2nd International ACM Workshop on Music Information Retrieval with User-Centered and Multimodal Strategies (MIRUM)

Cynthia C.S. Liem
Delft University of Technology
Delft, The Netherlands
c.c.s.liem@tudelft.nl

Meinard Müller
Bonn University
and MPI Informatik
Saarbrücken, Germany
meinard@mpi-inf.mpg.de

Steven K. Tjoa
iZotope, Inc.
San Francisco, USA
stjoa@izotope.com

George Tzanetakis
University of Victoria
Victoria, Canada
gtzan@cs.uvic.ca

ABSTRACT

The International ACM Workshop on Music Information Retrieval with User-Centered and Multimodal Strategies (MIRUM) at ACM Multimedia was proposed in order to gather experts from the Music and Multimedia Information Retrieval communities, as well as other neighboring fields, and to provide a high-profile platform for presenting current work on Music Information Retrieval with a strong focus on user-centered and multimodal approaches. Following a successful first edition at ACM Multimedia 2011, a second edition of MIRUM was held at ACM Multimedia 2012, which is the focus of this overview. After a description of the rationale and focus areas of the workshop, the accepted submissions and other program elements are summarized.

Categories and Subject Descriptors

H.5.5 [**Information Interfaces and Presentation**]: Sound and Music Computing

General Terms

Algorithms, Design, Experimentation, Human Factors

Keywords

music information retrieval, multimodal music processing, user-centered design, cross-domain methodology transfer

1. INTRODUCTION

Music is an outstanding example of a content type with many different representations, including audio recordings, symbolic scores, folksonomy descriptions and accompanying video material. Due to the existence of such complementary musical representations and information sources, music content is multimedia data by definition. However, research on music content is typically much less represented at premier multimedia venues than work focusing on visual content only. This led to the International ACM Workshop on Music Information Retrieval with User-Centered and Multimodal Strategies (MIRUM), which was first proposed for ACM Multimedia 2011.

MIRUM 2011 was a successful workshop, being highly visible and attracting the second highest submitted paper volume of all ACM Multimedia 2011 workshops. In order to sustain this increased visibility of music work in the Multimedia community, a renewed edition of the workshop at ACM Multimedia 2012 was desirable, leading to the second edition of MIRUM being proposed and accepted.

Once again, MIRUM 2012 managed to attract a considerable number of papers. With 27 submissions, MIRUM 2012 received more submissions than any other ACM Multimedia workshop, being the only workshop with more than 20 submissions. The growing popularity of MIRUM underlines its relevance, and allows for a balanced and high-quality paper selection which will be discussed later in this overview.

2. FOCUS AREAS

As outlined in the workshop theme paper of MIRUM 2011 [4], the MIRUM workshop especially focuses on user-centered and multimodal strategies, in order to stimulate publications on current open challenges shared by the Music and Multimedia Information Retrieval communities. As in the first edition, the concrete focus areas of MIRUM 2012 published in the Call for Papers included music multimedia content analysis, visual and sensory information for music processing, multimodal music search, retrieval and recommendation, social networks and indexing for music applications, music similarity measures at different specificity levels, fusion of multimodal music information sources, music knowledge representation and reasoning, interactive music systems and retrieval, (adaptive) user interaction and interfaces, user (context) models and personalization, real-world issues, evaluation methods and data understanding, and cross-domain methodology transfer.

3. PRESENTED PAPERS

As mentioned above, 27 technical papers were submitted for review to the MIRUM workshop. After a double-blind review process, 12 of these papers were selected for presentation at the workshop, yielding an acceptance rate of 44.4%. In the paper selection, which will be summarized below, multimodal and user-centered approaches are strongly featured. In many cases, the papers are

MM'12, October 29–November 2, 2012, Nara, Japan.
ACM 978-1-4503-1089-5/12/10.

motivated by real-life use cases and concrete application contexts, indicating practical relevance and applicability.

An example of a paper dealing with a concrete application context is the work by Kaminskas et al. [2], focusing on music recommendation for geographic places of interest. The authors propose a knowledge-based approach to deal with this type of application-specific music context. Another work focusing on music context is contributed by Liu and Yang [5], in which it is examined to what extent personal traits (age and gender) can be inferred from features related to music listening history.

If music is performed by musicians, different musicians will perform it in different ways. For audio matching use cases, in which musicial excerpts are to be matched robustly within and across performances, Thomas et al. [12] propose an indexing approach which allows for fast audio matching of musical excerpts in a music collection. Focusing on performance-specific features, Perez [8] presents work on extracting violin instrumental controls from audio data, which empirically was learned from performance sensor data.

Moving towards personalization, Su and Fung [11] employ active learning techniques to achieve personalized music emotion classification that will reduce required user labeling efforts. Another work focusing on personalization and the reduction of required user interaction is contributed by Pardo et al. [6]. Here, active learning and transfer learning are combined to obtain personalized results from an audio equalization interface.

In a qualitative study focusing on user insights for potential music recommender system applications, Laplante [3] investigates how real-life social networks influence the circulation of music information and the music taste of adolescents. Cartwright and Pardo [1] present a user study investigating the appropriateness of audio novelty measures to estimate temporal saliency, and the usefulness of temporal saliency for the improvement of audio similarity measures.

While tempo may seem a well-established and understood concept, the task of automatically estimating tempo in music audio signals is not solved yet. Peeters and Flocon-Cholet [7] address this problem by considering perceptual tempo, and proposing a GMM regression approach motivated by perceptual and musical notions. Another work departing from a conventional perspective, this time on classes of music emotion, is that of Wang et al. [13], who unify the typically separated categorical and dimensional music emotion classes under a probabilistic framework.

Challenging the practical success of mainstream approaches in genre recognition systems, Sturm [10] re-implements two established systems and examines their performance results in more detail, concluding that it is far from obvious that music genre is actually being recognized. In addition, Sturm presents an analysis of a frequently used genre recognition dataset [9], and lists aspects of the dataset that should explicitly be taken into account when interpreting results achieved on its data.

4. OTHER PROGRAM ELEMENTS

The keynote speech at MIRUM 2012 is given by Dr. Ye Wang, who works as an Associate Professor at the National University of Singapore, and established and directs the Sound and Music Computing Lab of this university. In his talk [14], Dr. Wang will present an ongoing collaborative research project between his lab and the Music, Neuroimaging, and Stroke Recovery Lab at the Beth Israel Deaconess Medical Center of Harvard Medical School. The project focuses on sound and music technology for mobile and cloud-based healthcare applications. More specifically, a cloud-based therapy delivery system is under development, using music and smart mobile devices to enhance limb function and speech production in patients with neurological impairments.

5. CONCLUSION

An overview of the MIRUM 2012 workshop was given. The growing popularity of the workshop led to a diverse selection of scientifically strong papers with clear practical relevance. Together with the keynote speech, the accepted papers will lead to a full-day workshop, providing a focused and relevant complement to the scientific program of ACM Multimedia 2012.

6. REFERENCES

[1] M. Cartwright and B. Pardo. Novelty Measures as Cues for Temporal Salience in Audio Similarity. In *Proceedings ACM MIRUM 2012*, Nara, Japan, November 2012.

[2] M. Kaminskas, I. Fernández-Tobías, I. Cantador, and F. Ricci. Knowledge-based Music Retrieval for Places of Interest. In *Proceedings ACM MIRUM 2012*, Nara, Japan, November 2012.

[3] A. Laplante. Who Influence the Music Tastes of Adolescents? A Study on Interpersonal Influence in Social Networks. In *Proceedings ACM MIRUM 2012*, Nara, Japan, November 2012.

[4] C. C. S. Liem, M. Müller, D. Eck, G. Tzanetakis, and A. Hanjalic. The Need for Music Information Retrieval with User-Centered and Multimodal Strategies. In *Proceedings ACM MIRUM 2011*, Scottsdale, Arizona, November 2011.

[5] J. Liu and Y.-H. Yang. Inferring Personal Traits from Music Listening History. In *Proceedings ACM MIRUM 2012*, Nara, Japan, November 2012.

[6] B. Pardo, D. Little, and D. Gergle. Building a Personalized Audio Equalizer Interface with Transfer Learning and Active Learning. In *Proceedings ACM MIRUM 2012*, Nara, Japan, November 2012.

[7] G. Peeters and J. Flocon-Cholet. Perceptual Tempo Estimation using GMM Regression. In *Proceedings ACM MIRUM 2012*, Nara, Japan, November 2012.

[8] A. Perez. Learning and Extraction of Violin Instrumental Controls from Audio Signal. In *Proceedings ACM MIRUM 2012*, Nara, Japan, November 2012.

[9] B. Sturm. An Analysis of the GTZAN Music Genre Dataset. In *Proceedings ACM MIRUM 2012*, Nara, Japan, November 2012.

[10] B. Sturm. Two Systems for Automatic Music Genre Recognition: What Are They Really Recognizing? In *Proceedings ACM MIRUM 2012*, Nara, Japan, November 2012.

[11] D. Su and P. Fung. Personalized Music Emotion Classification via Active Learning. In *Proceedings ACM MIRUM 2012*, Nara, Japan, November 2012.

[12] V. Thomas, S. Ewert, and M. Clausen. Fast Intra-Collection Audio Matching. In *Proceedings ACM MIRUM 2012*, Nara, Japan, November 2012.

[13] J.-C. Wang, Y.-H. Yang, K. Chang, and H.-M. Wang. Exploring the Relationship between Categorical and Dimensional Emotion Semantics of Music. In *Proceedings ACM MIRUM 2012*, Nara, Japan, November 2012.

[14] Y. Wang. When Music, Information Technology, and Medicine Meet. In *Proceedings ACM MIRUM 2012*, Nara, Japan, November 2012.