A Retrospective on Optical Music Recognition Research

Ichiro Fujinaga

Music Technology Area, Schulich School of Music

McGill University

The Pioneers

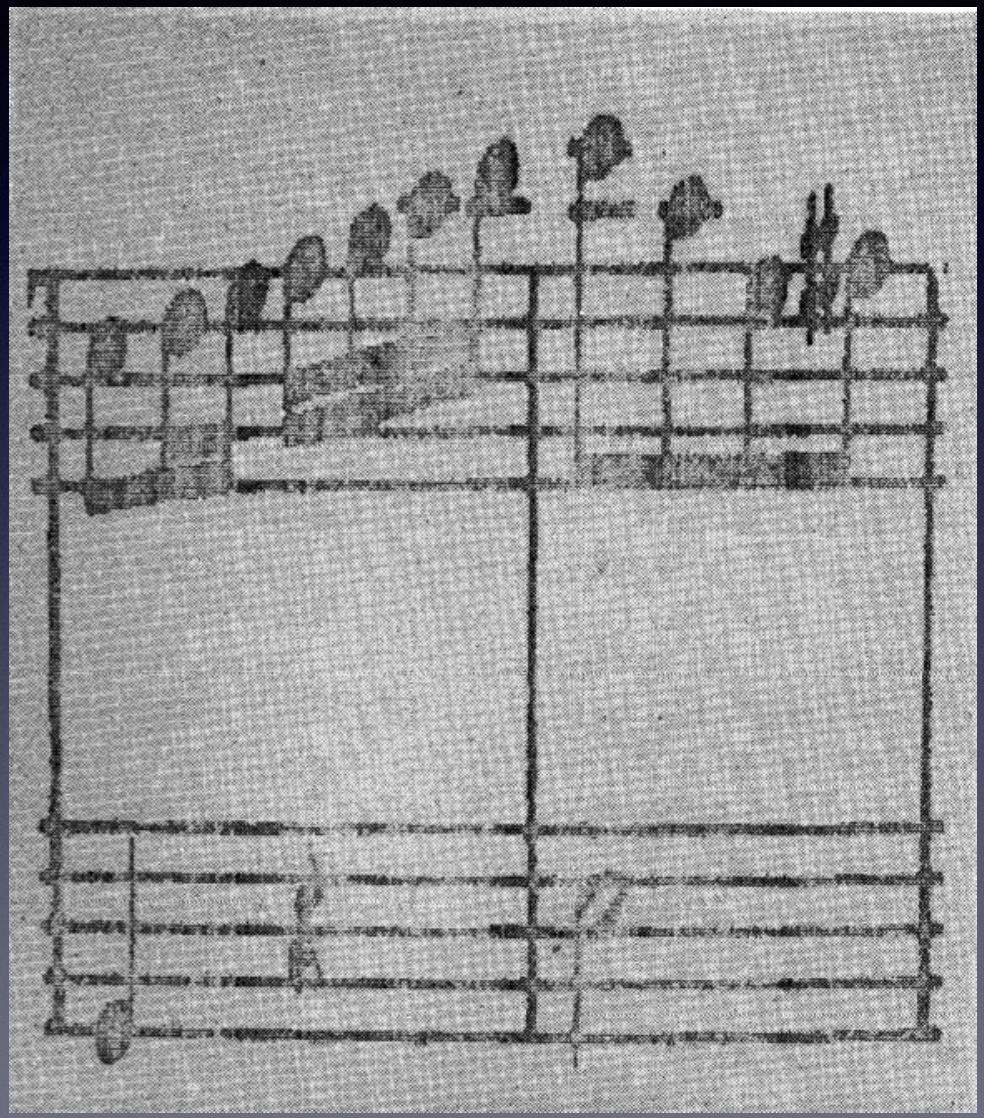
Denis Pruslin AKA The Tool

David Prerau

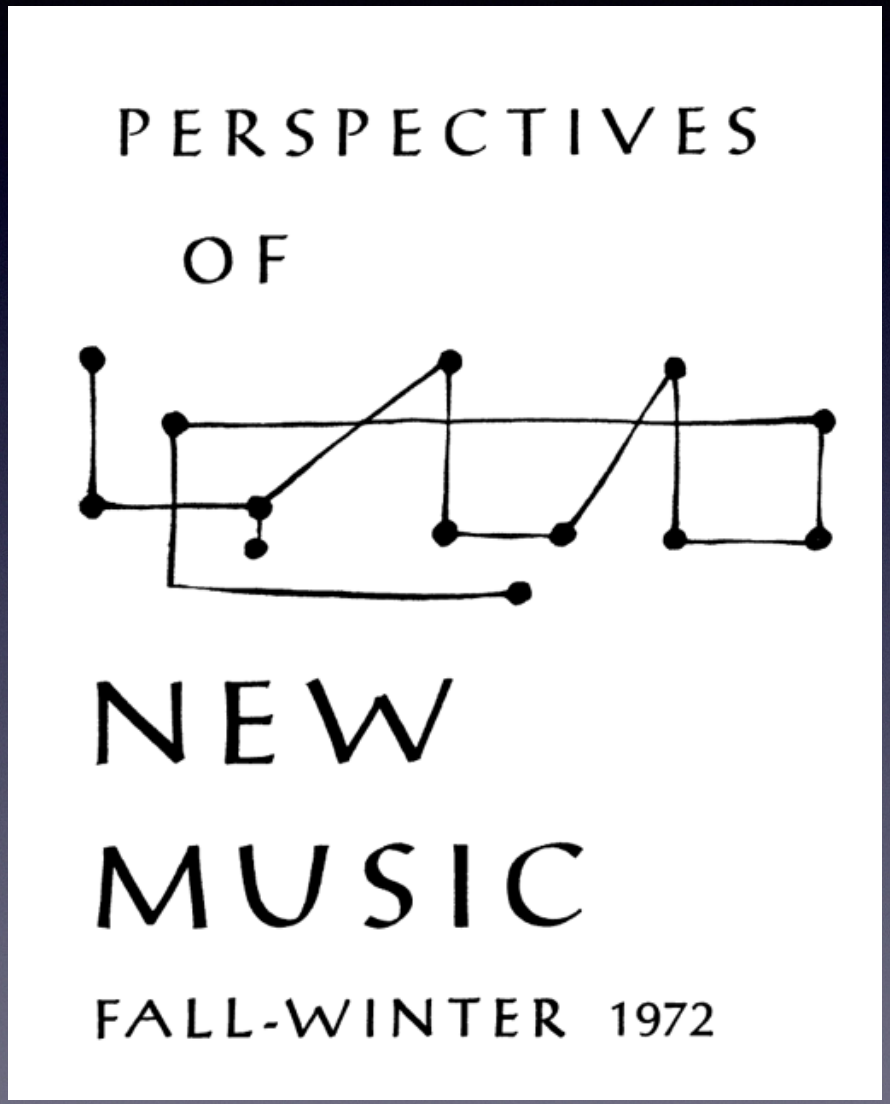
Baker House, c. 1959

Denis Pruslin with grandson Kevin, 2010/11

The first published digital scan of music (1970)



Review by Michael Kassler (1972)



Review: Optical Character-Recognition of Printed Music: A Review of Two Dissertations
Michael Kassler
Page 250 of 250-254

OPTICAL CHARACTER-RECOGNITION OF PRINTED MUSIC: A REVIEW OF TWO DISSERTATIONS

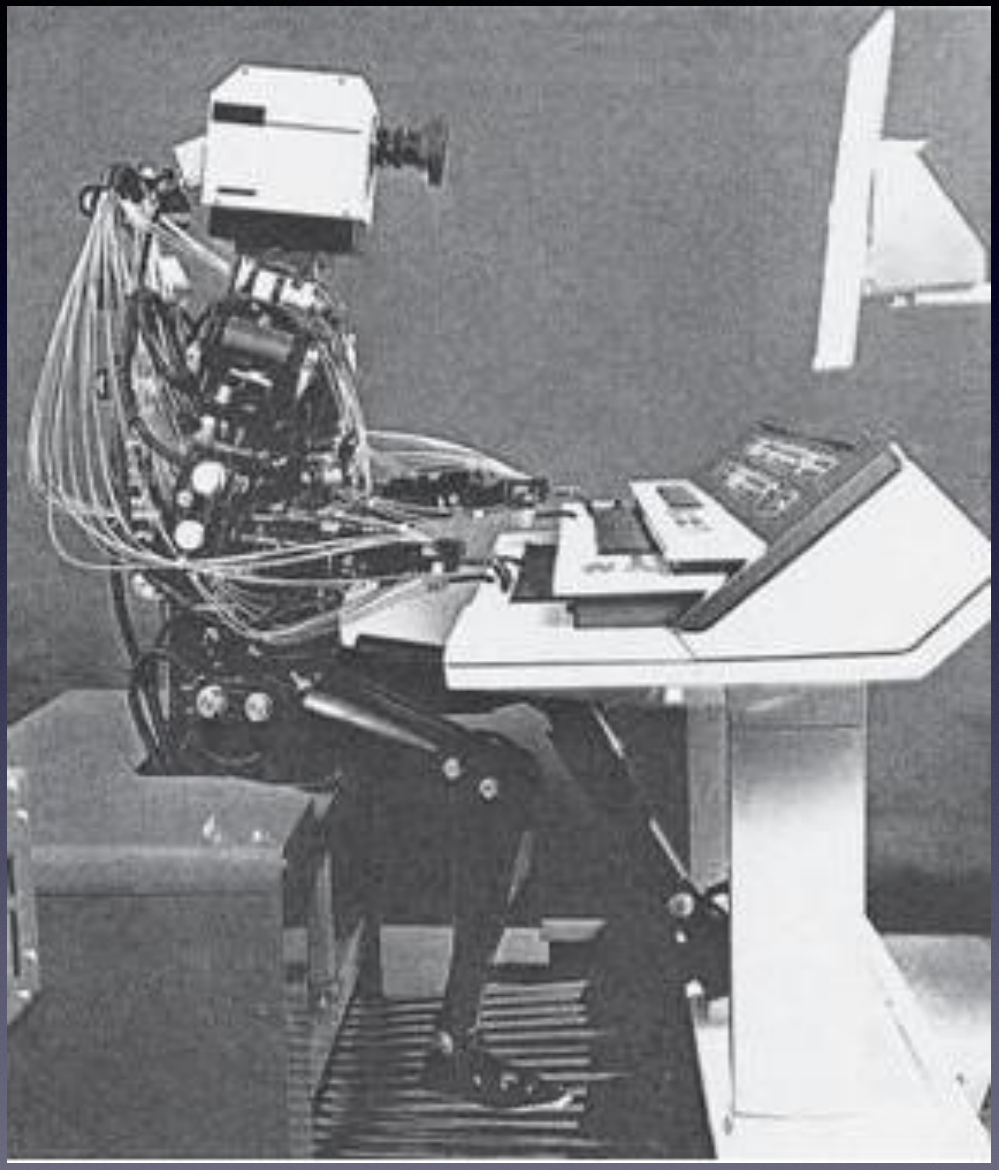
AUTOMATIC RECOGNITION OF SHEET MUSIC. By Dennis Howard Pruslin. Sc. D. Dissertation, Massachusetts Institute of Technology, 1966.

COMPUTER PATTERN RECOGNITION OF STANDARD ENGRAVED MUSIC NOTATION. By David Stewart Prerau. Ph. D. Dissertation, Massachusetts Institute of Technology, 1970.

Readers of *Perspectives* scarcely need be reminded of the pre-eminence of the written-musical domain (i. e., that domain of musical experience in which music is presented visually in one or another system of musical notation) in musicology: before Edison composers could not produce records of their work in the sounded-musical domain, and other domains of musical experience such as the tactile domain utilized in the Braille system have been employed comparatively infrequently; and even after Edison various extra-musical considerations (such as copyright law and the relatively high cost of sound-processing machinery) have joined with tradition to keep the written-musical domain a principal mode of non-transient musical communication. Within this domain various systems of musical notation have achieved various degrees of currency at various places and times, but of all these systems one—the current common musical notation ('CCMN' for short)—has dominated: virtually all music printed has been printed in one or another 'dialect' of CCMN: even music originally noted in another system generally has been transcribed into CCMN before printing.

In recent years digital computers have become more efficient and more prevalent, so that today, at least in computationally well-developed parts of the world, it no longer is unreasonable to delegate, or to plan to delegate, musical processes to electronic computing machinery. Of course, many musical processes do not involve previously recorded musical compositions: perhaps it is to the comparatively early success of a few such computer-mediated processes that an unfortunate synecdochic misidentification of 'computer music' with 'synthesizing sound through the use of a digital computer' has arisen.[1] But (and of this too readers will be well informed) central to musicology are processes that do involve prior musical compositions, and for the full delegation of these processes to computing machinery the relevant compositions must be put into computer-acceptable form. Human key-punchers can transcribe from CCMN onto (say) punch cards (at Princeton University the Masses of Josquin were so transcribed, at a rate of approximately 20 minutes per printed CCMN page), but as this task clearly requires no intelligence beyond that with which machines can be endued it is only natural to consider

1984: Wabot-2



https://www.scaruffi.com/mind/ai/wabot.jpg


茨城県映画『EXPO'85 科学の祭典』（1985年（昭和60年度）制作）　つくば科学万博 '85
https://www.youtube.com/watch?v=4Ixq9s2iDr0

OMR Thesis

- 1997 David Bain
- 2006 Laurent Pu
- 2009 Alicia Forn
- 2012 Ana Rebelo
- 2014 Andrew Ha
- 2016 Jorge Calv

2000: Gamera

- Fra
  rec
- De
- Im .)
- Do
- Sy
- Po atch

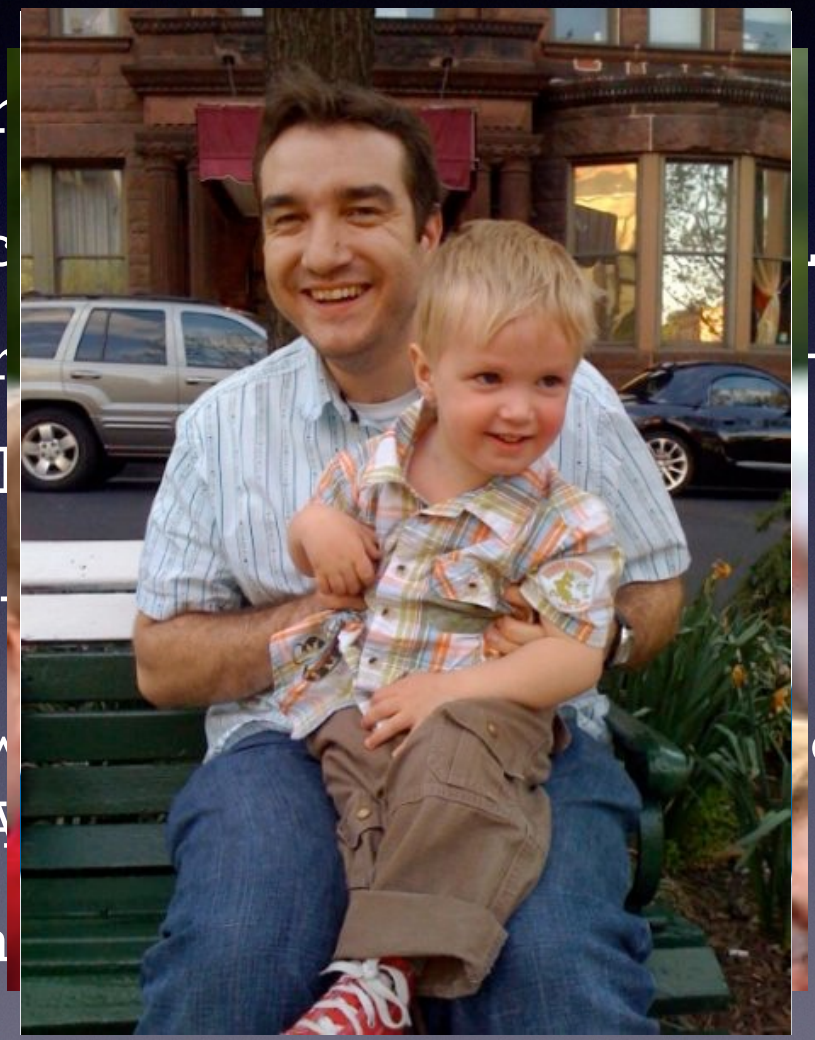
Generalized Algorithms and Methods for Enhancement and Restoration of Archives

Gamera @ Peabody

- Designed and built by Karl McMillan and Michael Droettboom (started fall 2000)
- Master's students at Computer Music Department at Peabody Conservatory of Music, Johns Hopkins University
- Both worked at Digital Knowledge Centre, Johns Hopkins University Library
- Both graduated in 2002

Gamera developers

- Karl McMillar
  - Expert of Sec …inux)
  - Worked at Tr …Hat
  - CTO at RAKI
- Michael Droe
  - Science softw …escope Science Institute (NA
  - Still active in



Birth of Gamera

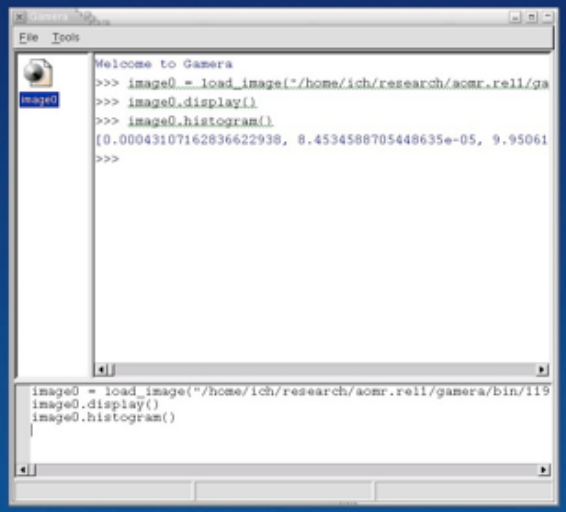
Gamera: A Structured Document Recognition Application Development Environment

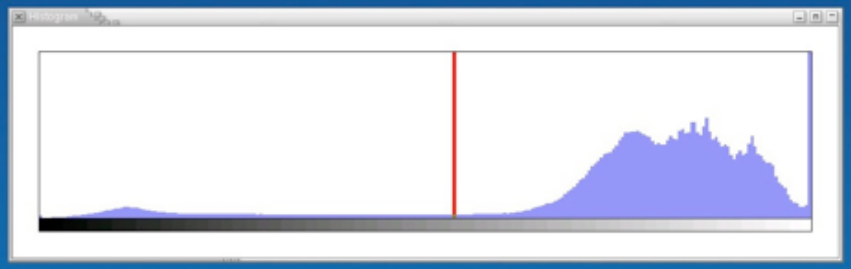
Karl MacMillan, Michael Droettboom, and Ichiro Fujinaga

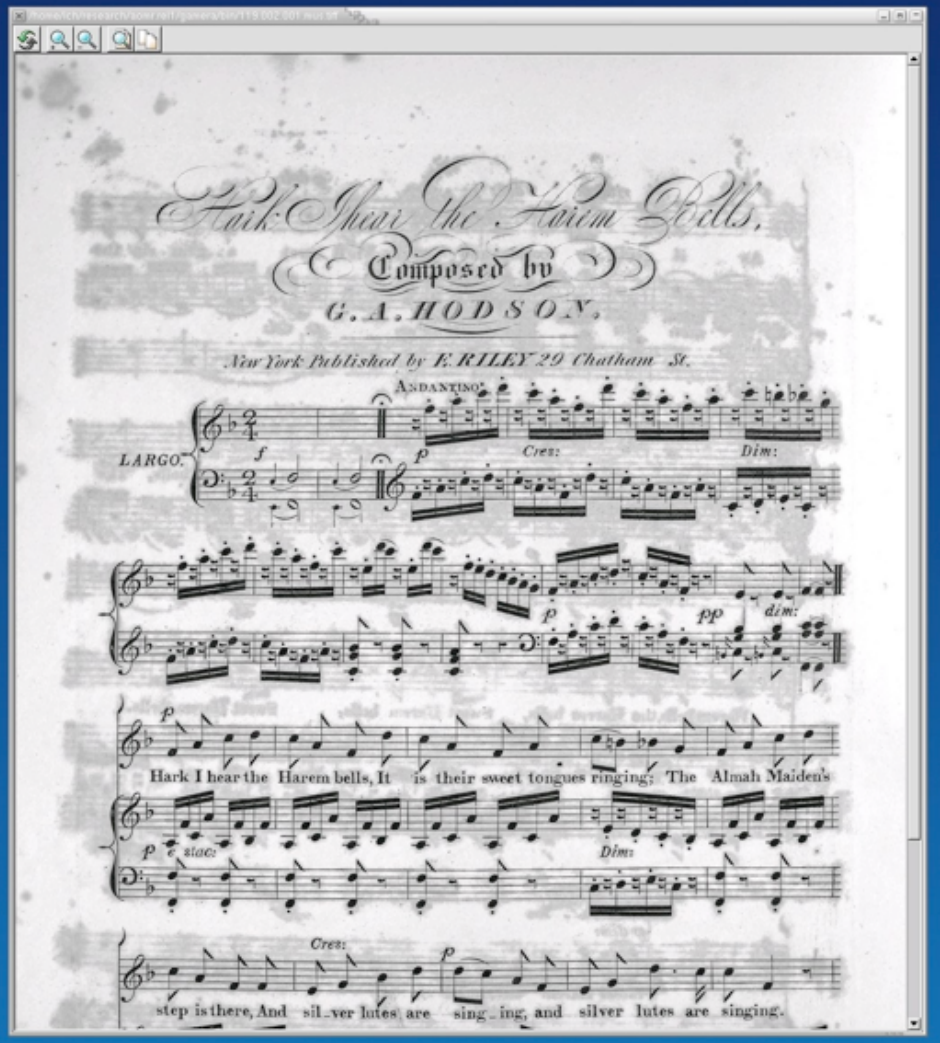
Peabody Conservatory of Music
Johns Hopkins University
1 East Mount Vernon Place, Baltimore MD 21202
email: {karlmac,mdboom,ich}@peabody.jhu.edu

- First paper presented at the 2nd International Symposium on Music Information Retrieval (ISMIR: October 2001) in Bloomington, IN

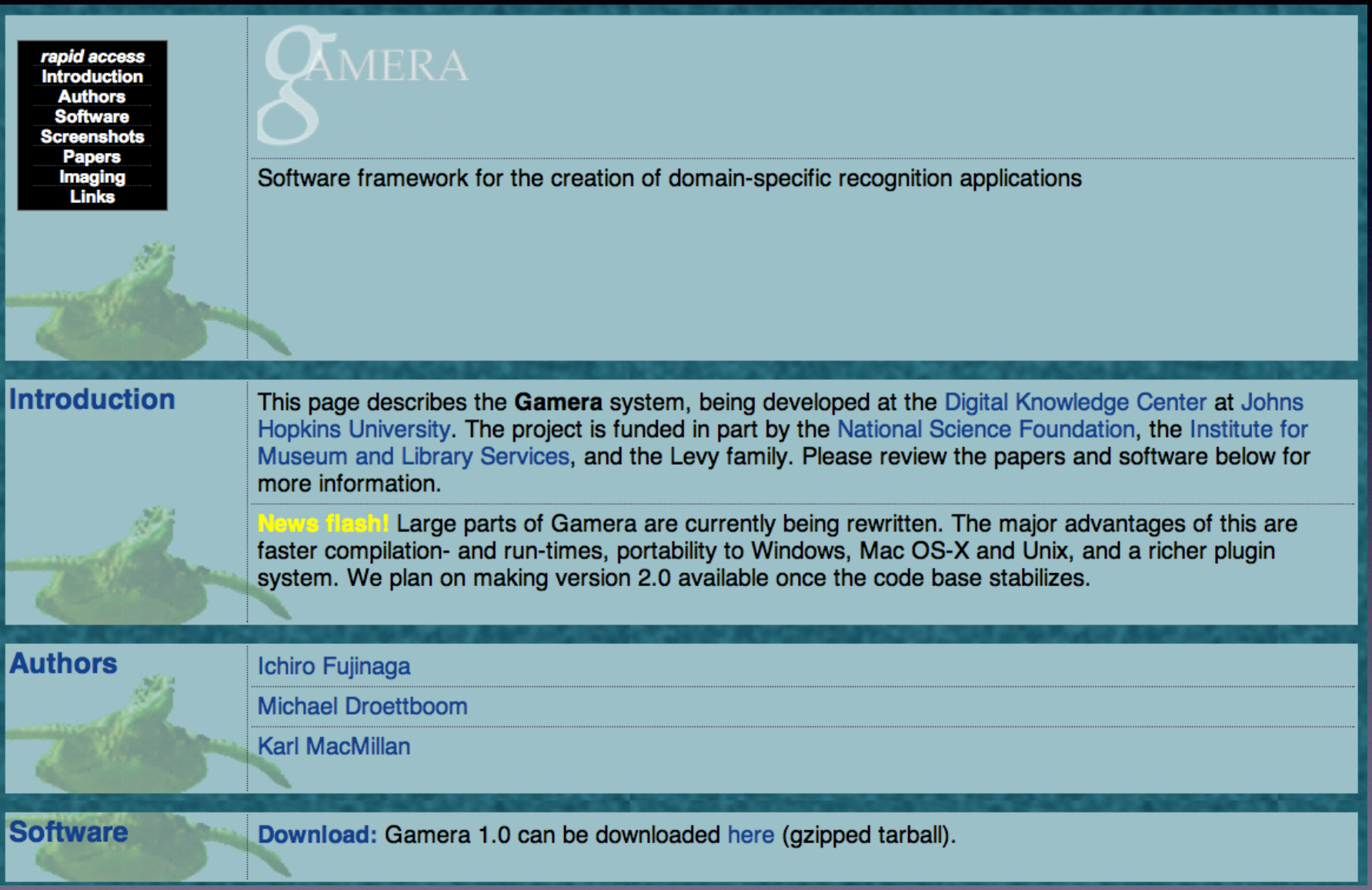
Early Gamera Screenshot (Linux) ca. June 2002







Original Gamera Website ca. December 2002



rapid access
Introduction
Authors
Software
Screenshots
Papers
Imaging
Links

GAMERA

Software framework for the creation of domain-specific recognition applications

Introduction

This page describes the **Gamera** system, being developed at the Digital Knowledge Center at Johns Hopkins University. The project is funded in part by the National Science Foundation, the Institute for Museum and Library Services, and the Levy family. Please review the papers and software below for more information.

News flash! Large parts of Gamera are currently being rewritten. The major advantages of this are faster compilation- and run-times, portability to Windows, Mac OS-X and Unix, and a richer plugin system. We plan on making version 2.0 available once the code base stabilizes.

Authors

Ichiro Fujinaga

Michael Droettboom

Karl MacMillan

Software

Download: Gamera 1.0 can be downloaded here (gzipped tarball).

2001–2008: Evolution of Gamera

- 20
  an
- 20
  re
- 20
  Ur
- 20

Project: Gamera: Summary

Summary | Admin | Home Page | Tracker | Bugs | Support | Patches | RFE | News | CVS | Files |

Gamera is a framework for the creation of structured document analysis applications by domain experts. It combines a programming library with GUI tools for the training and interactive development of recognition systems.

Foundry Member :Python Foundry

- Development Status: 4 - Beta
- Environment: Cocoa (MacOS X), Win32 (MS Windows), X11 Applications
- Intended Audience: Developers, Education, Information Technology, Science/Research
- License: GNU General Public License (GPL)
- Natural Language: English
- Operating System: MacOS X, Windows NT/2000, Linux
- Programming Language: C++, Python
- Topic: Graphics Conversion, Artificial Intelligence, Information Analysis

Developer Info

Project Admins:
abrzeczk
mdboom

Developers: 4
[View Members]

Project UNIX name: gamera
Registered: 2004-01-12 10:13

Activity Percentile (last week): 63%
View project activity statistics
View list of RSS feeds available for this project

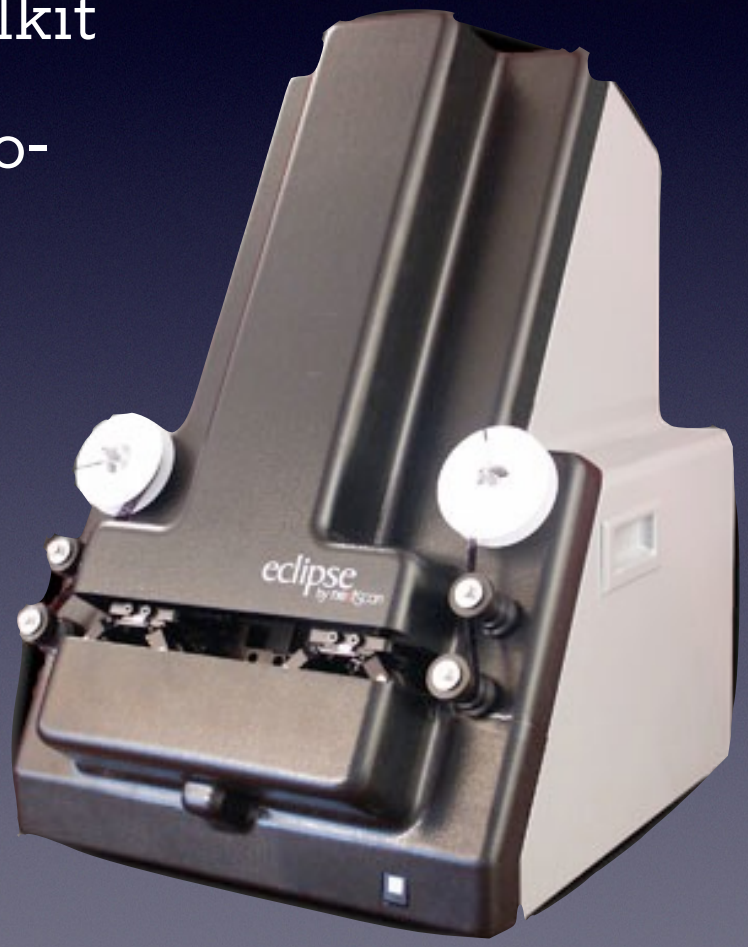
Latest File Releases

Package	Version	Date	Notes / Monitor	Download
gamera	gamera-2.2.0pre2	August 17, 2004	-	Download
gamera-video	1.0	January 13, 2004	-	Download
toolkit-skeleton	2.2.0pre2	September 8, 2004	-	Download
wxPython-Gtk-X11	wxPython-Gtk-X11-2.4.2.4	July 21, 2004	-	Download

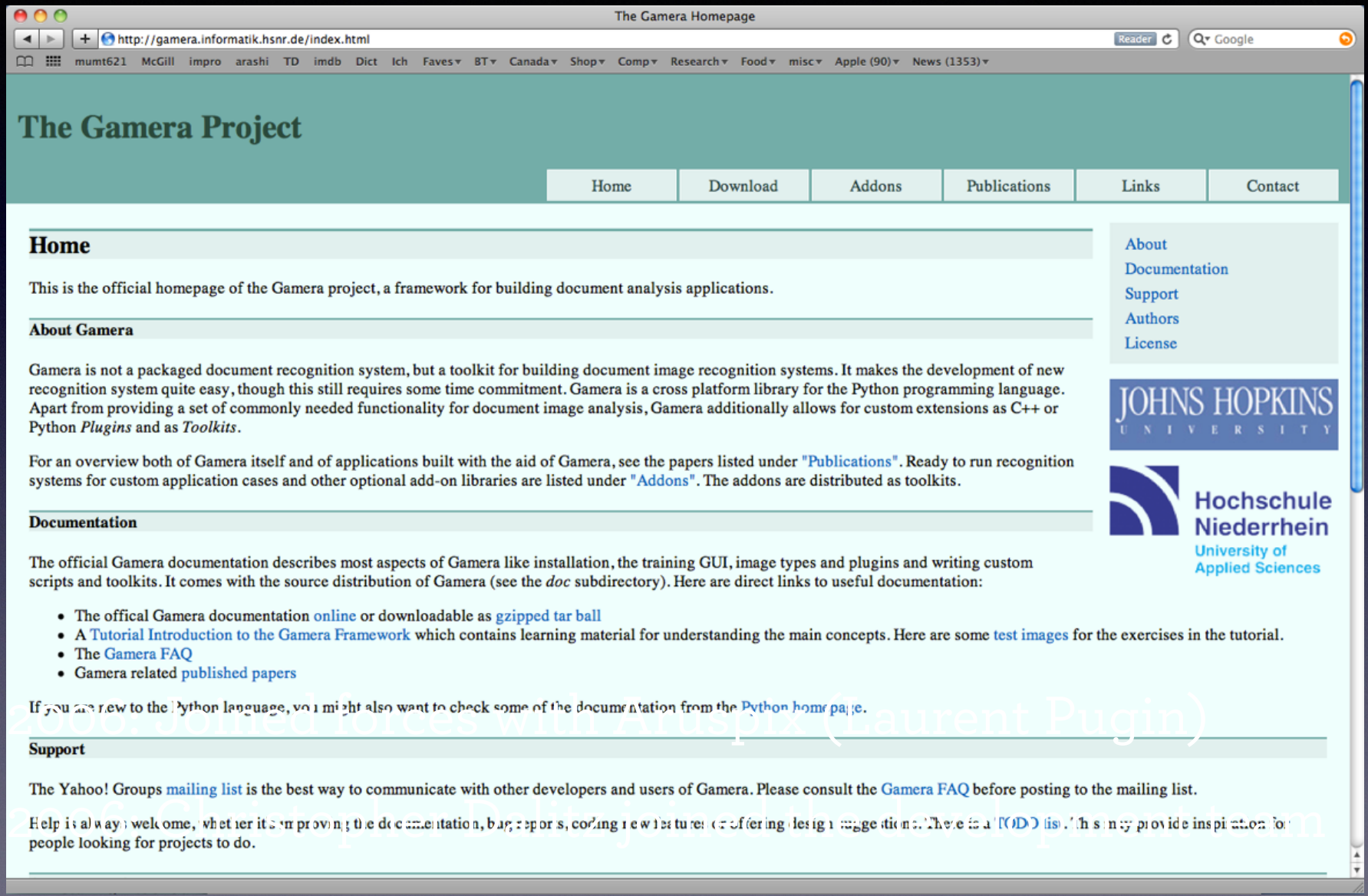
[View ALL Project Files]

2005: GEMM (Gamut for Early Music on Microfilms)

- Based on GAMUT: Gamera-based Automatic Music Understanding Toolkit
- Possibility of OMR for music on micro-films
- Almost all old Western music are on microfilms
- Efficient digitization using automatic microfilm scanner
  (Eclipse 500: 590ppm)



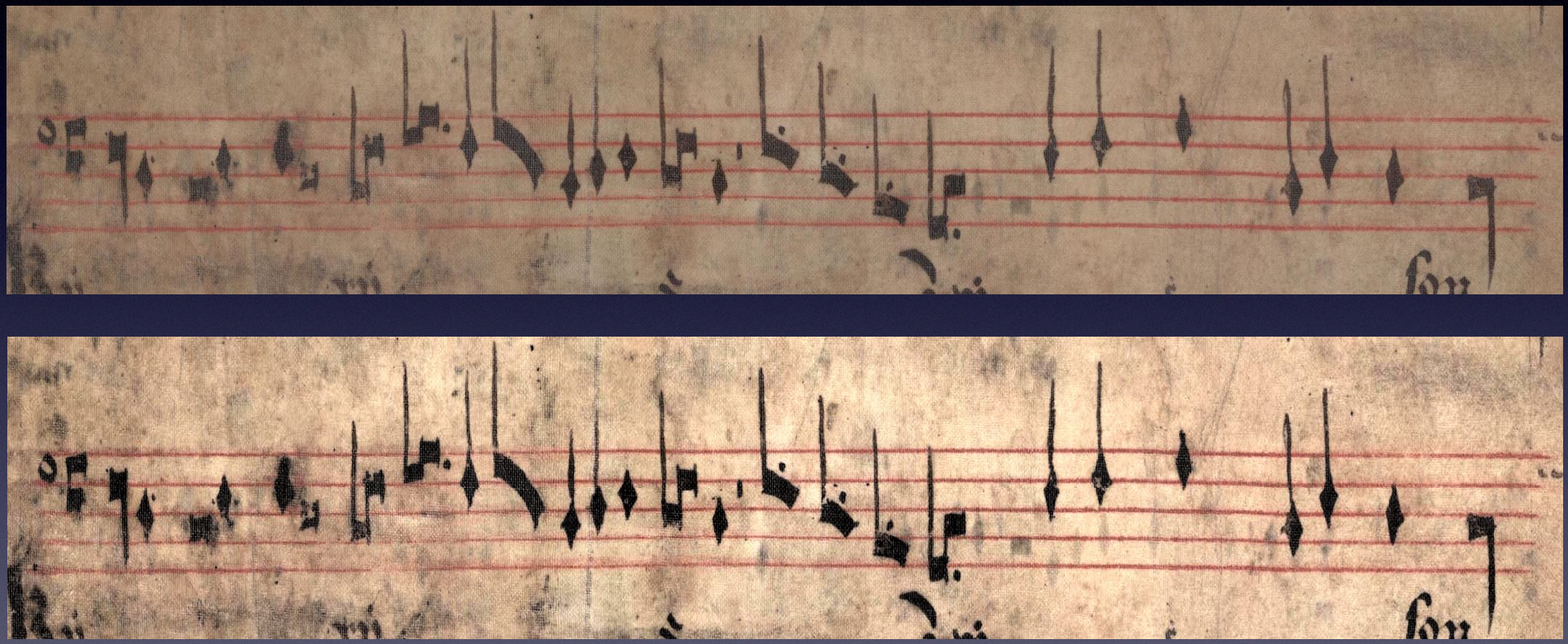
2001–2008: Evolution of Gamera



- 2008: Website moved to: http://gamera.informatik.hsnr.de

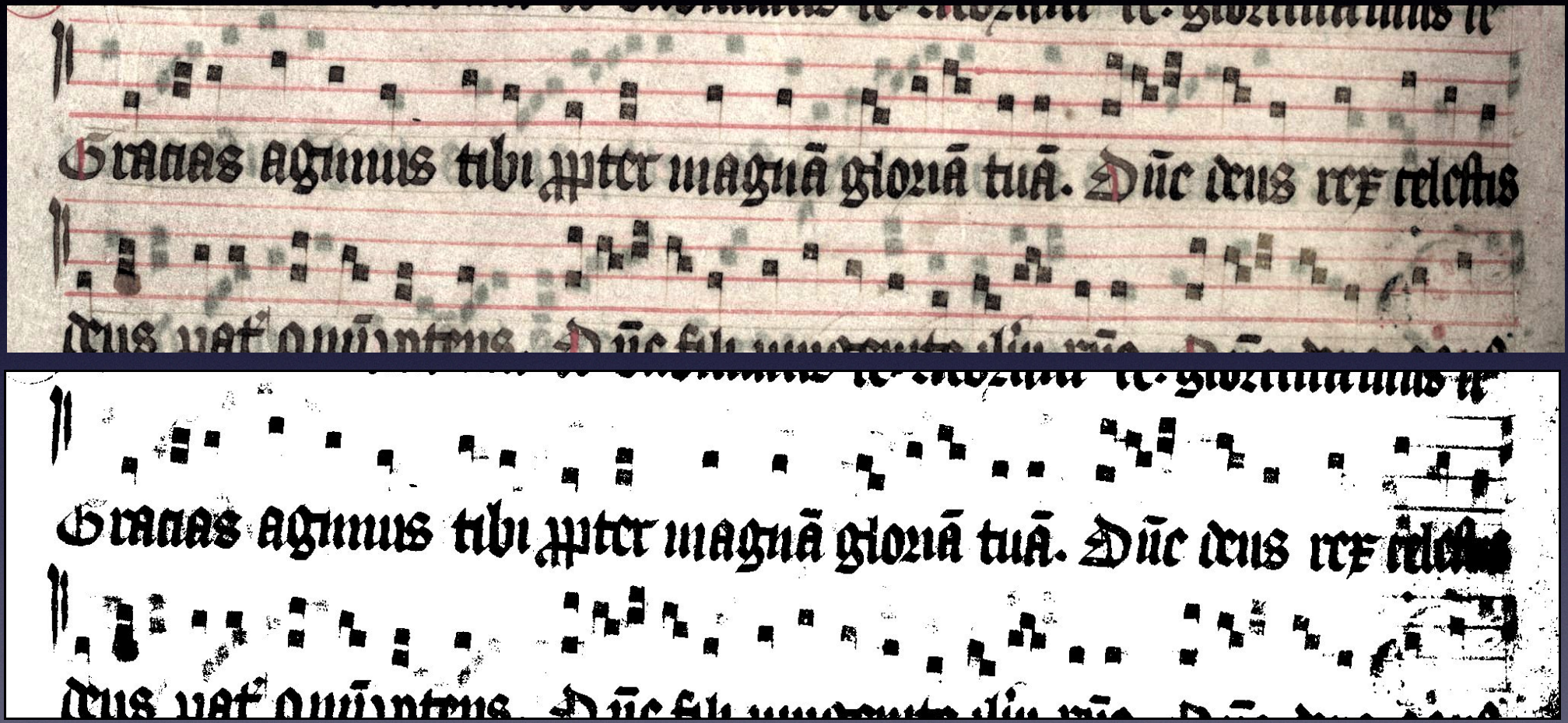
Some Features of Gamera c. 2008

Preprocessing



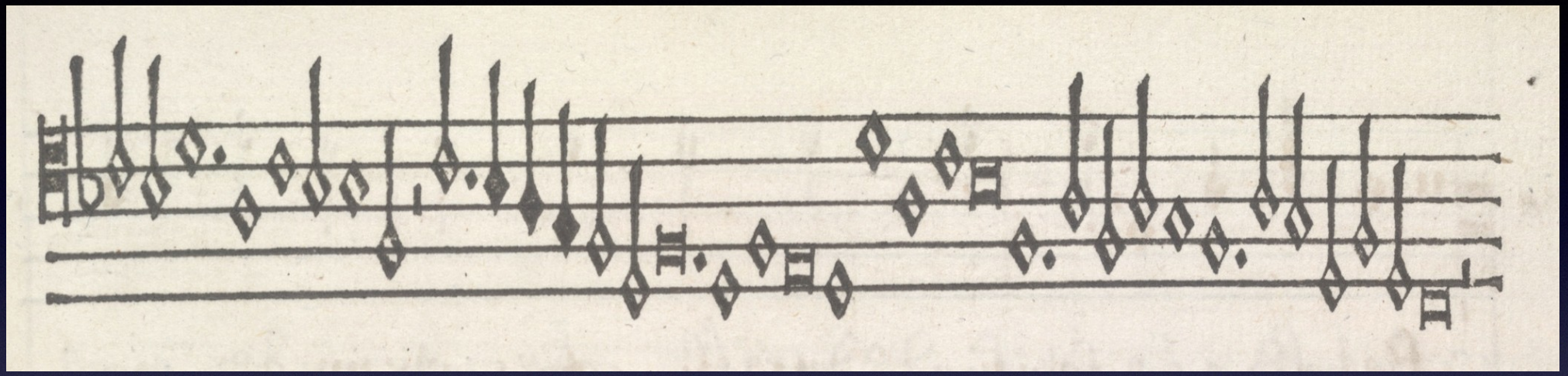
Brightness Enhancement

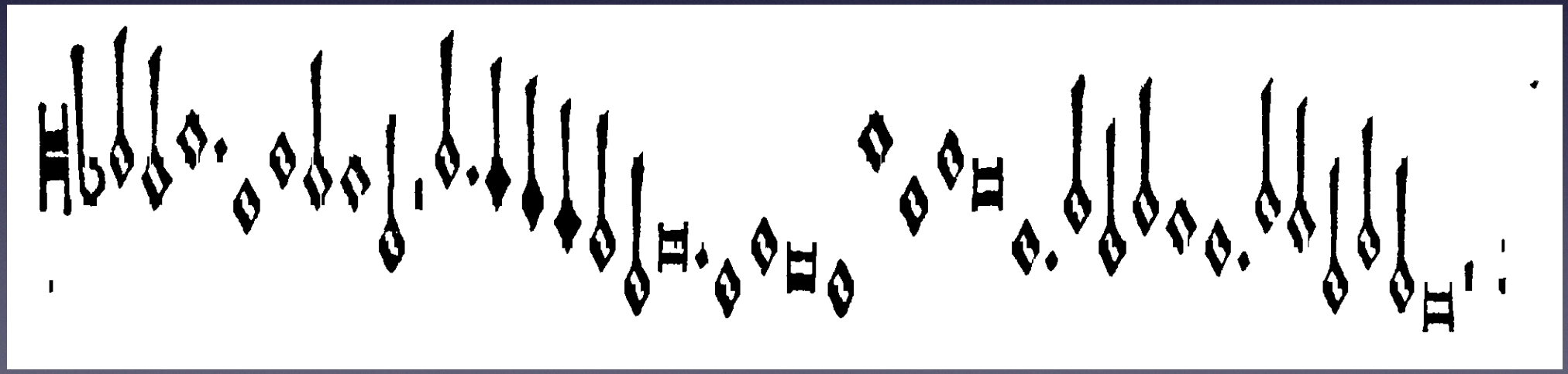
Preprocessing



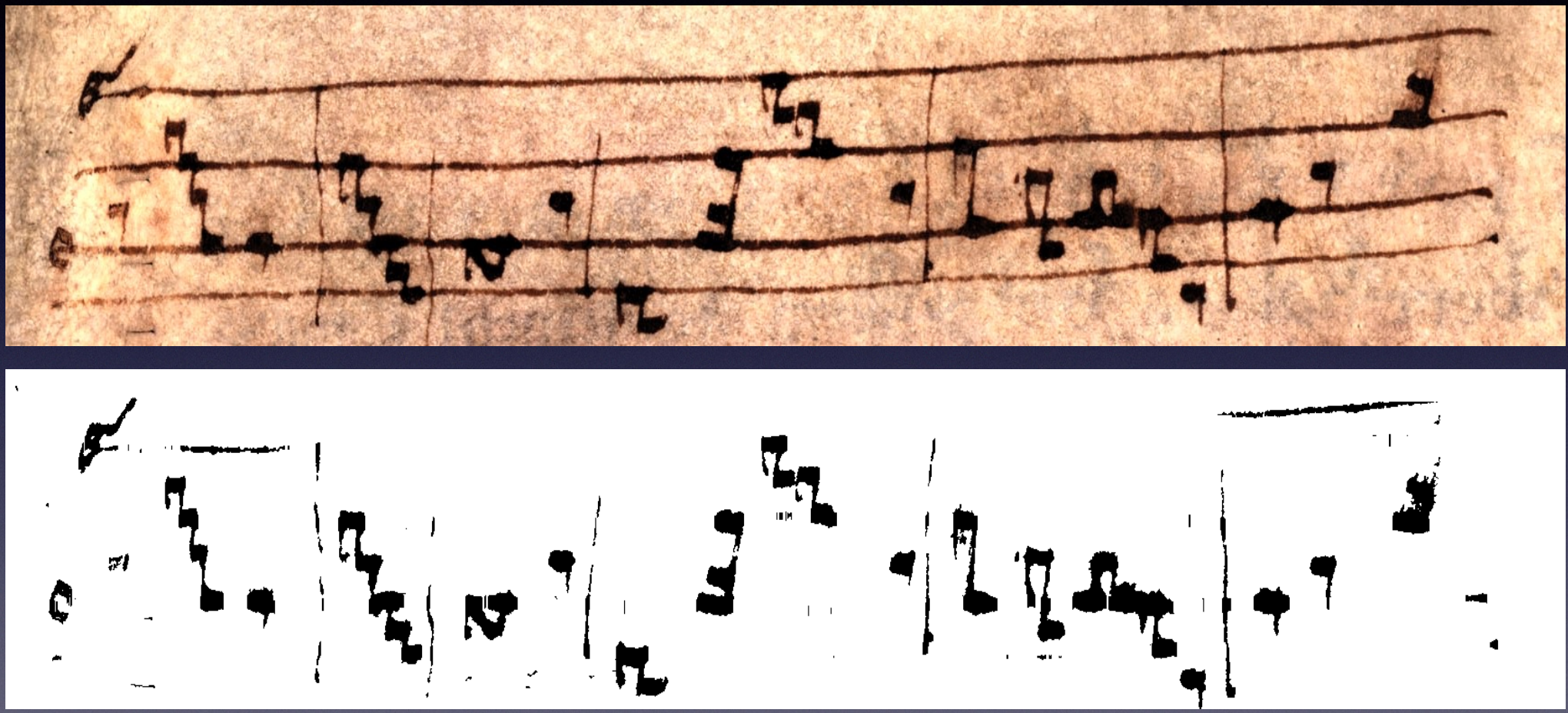
Thresholding

Staffline Removal



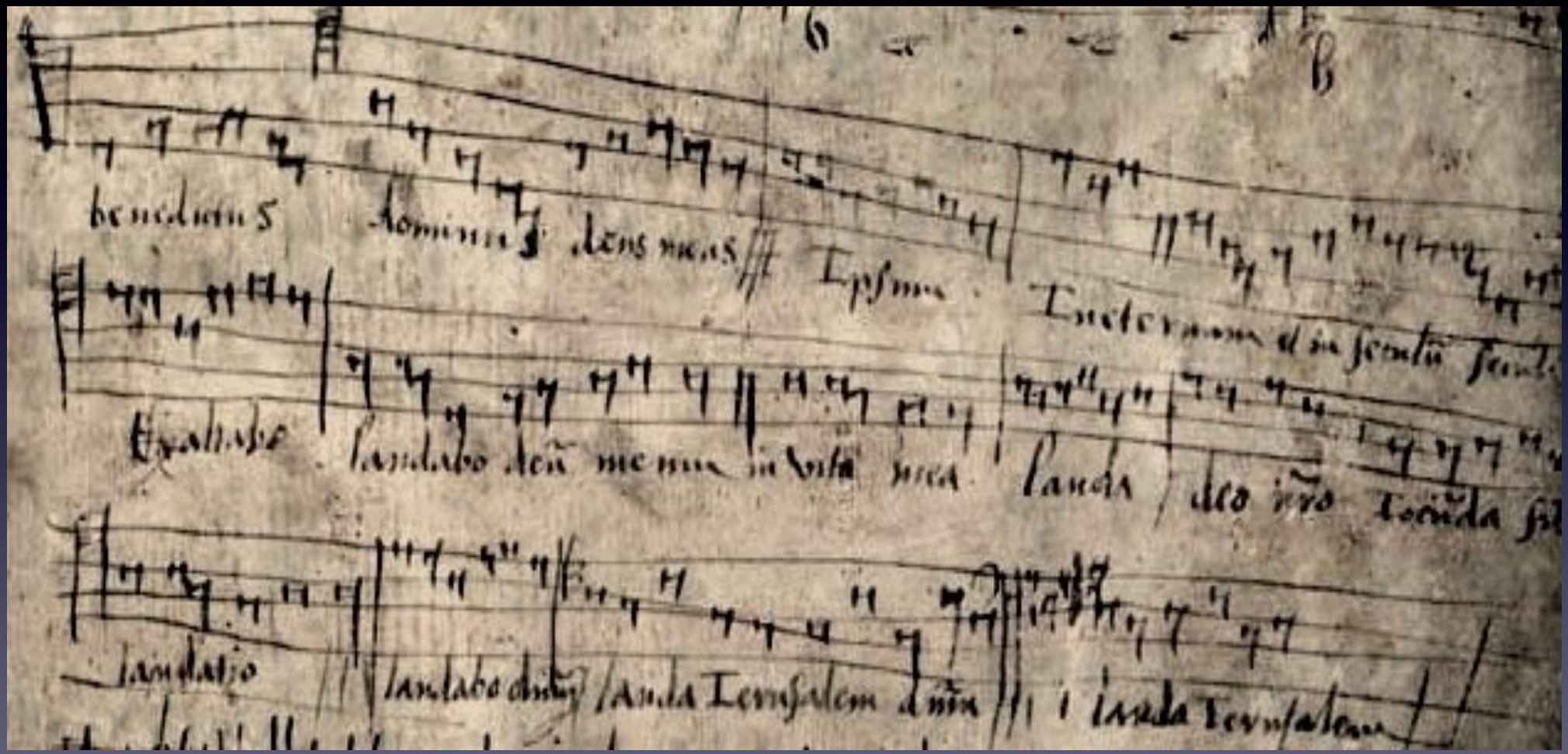


Staffline Removal



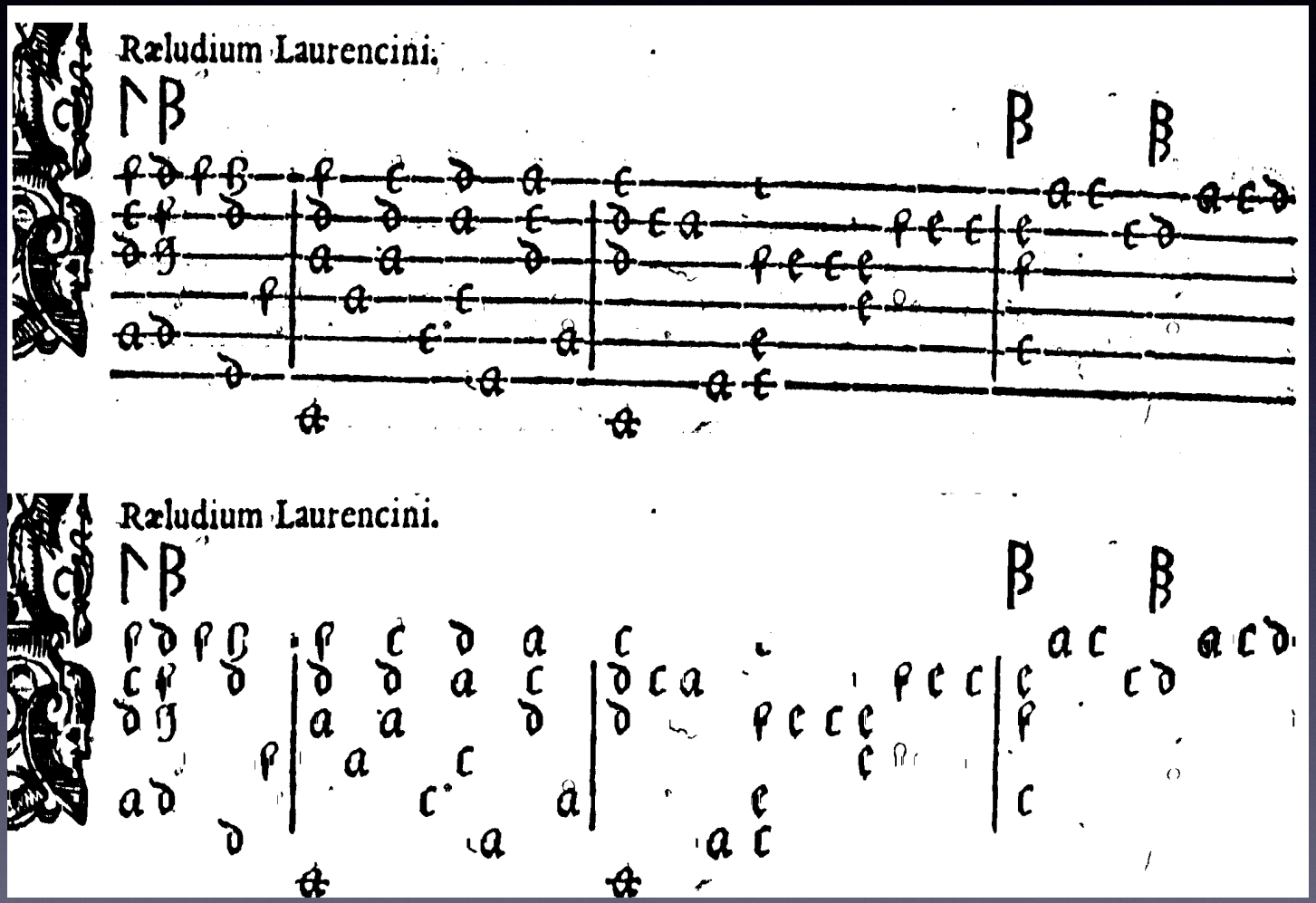
Four-line hand-drawn staff example

Staffline Removal



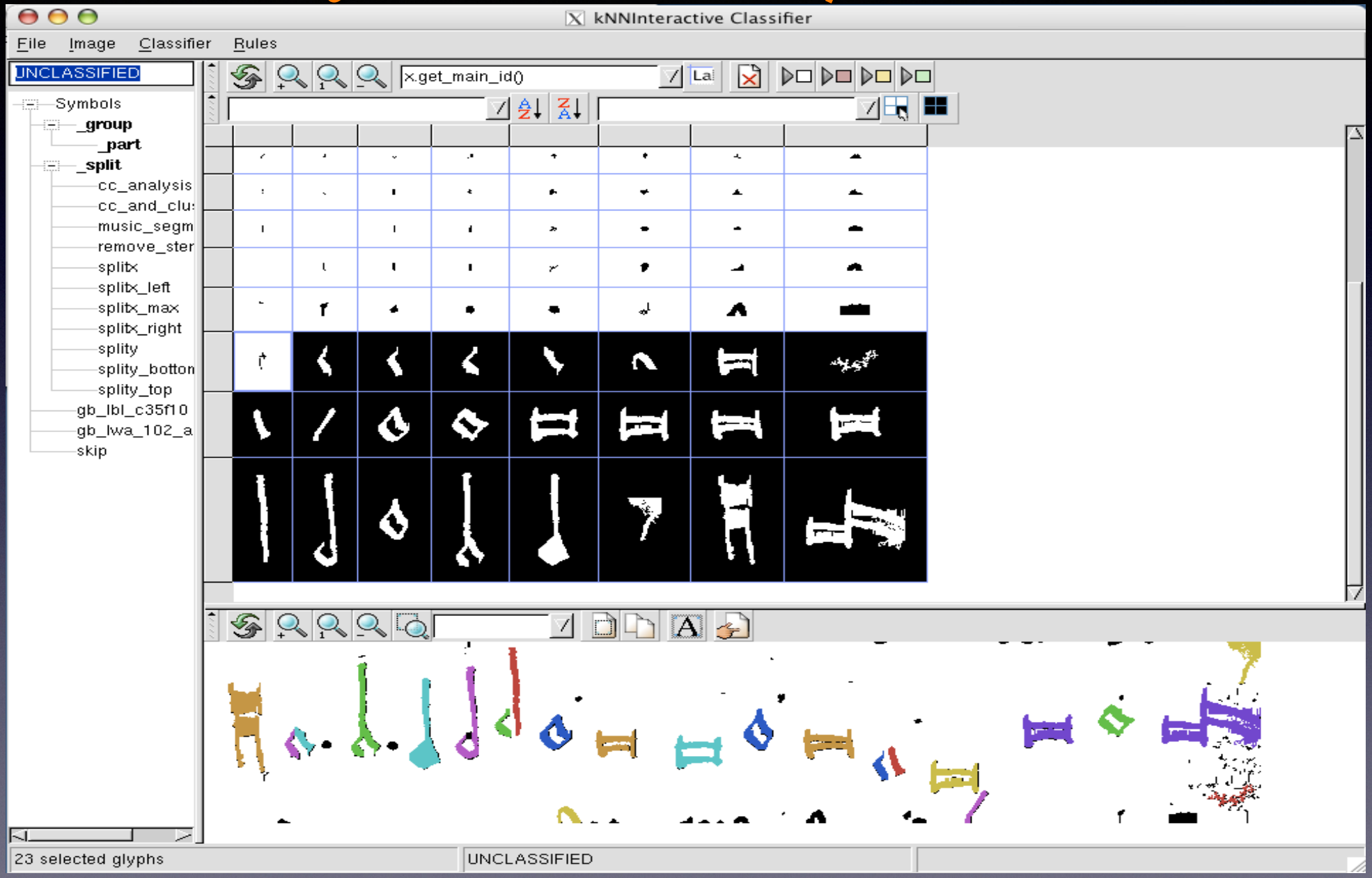
Difficult

Staffline Removal

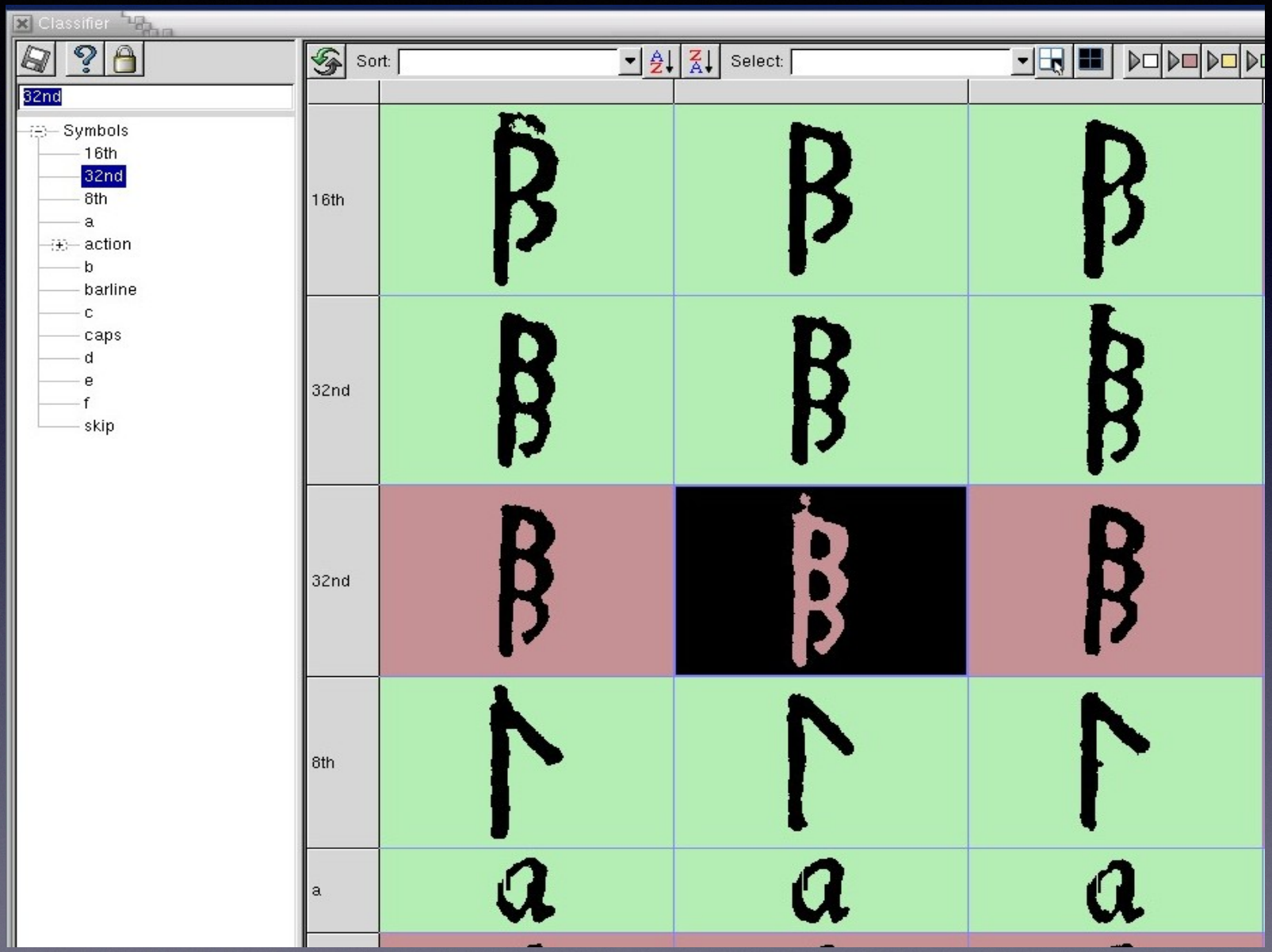


Lute tablature

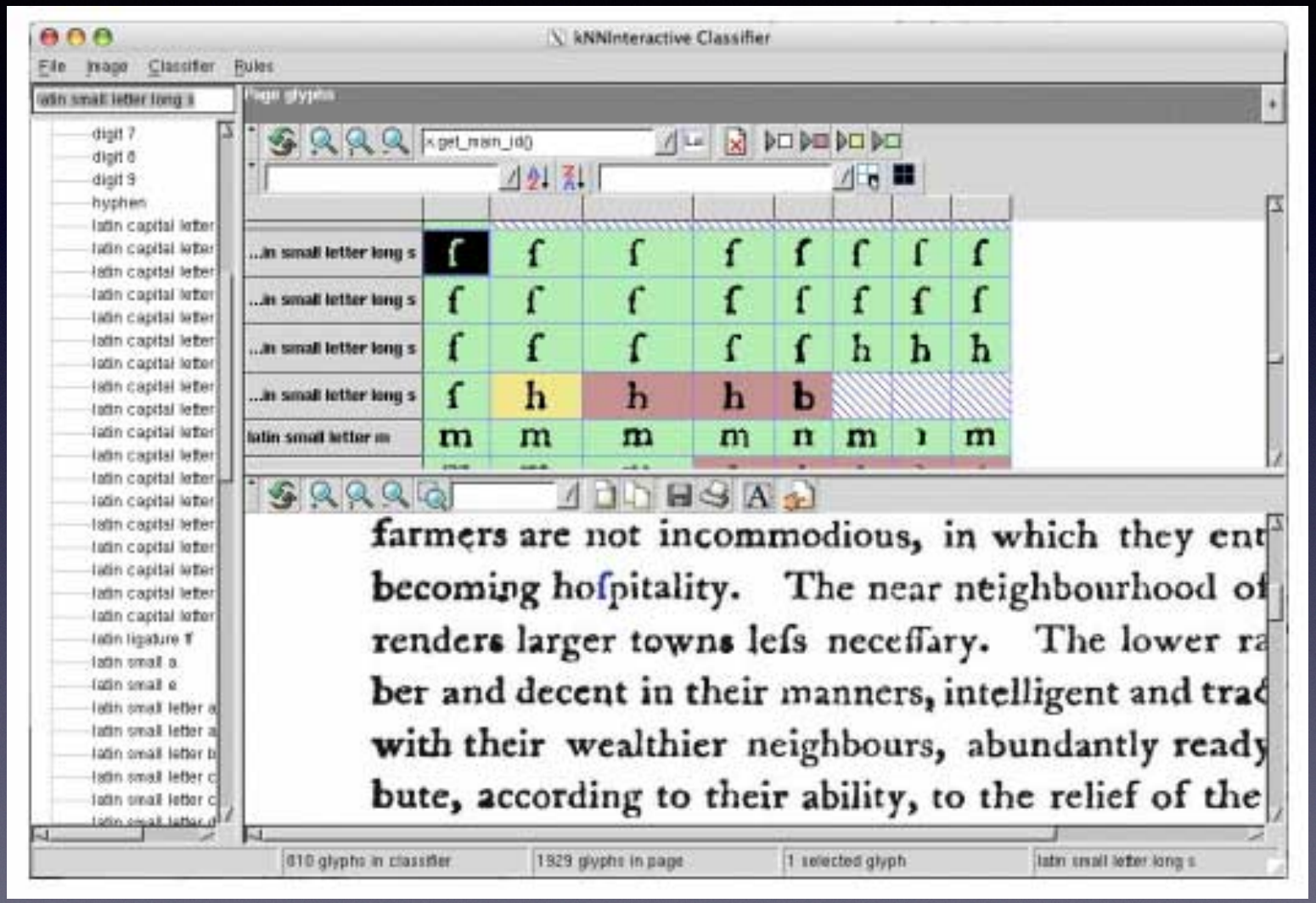
Symbol classifier / Gamera



Lute tablature symbol recognition

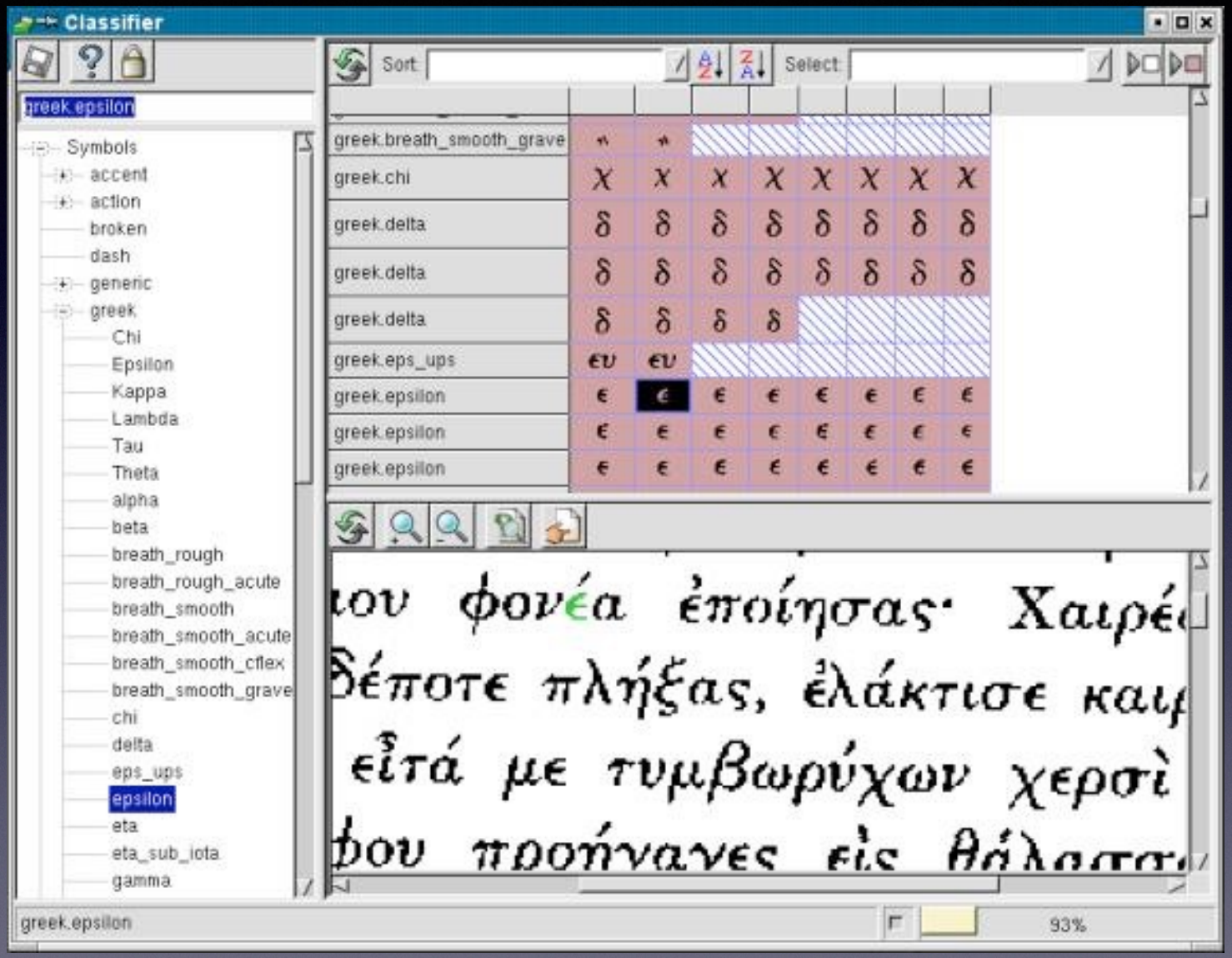


Other Applications



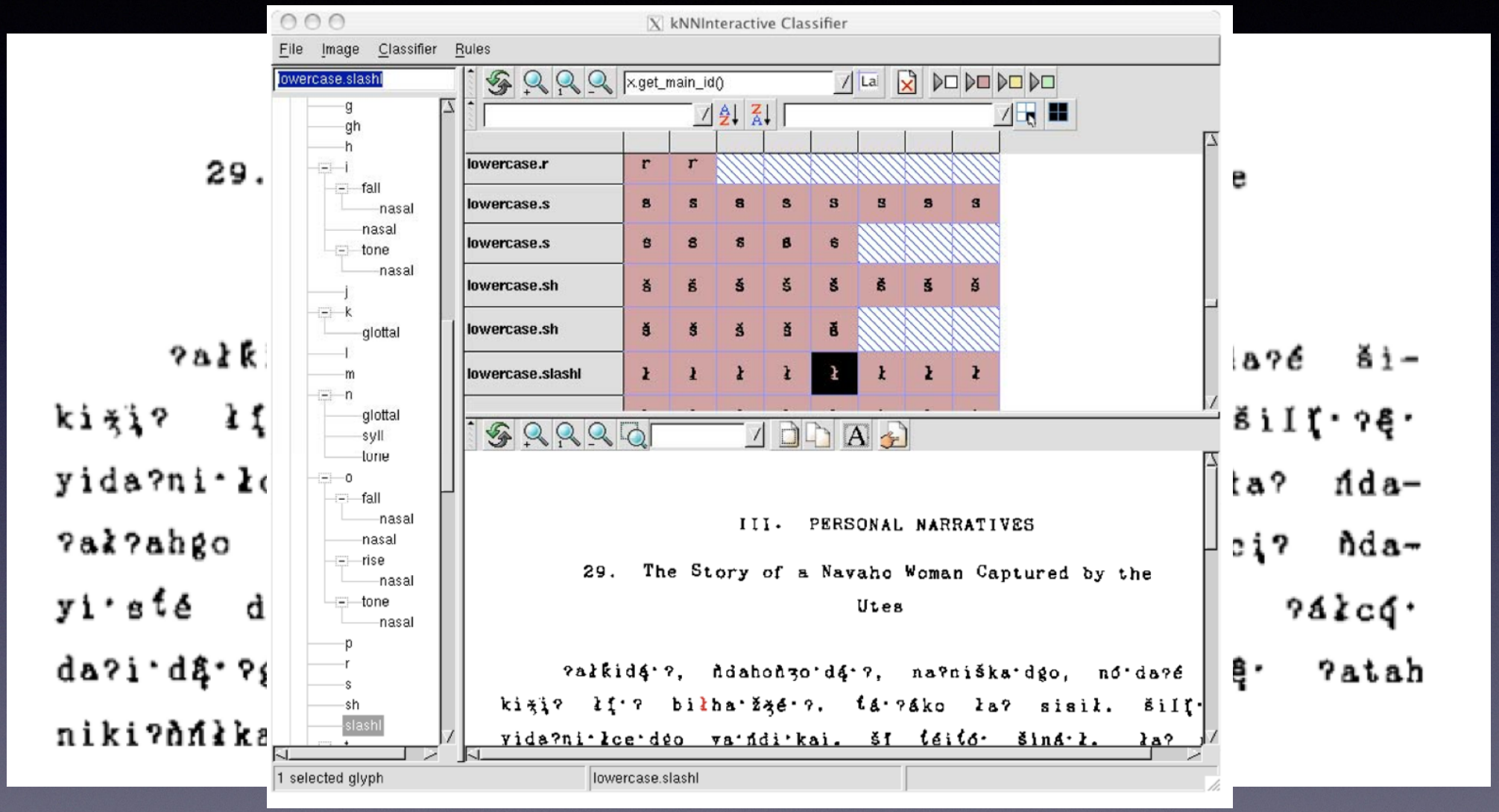
Early Modern English

Other Applications



Greek

Other Applications

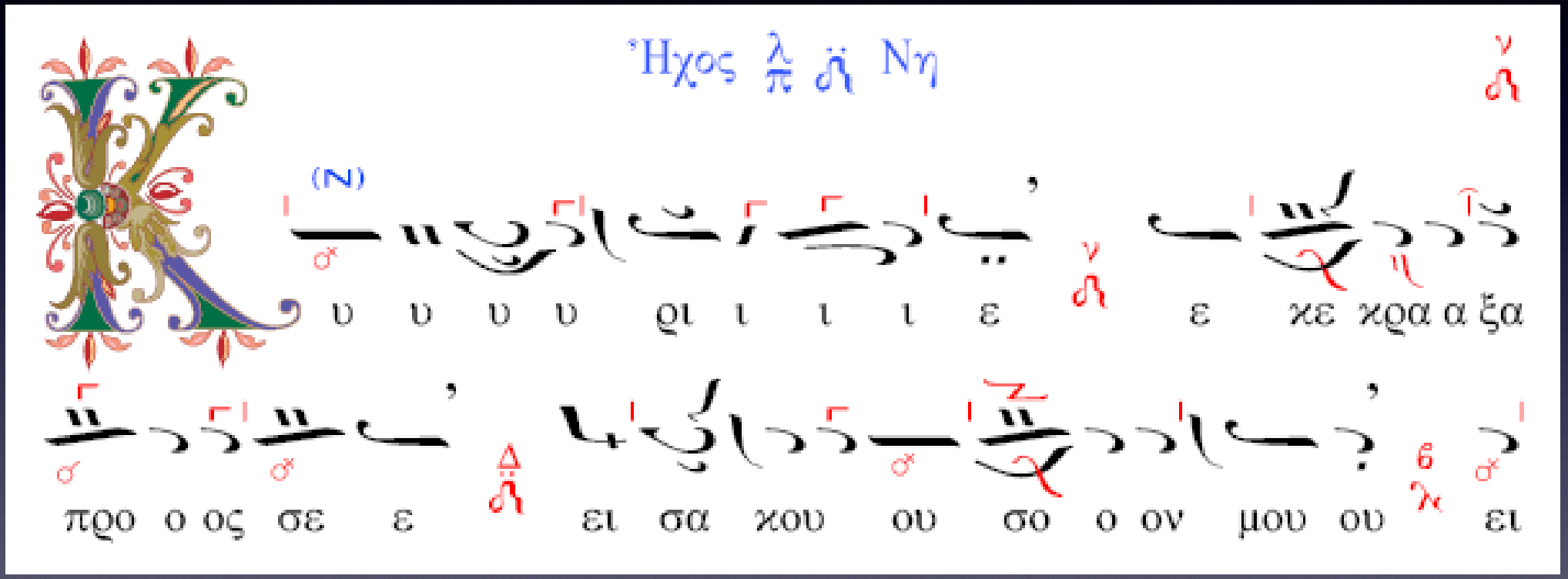


Navajo language recognition

Other Applications

Roman de la Rose (Bodelian MS Douce 195, 90v, 15th C.)

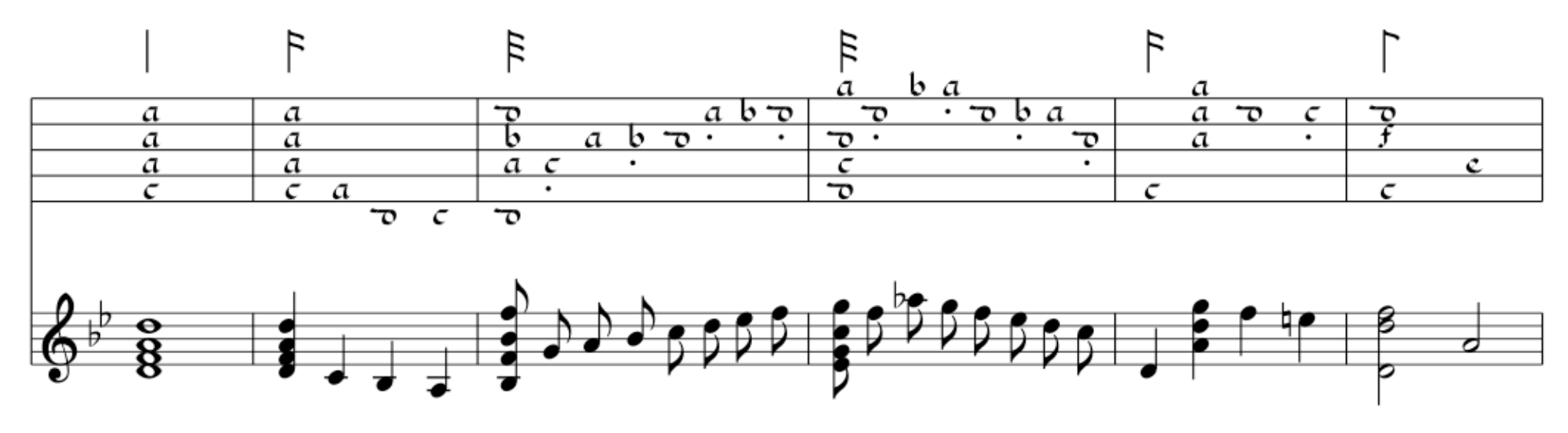
2008: Other Applications



Optical Recognition of Psaltic Byzantine Chant Notation

Christoph Dalitz · Georgios K. Michalakis · Christine Pranzas

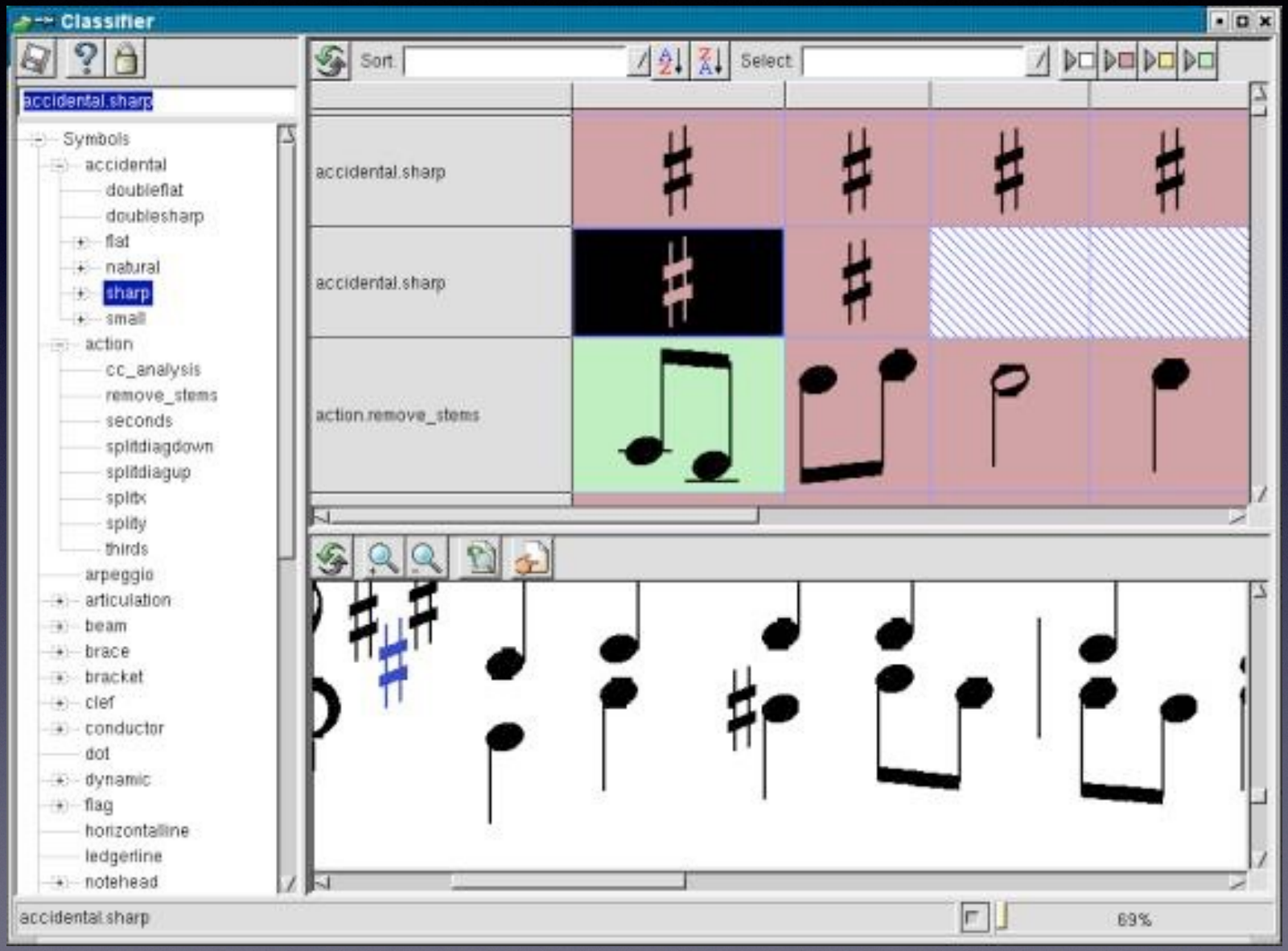
2009: Other Applications



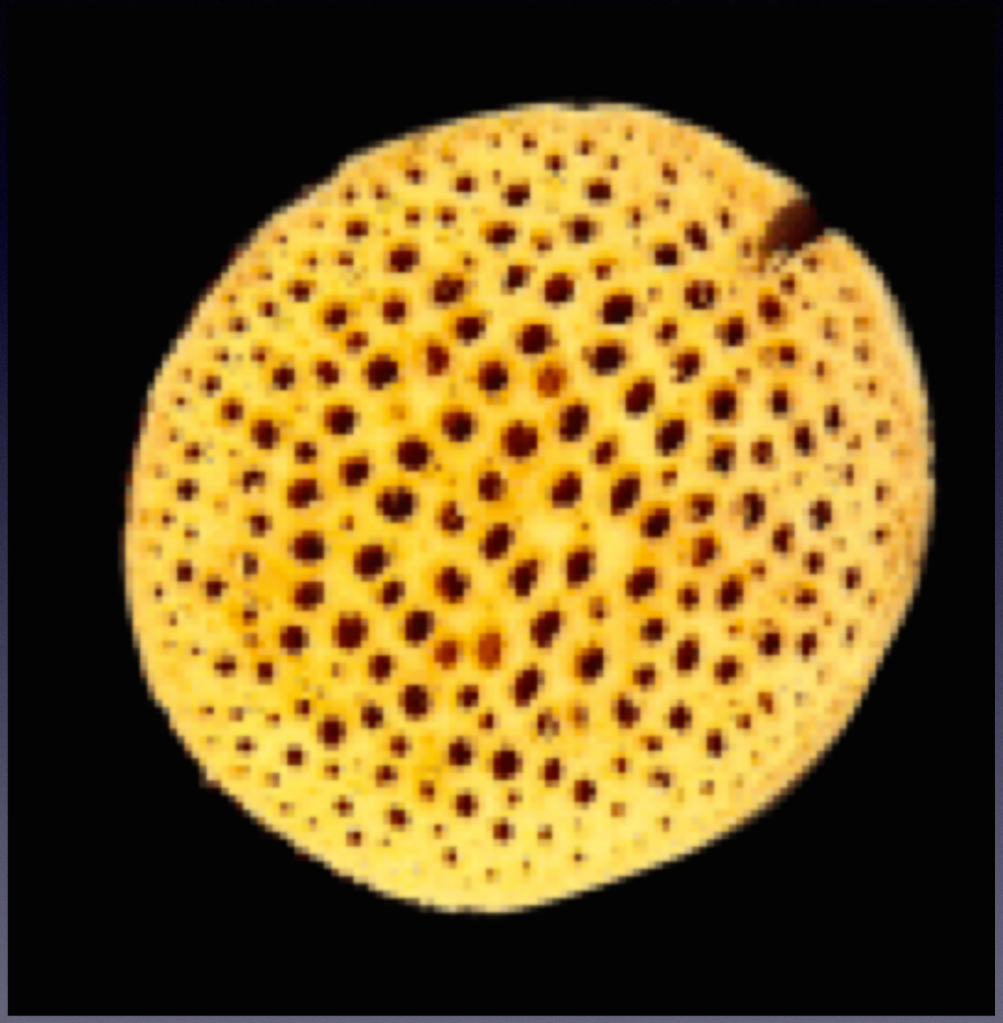
Optical Recognition of Lute Tablature

Christoph Dalitz · Thomas Karsten

Other Applications

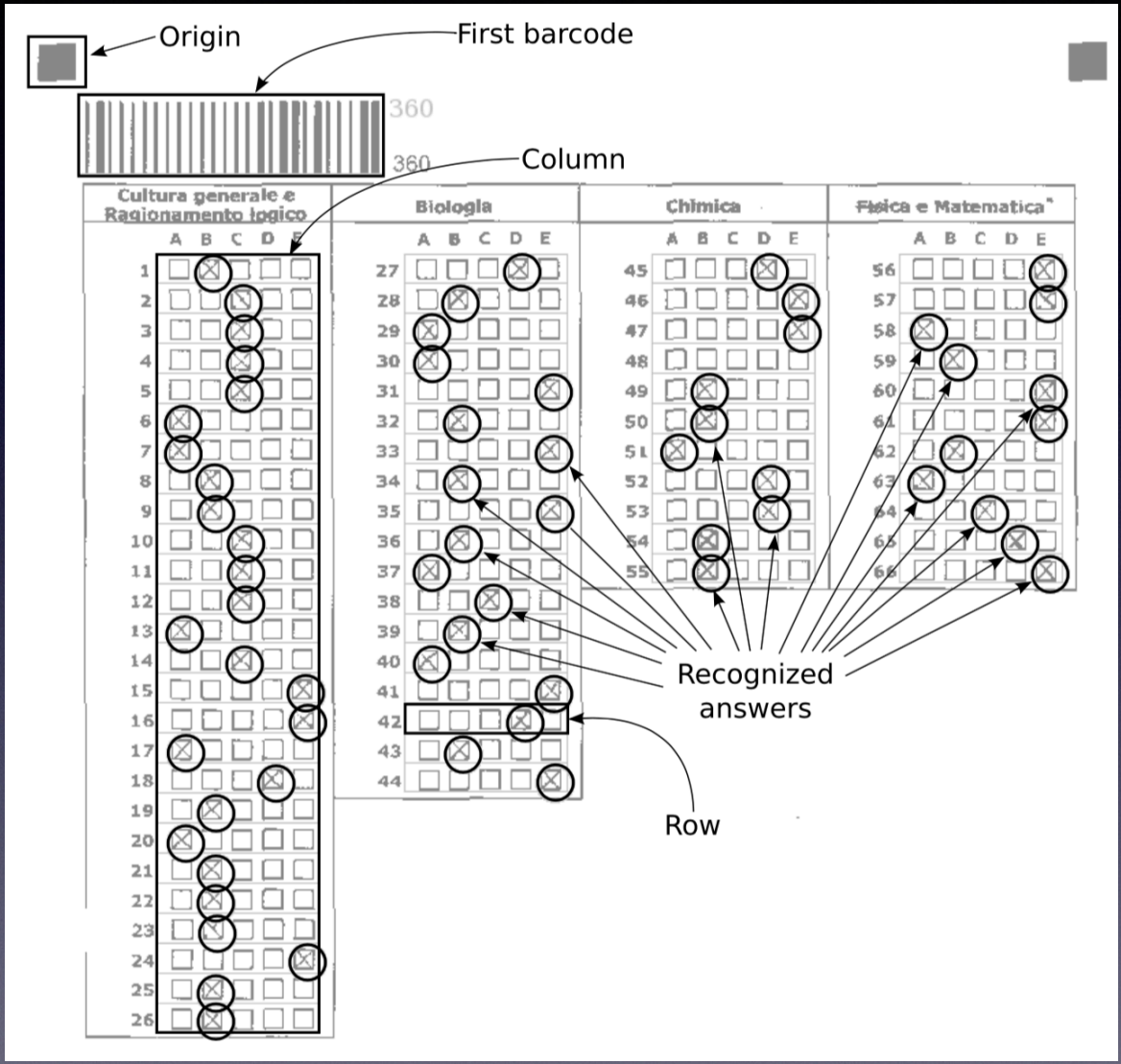


2005: Other Applications



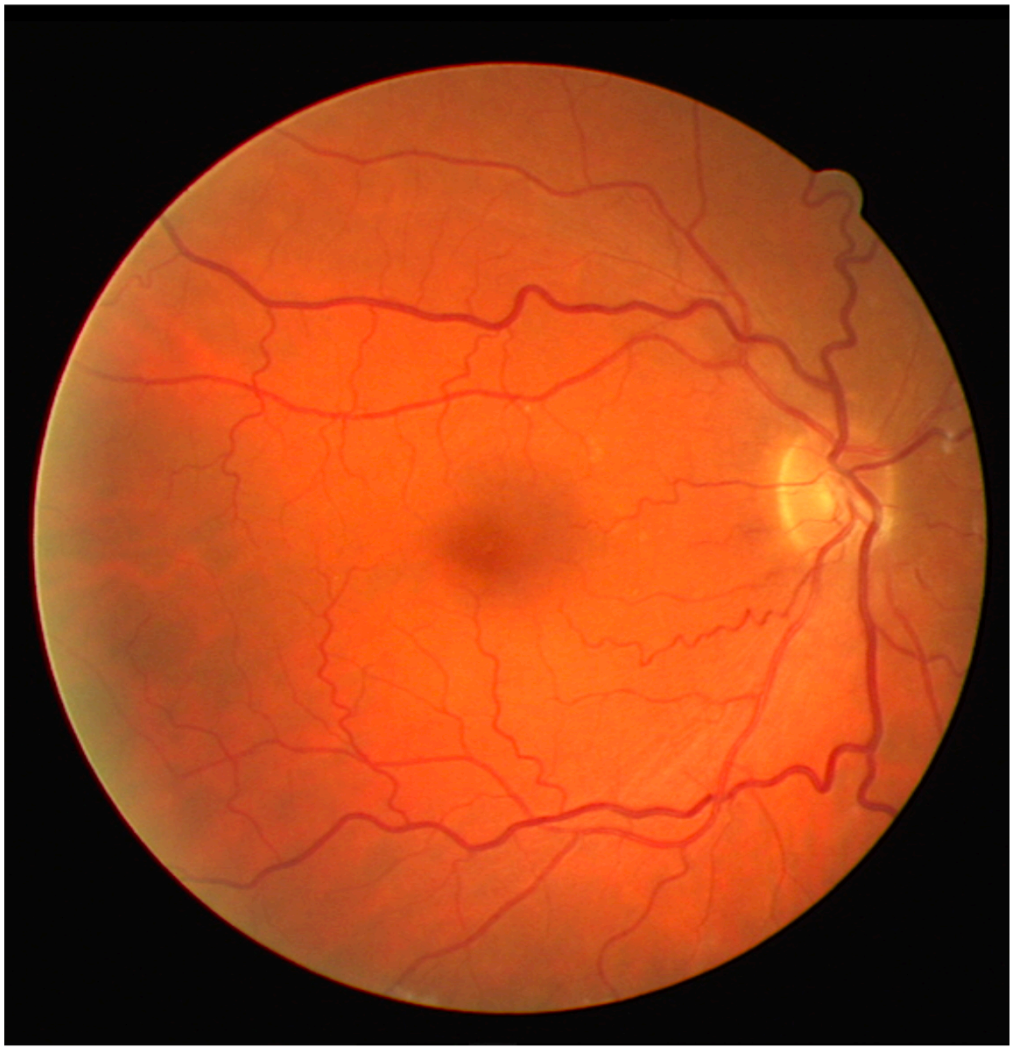
Vascular Anatomy of Plants (Alex Cobb, Harvard)

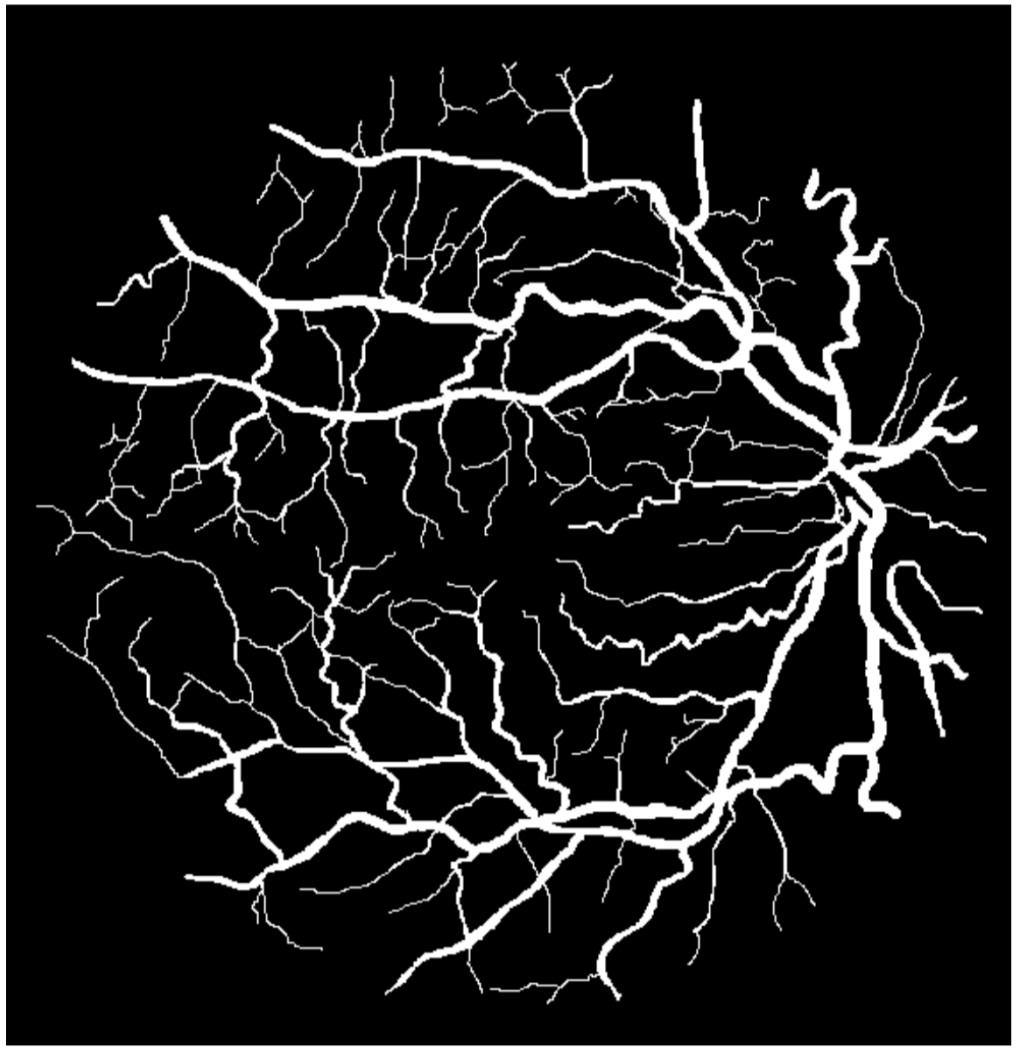
2009: Other Applications



Multiple-choice Test Recognition System (Spadaccini & Rizzo)

2015: Other Applications





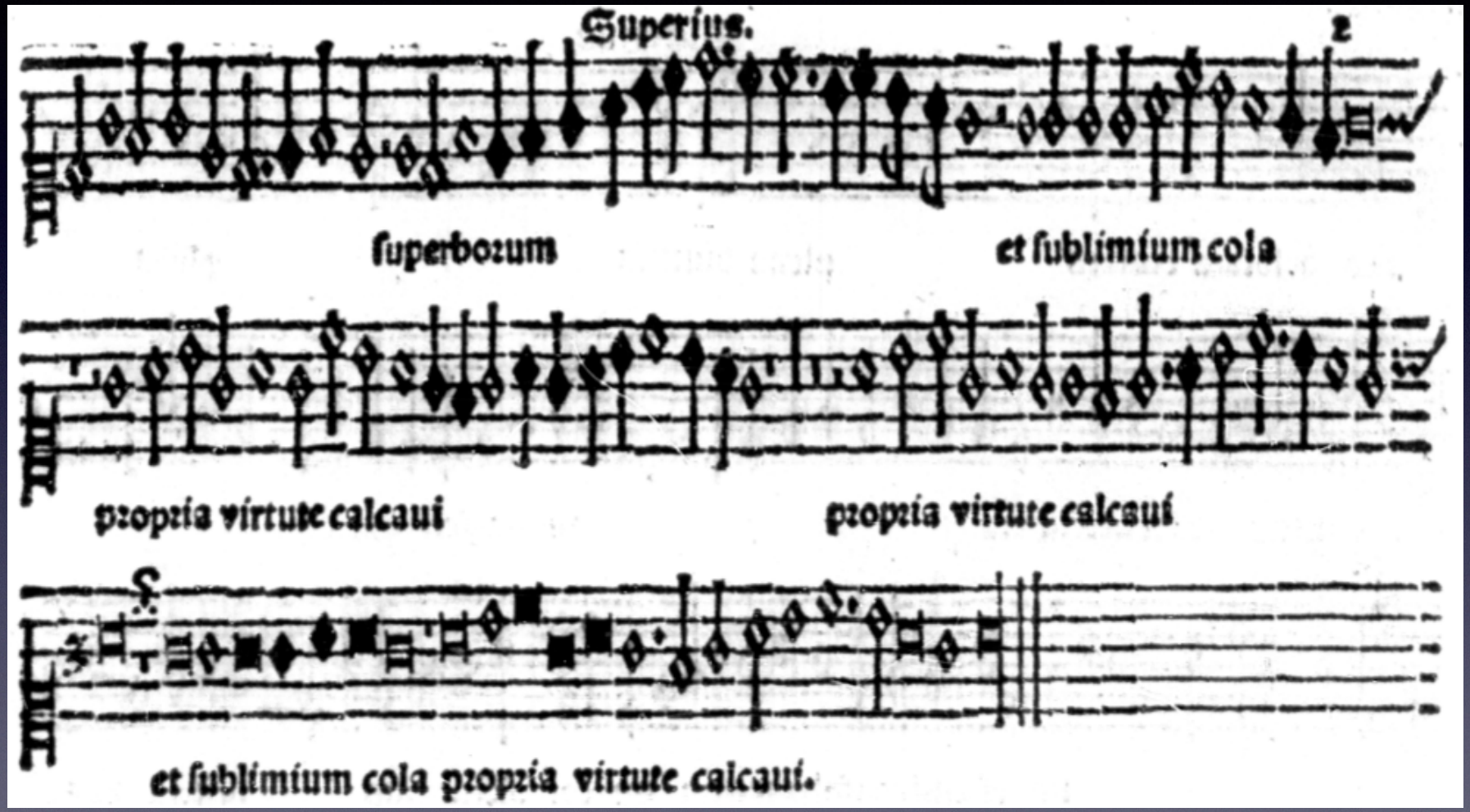
Blood vessel extraction (Dalitz et al.)

2005–2016: VIPs



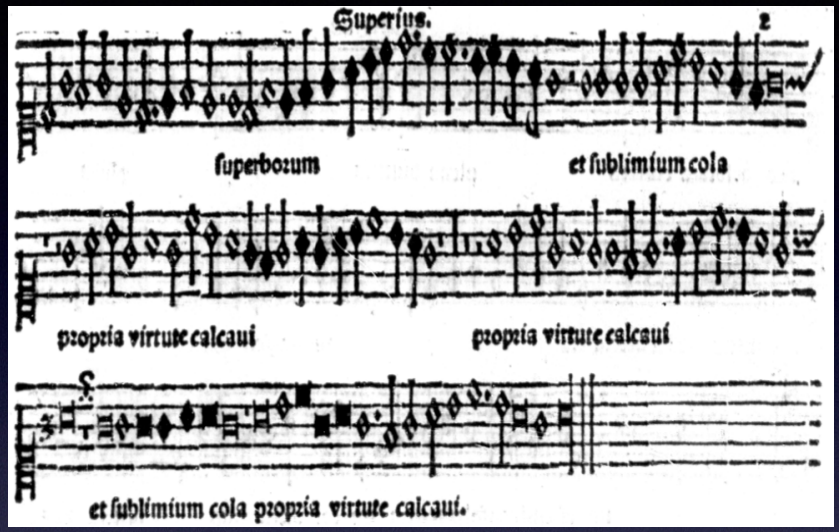
- 2005–2012 Ashley Burgoyne (PhD)
- 2006–2008 Laurent Pugin (Postdoc)
- 2007-2016 Andrew Hankinson (PhD/Postdoc)
- 23 publications on OMR between 2007–2016
- 18 publications on OMR between 2007–2012 (3 per year)

2002: Aruspix

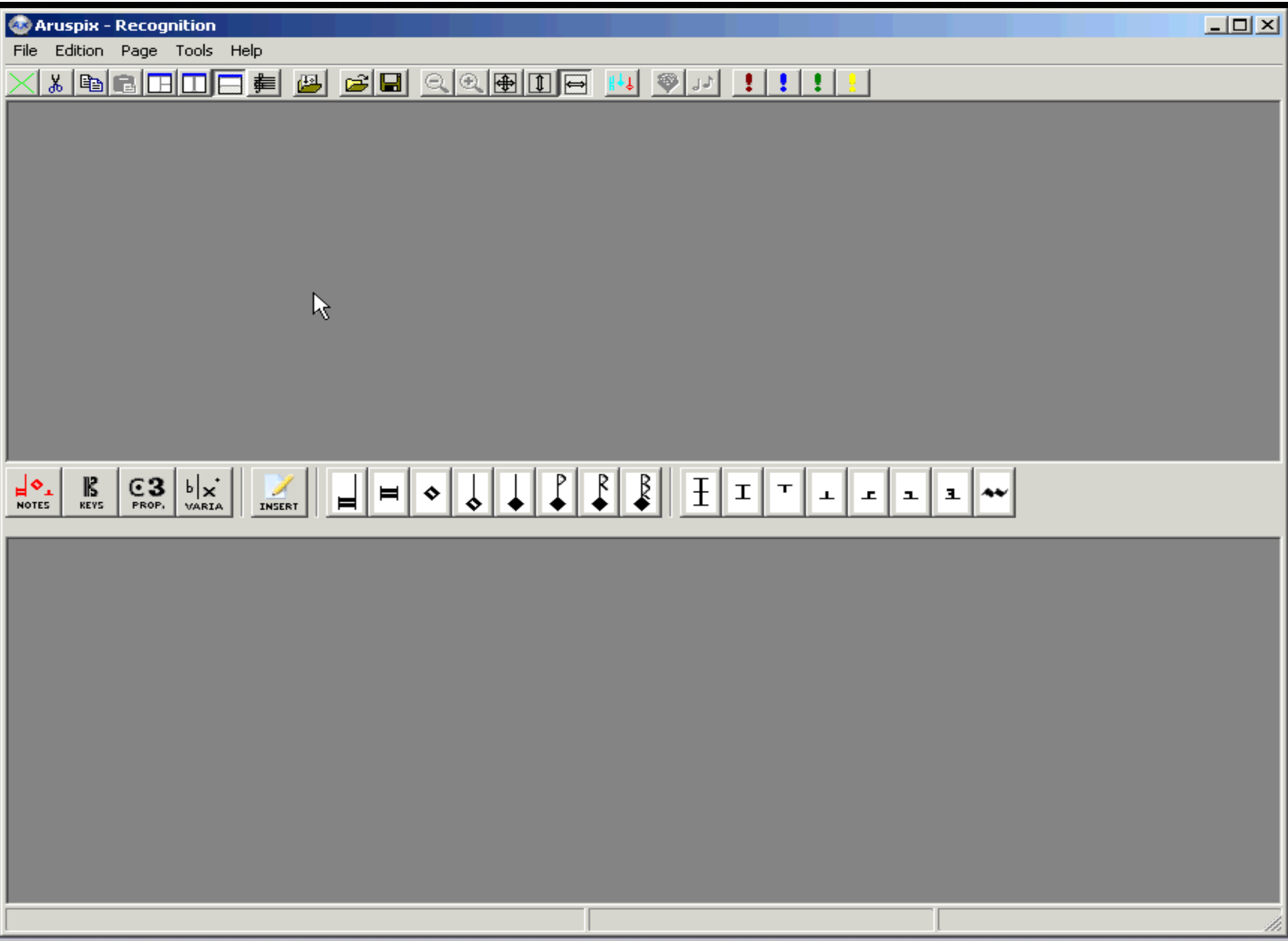


Typographic Music: music set with type (single-impression)

2002: Aruspix



- Developed by Laurent Pugin
- Specialized for typographic music
- Uses HMM (Hidden Markov Model)
- Does not remove staff lines


Aruspix - Recognition
File
Edition
Page
Tools
Help
NOTES
KEYS
PROP.
VARIA
INSERT

- Similar to “Google Books” minus Google
  - OMR (optical music recognition) to enable full-text search
  - Sophisticated music analysis and query
- Access to digitized scores world-wide from a single website
- SSHRC-funded 11-year project: 2011–2021: $4.4M

What would SIMSSA provide?

- Web-based OMR system with score editors
  - Rodan (Remote Online Document Analysis Network)
    - Gamera + Aruspix (a combination of existing OMR software)
  - Verovio (open-source music engraver) by Laurent Pugin
  - "Gradsourcing" to correct errors
  - Early music
- Web-based user interface to view, search, annotate, and analyze scores
  - MEI (Music Encoding Initiative) by Perry Roland and others
  - Diva.js (web-based IIIF-compatible document delivery system)
  - Humdrum / music21 (analytical tools)

The Vision: Global Music Library

Tools

diva.js

Music Encoding Initiative

Humdrum

music21

DIAMM
Digital Image Archive of Medieval Music

IMSLP

Early Music Online

Music Treasures Consortium

Metadata & Text

Cantus

RISM

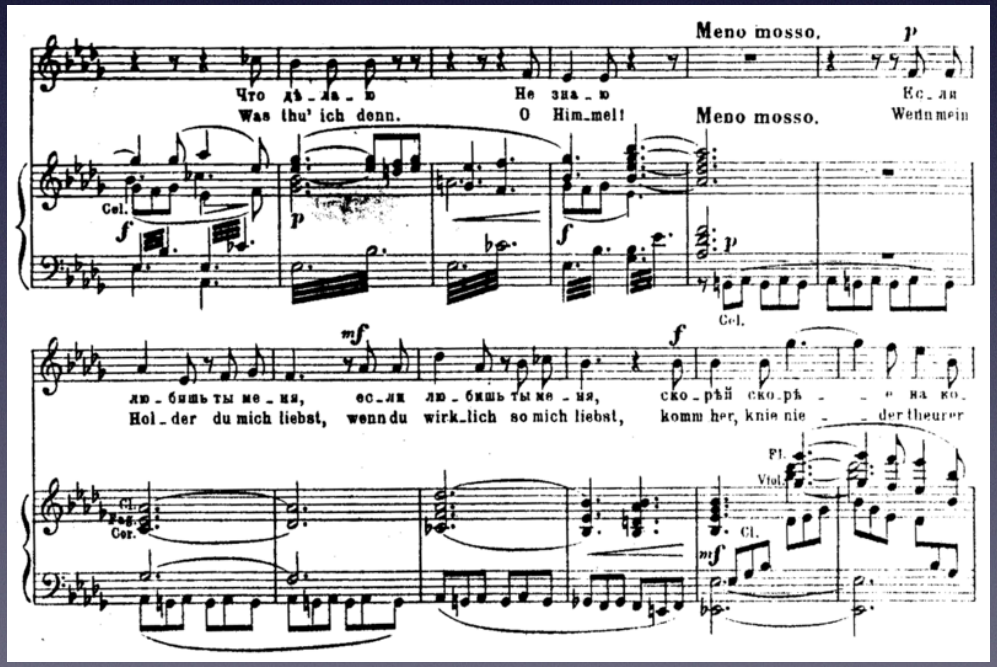
RILM

SIMSSA Team

- Musicologists (20)
- Music Librarians (8)
- Music Technologists (11)
- Partners (23) including:
  - Bavarian State Library
  - Bibliothèque nationale de France
  - British Library
  - Harvard University Music Library
  - HathiTrust Research Center
  - New York Philharmonic Archives

Optical Music Recognition (OMR)

A process of converting images of music scores into a symbolic computer representation, such as MIDI, MusicXML, or MEI (Music Encoding Initiative).



OMR

Steps Involved in OMR

Digitized Score

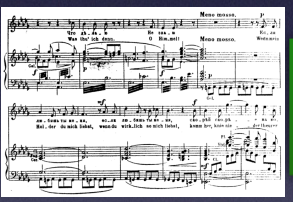


Image Preprocessing

- Binarization
- Noise Removal
- Structural Analysis
- Image Segmentation

Music Symbol Recognition

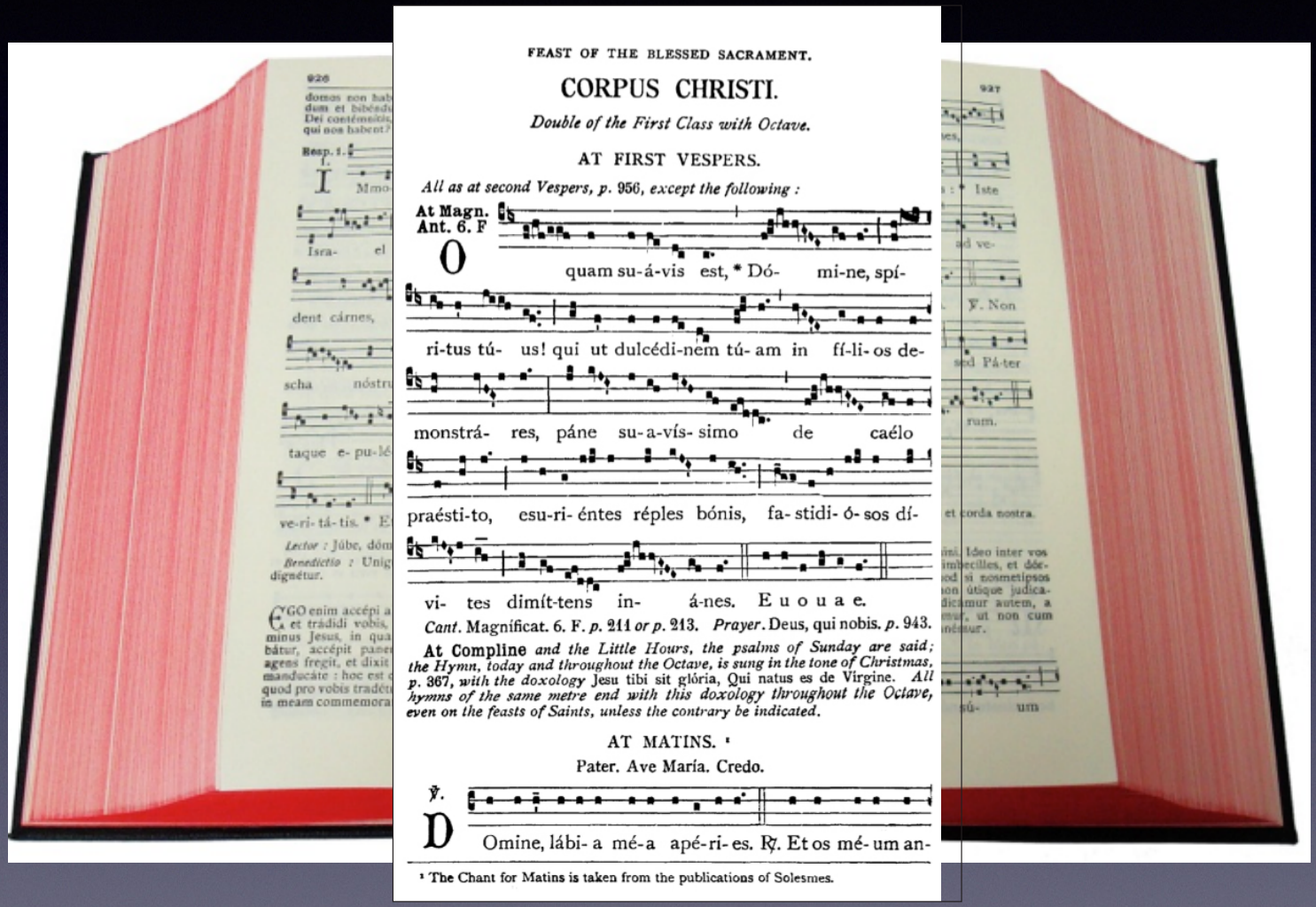
- Staves Processing
- Symbol Segmentation
- Symbol Classification

Music Notation Reconstruction

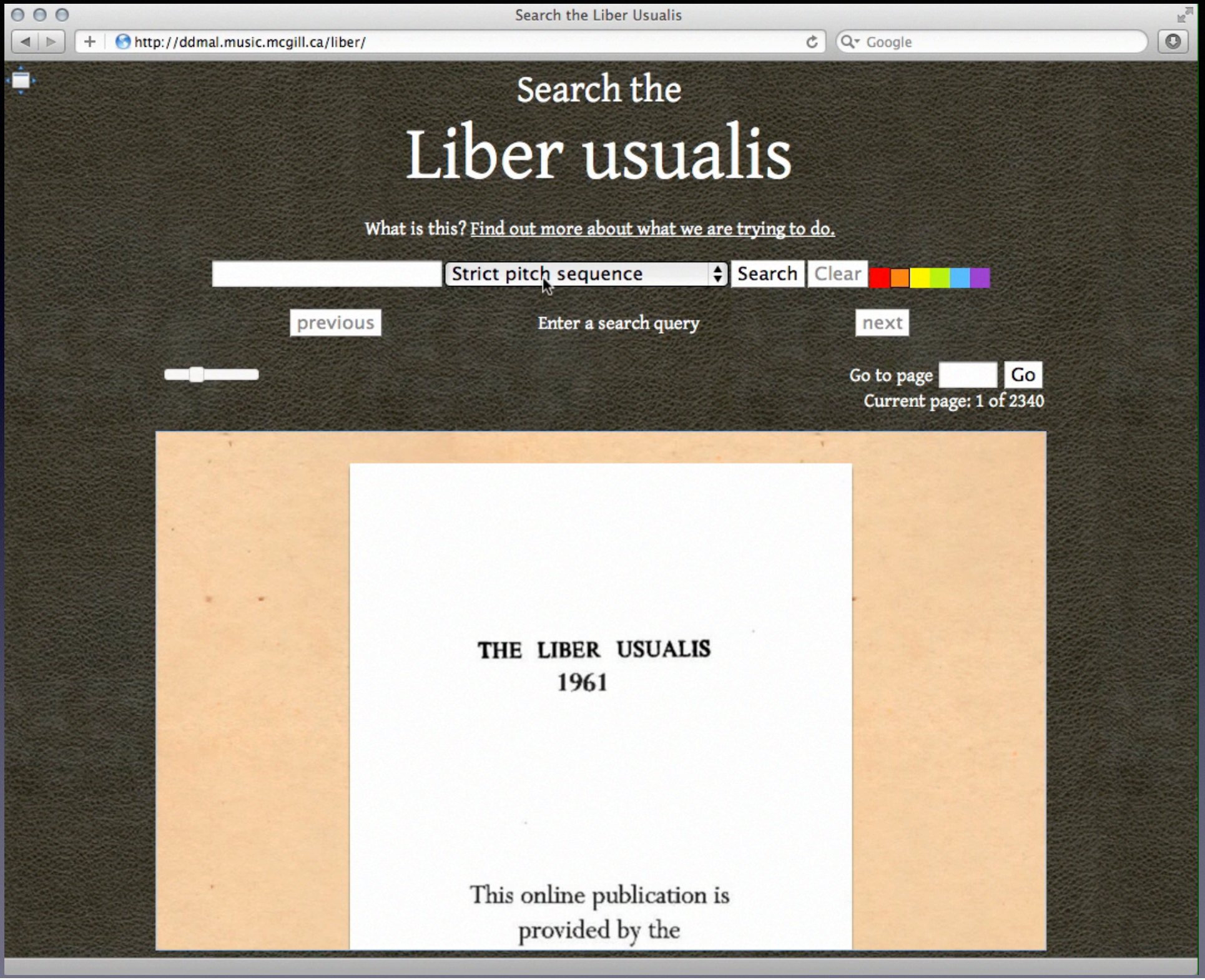
- Symbol Combination
- Semantic Assignment (pitch, value)
- Musical Structure Reconstruction

Final Output

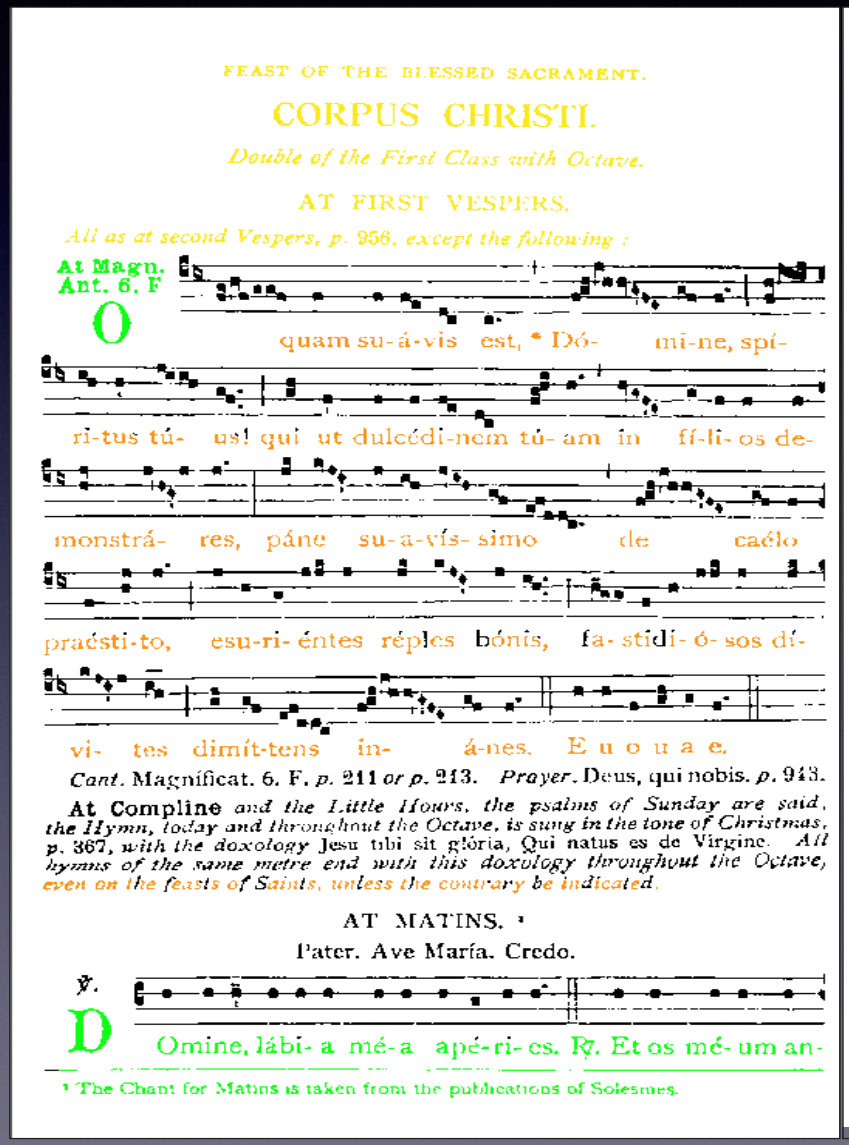
2011: *Liber Usualis* Project

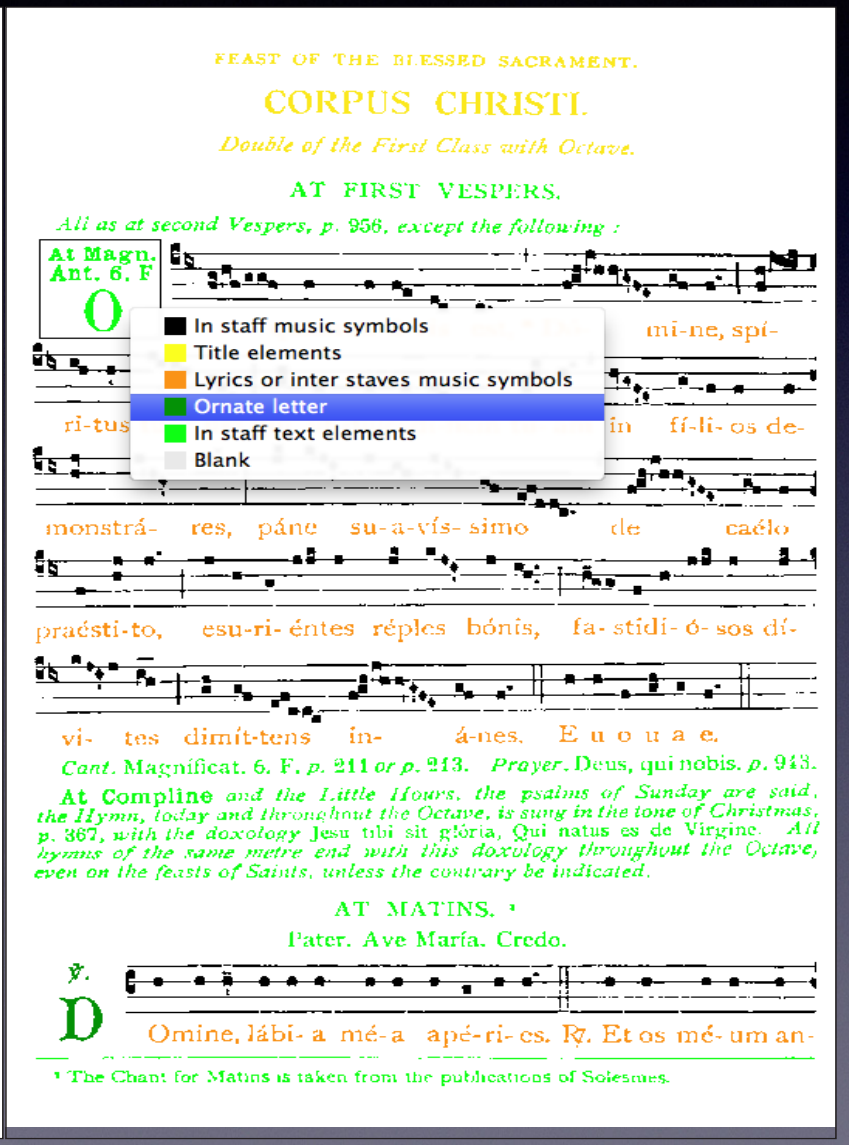


Full-text search of 2,000 pages of Latin text and square notation

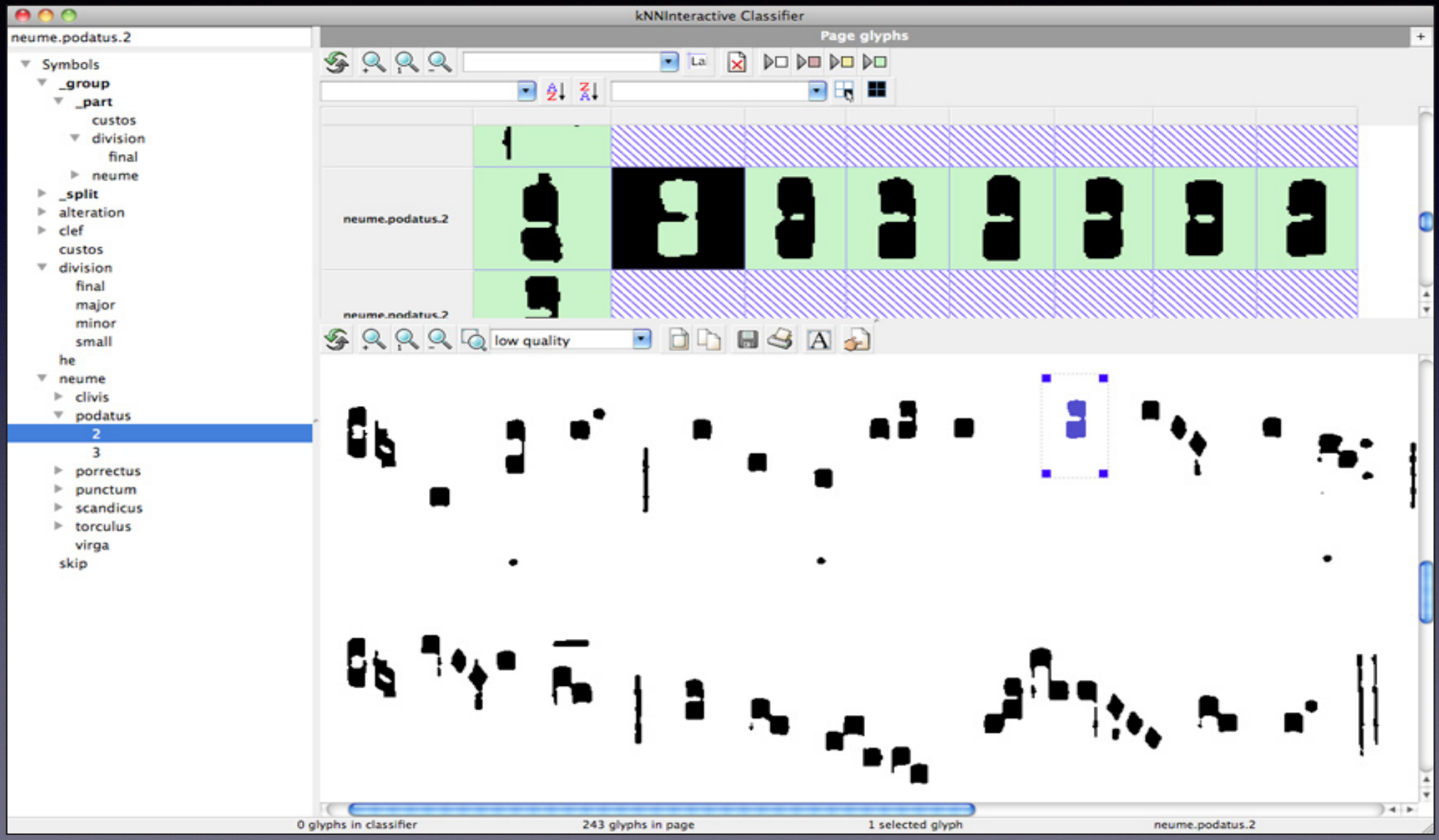

Search the Liber Usualis
http://ddmal.music.mcgill.ca/liber/
Google
Search the
Liber usualis
What is this? Find out more about what we are trying to do.
Strict pitch sequence
Search
Clear
previous
Enter a search query
next
Go to page
Go
Current page: 1 of 2340
THE LIBER USUALIS
1961
This online publication is
provided by the

Preprocessing: *Aruspix*

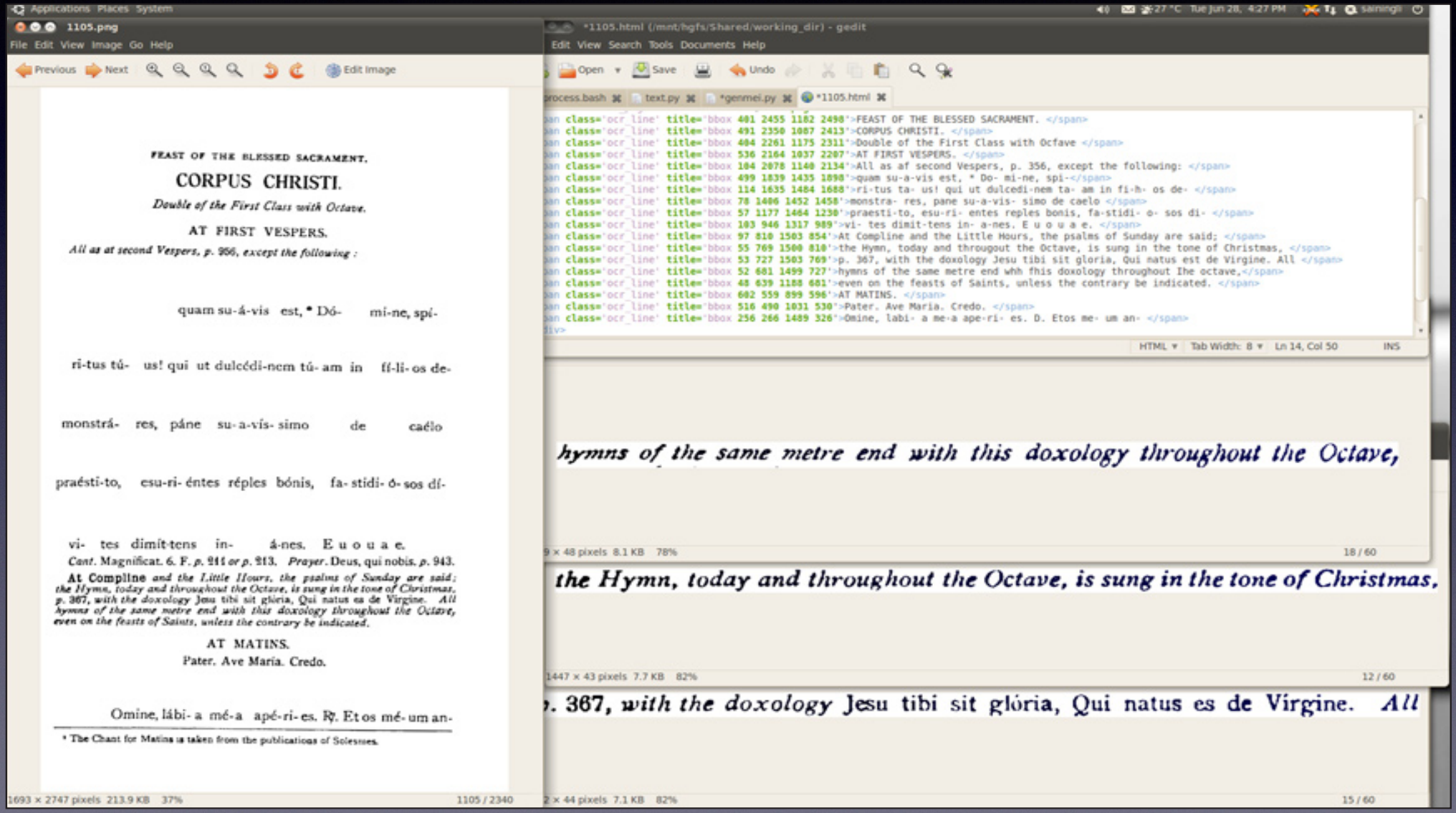




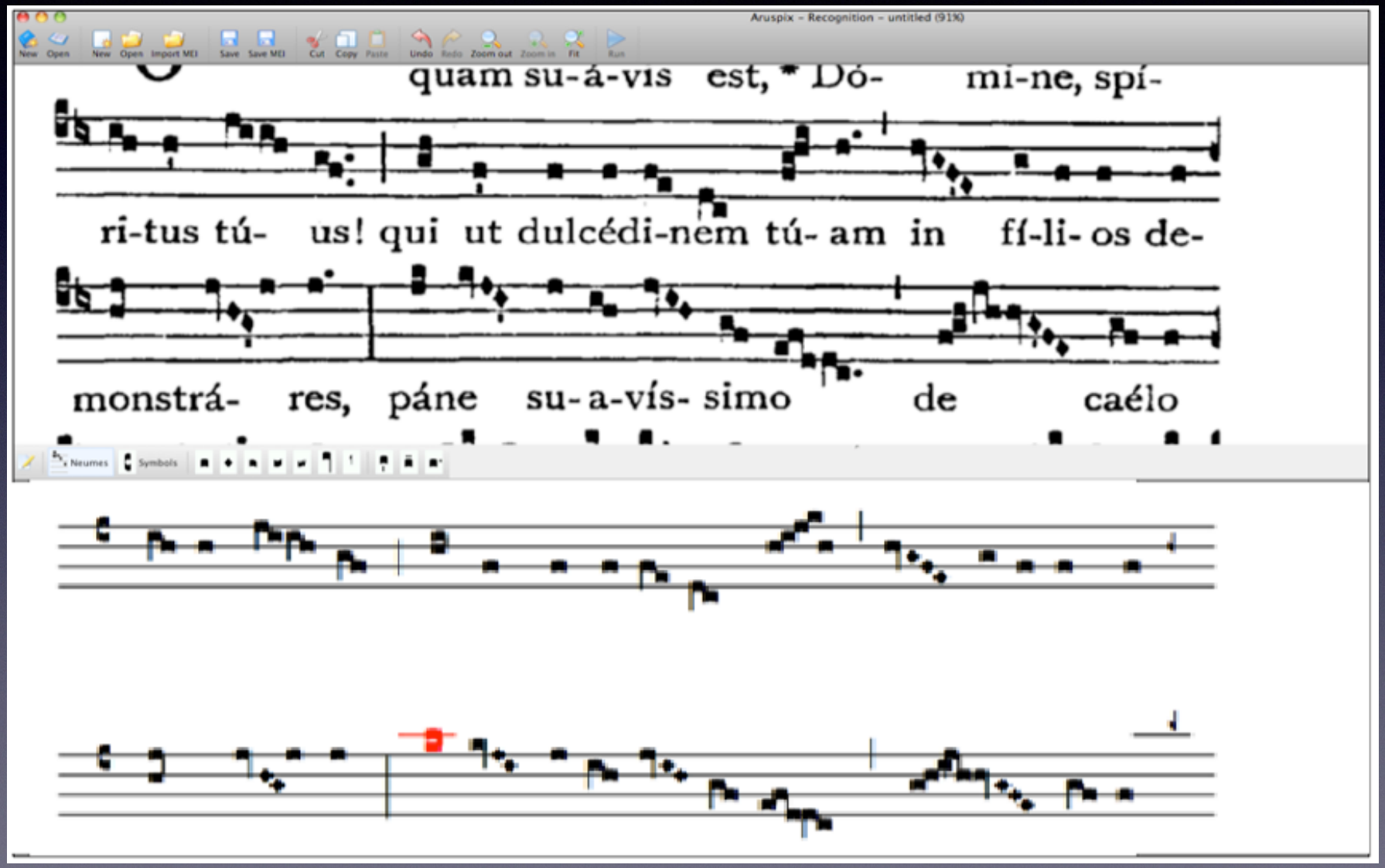
Music recognition: *Gamera*



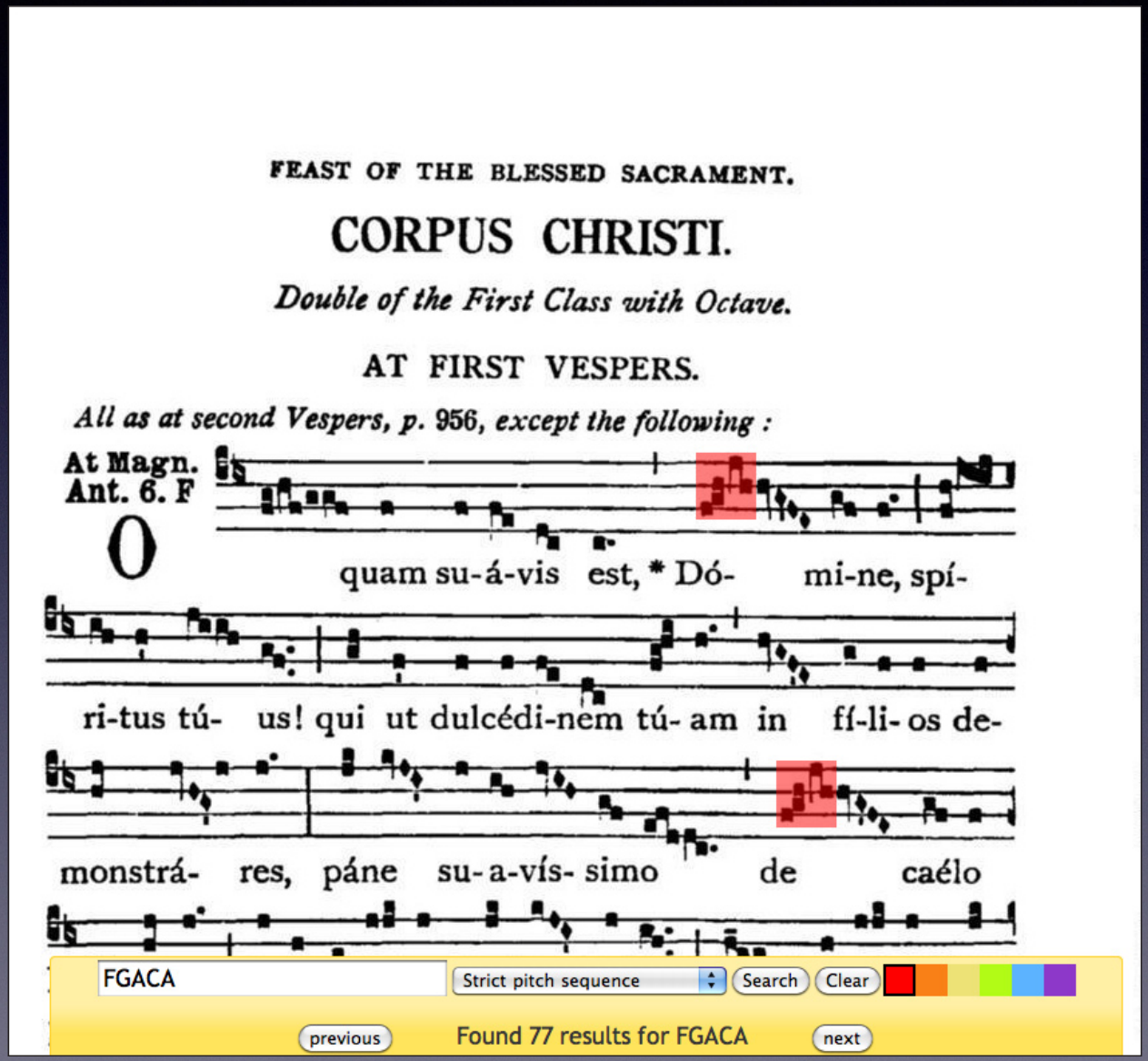
Text recognition: *Ocropus*



Pitch correction: *Aruspix*

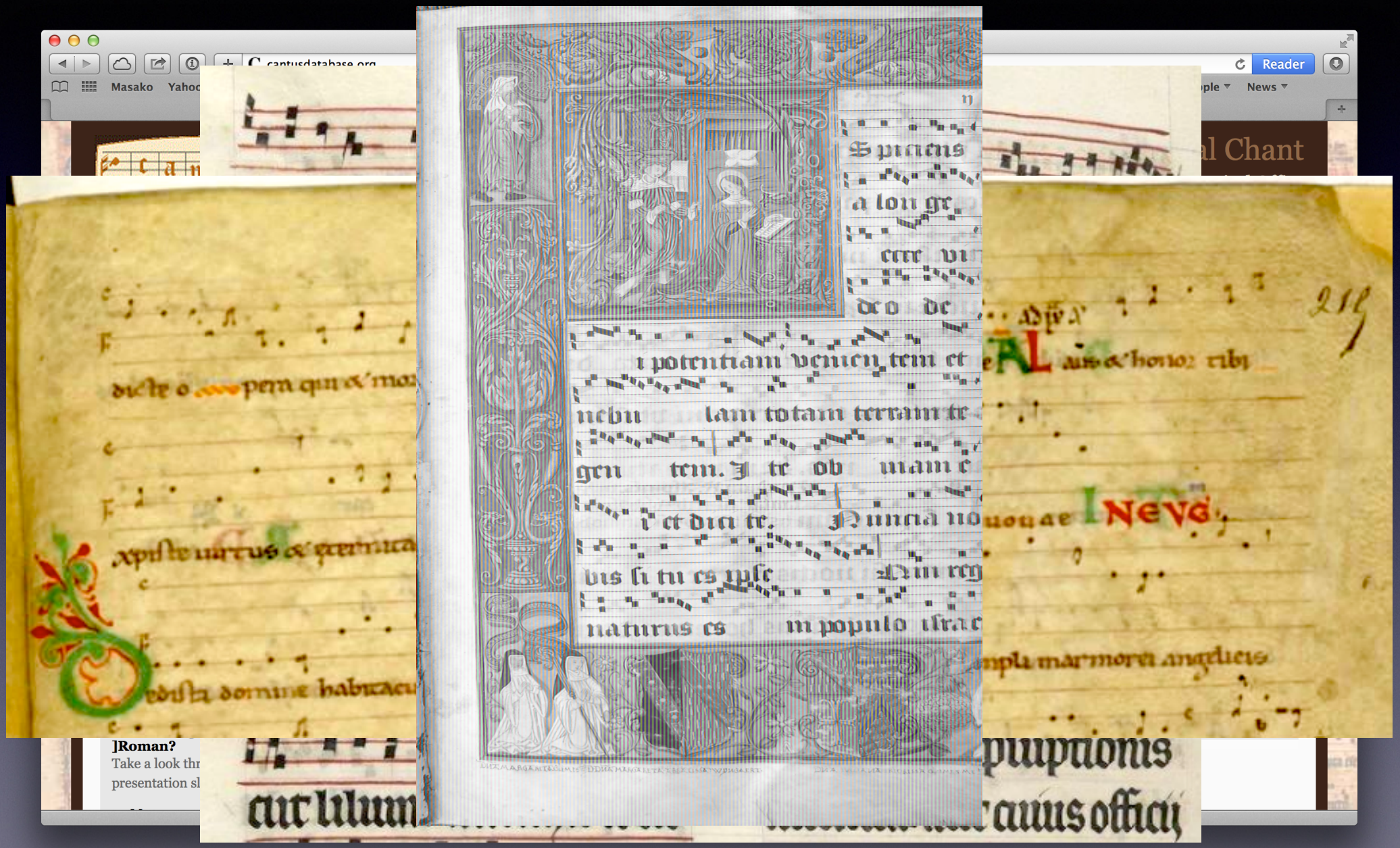


Web interface: *Diva.js*



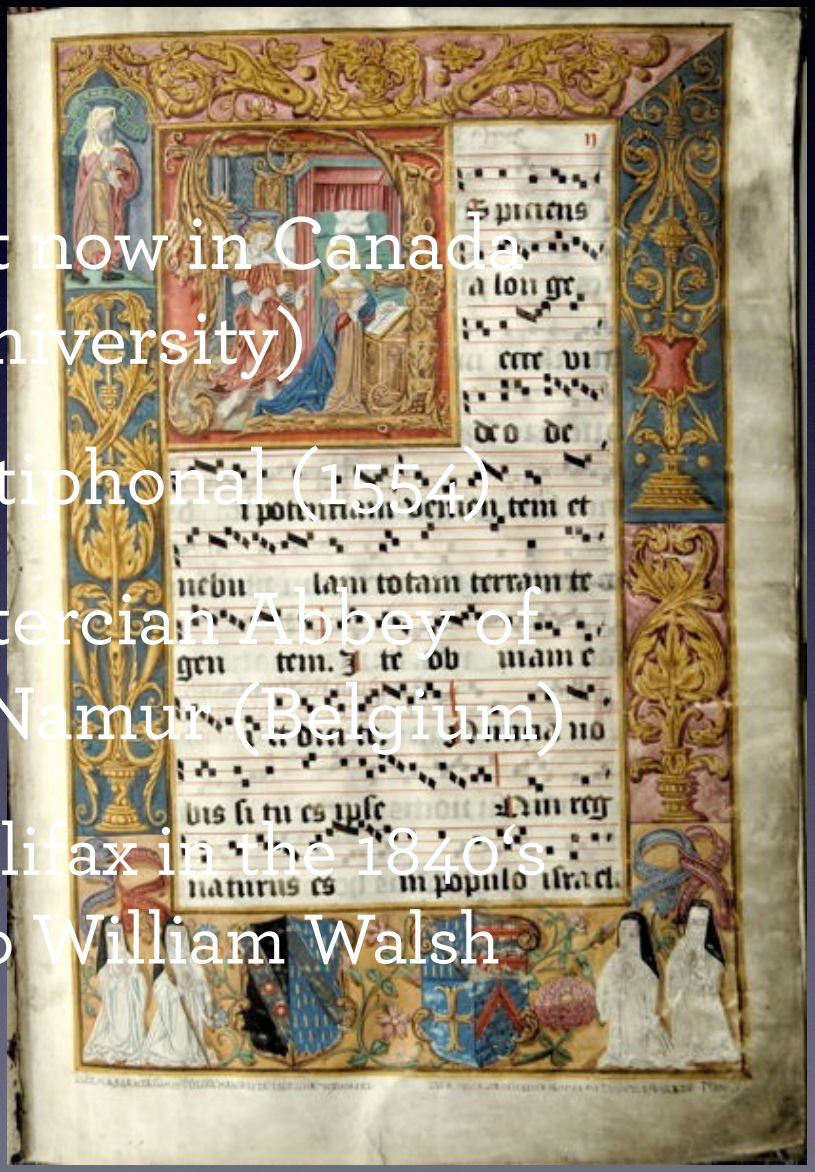
Cantus Ultimus

Main Goals of Cantus Ultimus



2012: *Salzinnes* Project

- A manuscript now in Canada (St. Mary's University)
- Salzinnes Antiphonal (1554)
- From the Cistercian Abbey of Salzinnes in Namur (Belgium)
- Brought to Halifax in the 1840's by Archbishop William Walsh



2012: Rodan

Andrew Hankinson

- Remote Online Document Analysis Network
- Wo
  rec
  -
  -
  -
  -
  -
  -
  -
  -
  -

Rodan: OMR Workflow Management System

Rodan Project Workflow Window Help

Status Users Pages Designer Jobs Results

Workflows
- Crop, greyscale, Niblack, despeckle
- try again: Crop, greyscale Niblack, de…
- Testing: crop, greyscale, BERNSEN, …

Runs
- Mon Jun 16 2014 10:33:39 G…

Pages
- csg-0390_017.jpg
- csg-0390_110.jpg

Workflow Run

URL: Mon Jun 16 2014 11:50:41 GMT-0400 (EDT)
Created on: Mon Jun 16 2014 10:33:39 GMT-0400 (EDT)
Updated on: http://rodan.simssa.ca/workflowrun/139ed692304d4402a34e8059c820f129/

Results Packages | Cancel Workflow Run

Run Jobs

Start Job | View Error Details | View Run Job Settings

Sequence	Job Name	Status	Error	UUID	Page Name
1	Crop Bord…	Has finished		9d65fc9012fe4…	csg-0390_110.jpg
2	To Greysc…	Has finished		30179f3e4e504…	csg-0390_110.jpg
3	Bernsen T…	Has finished		01ebd132cb88…	csg-0390_110.jpg
4	Rdn Desp…	Has finished		bf17ef9e78524…	csg-0390_110.jpg
6	Pixel Seg…	Has finished		478ba796135d…	csg-0390_110.jpg

Results

View Result

Run Job name Bernsen Threshold
URL: /data/projects/c50c031df54e4e4a986852e715f06ad8/workflows/837f52c
Created on: Mon Jun 16 2014 10:43:07 GMT-0400 (EDT)
Updated on: Mon Jun 16 2014 10:50:01 GMT-0400 (EDT)

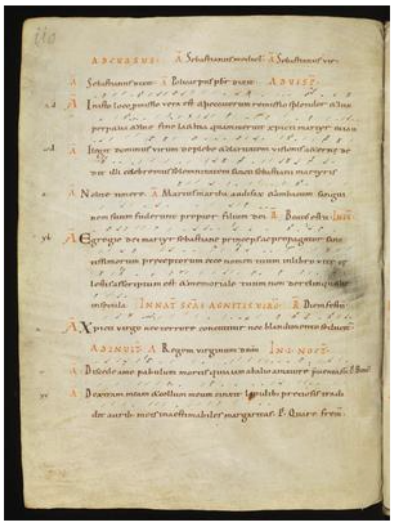
View Result

Run Job name Rdn Despeckle
URL: /data/projects/c50c031df54e4e4a986852e715f06ad8/workflows/837f52c
Created on: Mon Jun 16 2014 10:49:52 GMT-0400 (EDT)
Updated on: Mon Jun 16 2014 10:50:08 GMT-0400 (EDT)

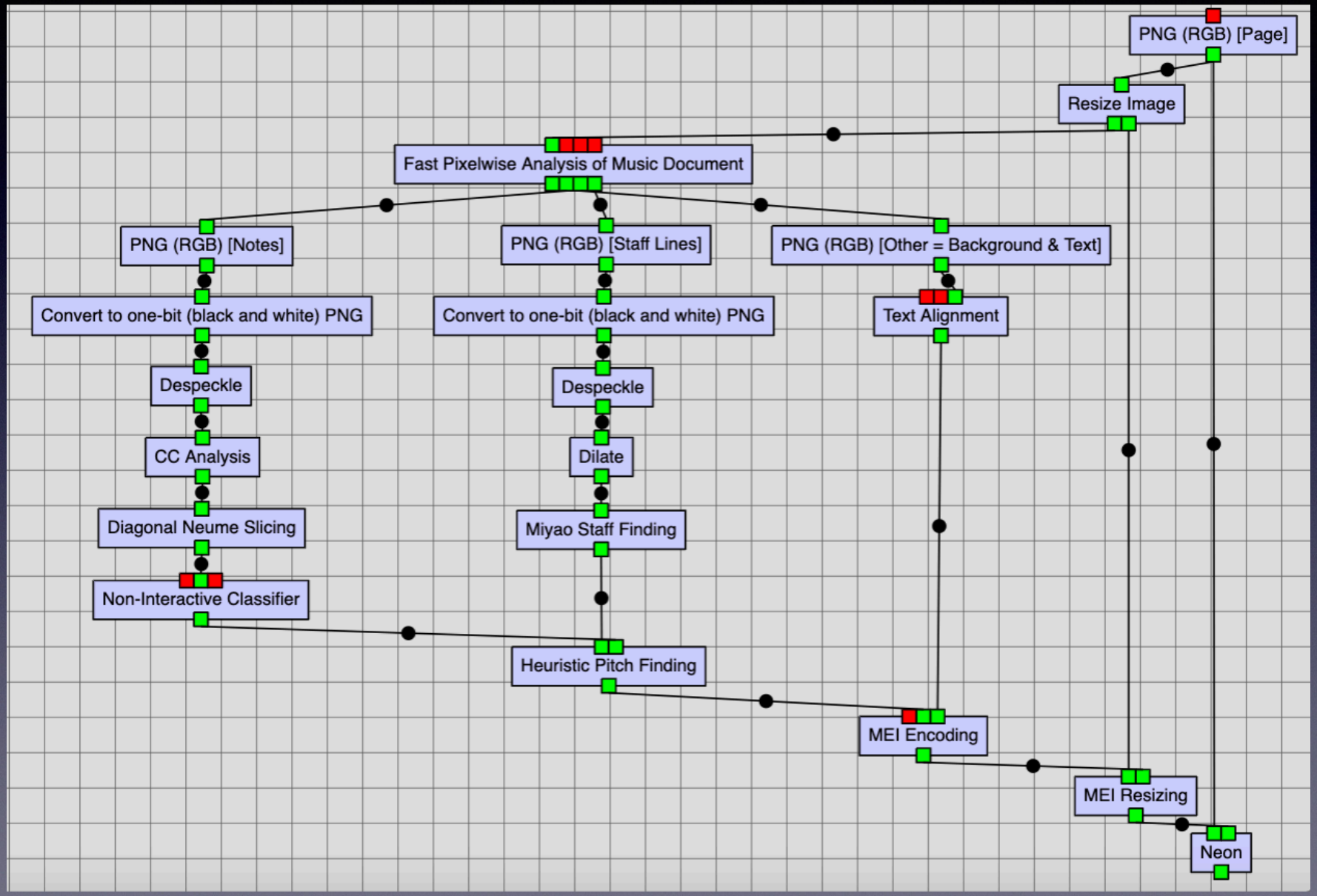
View Result

Page

Page name: csg-0390_110.jpg
URL: http://rodan.simssa.ca/page/f7603e9fb397461db367f77849771e5b/



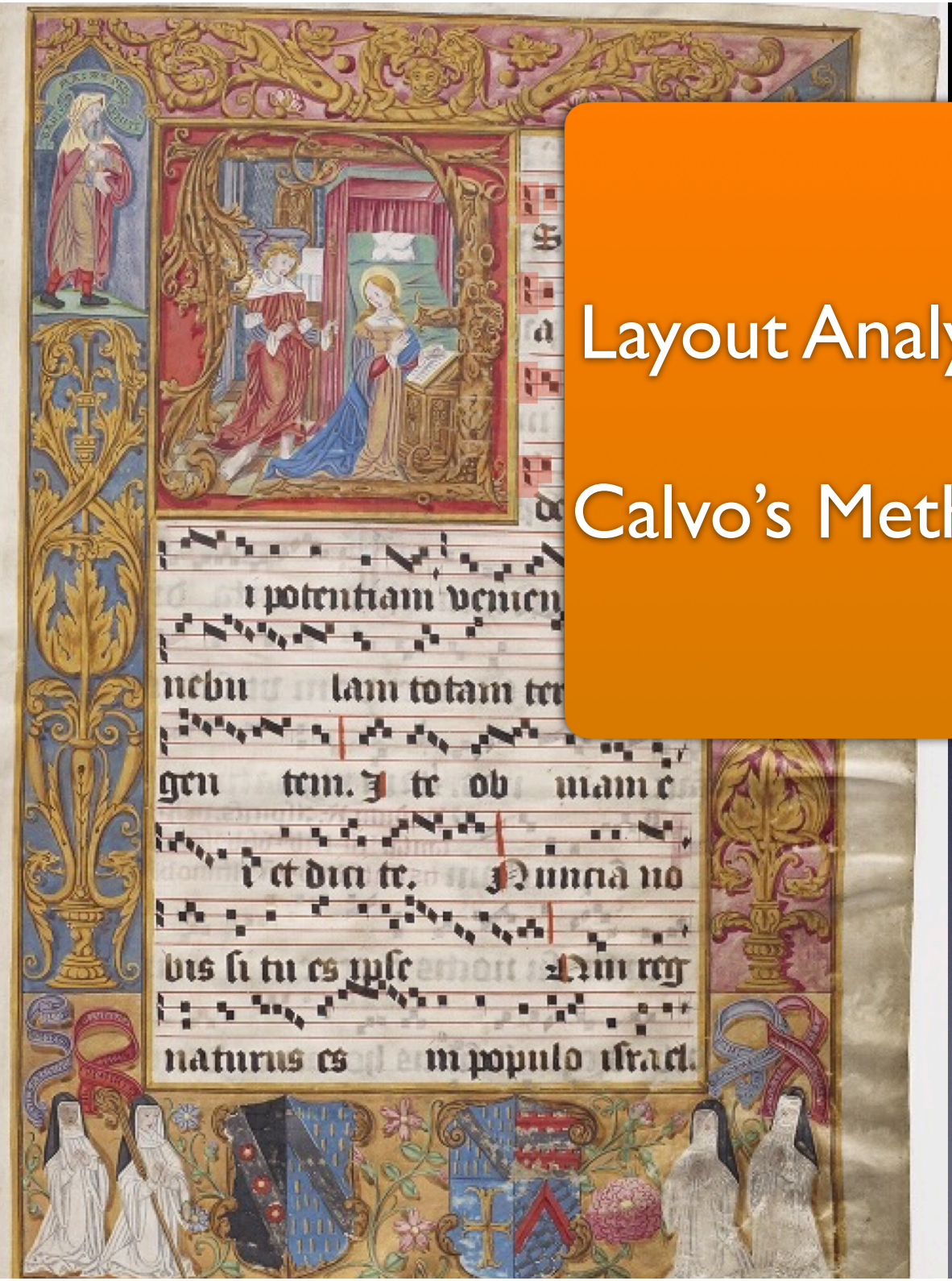
Rodan Workflow



2016: Breakthrough in OMR preprocessing!

- Pixel-level classification
  - Background
  - Text
  - Staff lines
  - Musical symbols
- Convolution Neural Network
- Jorge Calvo Zaragoza: “Calvo’s Method”

- Greyscale
- Binarization
- Border Removal
- Lyric Removal
- Staff Removal
- Shape Classification
- Music Reconstruction
- Shape/Image Alignment

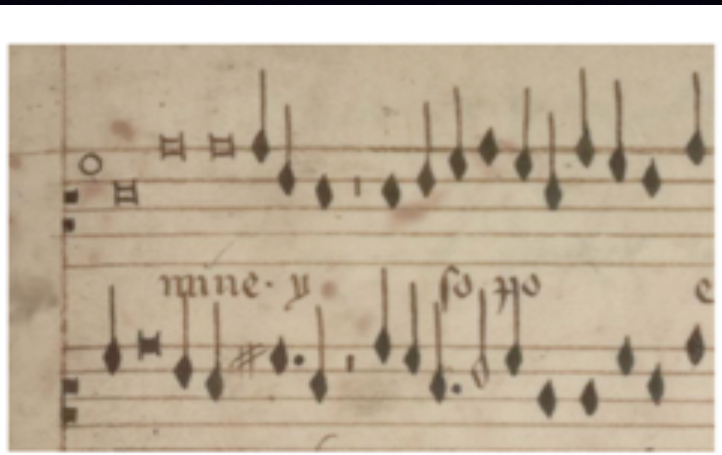


Layout Analysis

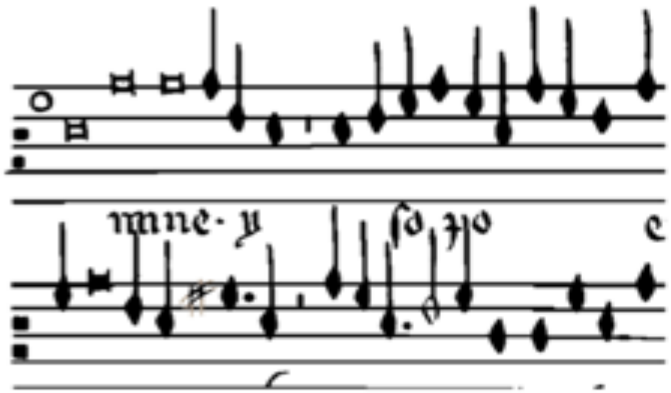
Calvo's Method

- Greyscale
- Binarization
- Border Removal
- Lyric Removal
- Staff Removal
- Shape Classification
- Music Reconstruction
- Shape/Image Alignment

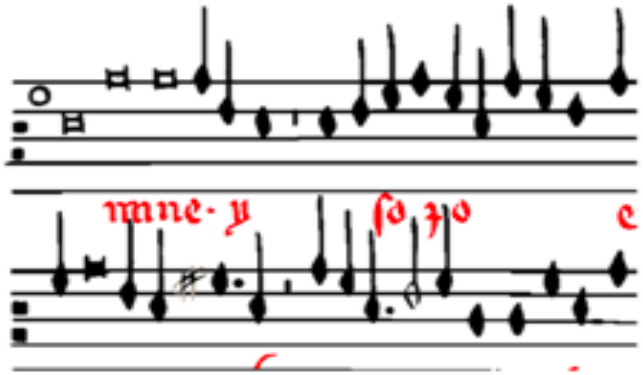
Three Different Outputs in One Step! Using Convolutional Neural Networks



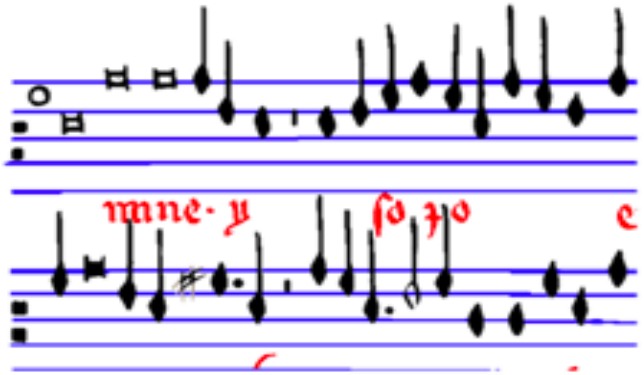
(a) Input image



(b) Binarization

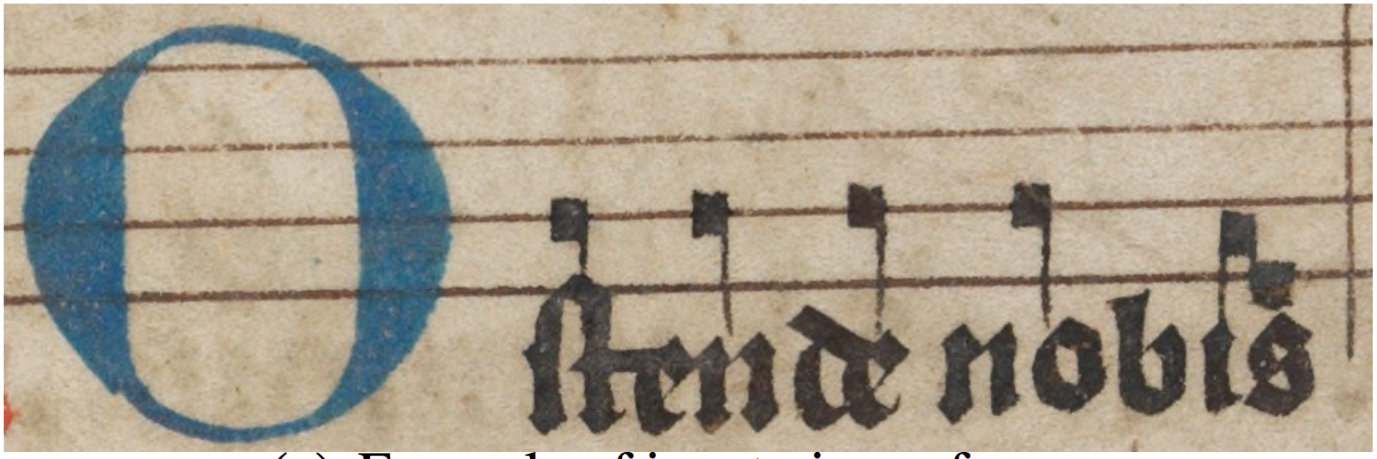


(c) Text detection

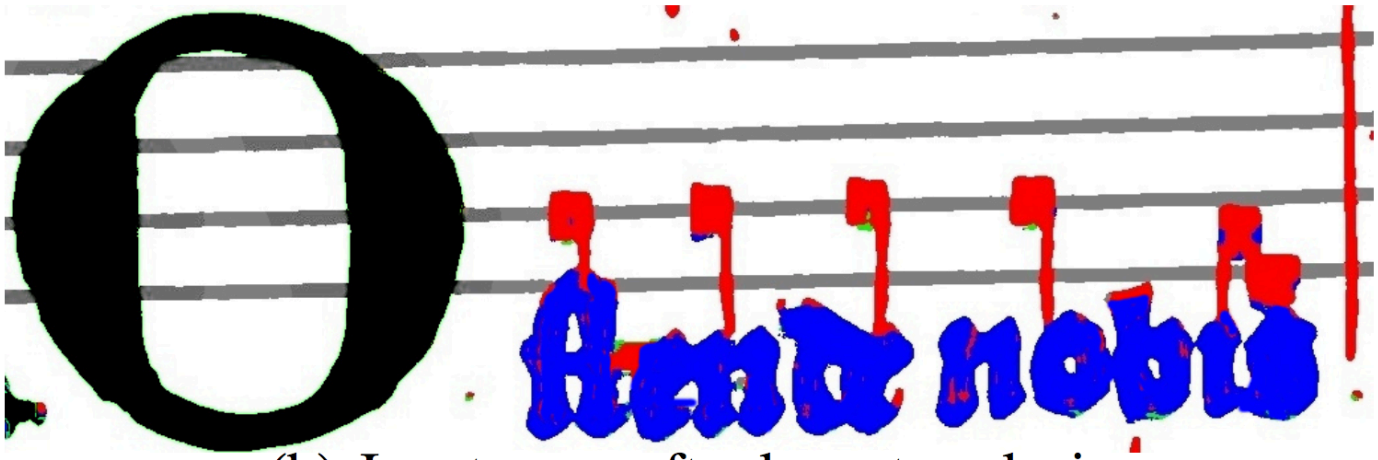


(d) Staff-line detection

Calvo's Method
Complete Layout Analysis



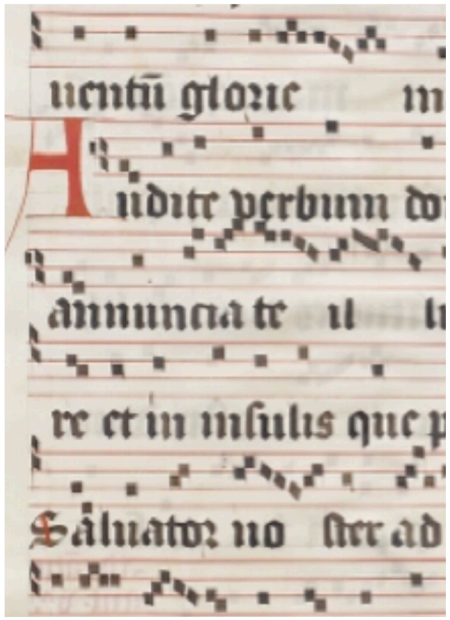
(a) Example of input piece of score



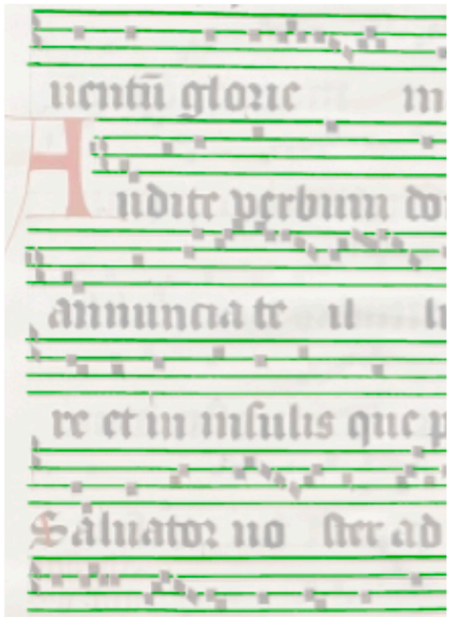
(b) Input score after layout analysis

Separation of Staff, Notes, & Text
Jorge Calvo Zaragoza

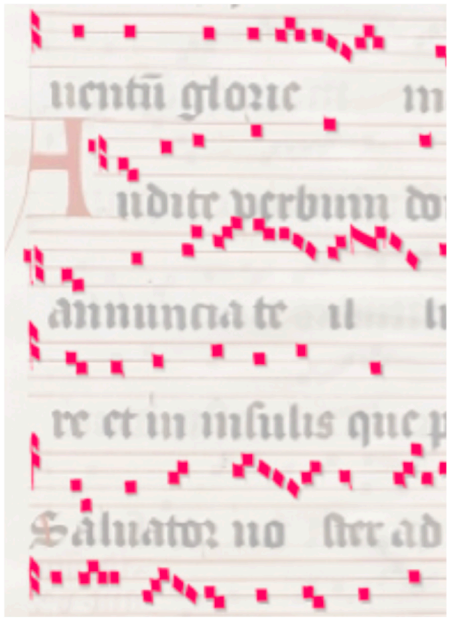
Actual



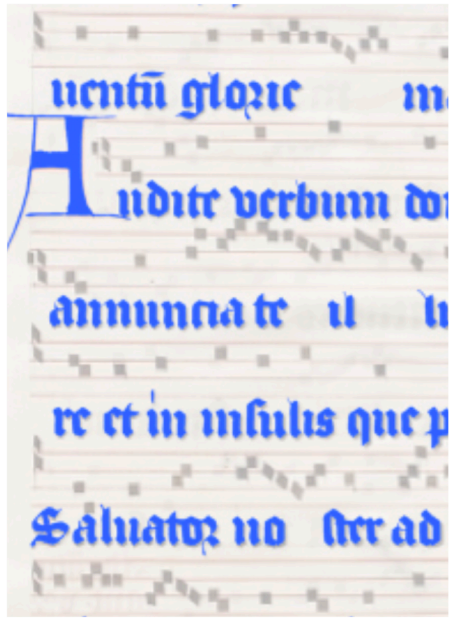
Staff



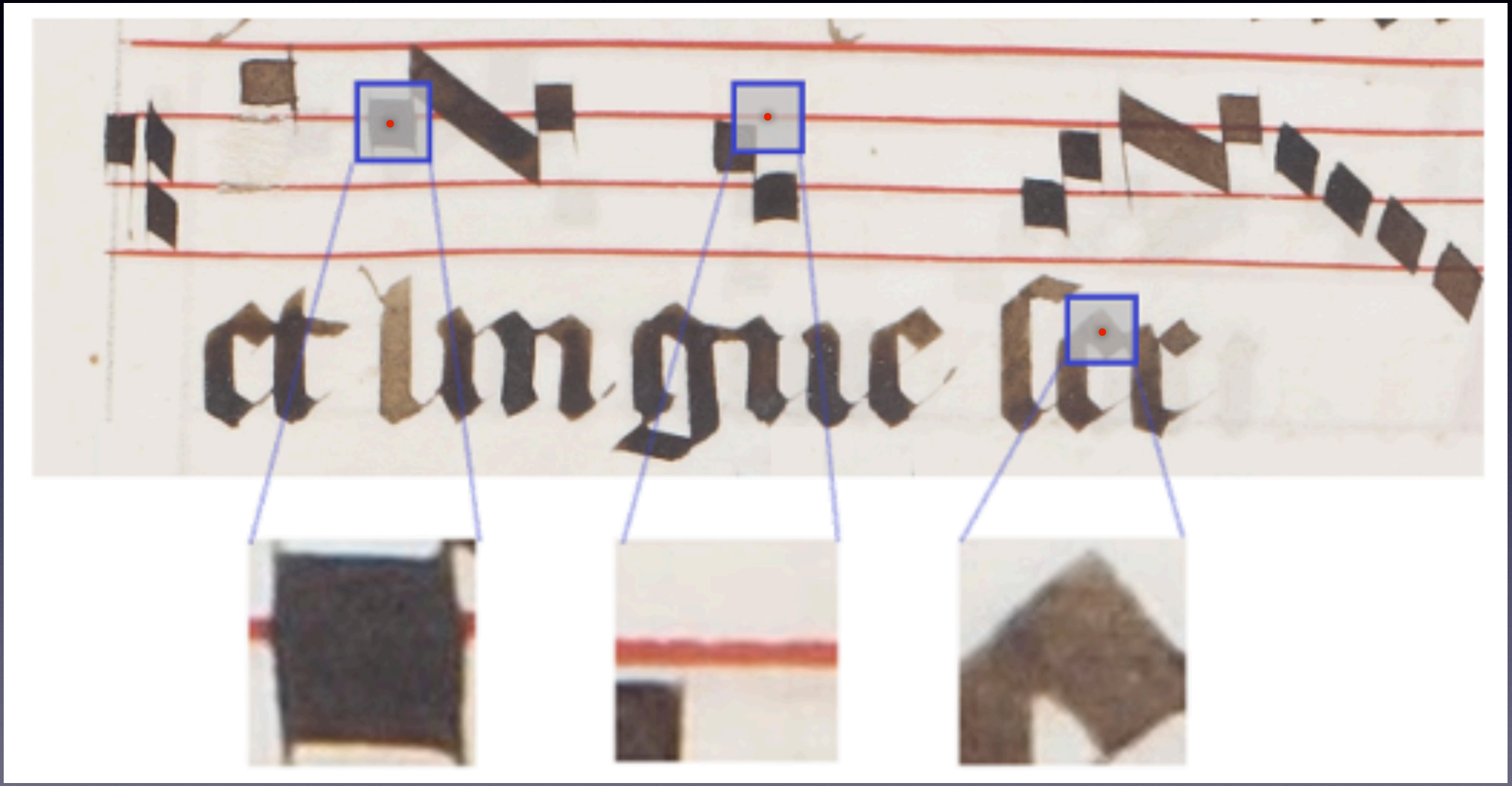
Note



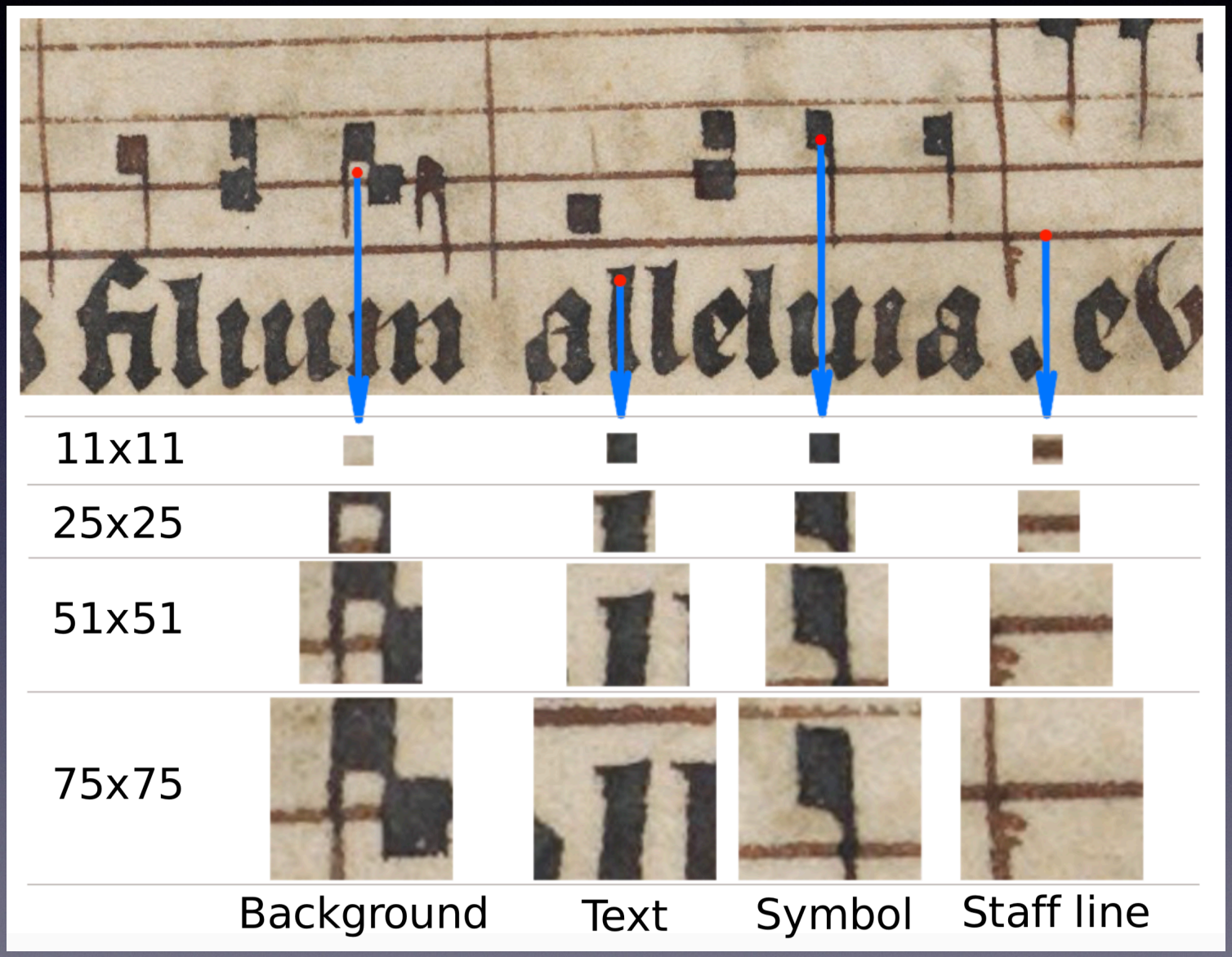
Text



Creating the Ground Truth

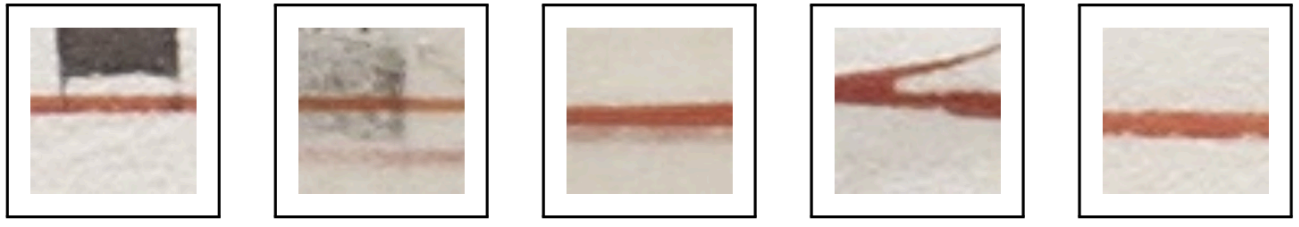


Examples of Different CNN Input Window Size



Samples of different classes

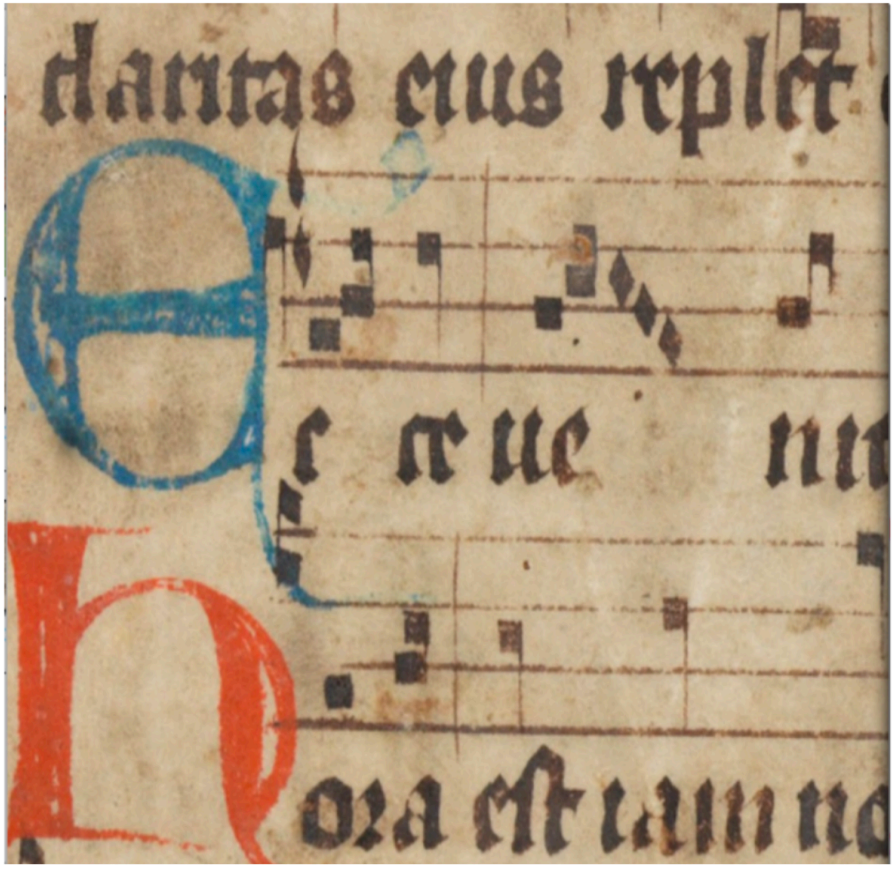
(a) Samples of *background* class

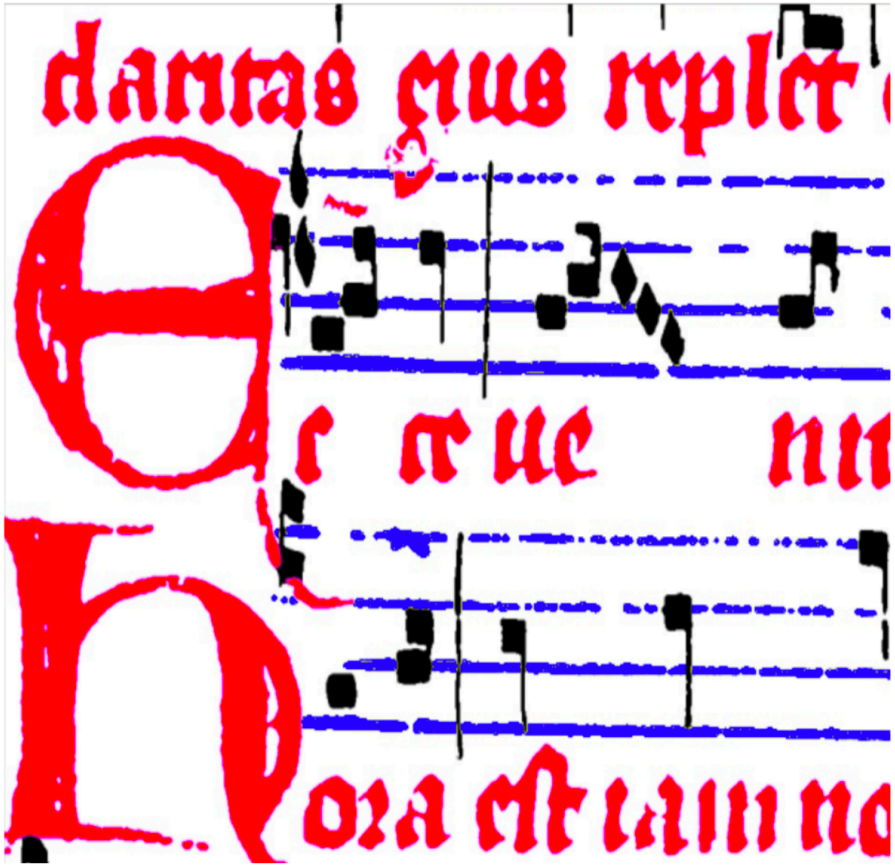


(b) Samples of *staff* class

(c) Samples of *text* class

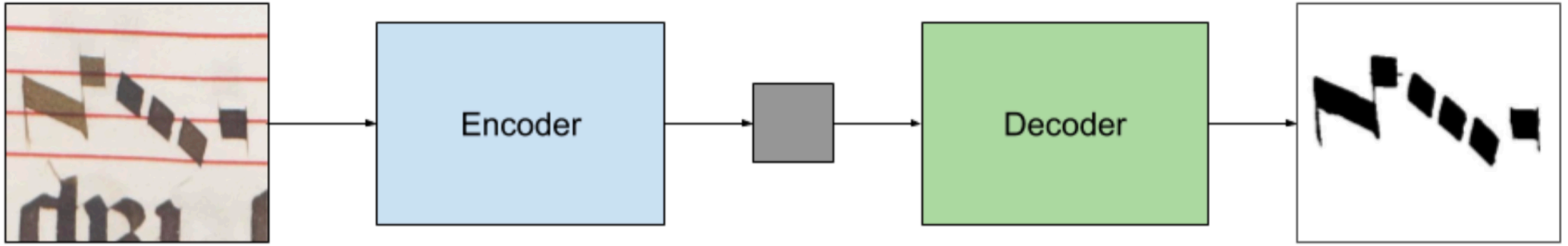
Separation of Staff, Notes, & Text

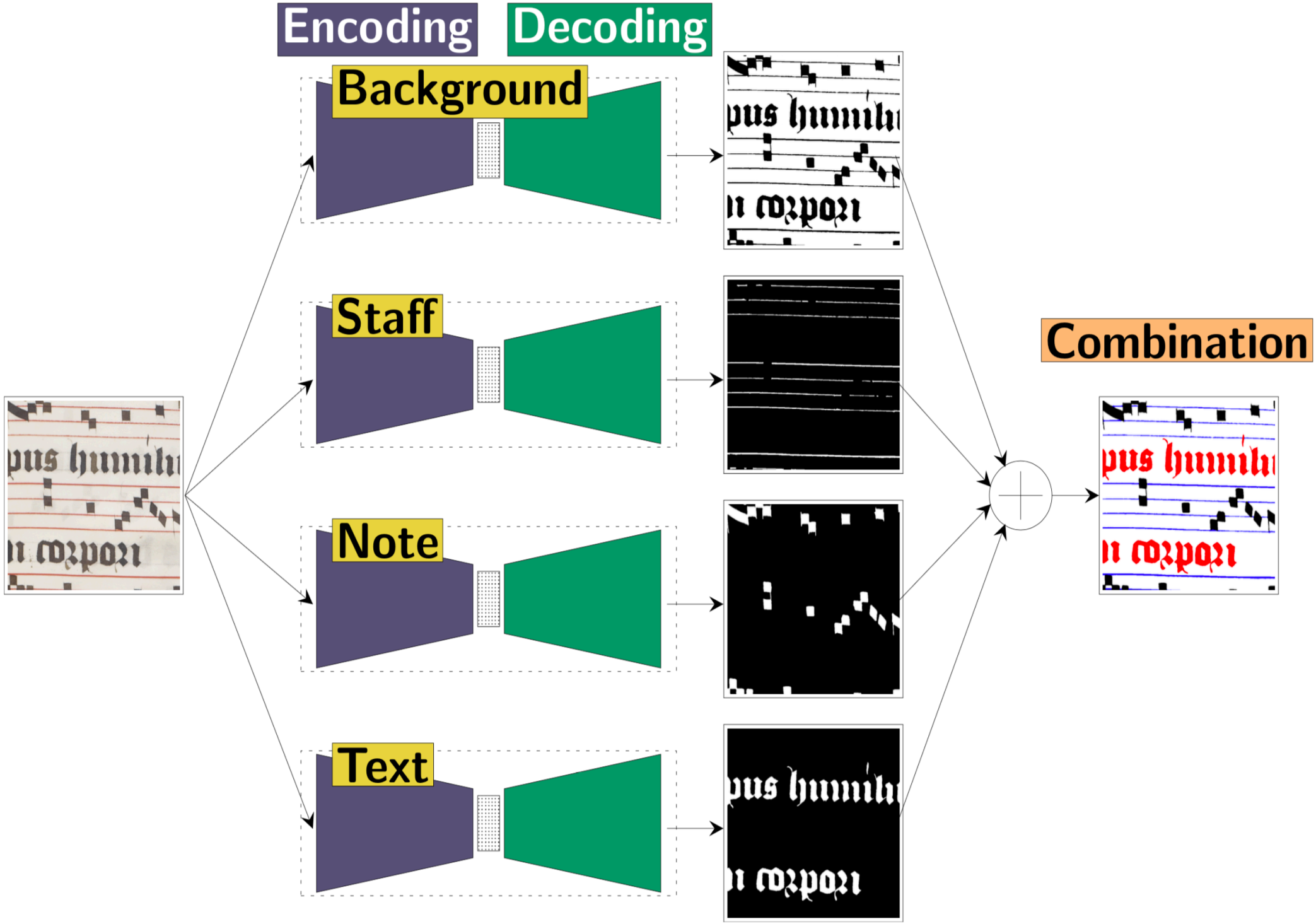




2018: A Different Neural Network Model
Selective Auto Encoders

Jorge Calvo Zaragoza




Encoding
Decoding
Background
Staff
Note
Text
Combination

Accuracy & Classification Time Comparison

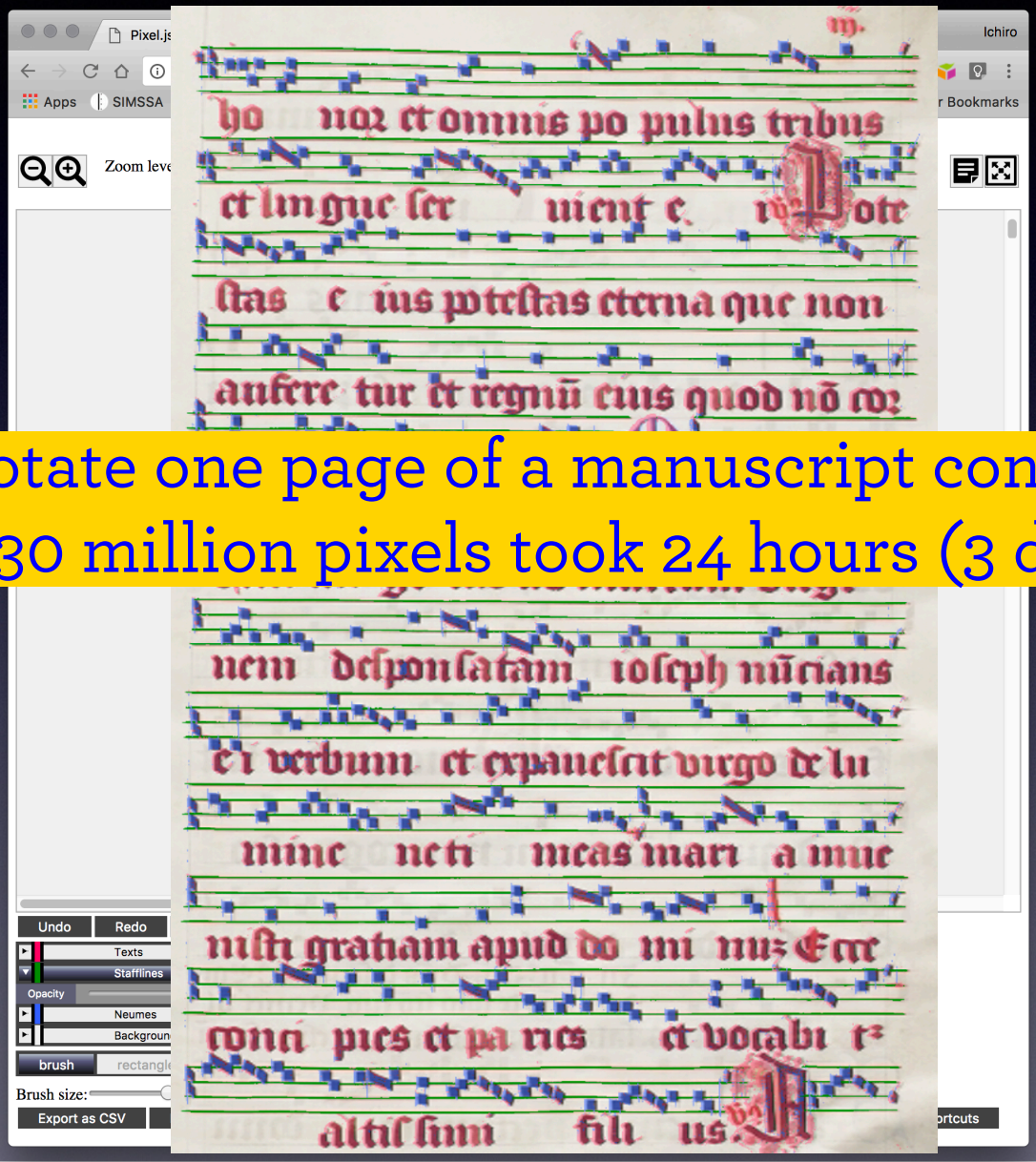
Selective Auto Encoders (SAE) vs Convolutional Neural Nets (CNN)

Two Medieval Manuscripts: Salzinnes & Einsiedeln

Strategy	Macro F_1		Time per page
	Salzinnes	Einsiedeln	
SAE	96.4	89.3	$\sim$ 1 minute
CNN	91.3	88.4	$\sim$ 6 hours

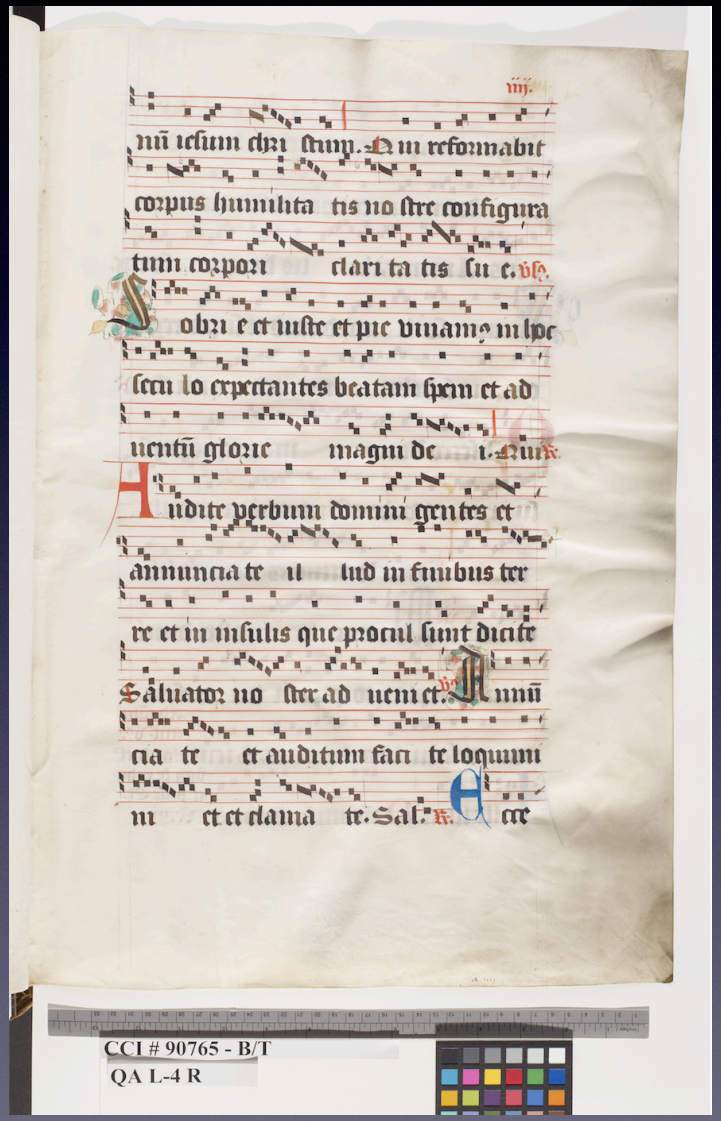
Pixel.js

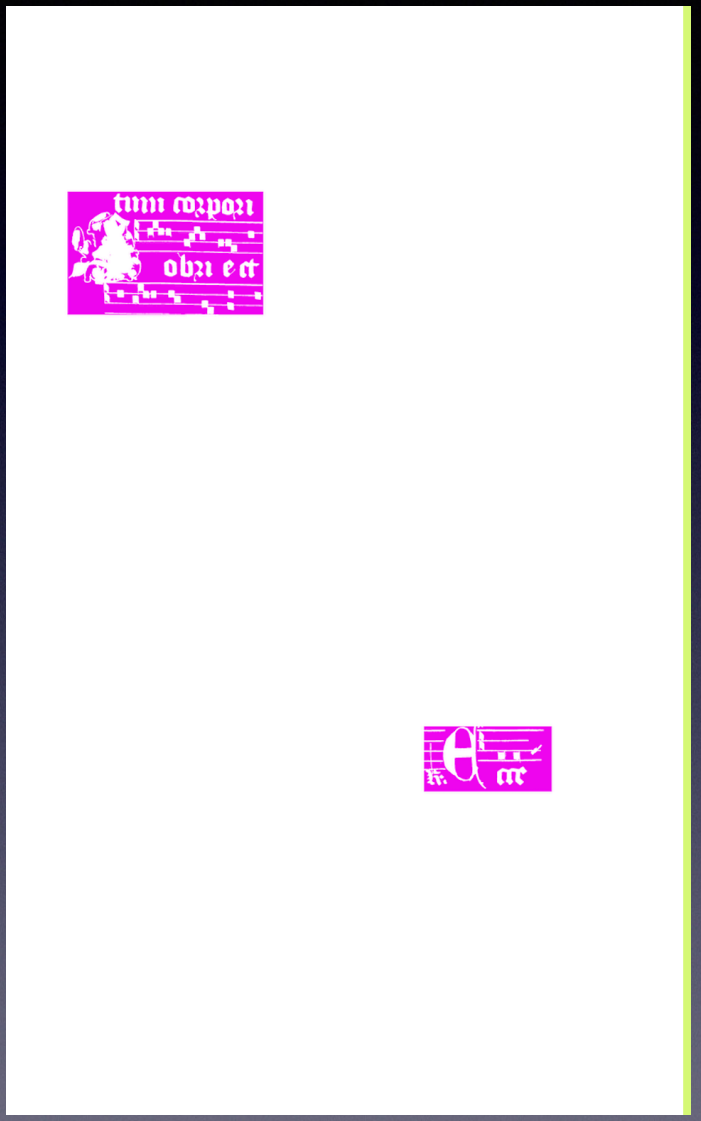
Zeyad Saleh, Ké Zhang, Eric Liu



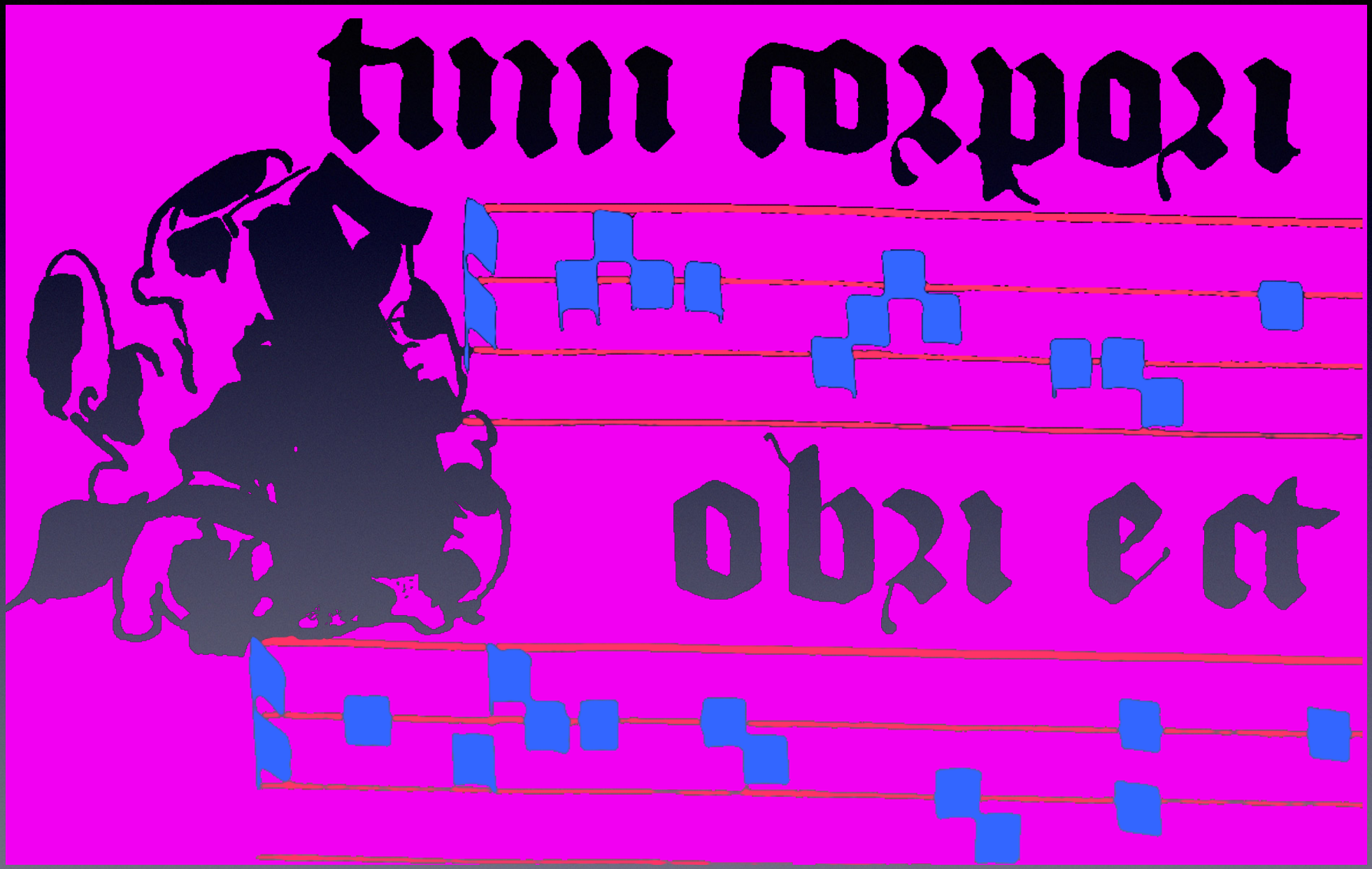
To annotate one page of a manuscript containing over 30 million pixels took 24 hours (3 days!)

Partial Creation of Ground Truth

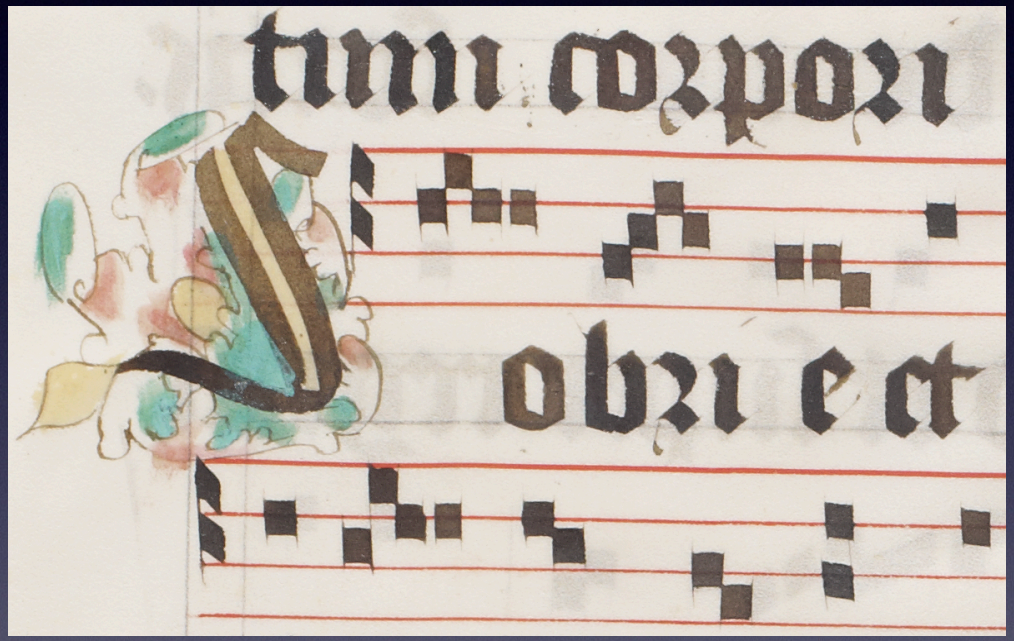




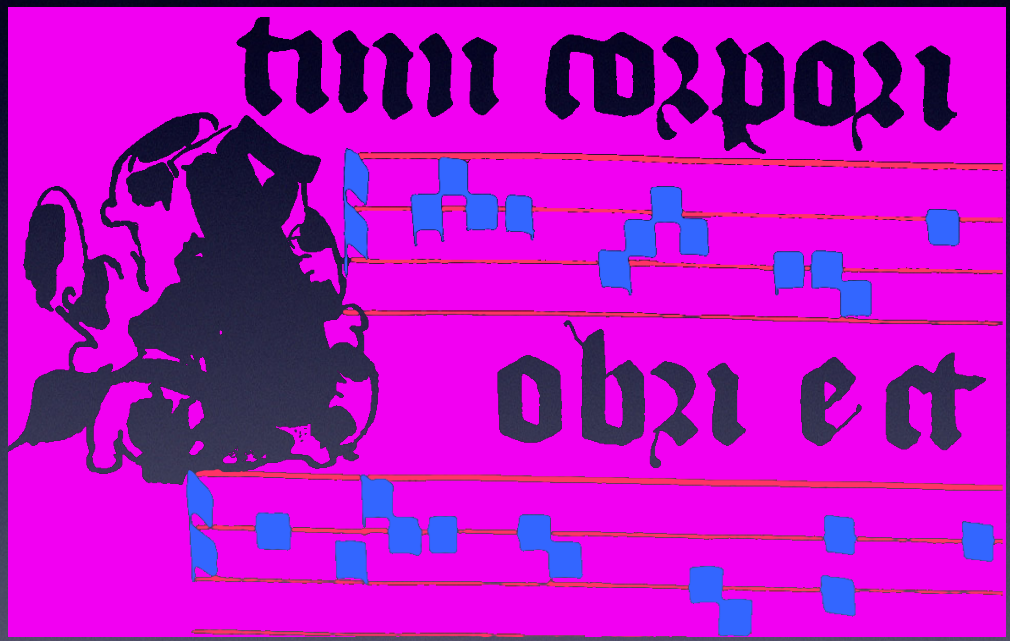
Ground Truth



Original Image & Ground Truth

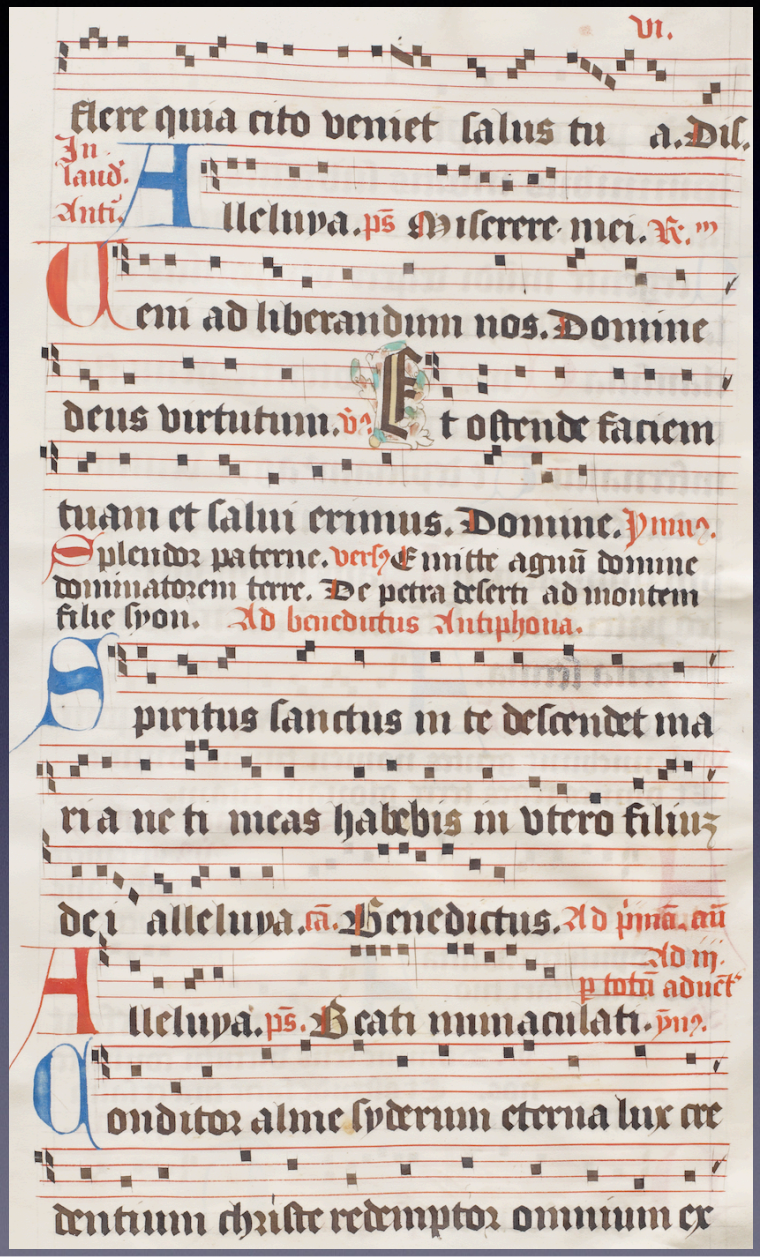


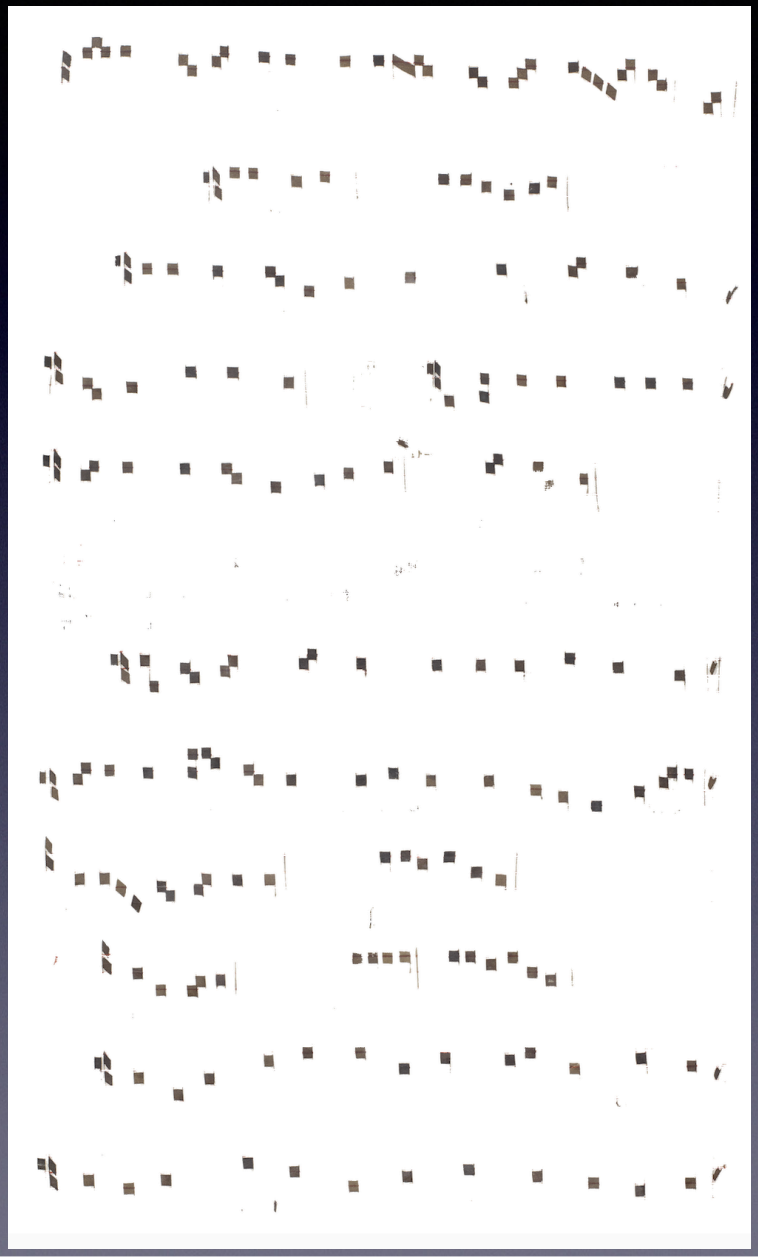
Original Image



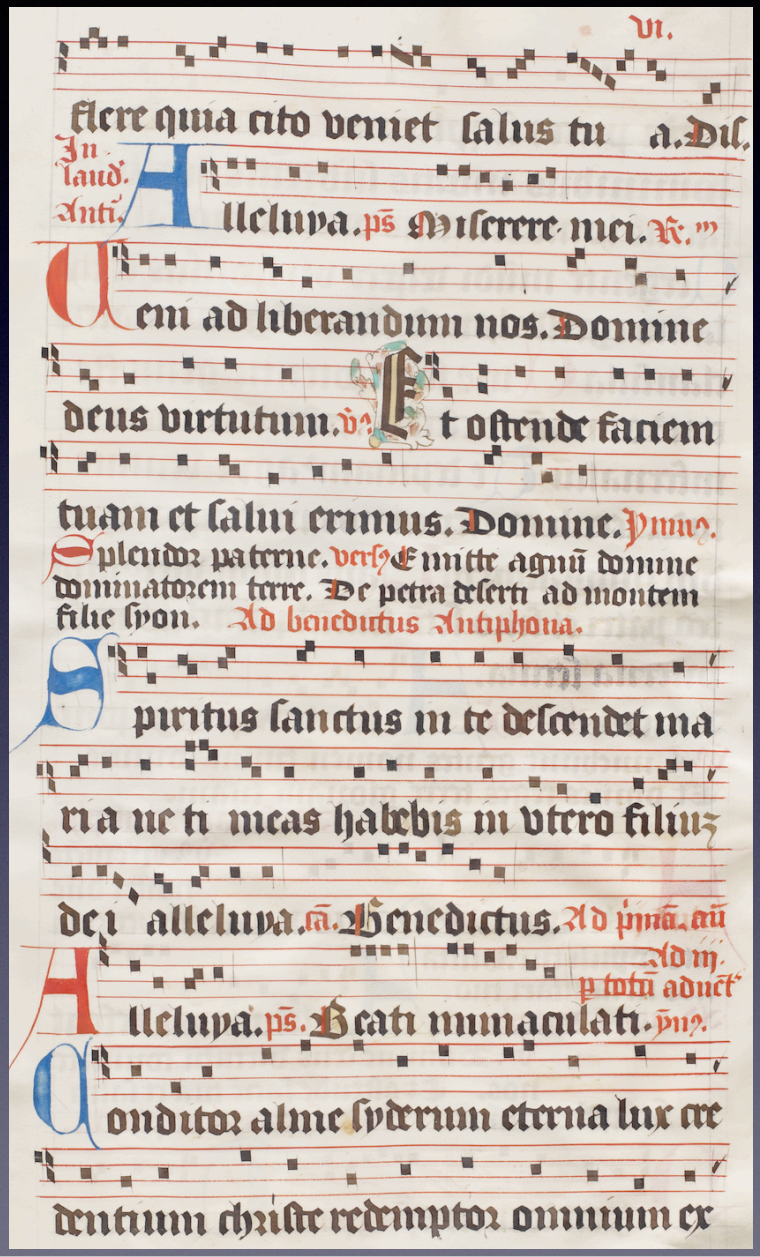
Ground Truth

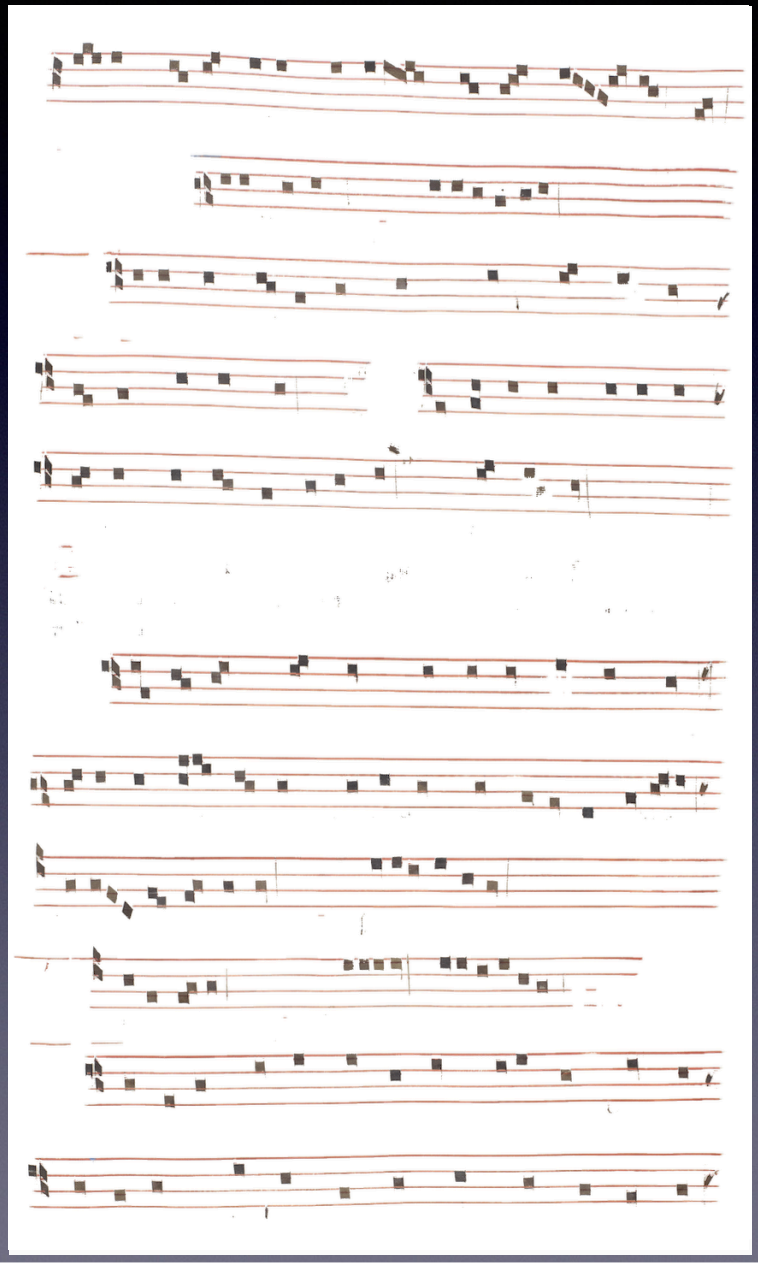
Classification of a Page: Notes



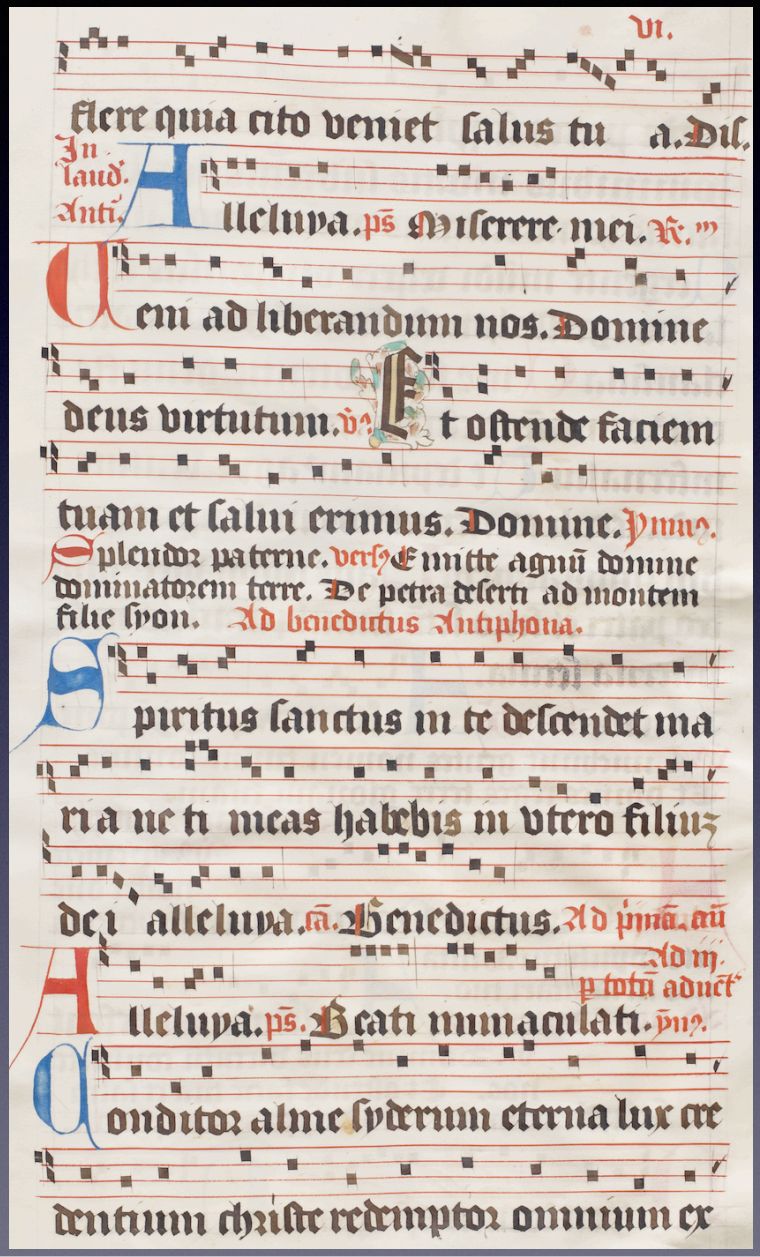


Classification of a Page: with Staves

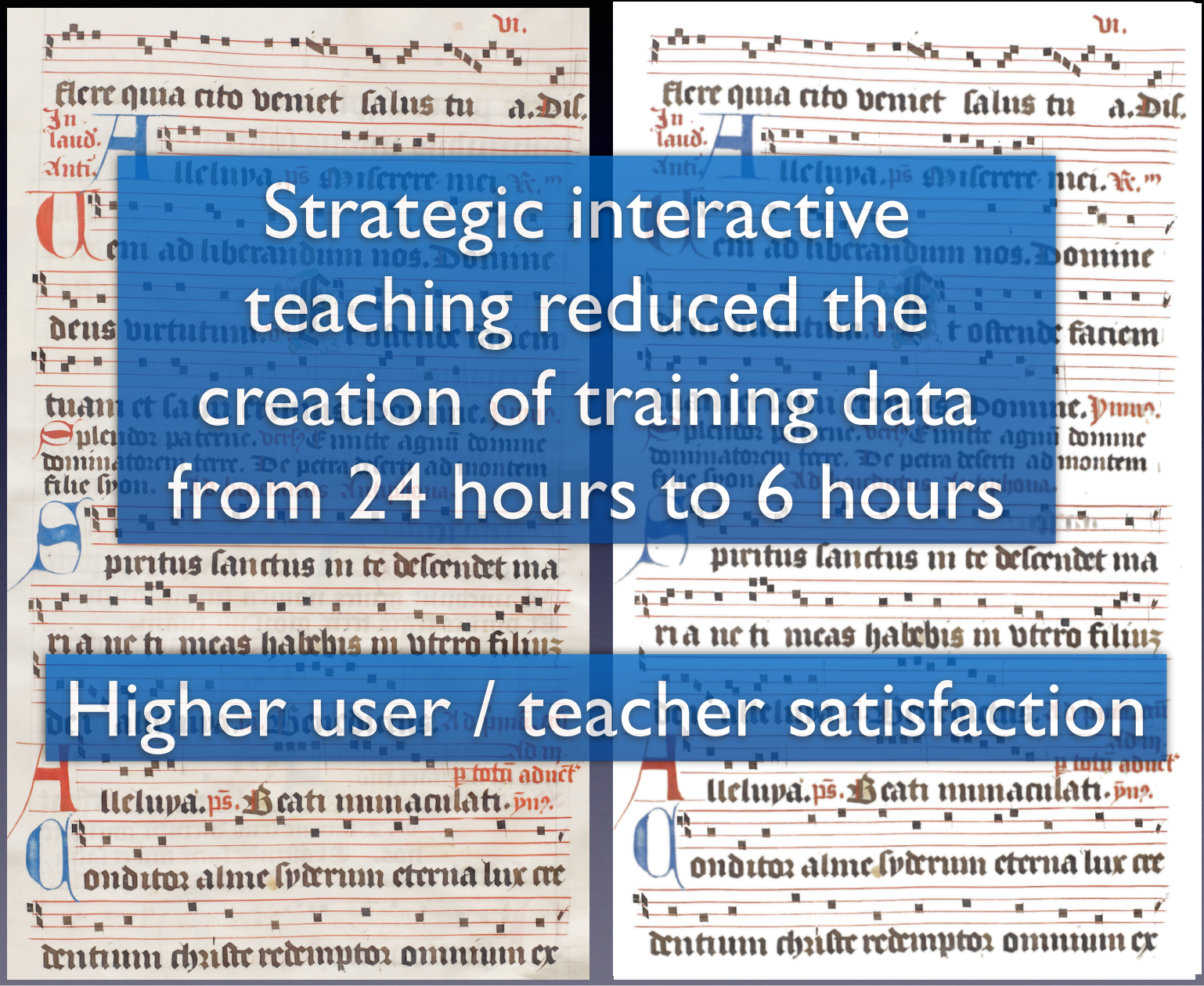




Classification of a Page



Results with a New Page



Strategic interactive teaching reduced the creation of training data from 24 hours to 6 hours

Higher user / teacher satisfaction

SIMMSA Workflow for Neume Notation

Digitized Manuscript

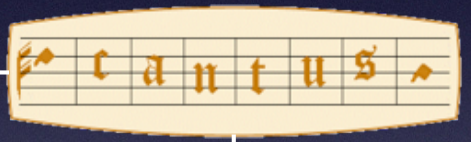
Layout Analysis & Correction
Pixel.js

Symbol Classification & Correction
InteractiveClassfier.js

Neume Mapping Tool

Automatic Pitch Finder

OCR & Text Aligner



Output Generator

MEI

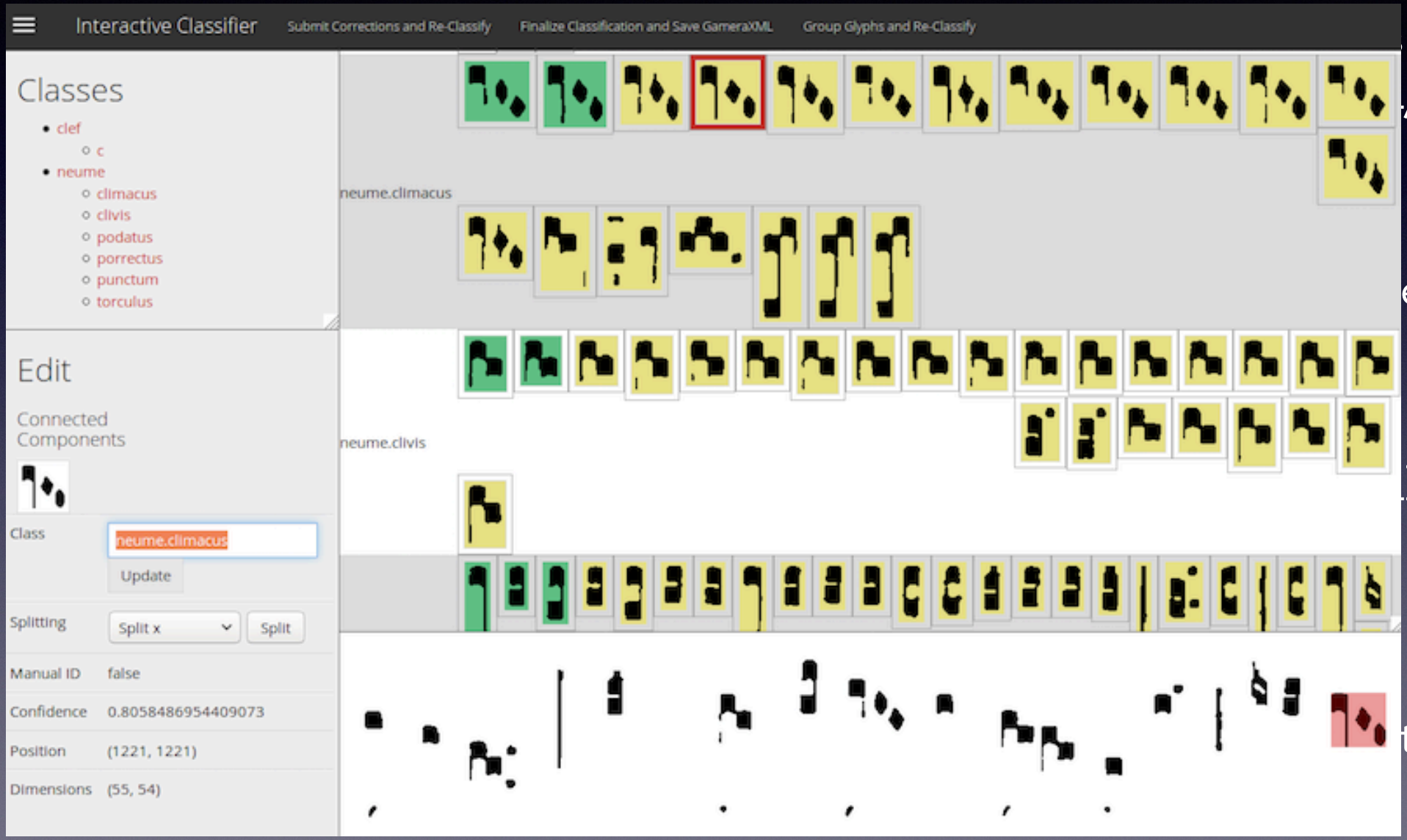
Corrections: Neon.js

Cantus Ultimus
Interface

RODAN
Workflow Management System

InteractiveClassifier.js

Minh Anh Nguyen



- The Music Encoding Initiative (MEI) is a community-driven effort to define a system for encoding musical documents in a machine-readable file format (XML).
- In development since 1999.
- MEI is based on Text Encoding Initiative (TEI).
- MEI is an alternative to MusicXML.

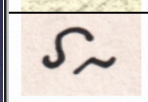
Neume Mapping Table to MEI

		Torculus 3		
3 pitches: n-h-l				

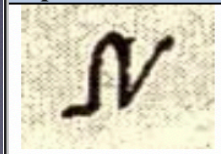
	21,5 (10r) **ple**nitudine m	**Torculus**	neume.torculus	`<neume>`
				`<nc tilt= "e" />`
				`<nc angled= "true" tilt= "n" intm= "u" />`
				`<nc tilt= "se" intm= "d" />`
				`</neume>`

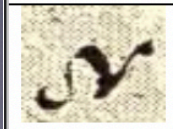
	21,5 (10r) **tua**	**Torculus**	neume.torculus	`<neume>`
				`<nc curve= "a" />`
				`<nc tilt= "nw" intm= "u" />`
				`<nc curve= "c" intm= "d" />`
				`</neume>`

	21,5 (10r) **et**	**Torculus melodic**	neume.torculus	`<neume>`
				`<nc tilt= "e" />`
				`<nc angled= "true" tilt= "n" intm= "u" />`
				`<nc tilt= "se" intm= "d" rel_len= "l" />`
				`</neume>`

		**Torculus**	neume.torculus	`<neume>`
				`<nc curve= "a" />`
				`<nc tilt= "n" intm= "u" curve= "c" />`
				`<nc con= "g" tilt= "ne" angled= true" intm= "d"`
				`/>`
				`</neume>`

		Torculus resupinus 4		
4 pitches: n-h-l-h				

	21,12 (10r) **nos**trum	**Torculus resupinus**	neume.torculus_resupinus	`<neume>`
				`<nc tilt= "e" />`
				`<nc tilt= "n" angled= "true" intm= "u" />`
				`<nc tilt= "s" intm= "d" />`
				`<nc angled= "true" tilt= "ne" intm= "u" />`
				`</neume>`

	22,13 (10v) **tri**buisti	**Torculus resupinus**	neume.torculus_resupinus	`<neume>`
				`<nc curve= "a" />`
				`<nc tilt= "n" intm= "u" />`
				`<nc tilt= "se" intm= "d" />`
				`<nc angled= "true" tilt= "ne" intm= "u" />`
				`</neume>`

McGill

CIRMMT

Neume Mapping Tool

Imane Chafi

Home | New neume | About us | Contact | Help | Log-out | User1

Punctum

Images:

Name:

Punctum

Folio:

1/r9

Description:

Punctum

Classification:

neume.punctum

Update

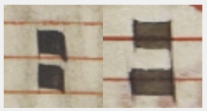
Delete

Mei Snippet:

```
1  //mei snippet for neume.punctum
2  <neume>
3      <nc>
4  </neume>
```

Pes

Images:



Name:

Pes

Folio:

1r/3, 1r/5

Description:

Classification:

neume.pes.b.3

Update

Delete

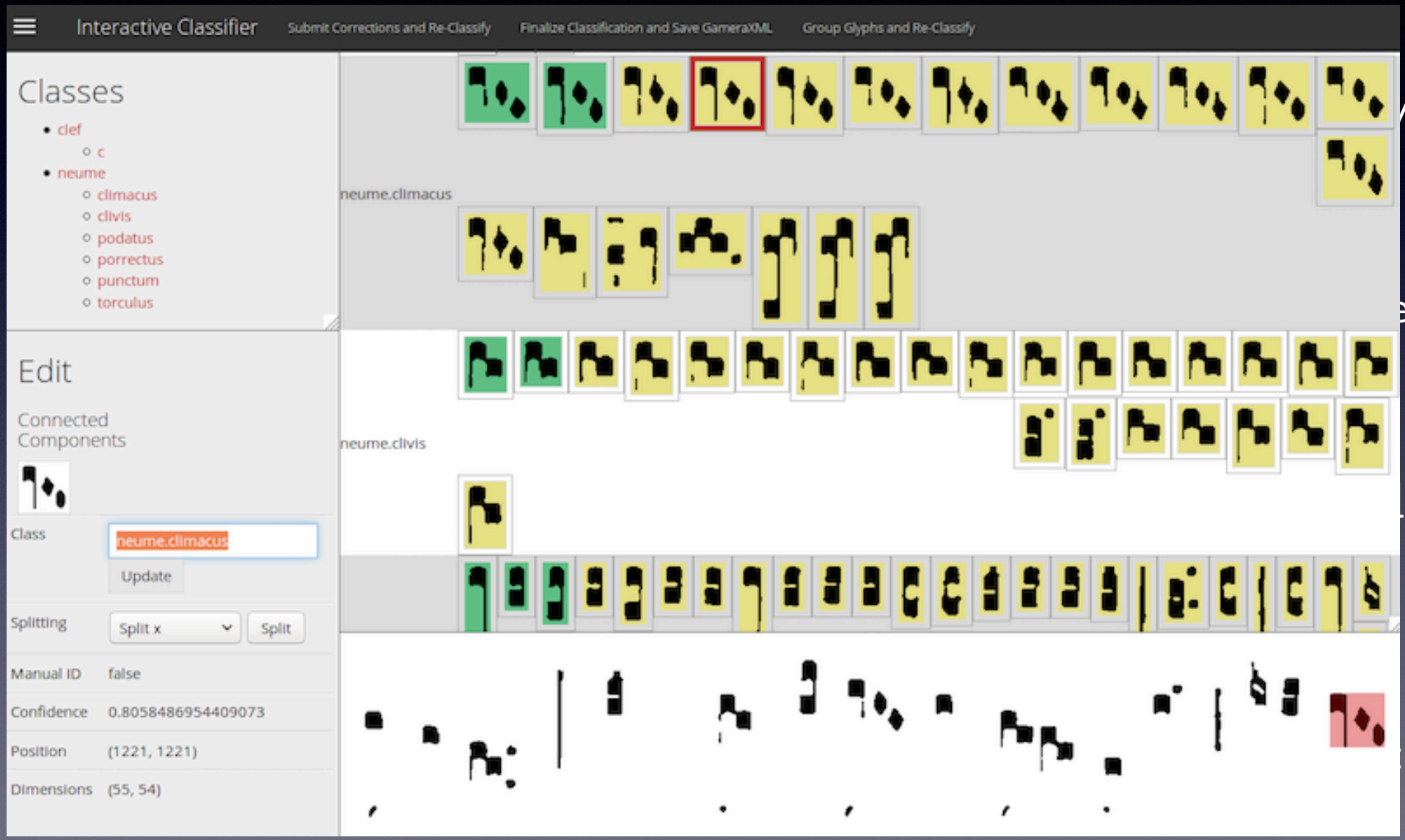
Mei Snippet:

```
1  //mei snippet for neume.pes.b.3
2  <neume>
3      <nc/>
4      <nc intm = "25"/>
5  </neume> <
6
```

element parse error: Error: invalid tagName:

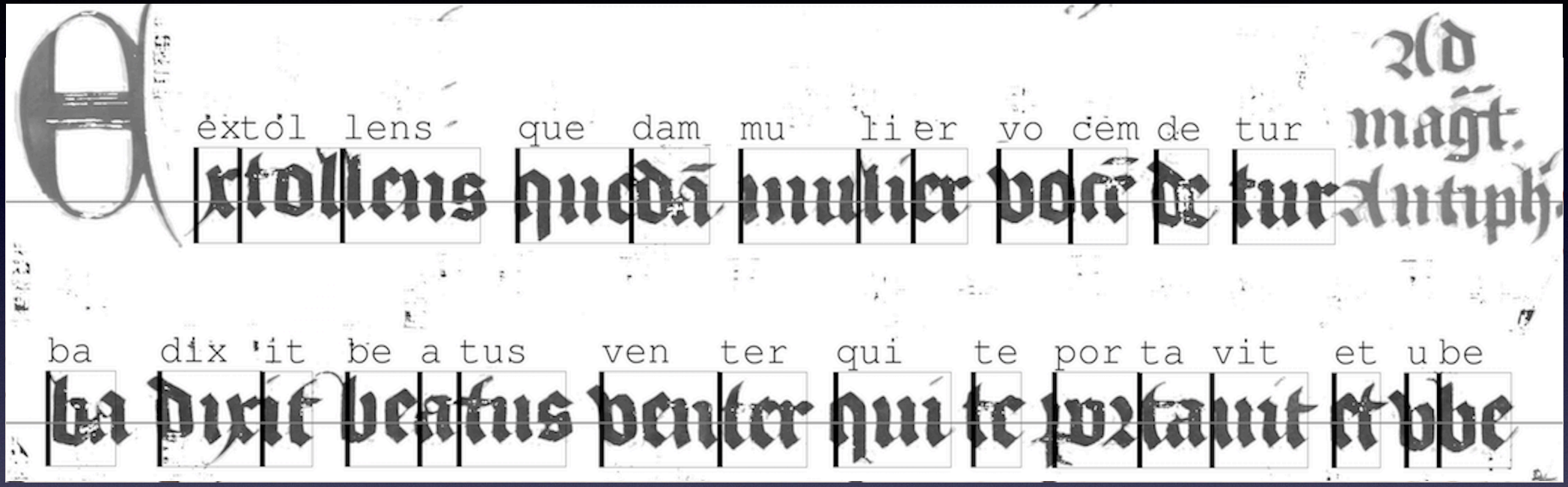
InteractiveClassifier.js (K-NN)

Minh Anh Nguyen



OCR & Text Aligner

Timothy de Reuse



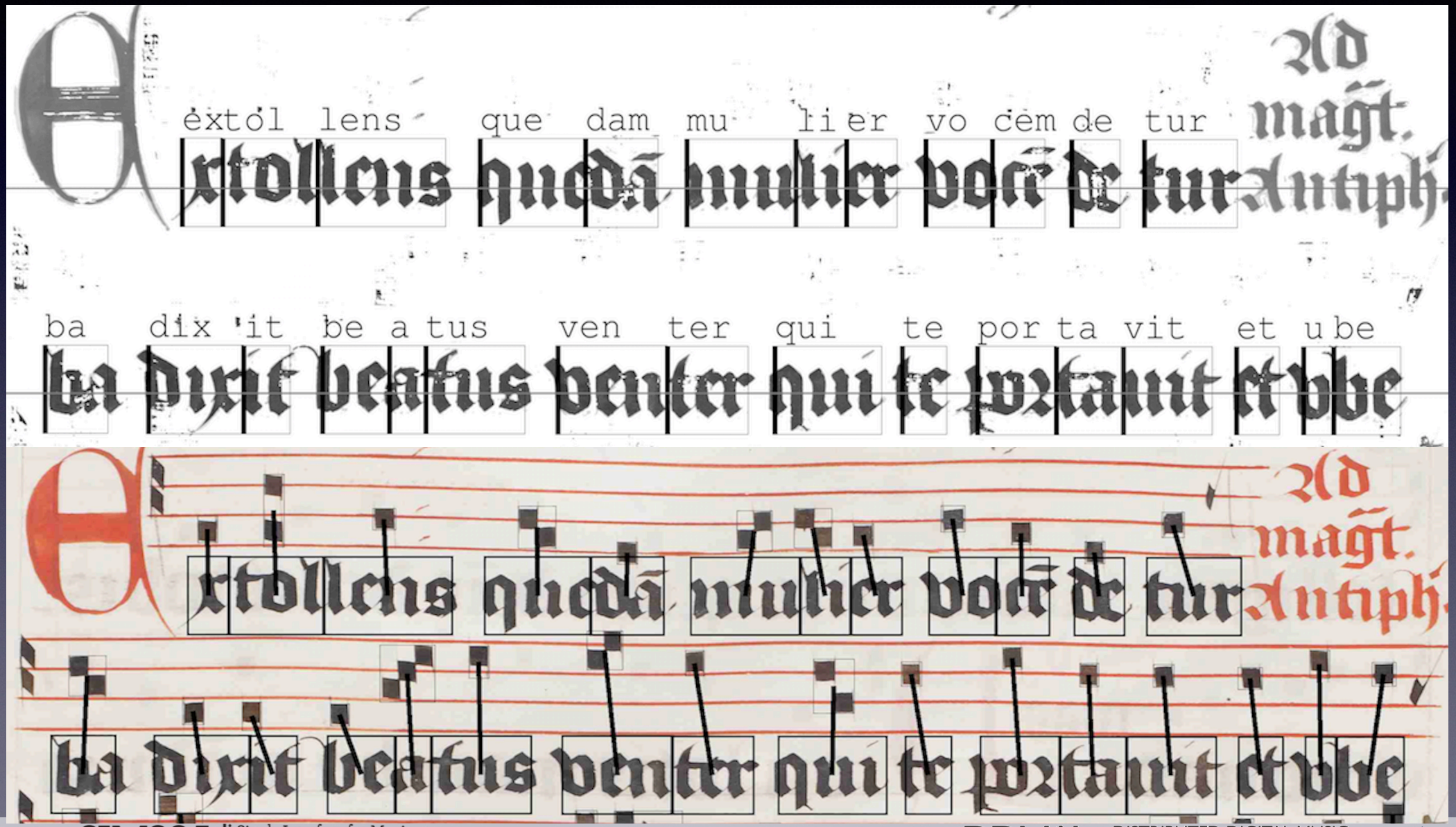
From Cantus Database

Extollens quaedam mulier vocem de tur-
ba dixit beatus venter qui te portavit et ube…

- OCR used: OCRopus (recurrent neural network: LSTM)
- Sequence alignment: Needleman-Wunsch algorithm

OCR & Text Aligner

Timothy de Reuse

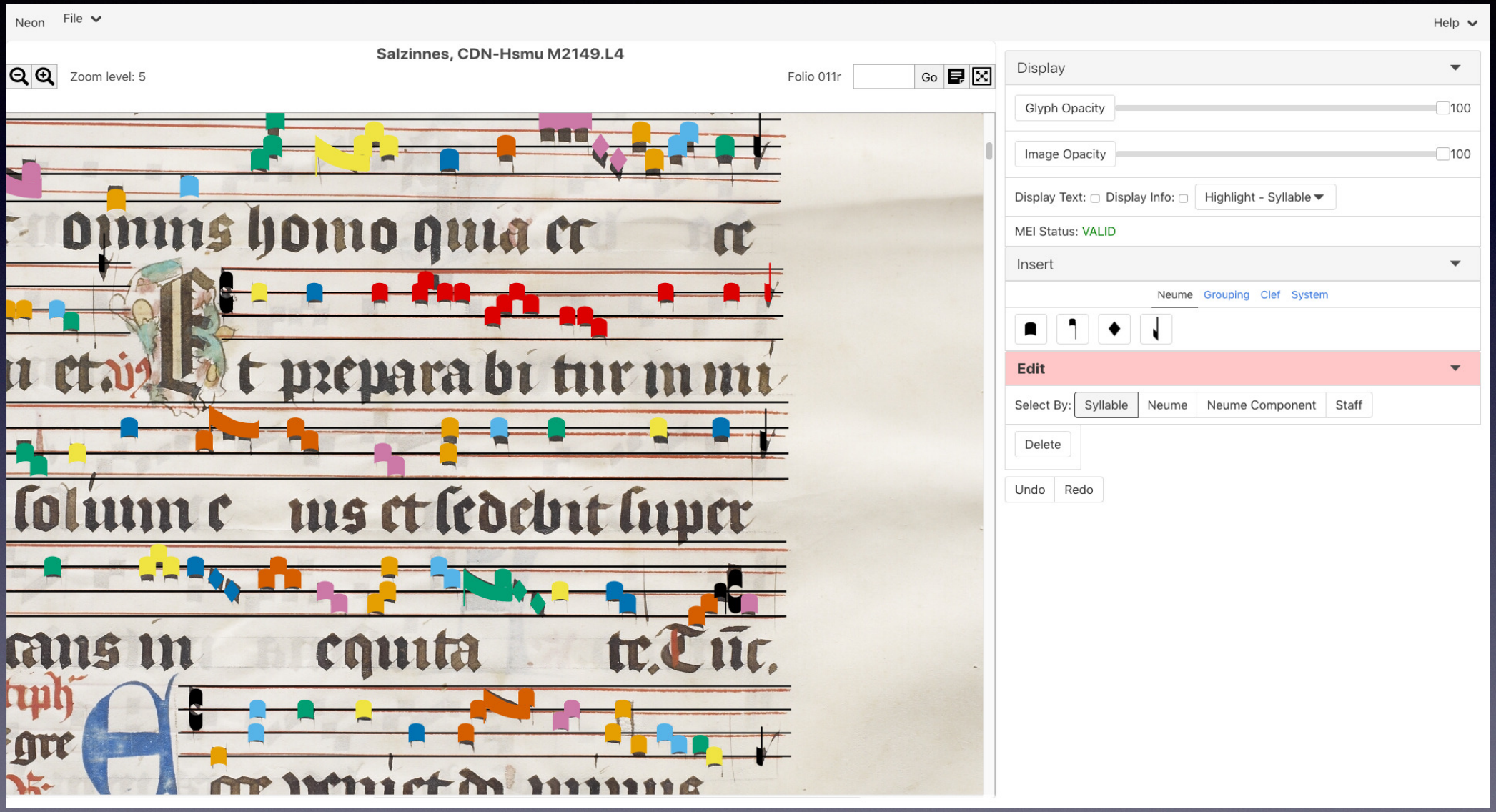


McGill

CIRMMT

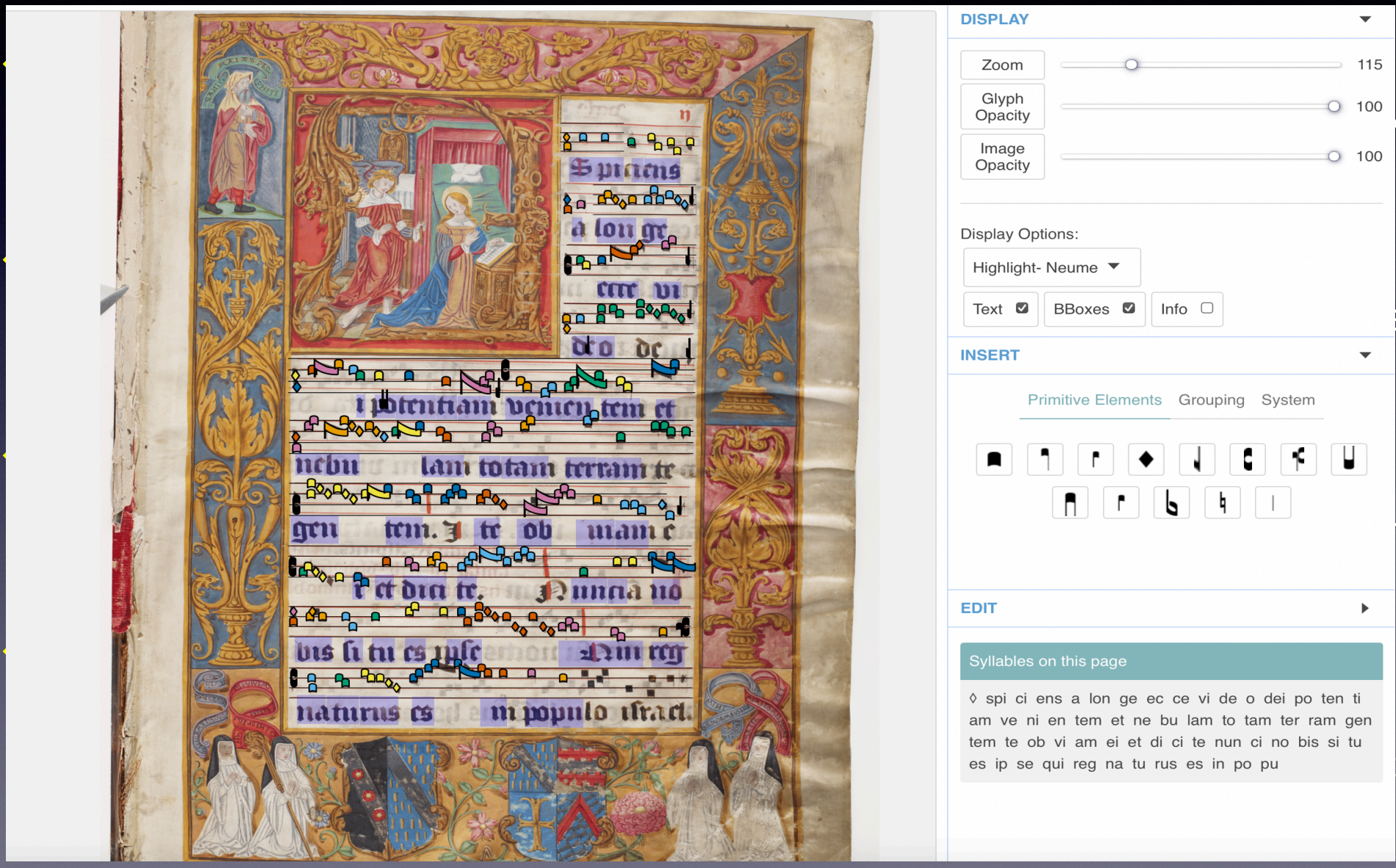
Neume Editor ONline: Neon.js

Juliette Regimbal



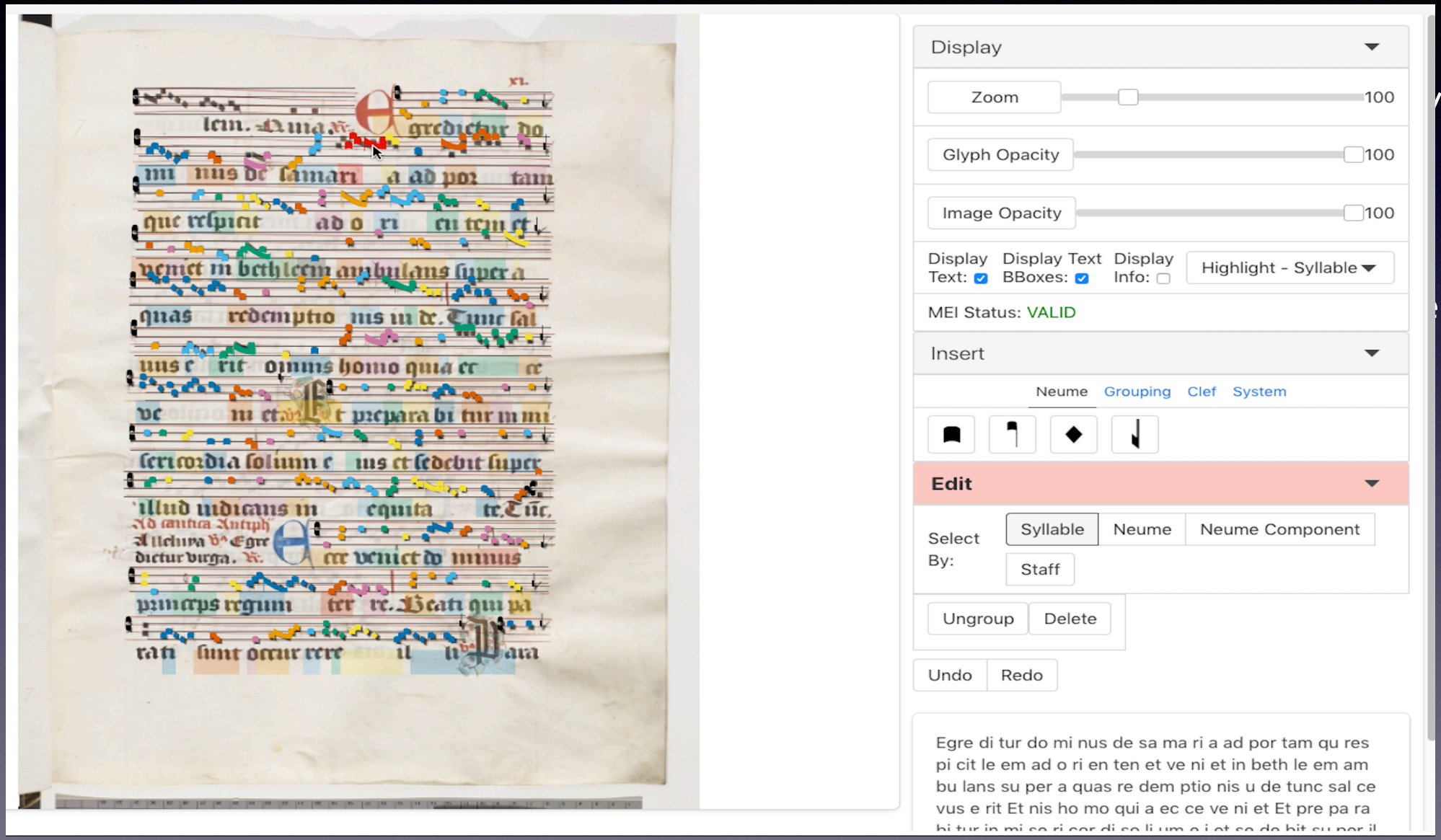
Neon: Neume Editor ONline

Juliette Regimbal, Caitlin Hutnyk, Gaby Halpin, Yinan Zhou

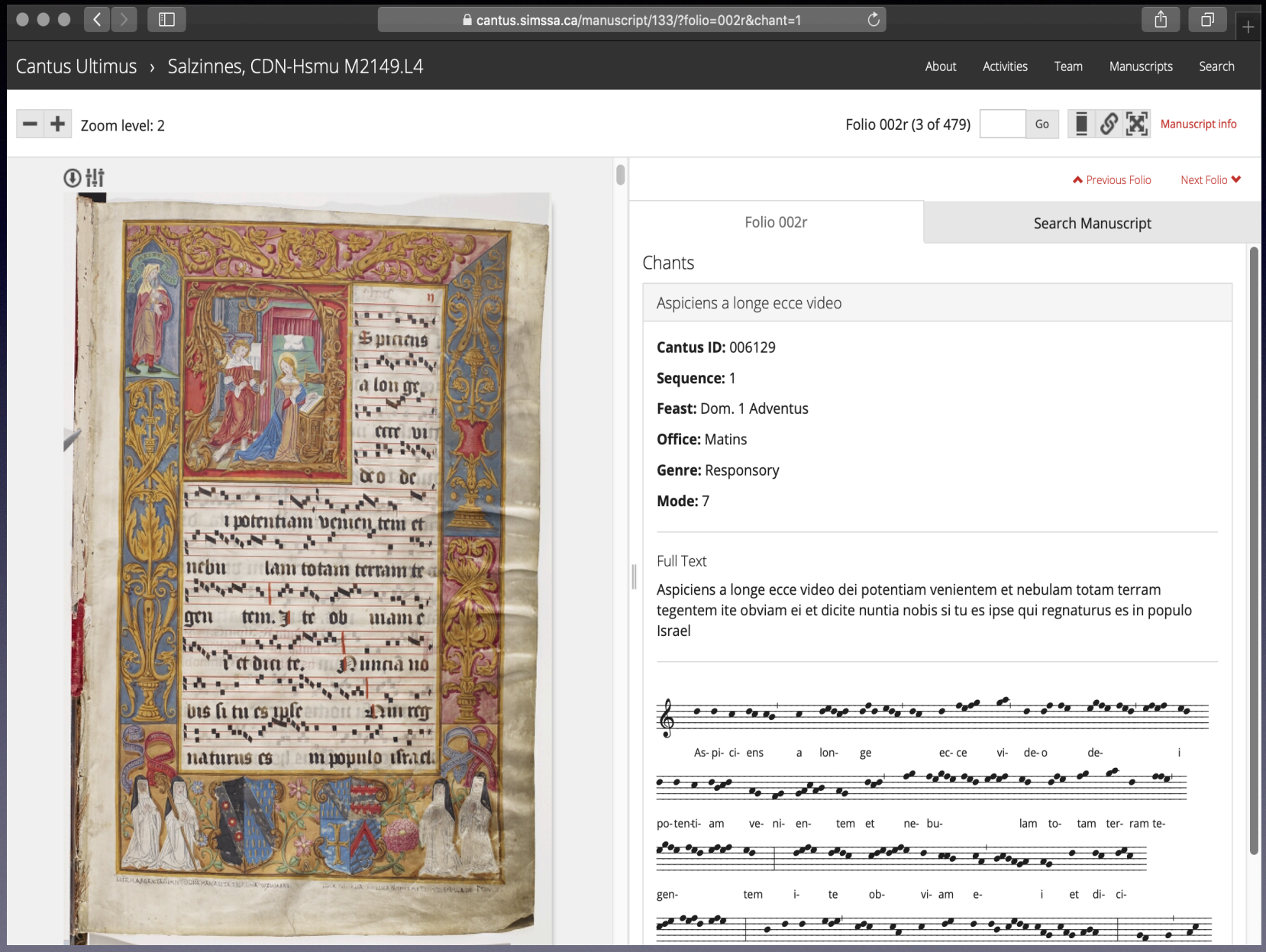


Neon.js: Text Editing

Caitlin Hutnyk

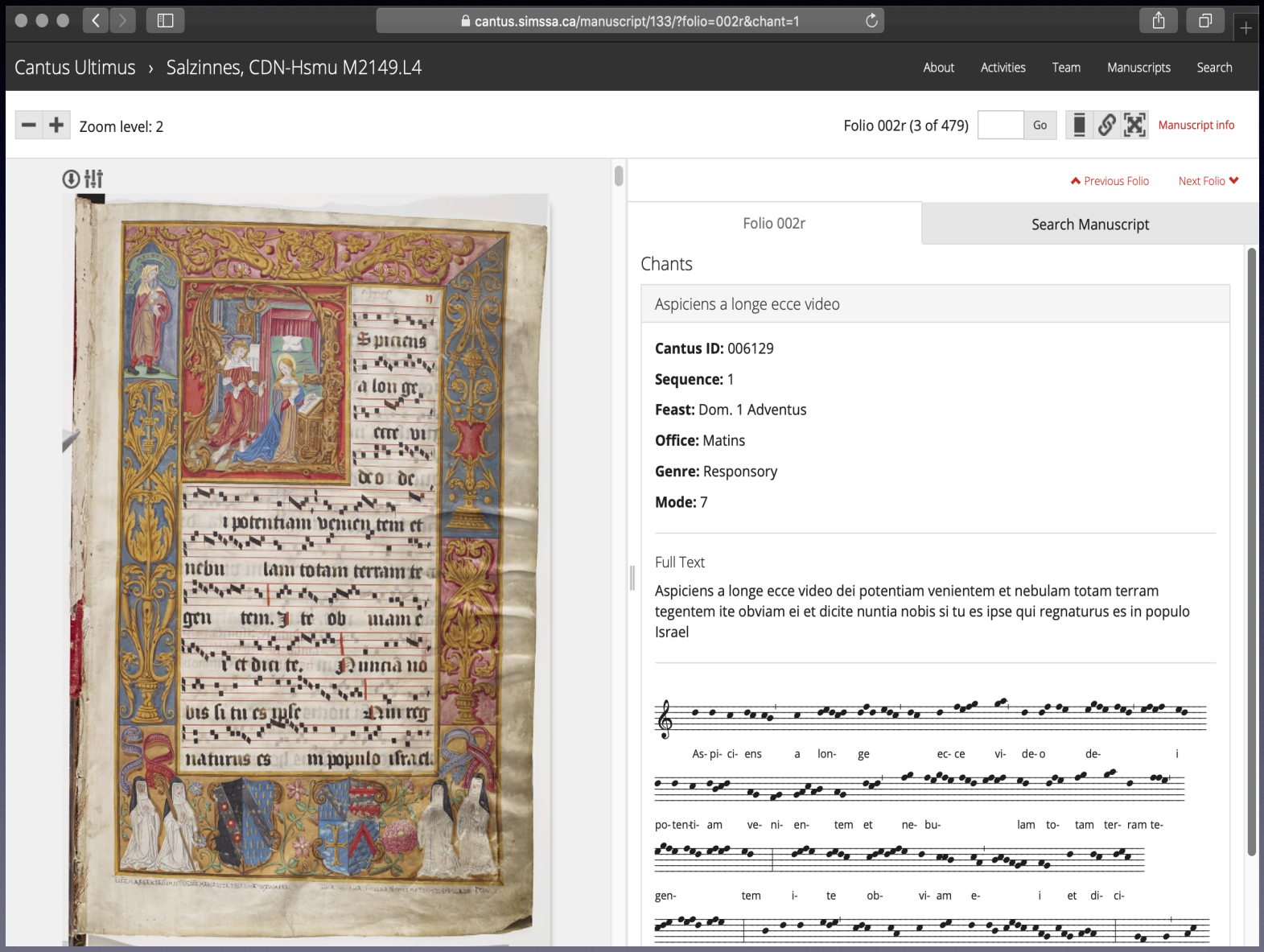


Cantus Ultimus Interface



Cantus Ultimus Interface

Andrew Fogarty, William Bain, Néstor Nápoles,Dylan Hillerbrand

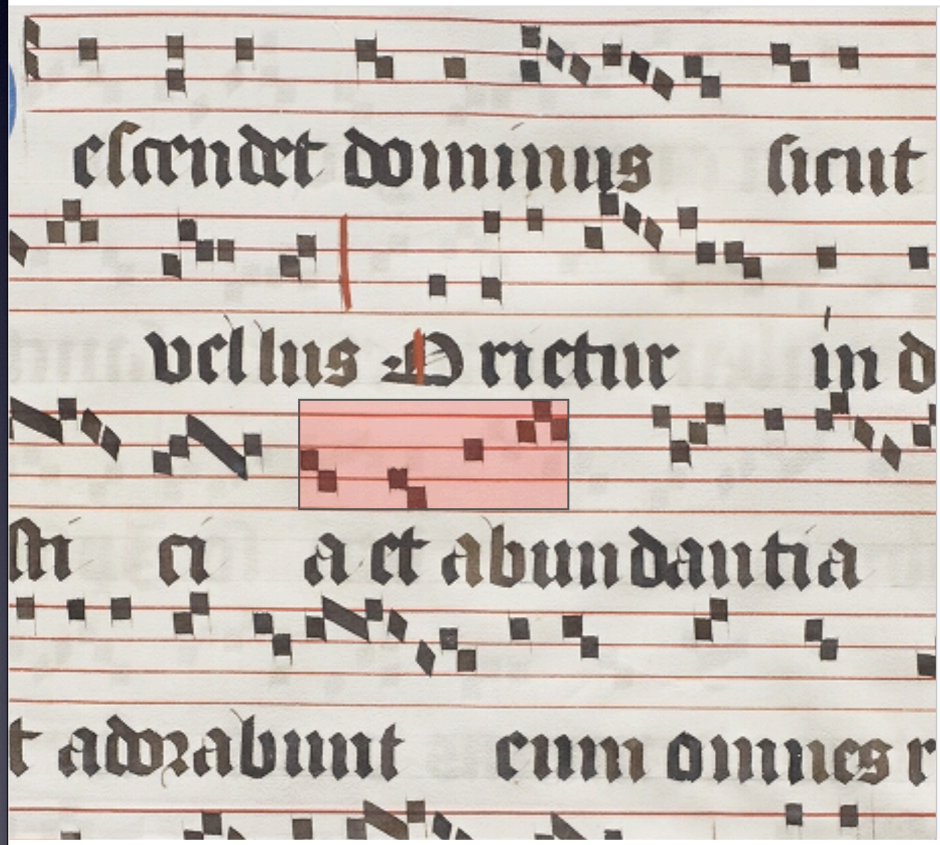


Neume search in Cantus Ultimus

cantus.archive.simssa.ca/manuscript/133/?folio=016r&search%5Btype%5D=ne

Cantus Ultimus › Salzinnes, CDN-Hsmu M2149.L4

Zoom level: 3.00 | Folio 016r (31 of 479) | Go | Manuscript info



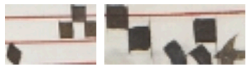
Previous Folio | Next Folio

Folio 016r | Search Manuscript

Neume | clivis clivis punctum torculus | Search

Neume search

29 results for query "clivis clivis punctum torculus"

Folio	Neumes	Pitches	Contour
016r		a g f e b d e d	
024r		b a g d a c d c	
034v		b a b g c e f e	

Summary

"A Retrospective on Optical Music Recognition Research"

- Early developments
- OMR Thesis
- Gamera
- SIMSSA (Single Interface for Music Score Searching and Analysis)
- Cantus Ultimus

McGill

CIRMMT

Acknowledgements

Rian Adamian
Christopher Antila
Claire Arthur
William Bain
Ryan Bannon
Laurier Baribeau
Noah Baxter
Laura Beauchamp
Justin Bell
Ruth Berkow
Marina Borsodi-Benson
John Ashley Burgoyne
Greg Burlet
Jorge Calvo-Zaragoza
Paco Castellanos
Imane Chafi
Jiali Chen
Remi Chiu
Morgane Ciot
Nat Condit-Schultz
Kemal Conger
Julie Cumming
Alex Daigle
Marie DeYoung
Anna de Bakker
Tim de Reuse
Natasha Dillabough
Daniel Donnelly
Neda Eshraghi
Meredith Evans
Wei Gao
David Garfinkel
Gen Gates-Panneon
Mahtab Ghamsari
Arielle Goldman
Ryan Groves
Jamie Klassen
Zhanna Klimanova
Gaby Halpin
Peter Henderson
Jason Hockman
Emily Hopkinson
Andrew Horwitz
Sam Howes
Catlin Hutnyk
Alessandra Ignesti
Yaolong Ju
Andrew Kam
Anton Khelou
Jamie Klassen
Suzuka Kokubu
Reiner Krämer
Véronique Lagacé
Andy Le
Saining Li
Katie Lin
Eric Liu
Ian Lorenz
Wendy Liu
Sylvain Margot
Evan Magoni
Cory McKay
Zoé McLennan
Nicky Mirfallah
Shahrad Mohammadzadeh
Lillio Mok
Alexander Morgan
Catherine Motuz
Maria Murphy
Néstor Nápoles
Clare Neil
Minh Anh Nguyen
Chris Niven
Rory O'Connor
Laura Osterlund
Phyllis Ouyang
Jérôme Parent-Lévesque
Alexandre Parmentier
Gustavo Pedro
Sacha Perry-Fagant
Alastair Porter
Laurent Pugin
Ravi Raina
Juliette Regimbal
Deepanjan Roy
Zeyad Saleh
Evan Savage
Peter Schubert
Harry Simmonds
Brian Stern
Tristano Tenaglia
Martha Thomae
Andrew Tran
Vi-An Tran
Finn Upham
Gabriel Viglensoni
Junhao Wang
Tim Wilfong
Mike Winters
Ling-Xiao Yang
Jinho Yoon
Ké Zhang
Yinan Zhou

SIMSSA Team @McGill: Summer 2023

Jacob deGroot-Maggetti
Ichiro Fujinaga
Lucas March
Anthony Tan
Dylan Hillerbrand
Taz Scott-Talib
Martha Thomae
Jacky Zhang
Yinan Zhou
Marisa Goldman
Sabrina Mansour
David Peterman
Tim de Reuse
Hong Van Pham
Max Zhang
Way Yi Lin
Andrew Hankinson
Anna de Bakker
Rebecca Mizrahi

Cantus Ultimus: cantus.simssa.ca/manuscripts
Project page: simssa.ca
Github sources: github.com/DDMAL

Funded by

Social Sciences and Humanities Research Council of Canada
Conseil de recherches en sciences humaines du Canada
Canada

McGill Schulich School of Music École de musique Schulich

CIRMMT Centre for Interdisciplinary Research in Music Media and Technology

cantus

Fonds de recherche Société et culture Québec

compute canada | calcul canada

WESTGRID