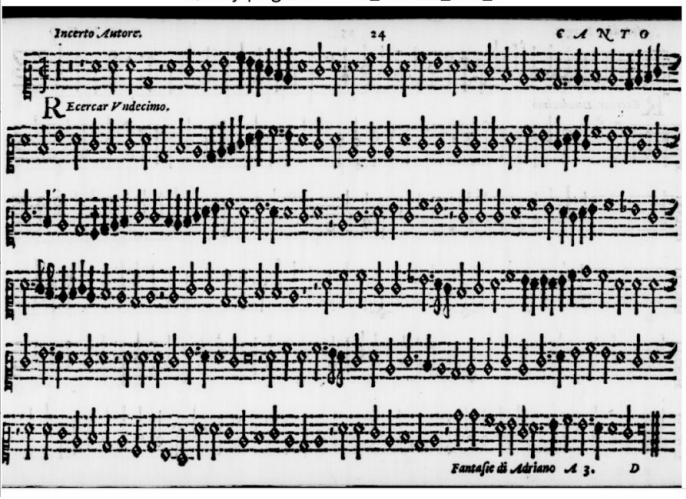
User-interface: Identify composer for a previously anonymous work

Early Music Online Search | Search within Corpus | Search with Image Upload

Show Examples | Help

Previous Book | Previous Page | Show a Random Page | Next Page | Next Book

Query page: GB-Lbl_A569c_013_1



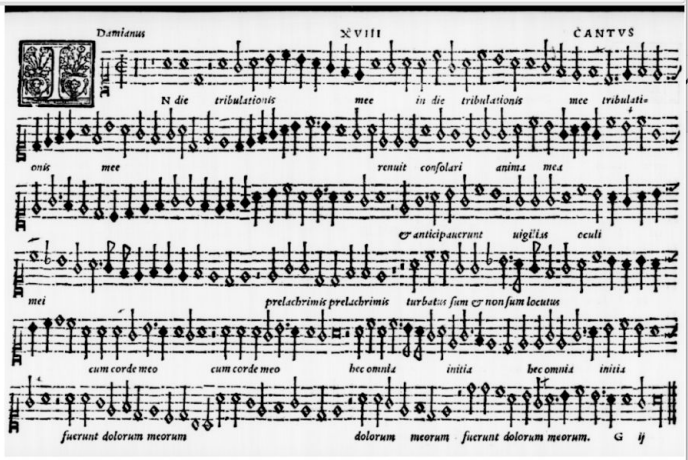
15 results - 400 words in query

Page ID	Match Score
GB-Lbl_A569c_013_1	100.00
GB-Lbl_K3d8_010_1	40.60
GB-Lbl_A569c_029_1	14.06
GB-Lbl_K4g2_005_1	12.92
GB-Lbl_A70n_021_0	12.69
GB-Lbl_K8k1_011_0	12.00
GB-Lbl_A103a_015_1	11.81
GB-Lbl_K3d9_008_1	11.61
GB-Lbl_K3d5_022_0	11.59
GB-Lbl_K3d2_002_0	11.59
GB-Lbl_A251b_093_1	11.48
GB-Lbl_K4d14_096_0	11.47
GB-Lbl_K3d2_056_1	11.38
GB-Lbl_A296d_007_0	11.37
PL-Wn_26567886_27525148	11.37

Result Ranking: Jaccard Distance

Results to display: 10

Provide judgements to help improve the system



216/348 words in page match the query

Instrumental recercar by ‘Unknown composer’ (*Incerto Autore*)

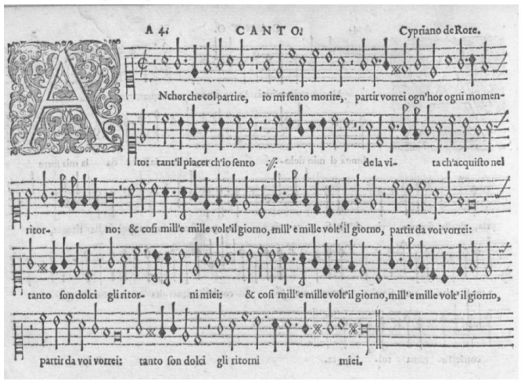
Best next match in database: ‘In die tribulationis’ by Damianus

A well-known madrigal by Cipriano da Rore, ‘Ancor che col partire’, *Cantus* part

Previous Book | Previous Page | Random Page

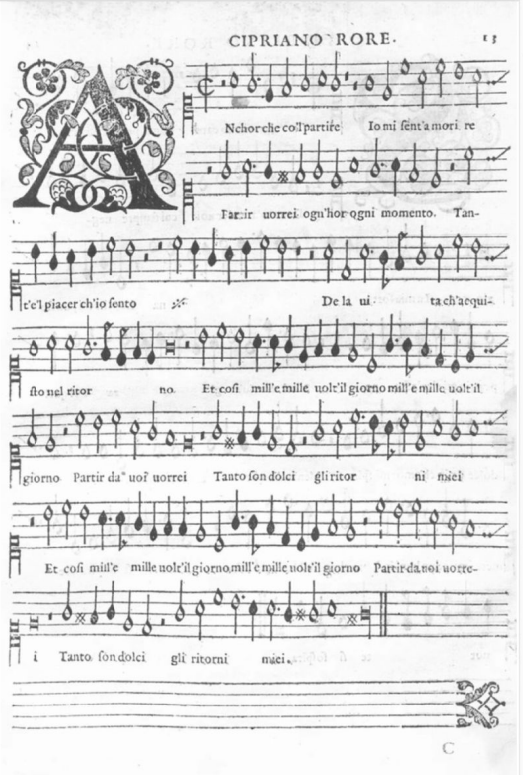
Next Page | Next Book

Query page: D-Mbs_bsb00074752_00018



20 results - 155 words in query

Page ID	Match Score
1. D-Mbs_bsb00074752_00018	100.00
2. D-Mbs_bsb00072033_00014	96.20
3. D-Mbs_bsb00072031_00014	96.20
4. GB-Lbl_A324c_008_0	95.00
5. GB-Lbl_A70n_009_0	93.83
6. D-Mbs_bsb00071785_00029	74.59
7. D-Mbs_bsb00071900_00004	69.47
8. PL-Wn_26567912_27022617	64.53
9. D-Mbs_bsb00071910_00030	61.62
10. PL-Wn_22116341_23420224	59.67
11. PL-Wn_27033479_27053649	50.9
12. cpdl_Cipriano_de_Rore_Ancor_che_col_partire_Rore_Soprano	46.7
13. cpdl_Cipriano_de_Rore_11-anchor_che_col_partire---0-score_Canto	44.4
14. D-Mbs_bsb00071899_00013	37.0
15. cpdl_Cipriano_de_Rore_De_Rore-Anchor_chi_col_partire_Canto	31.
16. D-Mbs_bsb00092950_00217	1
17. D-Mbs_bsb00074221_00333	1
18. PL-Wn_26426907_26770184	1
19. GB-Lbl_K3g4_036_0	1
20. D-Mbs_bsb00094226_00020	1



135/161 words in page match the query

Extending this

- Index whole library collections directly through IIIF
- Index digital scores (such as CPDL)
- Index polyphonic and embellished sources (tablatures) - currently hard

Goldsmiths
UNIVERSITY OF LONDON

III. Musicological close readings

Working with many people (Kevin Page, David Weigl, Johannes Kepper, Mark Saccamano, …)

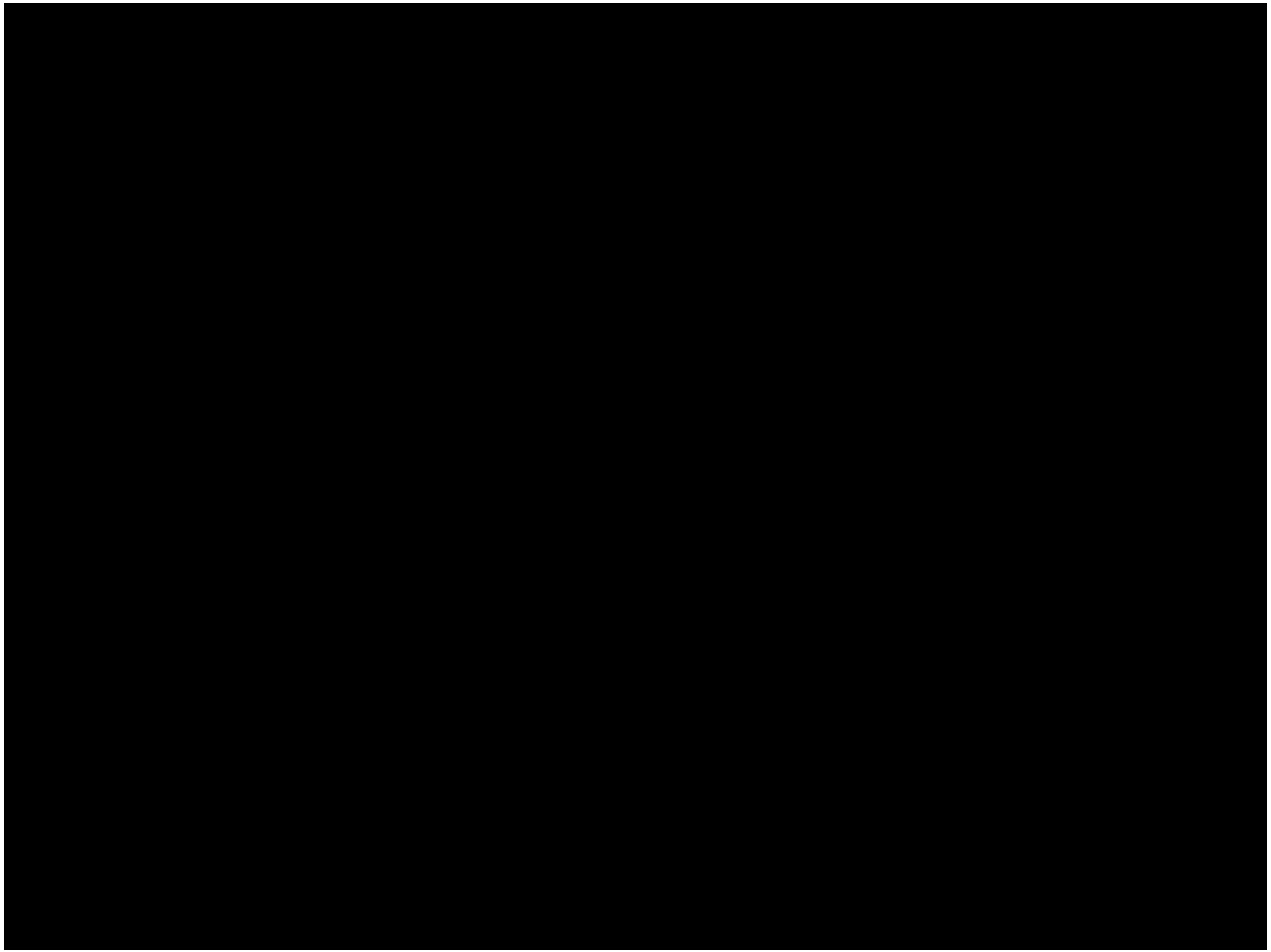
Utopian approaches?

Musicological research...

- ...may start with no clearly-defined corpus
- ...may build up incrementally
- ...may work with an expanding, changing corpus
- ...may work with evolving methodologies

Incremental approaches to tooling...

- Support tool chaining via data standards
- Design data assembly into the research process
- Accept mixed, imperfect, changing data



Musicological research...

- ...may start with no clearly-defined corpus
- ...may build up incrementally
- ...may work with an expanding, changing corpus
- ...may work with evolving methodologies

Goldsmiths
UNIVERSITY OF LONDON

Pragmatic Digital Musicology

Tim Crawford, David Lewis
{ t.crawford | d.lewis } @gold.ac.uk