

# Correcting Large-Scale OMR Data with Crowdsourcing

Charalampos Saitis, Andrew Hankinson, and Ichiro Fujinaga

Schulich School of Music, McGill University  
Montreal, QC H3A 1E3, Canada

{charalampos.saitis, andrew.hankinson}@mail.mcgill.ca, ich@music.mcgill.ca

## ABSTRACT

This paper discusses several technical challenges in using crowdsourcing for distributed correction interfaces. The specific scenario under investigation involves the implementation of a crowd-sourced adaptive optical music recognition system (Single Interface for Music Score Searching and Analysis project). We envisage the distribution of correction tasks beyond a single workstation to potentially thousands of users around the globe. This will have the effect of producing human-checked transcriptions, as well as significant quantities of human-provided ground-truth data, which may be re-integrated into an adaptive recognition process, allowing an OMR system to “learn” from its mistakes. Drawing from existing crowdsourcing approaches and user interfaces in music (e.g., Bodleian Libraries) and non-music (e.g., CAPTCHAs) applications, this project aims to develop a scientific understanding of what makes crowdsourcing work, how to entice, engage and reward contributors, and how to evaluate their reliability. While results will be considered based on the specific needs of SIMSSA, such knowledge can be useful to a variety of musicological investigations that involve labour-intensive methods.

## Categories and Subject Descriptors

H.3.4 [Information Storage and Retrieval]: Systems and Software—*Distributed systems*; H.5.3 [Information Interfaces and Presentation]: Group and Organization Interfaces—*Collaborative computing*; H.5.5 [Information Interfaces and Presentation]: Sound and Music Computing—*Systems*

## Keywords

Optical music recognition, music notation, crowdsourcing, web applications, music score searching

## 1. INTRODUCTION

The Single Interface for Music Score Searching and Analysis (SIMSSA) project<sup>1</sup> is a recently funded initiative to investigate the development of a fully searchable digital musical document collection, including document digitization, high-volume optical music recognition (OMR), large-scale symbolic search systems, and digital document display. The goal of SIMSSA is to teach computers to recognize the musical symbols in digitized notation and assemble the data on a single website, making online search and analysis of musical scores possible for the first time. Scholars and the general public will be able to consult millions of musical works in the blink of an eye, a task that would previously have required many lifetimes. Central to the SIMSSA project is the use of Web-based technologies to facilitate the development of online *crowdsourcing* tools for the collection of large amounts of human-checked transcriptions.

SIMSSA aims to integrate Gamera [7] and Aruspix [19], two major desktop open-source OMR software platforms, into a cloud-based environment. These systems are unique for their capacity to “learn” from their mistakes by using human corrections of misrecognized symbols to improve their recognition abilities over time, resulting in highly cost-effective digitization and recognition workflows [18]. By making them browser-based, distributed proofreading and correction can be potentially performed by anyone with a web browser and an Internet connection, anywhere in the world, thereby allowing for a constantly-improving search and retrieval system, as well as for the collection of human-provided ground-truth data to further reduce the amount of error in the final product while maintaining a level of quality and efficiency beyond simple automatic recognition [11].

In the following sections we outline the prime challenges facing the design of the crowd-sourced, adaptive OMR system envisioned by the SIMSSA initiative, focusing on those related to the user engagement and assessment. This position paper is part of an ongoing research programme that aims to provide SIMSSA with a theoretical framework for developing effective crowdsourcing tools.

## 2. BACKGROUND

Humans are critical components for performing quality control in a pattern recognition process, correcting the inevitable errors that automated systems make and ensuring these errors do not compound themselves in subsequent workflow steps [12, 4]. Correction, however, can be very time, labour and cost intensive. In the case of optical character recognition (OCR)—the textual counterpart of OMR—some unique solutions have been developed, lever-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.  
*DLJ/M'14*, September 12 2014, London, United Kingdom

Copyright is held by the owner/author(s). Publication rights licensed to ACM. ACM 978-1-4503-3002-2/14/09 \$15.00.

<http://dx.doi.org/10.1145/2660168.2660186>.

<sup>1</sup><http://simssa.ca>

aging collective intelligence and the Web to help offset the costs of this task. The reCAPTCHA project [23] has produced over 5 billion human-corrected OCR words by presenting the correction task as a spam-fighting challenge to prove that the user is a human and not an automated system. The Australian Newspapers Digitisation Project (ANDP) [13] has created a distributed correction system, where more than 9,000 volunteers have now corrected more than 12.5 million lines of text, with more corrections added all the time.

Both these examples are manifestations of crowdsourcing, a Web-enabled, distributed problem-solving model that has emerged during the past decade [2]. Its technical and conceptual underpinnings of crowdsourcing lie at the intersection of collective intelligence, outsourcing and the Internet [3, 20]. A crowdsourcing system is a system where a large number of people, known as contributors, are enlisted to help solve a problem defined by the system owners (corporations, libraries, researchers, etc.), which would normally require intensive (and often tedious), costly labour. Apart from helping reaching out to potentially thousands of users around the globe, the Internet offers a high degree of automation and unique possibilities for user management (e.g., wiki, discussion groups) [6]. Ever since the term was coined in 2006 [15], many names (citizen science, crowd wisdom, human computation, etc.), definitions [8] and typologies [10] of crowdsourcing have appeared in an attempt to provide this new concept with a theoretical framework.

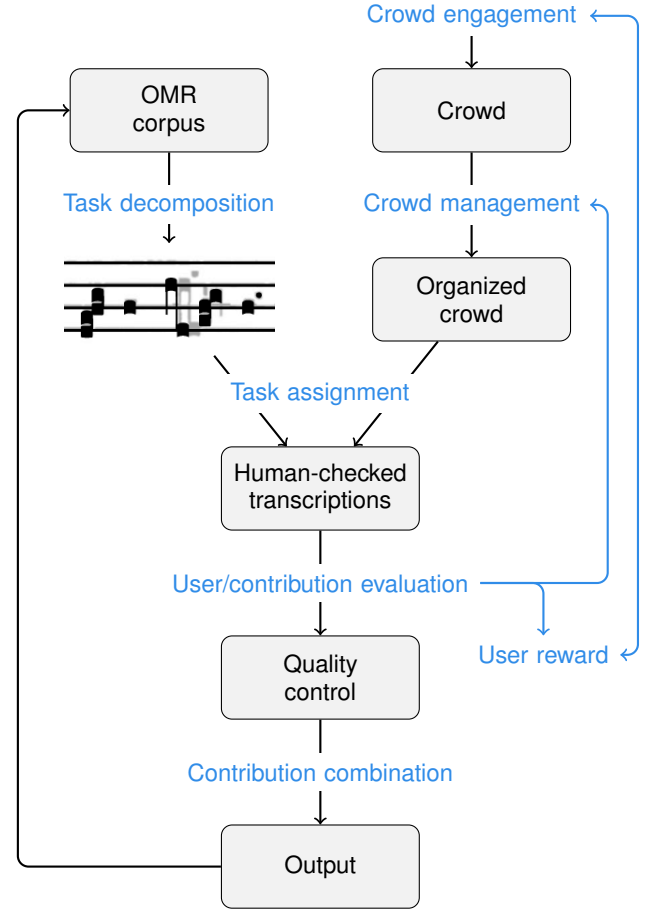
### 3. CHALLENGES

Unlike small-scale public engagement activities such as tagging or rating a song, crowdsourcing requires a greater level of commitment from the contributor in terms of effort, time and intellectual input, as well as continuous assessment and reward by the system owners [14].

**How to engage contributors?** According to Doan, Ramakrishnan, and Halevy [6], one can either *require users* (e.g., students may be required to help correct transcriptions as part of the class curriculum), *make users “pay” for service* (i.e., *implicit* contribution; e.g., reCAPTCHA [23]), *piggyback on the user traces* (e.g., search query logs may be used for spelling corrections [1] or keyword generation [9]), or simply *ask for volunteers*. For the latter a strategy to motivate individuals to contribute is required. To this end, methods and topics from the uses and gratifications theory can provide a useful framework for understanding intrinsic (e.g., fun, challenge, community) and extrinsic motivating factors (e.g., class assignment, practical reward) [3].

**What type of tasks can users perform?** Many contributors may be hesitant at the prospect of re-editing an entire score, therefore it is necessary to “chunk” up error correction of an entire OMR corpus into much smaller units of work such as a single page, line or bar. In addition, different tasks may require different levels of musicological knowledge (e.g., notation systems) and thus be assigned to appropriate subgroups of contributors. For example, musicologists specializing in early music notation may perform more complex correction tasks (e.g., replacement of incorrect neumes) than volunteers with little or no relevant background (e.g., position and pitch of single notes). A two-stage strategy for crowd management is therefore needed, both prior to the assignment of the first task (using test tasks and/or questionnaires) and following the evaluation of previously assigned tasks (which may gradually increase the abilities of a contributor [21]).

**How to evaluate users and their contributions?** Crowdsourcing, as with any open call to an undefined group of people, is prone to incompetent and sometimes



**Figure 1: A framework for correcting large-scale OMR data with crowdsourcing. Adapted from Kit-tur et al. [16, Fig. 2]. Workflow- and user-specific challenges are marked in light blue.**

malicious use. Continuous evaluation of users and their contributions is essential in monitoring individuals’ abilities and thus assuring effective crowd management. A manual approach to this challenge includes the further distribution of contributions by ordinary users to an independent team of trusted users, whereby “bad” work is flagged on message boards [6]. For example, the same task can be assigned to three contributors. If one result differs substantially from the other two, then that contributor gets flagged. The less flags a contributor has, the more trustworthy he is. Such reliability testing can also be done on the fly; a contributor can be given a correction task for which the system already knows the solution, then flagged as reliable or unreliable based on their response [6].

**How to reward and retain users?** Provided that the contributions of a user are reliable, a strategy to reward and retain them should be encouraged. Such a strategy should be part of the overall crowd engagement approach (e.g., extrinsic motivation). For example, contributors may be offered exclusive beta versions of certain (possibly relevant) software or free access to digital libraries. Ideas may also be drawn from the emerging area of *gamification* [5]: contributors can earn badges that they can then redeem as points in online games; high-volume data collection and verification can be turned into a game where participants are rewarded points for their work (e.g., the ESP game [22]; MajorMinor [17]).

Based on these considerations, Fig. 1 presents a frame-

work for crowdsourcing OMR transcription corrections, adapted from the more general one proposed by Kittur et al. [16, Fig. 2]. We have identified two major types of challenges in implementing such a system (light blue colour): workflow-specific (task decomposition, task assignment, contribution combination [6]) and crowd-specific (engagement, management, evaluation, reward). We aim to explore solutions to these issues based on both hypothetical and real-world scenarios such as the Bodleian Libraries at the University of Oxford (currently using the Zooniverse platform<sup>2</sup> for crowdsourcing descriptions of newly digitized scores<sup>3</sup>) and Neon.js (a browser-based neume editor<sup>4</sup> [4]).

Large-scale music databases (e.g., songs, scores) are becoming increasingly central to musicological scholarship. We hope that our ongoing survey of crowdsourcing experiences and tools will result in essential infrastructure components for creating next-generation globally accessible digital music libraries for the future musicologist.

## 4. ACKNOWLEDGMENTS

We would like to thank Gabriel Vigliensoni and Esteban Maestre for fruitful discussions. The SIMSSA project is funded by the Social Sciences and Humanities Research Council (SSHRC) of Canada.

## 5. REFERENCES

- [1] F. Ahmad and G. Kondrak. Learning a spelling error model from search query logs. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language*, pages 955–962, Vancouver, Canada, 2005.
- [2] D. C. Brabham. Crowdsourcing as a model for problem solving: An introduction and cases. *Convergence: The International Journal of Research into New Media Technologies*, 14(1):75–90, 2008.
- [3] D. C. Brabham. *Crowdsourcing*. MIT Press, Cambridge, MA, 2013.
- [4] G. Burlet, A. Porter, A. Hankinson, and I. Fujinaga. Neon.js: Neume editor online. In *Proceedings of the 13th International Society for Music Information Retrieval Conference*, pages 121–126, Porto, Portugal, 2012.
- [5] S. Deterding, M. Sicart, L. Nacke, K. O’Hara, and D. Dixon. Gamification: Using game-design elements in non-gaming contexts. In *CHI’11 Extended Abstracts on Human Factors in Computing Systems*, pages 2425–2428, Vancouver, Canada, 2011.
- [6] A. Doan, R. Ramakrishnan, and A. Y. Halevy. Crowdsourcing systems on the World-Wide Web. *Communications of the ACM*, 54(4):86–96, 2011.
- [7] M. Droettboom, K. MacMillan, and I. Fujinaga. The Gamera framework for building custom recognition systems. In *Symposium on Document Image Understanding Technologies*, pages 275–286, Greenbelt, MD, 2003.
- [8] E. Estellés-Arolas and F. G.-L. de Guevara. Towards an integrated crowdsourcing definition. *Journal of Information Science*, 38(2):189–200, 2012.
- [9] A. Fuxman, P. Tsaparas, K. Achan, and R. Agrawal. Using the wisdom of the crowds for keyword generation. In *Proceedings of the 17th international conference on World Wide Web*, pages 61–70, Beijing, China, 2008.
- [10] D. Geiger, S. Seedorf, T. Schulze, R. Nickerson, and M. Schader. Managing the crowd: Towards a taxonomy of crowdsourcing processes. In *Proceedings of the Seventeenth Americas Conference on Information Systems*, page paper 430, Detroit, MI, 2011.
- [11] A. Hankinson, J. A. Burgoyne, G. Vigliensoni, and I. Fujinaga. Creating a large-scale searchable digital collection from printed music materials. In *Proceedings of the 21st International Conference Companion on World Wide Web*, pages 903–908, Lyon, France, 2012.
- [12] A. Hankinson, J. A. Burgoyne, G. Vigliensoni, A. Porter, J. Thompson, W. Liu, R. Chiu, and I. Fujinaga. Digital document image retrieval using optical music recognition. In *Proceedings of the 13th International Society for Music Information Retrieval Conference*, pages 577–582, Porto, Portugal, 2012.
- [13] R. Holley. Many hands make light work: Public collaborative OCR text correction in Australian historic newspapers. *National Library of Australia Staff Papers*, 2009.
- [14] R. Holley. Crowdsourcing: How and why should libraries do it? *D-Lib Magazine*, 16(3/4), 2010.
- [15] J. Howe. The rise of crowdsourcing. *WIRED magazine*, 2006.
- [16] A. Kittur, J. V. Nickerson, M. S. Bernstein, E. M. Gerber, A. Shaw, J. Zimmerman, M. Lease, and J. J. Horton. The future of crowd work. In *Proceedings of the 16th ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW)*, pages 1301–1317, San Antonio, TX, 2013.
- [17] M. I. Mandel and D. P. W. Ellis. A web-based game for collecting music metadata. *Journal of New Music Research*, 37(2):151–165, 2008.
- [18] L. Pugin, J. A. Burgoyne, and I. Fujinaga. Reducing costs for digitising early music with dynamic adaptation. In L. Kovács, N. Fuhr, and C. Meghini, editors, *Research and Advanced Technology for Digital Libraries: Proceedings of the 11th European Conference, ECDL 2007, Budapest, Hungary*, pages 471–474. Springer, Berlin Heidelberg, 2007.
- [19] L. Pugin, J. Hockman, J. A. Burgoyne, and I. Fujinaga. Gamera versus Aruspix: Two optical music recognition approaches. In *Proceedings of the 9th International Conference on Music Information Retrieval*, pages 419–424, Philadelphia, PA, 2008.
- [20] G. D. Saxton, O. Oh, and R. Kishore. Rules of crowdsourcing: Models, issues, and systems of control. *Information Systems Management*, 30:2–20, 2013.
- [21] M. Shen, G.-H. Tzeng, and D.-R. Liu. Multi-criteria task assignment in workflow management systems. In *Proceedings of the 36th Hawaii International Conference on System Sciences*. IEEE, 2003.
- [22] L. von Ahn and L. Dabbish. Labeling images with a computer game. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 319–326, Vienna, Austria, 2004.
- [23] L. von Ahn, B. Maurer, C. McMillen, D. Abraham, and M. Blum. reCAPTCHA: Human-based character recognition via web security measures. *Science*, 321:1465–1468, 2008.

<sup>2</sup><http://www.zooniverse.org/>

<sup>3</sup><http://www.whats-the-score.org/>

<sup>4</sup><http://ddmal.music.mcgill.ca/neon>