

Book-Adaptive and Book-Dependent Models to Accelerate Digitization of Early Music

Laurent Pugin, John Ashley Burgoyne,
Douglas Eck & Ichiro Fujinaga

NIPS MBC Workshop, Whistler
December 2007



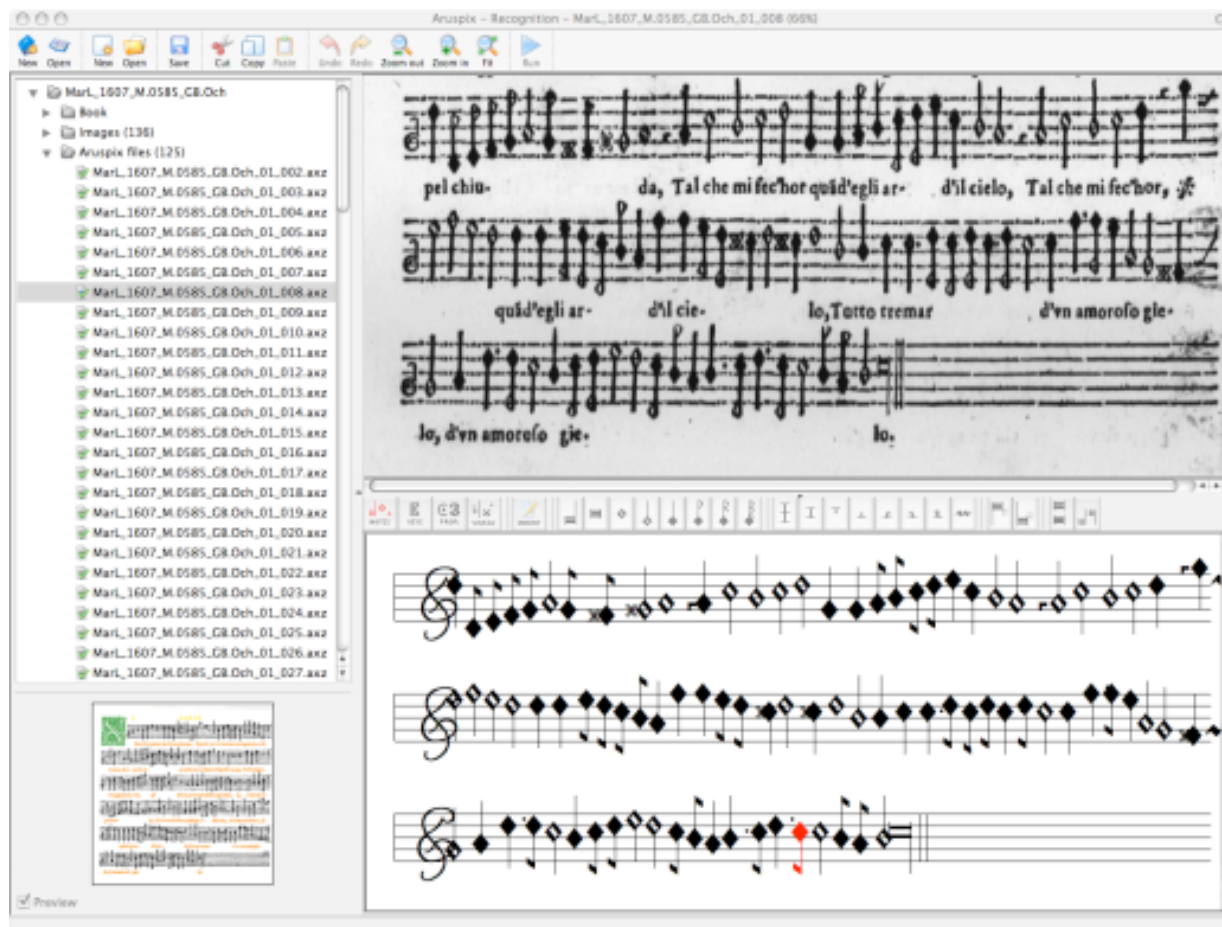
McGill



Centre for Interdisciplinary Research
in Music Media and Technology

Université 
de Montréal

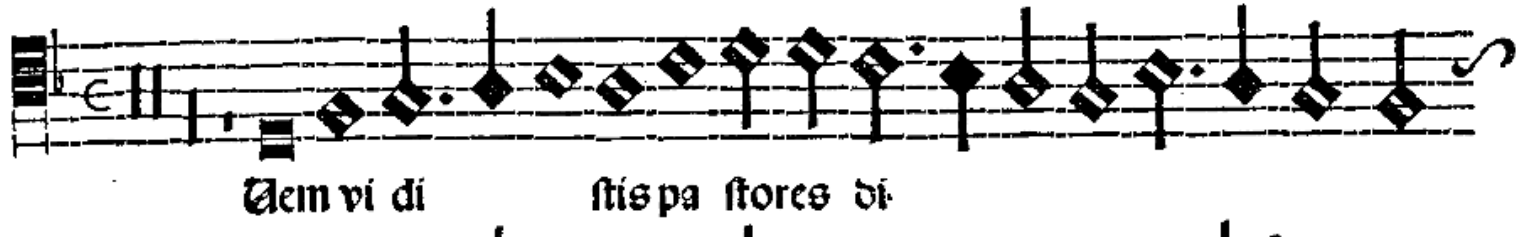
OMR on early music sources



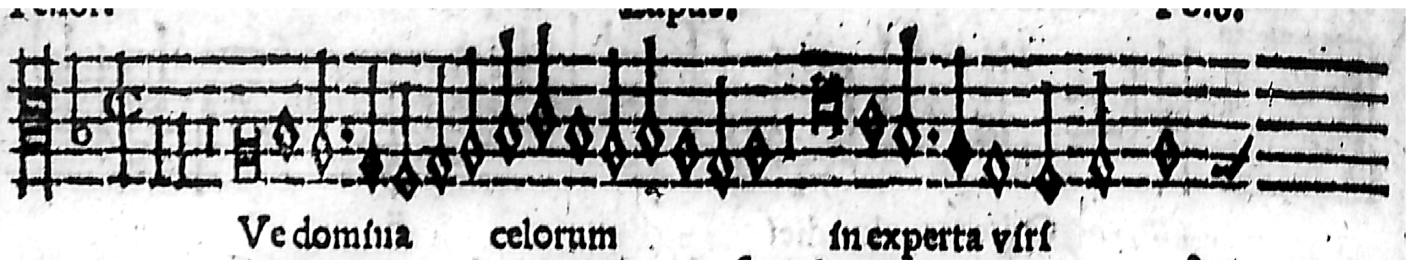
Aruspix

Font shape variability

TENOR

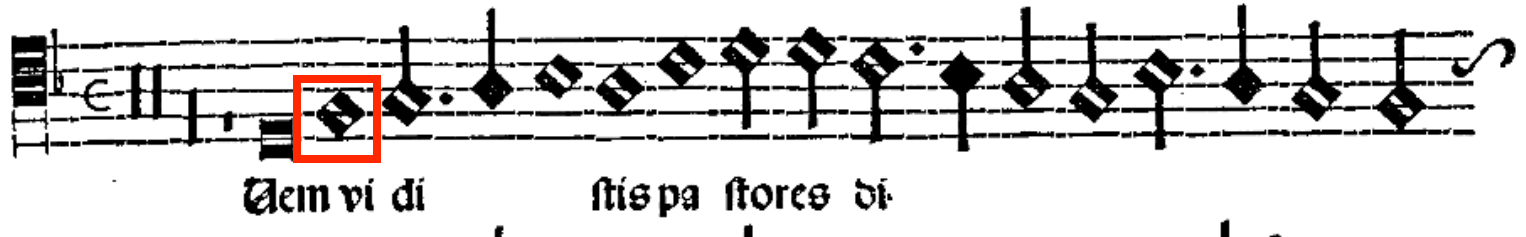


Secūda
pars.

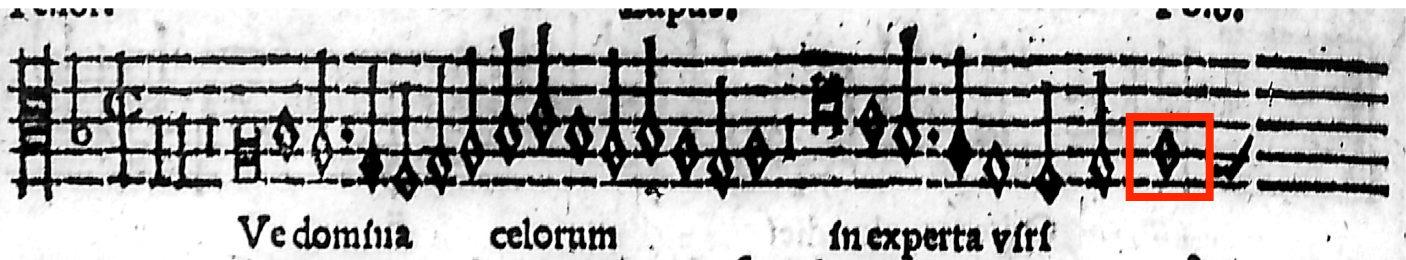


Font shape variability

TENOR

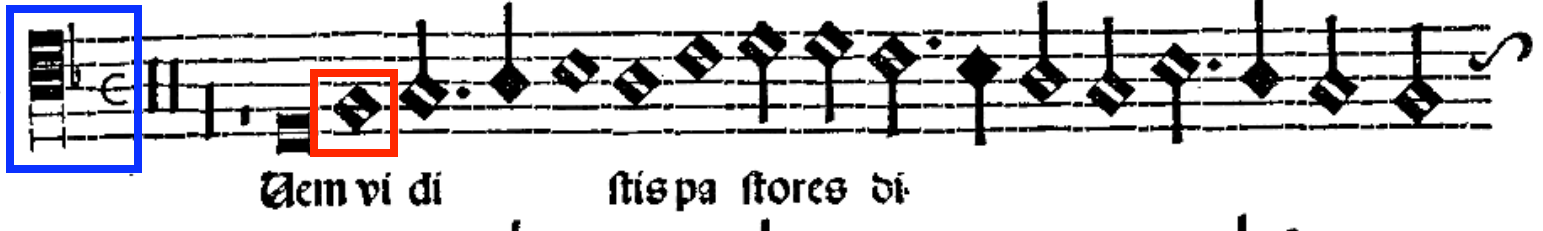


Secūda
pars.

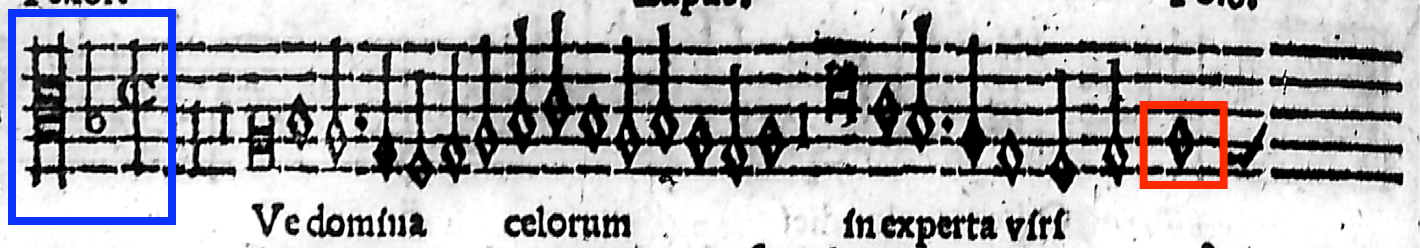


Font shape variability

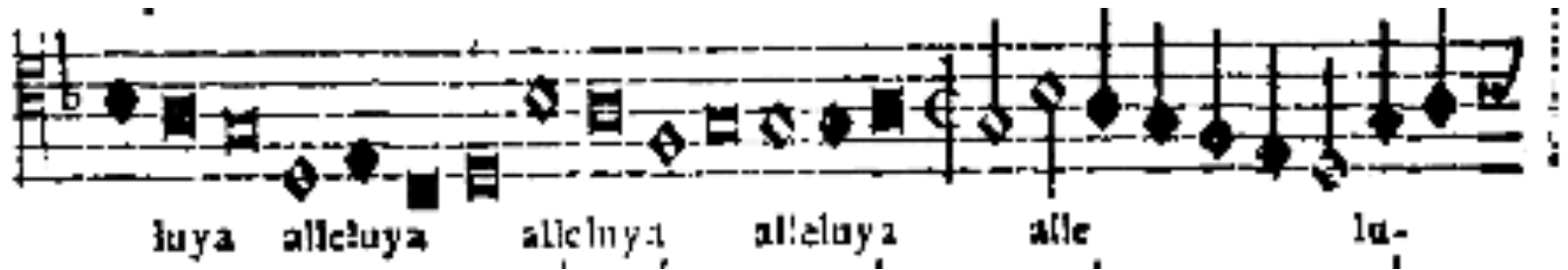
TENOR



Secūda
pars.



Document degradation variability



Goal of this research

What?

- Enable Aruspix to be able to handle variabilities more efficiently

How?

- Make it adaptable to each book
- Use a supervised adaptation of trained HMMs
- Does involve manual correction

Why?

- OMR output has to be corrected anyway in most applications

MAP (maximum a posteriori) adaptation

MAP adaptation in speech

- A speaker-independent model (SI) is built off-line on a large set of data
- During recognition, the SI model is optimized to obtain the speaker-dependent (SD) model using a small set of data

MAP adaptation in OMR

- A book-independent model (BI) is built off-line on a large set of pages
- During recognition, the BI model is optimized to obtain the book-dependent (BD) model using a small set of pages

MAP (maximum a posteriori) adaptation

MAP adaptation in speech

- A speaker-independent model (SI) is built off-line on a large set of data
- During recognition, the SI model is optimized to obtain the speaker-dependent (SD) model using a small set of data

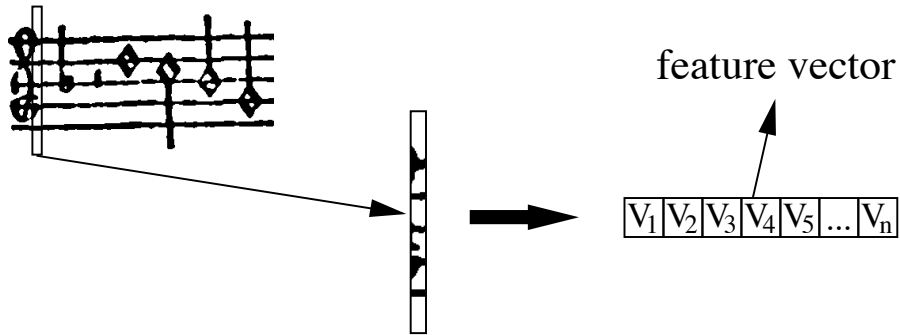
MAP adaptation in OMR

- A book-independent model (BI) is built off-line on a large set of pages
- During recognition, the BI model is optimized to obtain the book-dependent (BD) model using a small set of pages

Experiment data

- 1 BI model trained from scratch on 457 pages from various printed books
- 5 books to experiment with MAP adaptation
 - 50 pages of ground-truth in each book
 - 5 training sets of 40 pages (first pages of the book)
 - 5 testing sets of 10 pages
- Cross-validated one of the books

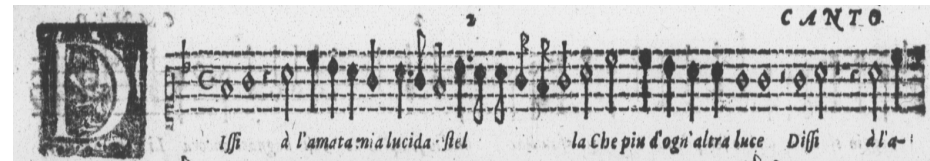
What We Observe



(a) RISM M.0579 (R. Amadino, Venice, 1587) – baseline rec. rate = 83.91%



(b) RISM M.0580 (G. Vincenti, Venice, 1587) – baseline rec. rate = 67.64%



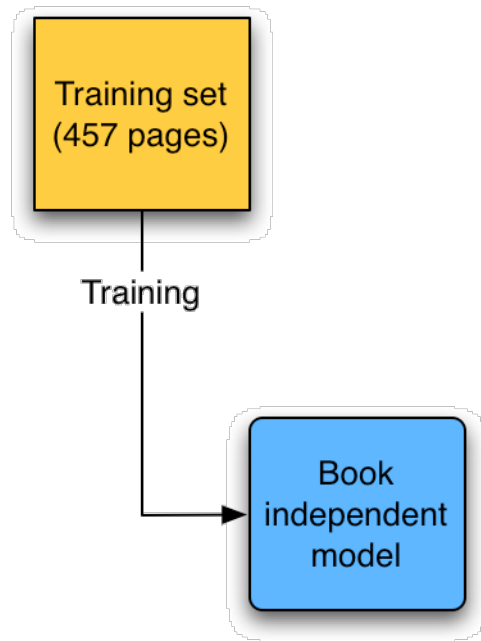
(c) RISM M.0583 (A. Gardano, Venice, 1603) – baseline rec. rate = 82.85%



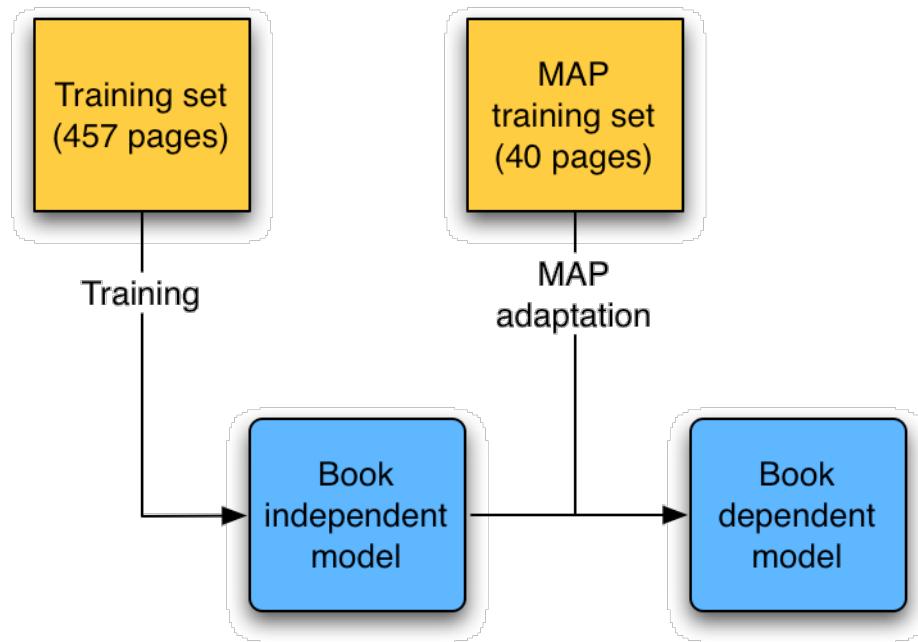
(d) RISM M.0585 (P. Phalèse, Antwerp, 1607) – baseline rec. rate = 86.07%

Experiment workflow

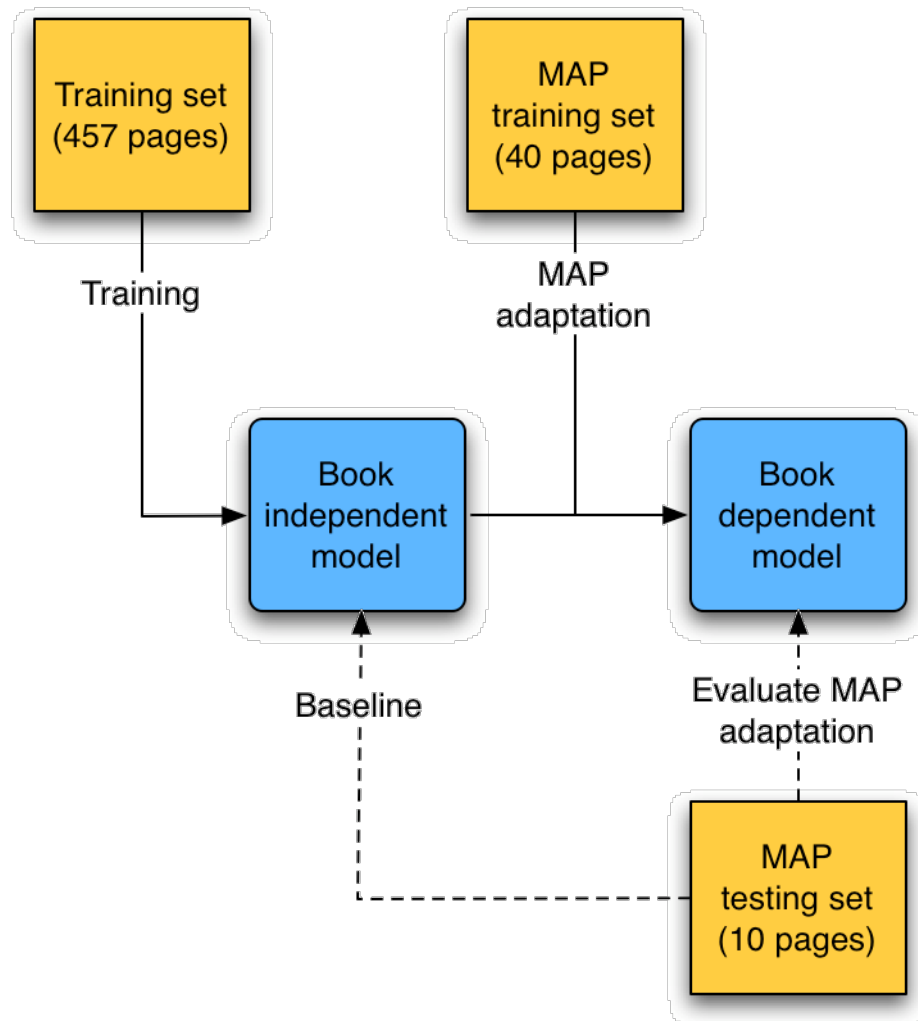
Experiment workflow



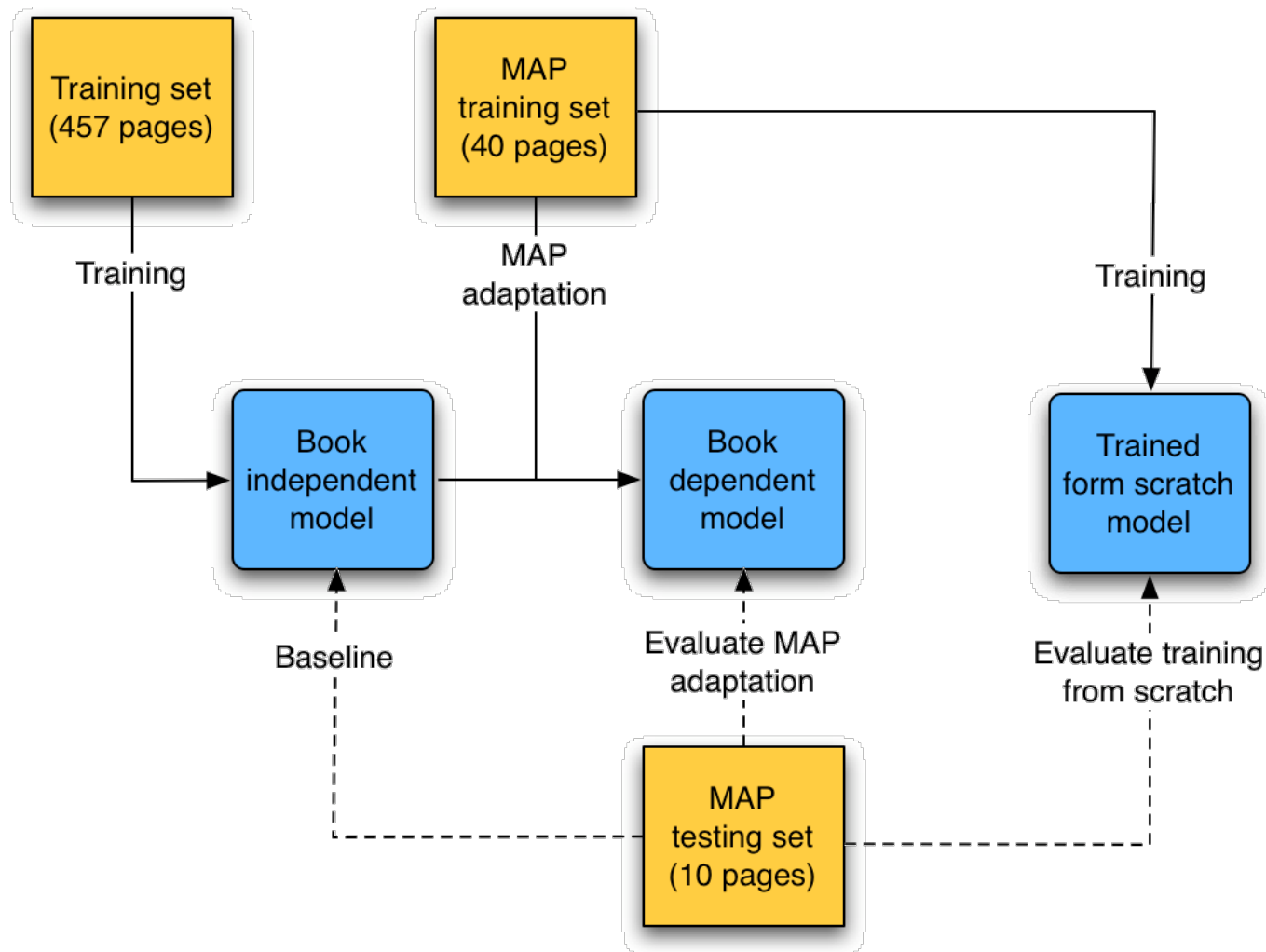
Experiment workflow



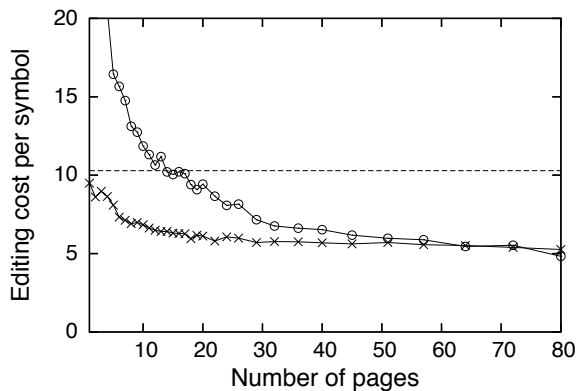
Experiment workflow



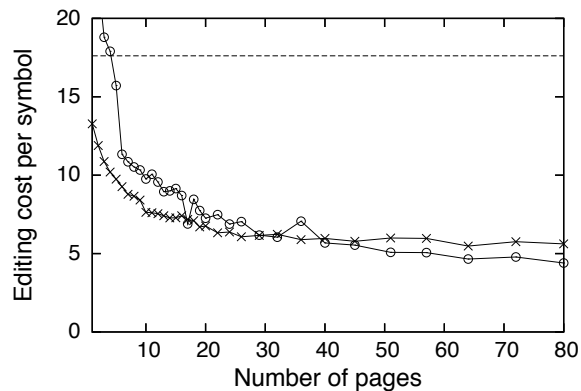
Experiment workflow



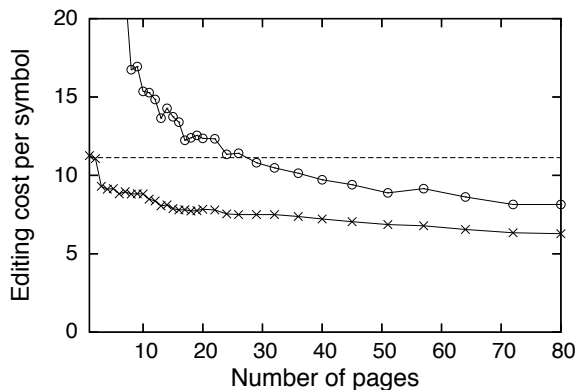
Results



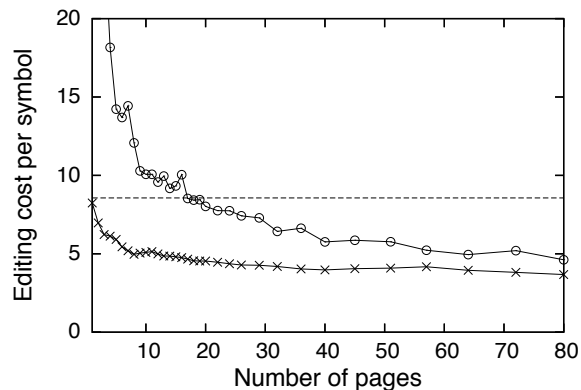
(a) RISM M.0579



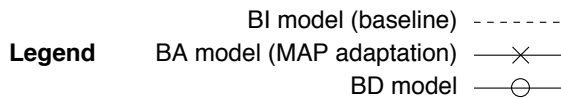
(b) RISM M.0580



(c) RISM M.0583



(d) RISM M.0585



Analysis of the results

- MAP adaptation improves both recall and precision
- MAP adaptation is faster than training from scratch in most cases
- Only 5 to 10 pages are needed to achieve optimal performance
- MAP adaptation can be used in real-time when alignment is good enough

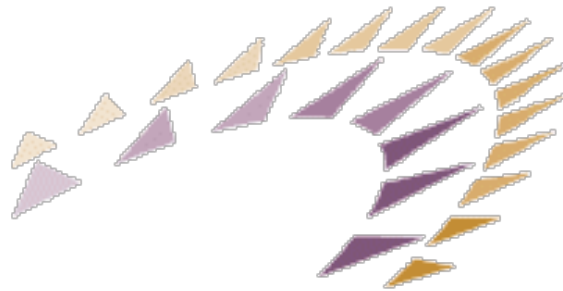
Conclusion and future work

- Exploit vertical as well as horizontal information
- This comes in the form of harmonic structure
- More complex graphical model combining information from multiple staves
- Appropriate for polyphonic music

Thank you!

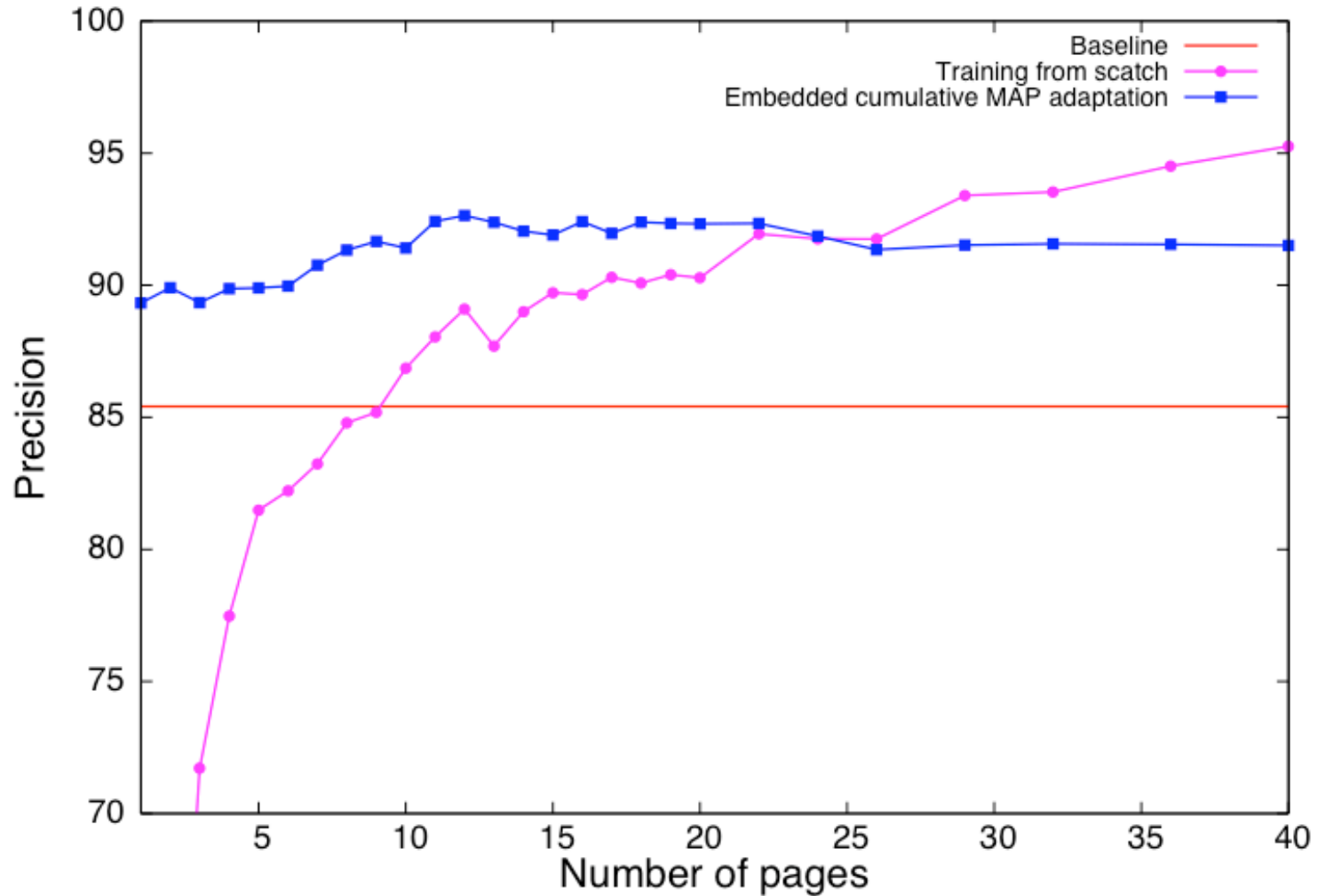


Social Sciences and Humanities
Research Council of Canada



Canada
Foundation for
Innovation

Results - Precision



Results - Recall

