Gamera versus Aruspix

Two Optical Music Recognition Approaches

Laurent Pugin
laurent@music.mcgill.ca

Jason A. Hockman
hockman@music.mcgill.ca

John Ashley Burgoyne
ashley@music.mcgill.ca

Ichiro Fujinaga
ich@music.mcgill.ca

Music Technology Area Schulich School of Music McGill University

abstract

Optical music recognition (OMR) applications are predominantly designed for common music notation and as such, are inherently incapable of adapting to specialized notation forms within early music. Two OMR systems, namely Gamut (a Gamera application) and Aruspix, have been proposed for early music.

We present a novel comparison of the two systems, which use markedly different approaches to solve the same problem, and pay close attention to the performance and learning rates of both applications.

In order to obtain a complete comparison of Gamut and Aruspix, we evaluated the symbol (core) recognition systems and the pitch determination processes separately. With our experiments, we were able to highlight the advantages of both approaches as well as causes of problems and possibilities for future improvements.

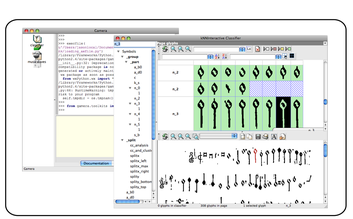
1 infrastructure

Gamera and Gamut

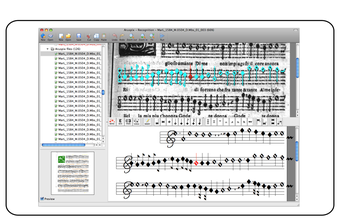
- open-source framework for the creation of structured document analysis applications by domain experts
- designed for use with any type of image
- composite of C++ and Python, with facilities to integrate plugins or toolkits
- MusicStaves toolkit for staff detection and removal
- connected-component analysis and k-nearest neighbor classifier (kNN)
- pitch detection in postprocessing phase

Aruspix

- cross-platform software program for OMR of early music prints
- may be used as both a research tool and end-user application
- fulfillment of entire OMR process: from preprocessing to score encoding
- unique OMR approach based on hidden Markov models (HMMs), without staff removal
- pitch detection integrated within the recognition process (postprocessing not required)



Gamera screenshot with MusicStaves plugin

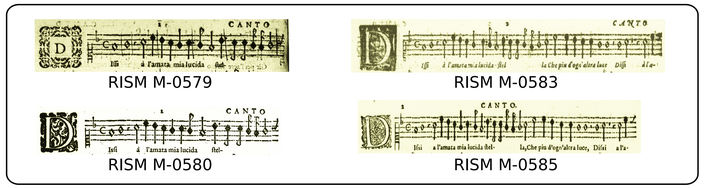


Aruspix screenshot

2 experiments

Data Set

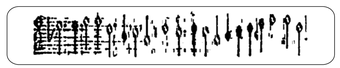
- built from 4 books of madrigals printed in Venice and Antwerp (1587–1607)
- all re-editions of same collection of madrigals and thus contain same musical content
- uniformity of content ensures that experimental performance variations across books are mainly related to fonts and document degredation
- from each book we select 40 pages for the training sets, 30 pages for the testing sets
- preprocessing in Aruspix (*skew-removal*, *binarization*, etc.)
- normalized to a staff height of 100 pixels



sample prints selected from the data set

Evaluation

- **How do the two core recognition systems (kNN vs. HMMs) perform?**
- **How can the two approaches be optimized?**
- **How do the two OMR approaches perform?**
- What is the best staff removal algorithm for early music?
- How problematic is it to retrieve the pitch once a symbol has been recognized?



typical staff removal error in Gamut (from print M-0583)

3 results

Evaluation of the core recognition systems

- Aruspix (HMM) outperforms Gamut (kNN) after 40 pages for each of the four books
- smaller variability across the results within HMMs than within the kNN classifiers
- best results obtained with Linetracking staff removal algorithm (for Gamut)

Evaluation of the optimization procedures

Optimization of the kNN with genetic algorithm (GA) in Gamut:

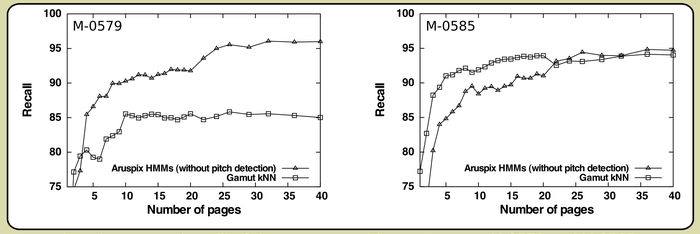
- substantial influence when number of pages is very small
- no significant improvement when using more than 20 pages

Integration of n-gram models into Aruspix HMMs:

- improvement of precision rates much more so than recall rates
- long stable growth period

Evaluation of the OMR approaches

- similar results for entire OMR process to those observed for core recognition systems
- significant decrease of recall and precision after pitch detection phase (in Gamut)



learning curves (recall) of the core recognition systems on M-0579 and M-0585

Book	Recall		Precision	
	kNN	HMMs	kNN	HMMs
M-0579	86.90	95.99	86.63	95.63
M-0580	86.76	93.51	89.36	97.32
M-0583	77.30	87.58	81.04	93.95
M-0585	93.35	94.72	93.47	96.80

recall and precision for **core recognition**

Book	Recall		Precision	
	Gamut	Aruspix	Gamut	Aruspix
M-0579	79.34	92.44	84.13	95.60
M-0580	76.18	92.52	84.55	96.44
M-0583	68.13	86.75	76.54	92.76
M-0585	81.61	92.77	88.34	96.23

recall and precision for **complete OMR** (with pitch recognition) results of the test pages

4 discussion

Need for Adaptability in OMR

- although systems varied in learning methods and performance, both under-performed on same book, M-0585 (most degraded in experimental set)
- difference between results of books with same content and quasi-similar fonts demonstrate effect of degradation upon the performance of the system

Staff Removal Issues

- our experiments with Gamut demonstrate that staff removal within early music sources cannot be considered a solved problem, as no tested algorithm performed extraordinarily well
- subsequent pitch determination from recognized symbols also proved difficult

Gamut Flexibility

- the extensibility of the Gamut environment allowed modifications to improve performance for a particular set of sources
- kNN classifiers learn quite rapidly, allowing an application to be generated on a handful of pages
- may be advantageous for projects with wider scope

Aruspix Robustness

- Aruspix distinguishes itself by its performance and robustness
- HMMs and integrated pitch detection are an excellent approach to handle noise due to printing irregularities and degradations of pages
- Aruspix benifits from optimization for early music prints
- may be used directly out-of-the-box

McGill
Schulich School of Music / École de musique Schulich
CIRMMT Centre for Interdisciplinary Research in Music Media and Technology
DDMAL Distributed Digital Music Archives & Libraries Lab