Optical Music Recognition to Digitise Early Music Collections on a Library Scale

Laurent Pugin, John Ashley Burgoyne, and Ichiro Fujinaga

McGill

Schulich School of Music
École de musique Schulich

CIRMMT Centre for Interdisciplinary Research in Music Media and Technology

Music Digitisation Workflow

Building indexed digital music libraries for intelligent retrieval is a laborious process. Optical music recognition (OMR), the musical analogue to optical character recognition (OCR), can speed the digitisation workflow. Human editing work is still necessary, however, to correct errors in the automatic recognition process. We present an innovative approach in OMR, dynamic adaptation, that benefits from this corrected data to improve the system dynamically.

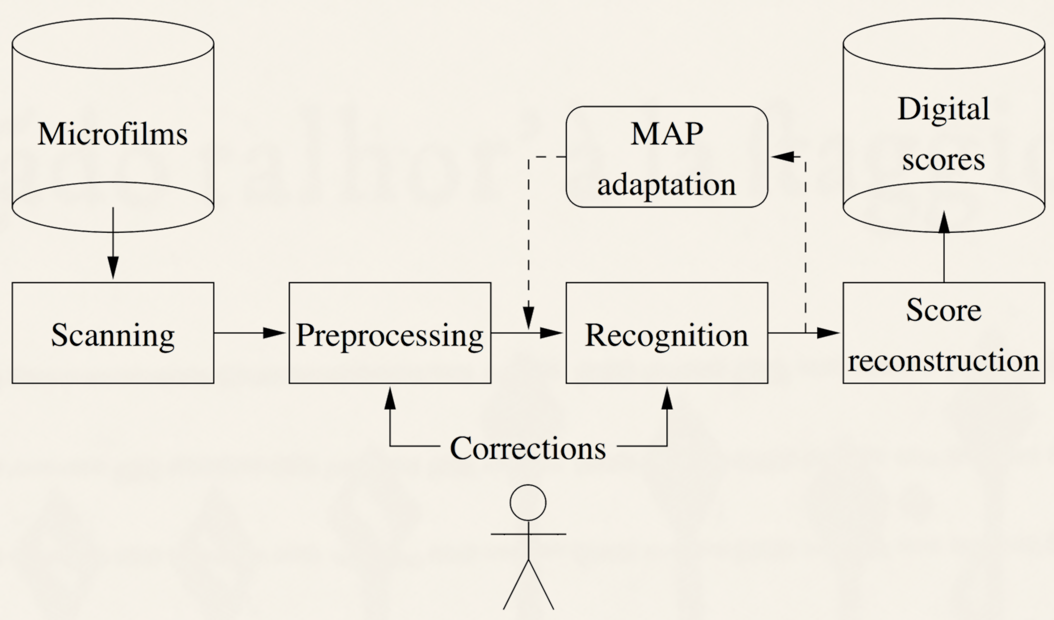

Fig. 1: MAP-adaptive workflow for music digitisation.

Aruspix Software for OMR

Aruspix is a cross-platform software program for OMR on early music prints. It performs the complete OMR process, from image pre-processing to music symbol recognition. It includes a unique integrated editor for early music, which make it usable not only as a research tool, but also as end-user application. The core of the system is based on hidden Markov models (HMMs), a widely-used adaptive statistical approach that requires training on annotated sample data, known as ground-truth.

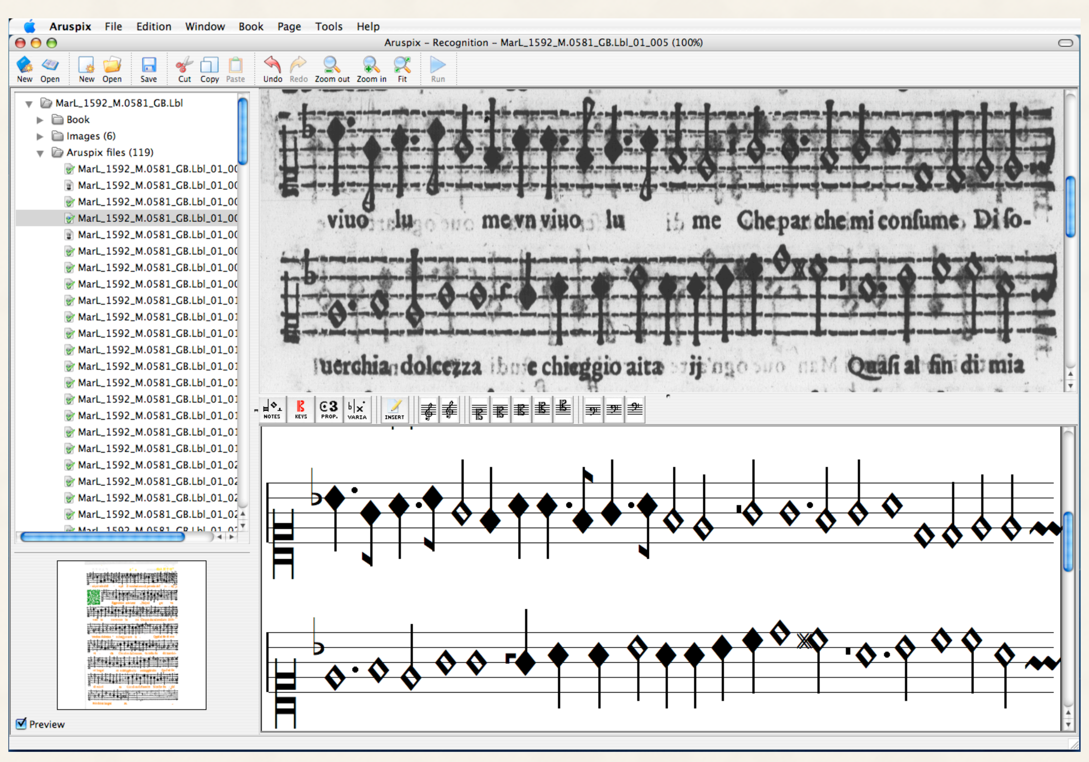

Fig. 2: The Aruspix system in action.

RISM 1528-2 (Pierre Attaignant, Paris, 1528)

RISM 1532-10 (Jacques Moderne, Lyon, 1532)

RISM V-1421 (Figliuoli di Gardano, Venezia, 1572)

RISM R-2512 (Angelo Gardano, Venezia, 1575)

RISM V-1433 (Domenico Basa, Roma, 1585)

Fig. 3: Five prints used for our experiments.

Early Prints

Early music prints exhibit great variability across sources (see fig. 3): font shape differs from one printer to another, original documents are often physically degraded, and scanning parameters are often inconsistent between digitisation runs. All of these factors can reduce the accuracy of the Aruspix system. Under these conditions, no single out-of-the-box solution can be expected to perform well for all books, but generating ground-truth data by hand for each new book in order to train new book-dependent (BD) systems from scratch is equally impractical.

Dynamic Adaptation

Dynamic adaptation avoids the above problems by using very small sets of new data to optimise previously trained recognition systems. In a preliminary phase, a book-independent (BI) system is trained using pages taken from a wide range of different books. The BI system gives acceptable results in general (baseline) but is not optimised for a particular source. During the editing process, we adapt the system as soon as any new page is corrected to generate a steadily improving BD system from the BI foundation.

Labour Cost Reduction

Dynamic adaptation consistently improved performance in our experiments, not only in terms of recogntion rate, but also in terms of a metric we have developed for evaluating the cost of human editing labour with Aruspix.

Book	Recognition rate (%)		Editing cost	
	Baseline	Dyn. Ad.	Baseline	Dyn. Ad.
RISM 1529-1	84.16	91.90	9.21	4.99
RISM 1532-10	74.95	89.39	13.52	6.05
RISM V-1421	92.35	94.50	5.26	4.08
RISM R-2512	95.10	96.97	3.46	2.56
RISM V-1433	91.31	95.72	5.56	3.08

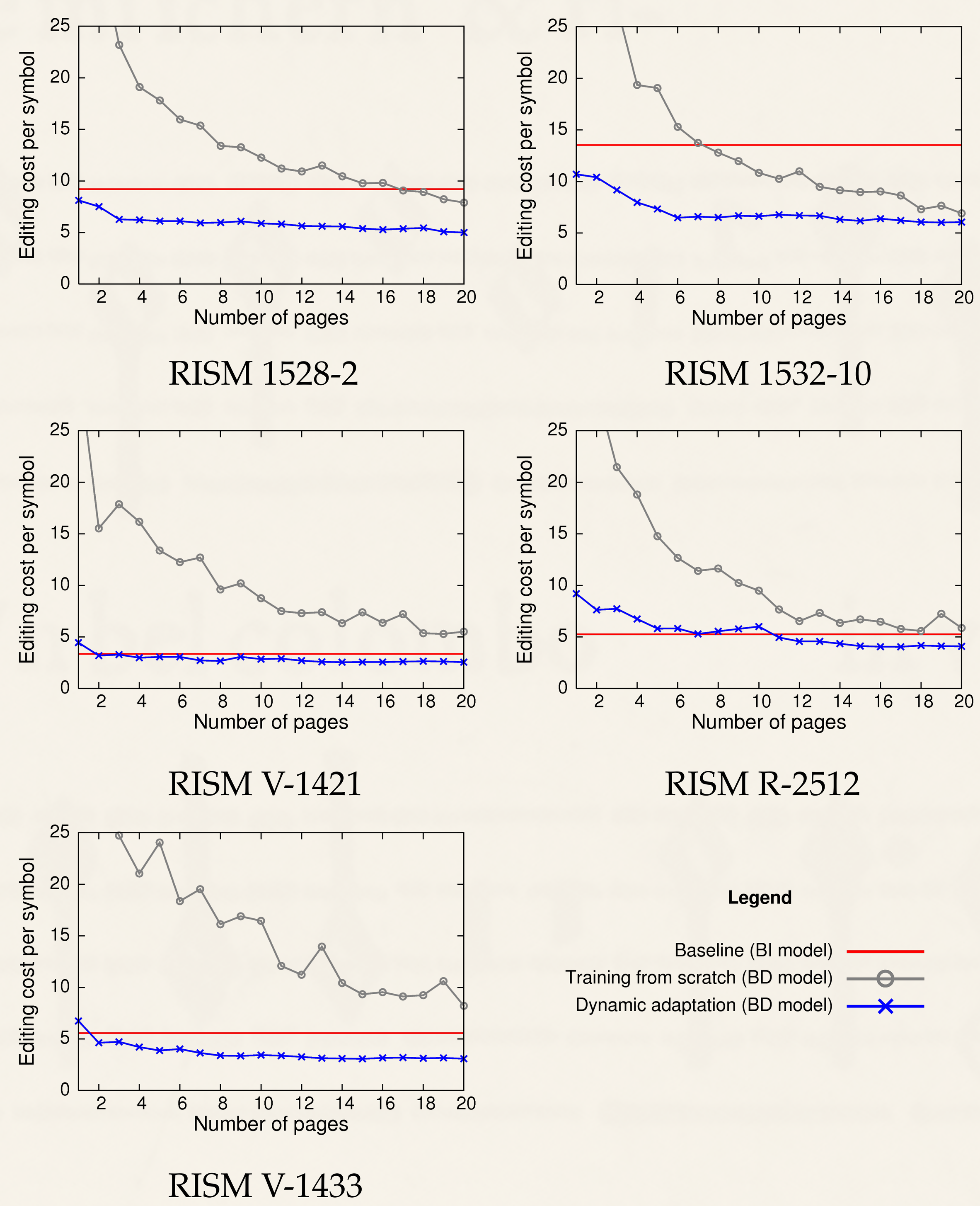

Fig. 4: Editing costs with and without dynamic adaptation

Summary

When digitising early music sources on microfilm, checking and correcting the OMR output is the most time-consuming – and thus most expensive – step in the workflow. To reduce the editing costs and deal with the high variability in the data, we experimented with adding dynamic adaptation to the Aruspix system. Our results show that this approach can improve the system after correcting only a couple of new pages, which means that human editors can very quickly glean time-saving benefits from their work while digitising new books.