SINGLE INTERFACE FOR MUSIC SCORE SEARCHING AND ANALYSIS (SIMSSA)

Ichiro Fujinaga
Schulich School of Music
CIRMMT, McGill University
ich@music.mcgill.ca

Andrew Hankinson
Schulich School of Music
CIRMMT, McGill University
andrew.hankinson@mail.mcgill.ca

ABSTRACT

Single Interface for Music Score Searching and Analysis (SIMSSA) project seeks to design tools and techniques for searching and analyzing digitized music scores. Specifically, we seek to provide researchers, musicians, and others to access the contents and metadata of a large number of scores in a searchable, digital format. In this project, we are developing prototypes for processing and accessing the scores by consulting closely music researchers, musicians, and librarians.

1. INTRODUCTION

While a growing number of digitized images of music scores are being made available on-line to a global audience, these digital images are only the beginning of true accessibility since the musical content of these images cannot be searched by computer. The goal of the Single Interface for Music Score Searching and Analysis (SIMSSA) project is to teach computers to recognize the musical symbols in these images and assemble the data on a single website, making it a comprehensive search and analysis system for online musical scores. SIMSSA is creating an infrastructure for processing music documents, transforming vast music collections into symbolic representations that can be searched, studied, analyzed, and performed anywhere in the world.

SIMSSA is made up of two research axes. The first axis, *Content*, is developing large-scale optical music recognition (OMR) systems for digital images to transforming them into searchable symbolic notation. The second axis, *Analysis*, is developing tools and techniques for large-scale search and analysis. We have assembled a diverse team of researchers and partners to accomplish this mission: music scholars, composers, and performers will ensure that we build tools to address real-world problems, librarians will provide expertise in collection management, metadata, and information-seeking behaviour, and music technologists will develop OMR systems, accessible web-based interfaces, and search and analysis software. Partner institutions including museums, national and research libraries, and universities will provide both digital images and expertise. Central to SIMSSA is the use of collaborative computing, which has been shown to reduce costs and increase accuracy. Musicians, students, and scholars from around the world will be provided with tools to correct and improve the results of the recognition process. They will correct the OMR output for music sources they care about, resulting in searchable music for their own work as well as for other musicians. The SIMSSA network will be a global network of digital music libraries and participant-users: anyone with a web browser will be able to search through vast amounts of music from anywhere in the world.

2. OBJECTIVES

Our objective is to develop a new approach to building globally-accessible digital music scores with a public online digital document analysis and retrieval system. Using OMR technology, we are working with partner institutions to automatically transcribe the contents of their large digital collections, and allow users to search music notation in millions of music scores.

The searchable symbolic content will make it possible to easily compare, analyze, study, arrange, and transpose musical material in new ways. Our tools will provide new kinds of access and exposure to the collections of our partner institutions, from document viewing technology to search engines. New access to large amounts of music and new tools will provide important fundamental materials for future scholarship, creation, and performance. The complexity and variety of musical styles and music notations will lead to important advances in information retrieval and digital document analysis with multiple uses beyond music. As the first project of its kind, we hope that SIMSSA will establish common standards and best practices for these types of music information retrieval and serve as a baseline for future work in this field.

3. BACKGROUND

OMR research began in the late 1960s and has seen limited but continuous interest with several commercial software packages available (e.g., SmartScore and SharpEye). Development of this technology has been slow, and most of the research on OMR has concentrated on Common Western Notation, the most widely used music notation system today (for a recent review, see [1]). In the development stage of this project, we have created a site for the *Liber usualis* (`http://liber.simssa.ca`), to experiment with the methods and procedures of performing OMR on entire books containing older music notation. The website allows a user to search a digitized edition of this book using pitch names, neume names, and OCR-transcribed text [2] (Figure 1). Our experience with the Liber project has reinforced the need for a robust and efficient workflow system for OMR.

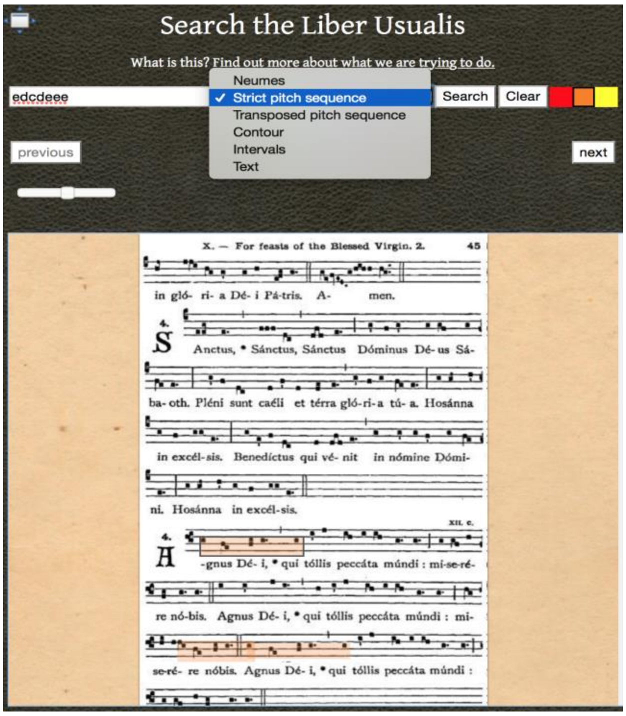

Figure 1. Search the *Liber usualis* website (`http://liber.simssa.ca`) with the *Mary had a little lamb* tune search.

4. THEORETICAL FRAMEWORK

The SIMSSA project is developing a new cloud-based OMR system that introduces a completely new paradigm for this class of software. Typically, OMR software is installed and operated on a single computer or workstation. Advanced techniques used to perform image restoration and automatic music transcription, however, are computationally intensive, sometimes requiring hours or even days to run on personal computers. Instead, we are developing systems where computationally intensive procedures can be distributed across many powerful server machines attached to the Internet to perform processing in parallel, meaning any computer or mobile device with a modern web browser and access to the Internet may act as a document recognition station, offloading the computationally-intensive recognition tasks to large clusters of computers in data centers. We see this as our most significant technological contribution: these techniques, known as distributed computing, are currently being explored in text-recognition research but have not yet been explored for music recognition systems.

There is a successful precedent for projects of this scope and scale. The IMPACT project [3] was a project funded by the European Union (€15.5 million, 2008–2012) that focused on digitization, transcription, and policy development for historical text documents. This project brought together national and specialized libraries, archives, universities, and corporate interests to advance the state of the art in automatic text document transcription, explicitly for the purposes of preserving and providing access to unique or rare historical documents. They have published significant advances in historical text recognition, tool development, policies, and best practices [4], [7].

At the core of the IMPACT project was a networked and distributed document recognition suite, providing a common document recognition platform for all their partners across Europe. As the computer vision and software engineering teams developed new tools and algorithms to improve recognition, these were made available immediately to all partners simply by updating the online suite of tools. All partners could then supply realtime feedback and evaluation on these updates, comparing them to previous techniques "in the field" and reporting their findings. The development teams then incorporated the feedback into further developments and refinements. This project has become self-sustaining and is now known as the IMPACT Centre of Competence, a not-for-profit organization that continues to build the technologies and best practices of the formally funded project. This represents a model that we hope to reproduce in the domain of music.

5. METHODOLOGY

5.1 Content and Analysis Axis

We are building a robust infrastructure with workflow management and document recognition systems, crowdcorrection mechanisms, networked databases, and tools for analyzing, searching, retrieving, and data-mining symbolic music notation. Responsibility for developing these tools within the project is shared between the *Content* and *Analysis* axes.

The Content axis is divided into three sub-axes: *Recognition*, *Discovery*, and *Workflow*. The Recognition sub-axis is responsible for developing the underlying technologies in machine learning and computer vision. The Discovery sub-axis is responsible for large-scale web crawling, finding, and identifying images of books that contain musical content. Finally, the Workflow sub-axis is responsible for developing user-friendly web-based tools that harness the technologies developed by the other two sub-axes.

The Analysis axis is divided into two sub-axes: *Search and Retrieval*, and *Usability*. Searching music is complex since, unlike text, it is not simply a string of characters: there are pitches, rhythms, text, multiple voices sounding simultaneously, chords, and changing instrumentation. The Search and Retrieval axis is responsible for developing ways of mining the notation data generated by the Content axis in all its complexity, building on the work done in the ELVIS Digging into Data Challenge project (http://elvisproject.ca). This axis is also developing techniques for computer-aided analysis of musical scores. The Usability sub-axis is responsible for studying retrieval systems and user behavior within the context of a symbolic music retrieval system, identifying potential areas where the tools may be improved to suit real-world retrieval needs.

5.2 Content: Discovery sub-axis

Mass digitization projects have been indiscriminately digitizing entire libraries' worth of documents—both text and musical scores—and making them available on individual libraries' websites. The Discovery sub-axis is developing a system that will automatically crawl millions of page images looking for digitized books with musical examples [8]. When it finds a document containing printed music it will use the OMR software to transcribe and index the music content for these documents.

5.3 Content: Recognition sub-axis

One of the major tasks of the Recognition sub-axis is the integration of two desktop open-source OMR software platforms: Gamera, a document analysis toolkit [9], and Aruspix, an advanced OMR system developed by Laurent Pugin [10]. These systems are unique for their ability to "learn" from their mistakes by using human corrections of misrecognized symbols to improve their recognition abilities over time. We have shown this to be cost-effective in digitization and recognition workflows [11]. The next logical step is to bring these systems to our cloud-based OMR platform. This will allow us to distribute the correction tasks to potentially thousands of users around the globe, thereby providing the means to collect large amounts of human correction data. This crowd-sourced adaptive recognition system will be the first of its kind [12].

5.4 Content: Workflow sub-axis

The Workflow sub-axis is primarily responsible for developing Rodan, the core platform for managing cloud-based recognition. Rodan is an automatic document recognition workflow platform. Its primary function is to allow users to build custom document recognition workflows containing document recognition tasks, such as, image pre-processing and symbol recognition (Figure 2). Rodan is capable of integrating many different recognition systems, such as Aruspix and Gamera, with other systems (e.g., integrating text recognition tasks for performing automatic lyric extraction) (see Figure 3). Once a workflow has been created, Rodan manages digital document images' progression through these tasks. Users interact with their workflows through a web application, allowing them to manage their document recognition on any Internet-connected device, but all tasks are actually run on the server-side. Storage and processing capabilities can be expanded dynamically, and new tasks can be seamlessly integrated into the system with no need for the users to update their hardware or software.

Moreover, as a web-based system, Rodan can incorporate many different methods for distributed correction or "crowd-sourcing" to provide human-assisted quality control and recognition feedback for training and improving recognition accuracy. This follows a similar model to that proposed by the IMPACT project where distributed proof-readers provide feedback. These proof-readers correct any misrecognized symbols, and their corrections will then be fed back into the recognition system, thereby improving the recognition for subsequent pages and documents. This type of crowd-sourced correction system is employed in many text-recognition projects [13], [14], but there are no such systems in development for musical applications. The success of crowd-sourcing as a viable means of collecting correction and verification data has been demonstrated by a number of projects, most notably the Australian newspaper (TROVE) [15], Zooniverse [16] and reCAPTCHA [17]. Along with developing the technical mechanisms for crowd-sourced musical corrections, the Workflow team is also working with the Usability sub-axis on creating new ways to entice users to participate. Some ways of doing this would be to create a game that rewards users with points or community credibility in exchange for performing work [18], or reframing musical tasks as simple non-musical tasks (e.g., shape or colour recognition) so that they become solvable by an untrained audience. By diversifying the number of approaches to collecting crowd-sourced correction data, we expect to appeal to a wide number of communities, from specialists to the general public.

Later in this project, we will experiment with optical character recognition (OCR) for print and manuscript sources of music. By this point in the project we will have collected a large number of written texts with human-transcribed ground-truth data. We will use this to train machine-learning algorithms to automatically recognize the various text scripts present in these sources. Our goal here is to automatically align text with the music above it, an important step that represents a significant challenge, and an avenue of research that has never before been explored. This will allow users to perform searches for recurring patterns that include music and text — to identify whether, for example, a particular musical idiom is frequently used when the text refers to "God" or "love" — a type of search that is not possible with current systems. When the text-alignment task is complete, the Recognition team will work with the Analysis team to design and implement a search interface so that the users can search music and text simultaneously.

Many musical documents, especially those that are hundreds of years old, pose difficulties for computer recognition due to faded inks, bleed-through, water, or insect damage. Each of these problems is a potential source of transcription errors. The Recognition team is working on integrating the latest document-imaging enhancement technologies, such as adaptive binarization, bleed-through reduction, colour adjustment, and distortion analysis and correction.

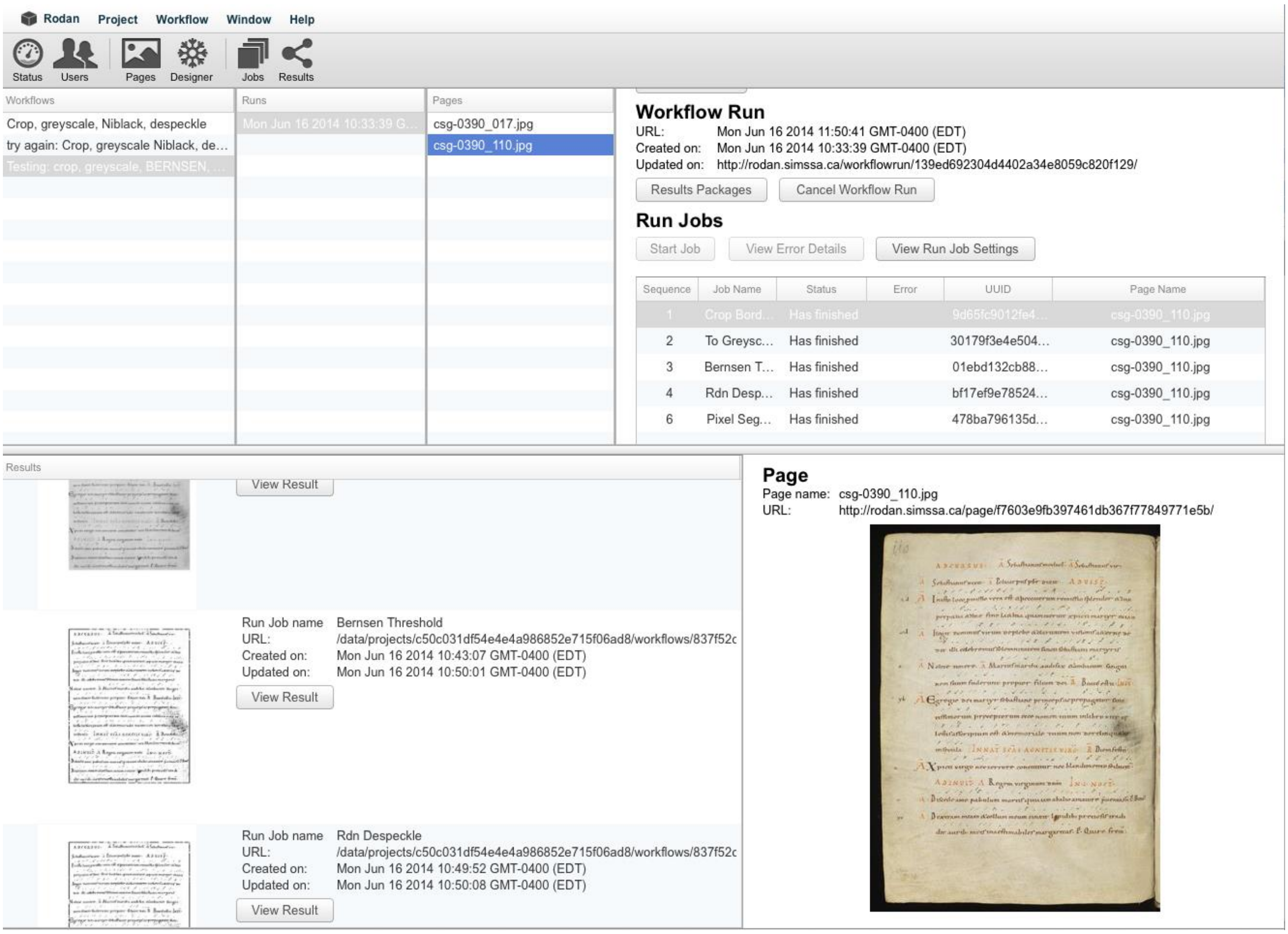

Figure 2. Rodan client interface.

It is also important to have a robust modern file format to store all of the symbolic data representations of these musical documents to meet our needs. Based on previous work we have chosen the MEI (Music Encoding Initiative) format [19]. As part of the SIMSSA project we will be forming a workgroup to enhance MEI support for digital encoding of early notation systems for chant and polyphonic music.

Figure 3 : Rodan's Gamera symbol classifier interface. The symbols are from a 10th-century St. Gallen music manuscript.

To evaluate Rodan and the accuracy of our OMR systems, we have selected several manuscripts and early printed scores that will be processed in order of increasing difficulty for our tools. We have started with a selection of Renaissance prints and late chant manuscripts, some of which are already available online (Figure 4).

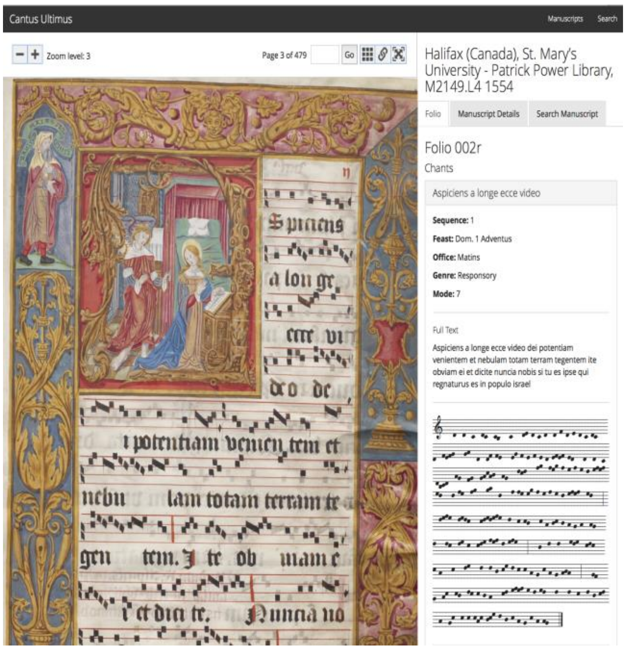

Figure 4 : Salzinnes manuscript website, `http://cantus.simssa.ca`.

As we proceed, we will evaluate the strengths and weaknesses of the workflow system, constantly adjusting our methods before moving on to the next source. For each document, we will create a human-created transcription of the music notation. This data will be the "ground truth" against which we will evaluate the performance and accuracy of the OMR system (and later for the OCR system). This will allow us to quantify any improvements we make in our OMR systems as we develop new recognition methods. By making incremental modifications using different types of sources, we hope to build a robust system capable of processing a wide range of musical documents.

5.5 Analysis: Search and Retrieval sub-axis

The Search and Retrieval component of the Analysis axis will involve music historians and music theorists to investigate how computerized searches of large collections of digital music can fundamentally change music history, analysis, and performance. They will develop new techniques for searching and analyzing digitized symbolic music. Searching music poses special challenges. A search interface must be able to search for strings of pitches, rhythms, or pitches and rhythms combined, search polyphonic music for multiple simultaneous melodies or chords, and search vocal music for both text and music. Searching and retrieving are only the beginning, however; members of the Analysis axis are developing software for many different types of computerized analysis of large amounts of music. This will allow scholars to describe style change over time, discovering which features of style stay the same and which change, or to describe what makes one composer's music unlike that of his or her contemporaries. Musicians and students will be able to find all the different ways composers have harmonized a specific melody from the Middle Ages to the present. Representation of search and analysis findings will be another focus of this axis, investigating new methods for searching and retrieving millions of digitized music documents.

Recent projects such as the Josquin Research Project (`http://jrp.ccarh.org`), Music Ngram Viewer (`http://www. peachnote.com`) [20], and ELVIS project (now part of the SIMSSA project) are already searching millions of notes. All these projects, however, have mostly depended on centralized, labour-intensive, manual processes to transcribe the sources into symbolic notation, append metadata to the resulting files, and arrange them in structured databases. SIMSSA will greatly streamline this process through automation and distributed labour, and enable the sophisticated automatic music analysis of very large corpora begun through ELVIS.

5.6 Analysis: Usability sub-axis

Librarians and information scientists are leading the Usability sub-axis. They are continually review the usability of our tools: Rodan, search interfaces, crowd-sourcing interfaces, and analysis and visualization interfaces, considering the needs and skillsets of many different types of users, from senior music scholars with little technical expertise, to computer-savvy amateur musicians, to choral directors and guitarists searching for sheet music.

6. CONCLUSIONS

The most important outcome of this project is to allow users—scholars, performers, composers, and the general public—to search and discover music held in archives and libraries around the world. We expect that this will fundamentally transform the study of music and allow a global audience of musicians and artists to discover previously unknown or overlooked pieces for performance, making undiscovered repertoires that extend beyond the classics available to the general public. We also expect the public availability of large amounts of musical data to lead to significant advances in the field of music theory and the birth of the long-awaited field of computational musicology. Lastly, we expect that the free and open-source tools we are developing will help lead significant advances in the following areas, all of which are either completely new or novel applications of existing technologies:

- Public, web-based tools for historical image restoration;
- Public, web-based distributed (“cloud”) processing tools for OMR and OCR;
- A large database of automatically transcribed music;
- Prototypes for a web-based editor for making corrections or comparative editions of digital sources;
- A music exploration interface allowing quick and efficient content-based search and retrieval across a large-scale notation database; and
- Advanced public, web-based music analytical tools.

Acknowledgments

This research is, in part, supported by the Social Sciences and Humanities Council of Canada and le Fonds de Recherche du Québec sur la Société et la Culture, and McGill University.

7. REFERENCES

[1] A. Rebelo, I. Fujinaga, F. Paszkiewicz, A. R. S. Marcal, C. Guedes, and J. S. Cardoso, “Optical music recognition: State-of-the-art and open issues.” *International Journal of Multimedia Information Retrieval* (March 2012), pp. 1–18, 2012.

[2] A. Hankinson, J. A. Burgoyne, G. Vigliensoni, A. Porter, J. Thompson, W. Liu, R. Chiu, and I. Fujinga. “Digital document image retrieval using optical music recognition,” in *Proceedings of the International Society for Music Information Retrieval Conference*, Porto, Portugal, 2012, pp. 577–582.

[3] H. Balk and L. Ploeger, “IMPACT: Working together to address the challenges involving mass digitization of historical printed text,” *OCLC Systems & Services*, vol. 25, no. 4, pp. 233–248, 2009.

[4] V. Kluzner, A. Tzadok, Y. Shimony, E. Walach, and A. Antonacopoulos, “Word-based adaptive OCR for historical books,” in *Proceedings of the 10th International Conference on Document Analysis and Recognition*, Barcelona, Spain, 2009, pp. 501–505.

[5] C. Neudecker, S. Schlarb, Z. Dogan, P. Missier, S. Sufi, A. Williams, and K. Wolstencroft, “An experimental workflow development platform for historical document digitisation and analysis,” in *Proceedings of the Workshop on Historical Document Imaging and Processing*, Singapore, 2011, pp. 161–168.

[6] Z. M. Dogan, C. Neudecker, S. Schlarb, and G. Zechmeister, “Experimental workflow development in digitisation,” in *Proceedings of the International Conference on Qualitative and Quantitative Methods in Libraries*, Chania, Greece, 2010, pp. 377–384.

[7] A. Antonacopoulos, D. Bridson, C. Papadopoulos, and S. Pletschacher, “A realistic dataset for performance evaluation of document layout analysis,” in *Proceedings of the International*

Conference on Document Analysis and Recognition, Barcelona, Spain, 2009, pp. 296–300.

[8] D. Bainbridge and T. Bell, “Identifying music documents in a collection of images,” in *Proceedings of the International Conference on Music Information Retrieval,* Victoria, BC, 2006, pp. 47–52.

[9] K. MacMillan, M. Droettboom, and I. Fujinaga, “Gamera: A structured document recognition application development environment,” in *Proceedings of the International Conference on Music Information Retrieval,* Bloomington, IN, 2001, pp. 15–16.

[10] L. Pugin, J. Hockman, J. A. Burgoyne, and I. Fujinaga, “Gamera versus Aruspix: Two optical music recognition approaches,” in *Proceedings of the International Conference on Music Information Retrieval,* Philadelphia, PA, 2008, pp. 419–424.

[11] L. Pugin, J. A. Burgoyne, and I. Fujinaga, “Reducing costs for digitising early music with dynamic adaptation,” in *Proceedings of the European Conference on Digital Libraries,* Budapest, Hungary, 2007, pp. 471–474.

[12] S. Charalampos, A. Hankinson, and I. Fujinaga, “Correcting large-scale OMR data with crowdsourcing,” in *Proceedings of the International Workshop on Digital Libraries for Musicology*, London, UK, 2014, pp. 88–90.

[13] H. Goto, “OCRGrid: A platform for distributed and cooperative OCR systems,” in *Proceedings of the International Conference on Pattern Recognition,* Hong Kong, 2006, pp. 982–985.

[14] G. Newby and C. Franks, “Distributed proofreading,” in *Proceedings of the Joint Conference on Digital Libraries*, Houston, TX, 2003, pp. 361–363.

[15] R. Holley, “Extending the scope of Trove: Addition of E-resources subscribed to by Australian libraries.” *D-Lib Magazine,* vol. 7, no. 11/12, 2011.

[16] K. D. Borne and Zooniverse Team, “The Zooniverse: A framework for knowledge discovery from citizen science data,” in *The Proceedings of the American Geophysical Union Fall Meeting,* 2011.

[17] L. Von Ahn, B. Maurer, C. McMillen, D. Abraham, and M. Blum, “reCAPTCHA: Human-based character recognition via web security measures,” *Science,* vol. 321, no. 5895, pp. 1465–1468, 2008.

[18] L. Von Ahn, “Games with a purpose,” *Computer,* vol. 39, no. 6, pp. 92–94, 2006.

[19] A. Hankinson, P. Roland, and I. Fujinaga, “MEI as a document encoding framework,” in *Proceedings of the International Society for Music Information Retrieval Conference,* Miami, FL, 2011, pp. 293–298.

[20] V. Viro, “Peachnote: Music score search and analysis platform,” in *Proceedings of the International Society for Music Information Retrieval Conference*, Miami, FL, 2011, pp. 359–362.