

# AES 53<sup>rd</sup> International Conference Semantic Audio

26–29 January 2014  
Barbican Centre  
London, UK

## CONFERENCE REPORT

**S**emantic Audio is concerned with content-based management of digital audio recordings. The rapid evolution of digital audio technologies, e.g., audio data compression and streaming, the availability of large audio libraries online and offline, and recent developments in content-based audio retrieval have significantly changed the way digital audio is created, processed, and consumed. New audio content can be produced at a very low cost, while also large audio archives at libraries or record labels are opening to the public. The amount of available audio data grows greater each day, which is why content-based management of audio recordings is essential to enable efficient use of these data. Aside from audio retrieval and recommendation techniques, the semantics of audio signals are also becoming more important in object-oriented and intelligent production, including audio coding, editing, and processing. Recent product releases demonstrate this but many more innovative functionalities are becoming available.

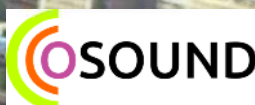
Conference partners





London's Barbican Centre, venue for the 53rd Conference

Sponsored by



Conference cochairs Mark Sandler (top) and Karlheinz Brandenburg

This was the motivation for the AES 53rd International Conference, Semantic Audio, held in London, from January 26 to 29, 2014. It was the second AES conference on the topic and follows the highly successful 42nd Conference from 2011. The aim of the 53rd Conference was to give a broad overview of the rapidly changing state of the art, address many of the new scientific disciplines involved in this still-emerging field, and continue to foster this line of interdisciplinary research. From the outset, it was agreed that the conference should feature a large number of invited talks. This brought together the luminaries in the field and allowed all attendees to get a big-picture view of the state of the art in semantic audio and where it is heading.

Although the conference was hosted by Queen Mary University of London (QMUL), it was a joint effort of the University and Fraunhofer IDMT. It was cochaired by Mark Sandler (QMUL) and Karlheinz Brandenburg (IDMT). Papers chairs were Christian Dittmar (IDMT), George Fazekas (QMUL), and Sebastian Ewart (QMUL). The local organization team was also led by George Fazekas.

The technical program was full and varied. There were 85 attendees for a program consisting of 17 technical talks, including keynote and invited talks, 34 poster and demonstration presentations, two tutorials, and one panel session. The schedule was arranged into sessions each dealing with particular aspects of semantic audio: music informatics and retrieval; semantic web, linked data and ontologies; automatic music transcription; semantic audio organization and retrieval; intelligent audio production; and speech processing and analysis.

## TUTORIAL DAY

The first day was dedicated to tutorials, and the venue was Queen Mary University of London's Media, Art and Technology Lab. Chris Cannam and Mark Plumbley gave a tutorial on reusable software, which inspired researchers to publish their code and data, and explained the means by which this can be done. This was followed by a tutorial overview given by Michael Terrell on the emerging field of semantic audio and music production.

## MAIN PROGRAM – DAY 1

The technical program began on the second day, and the venue for this was the prestigious Barbican, the largest performing arts center in Europe. Introductory remarks on the first day were given by Mark Sandler and George Fazekas, and the program kicked off with a keynote

talk on music structure analysis by Meinard Müller. The technical papers this Monday morning set the scene for the conference, with overview talks on creating research corpora and bridging the gap between audio and symbolic notations, as well as a thought-provoking talk on analysis of piano recordings that showed some important ways in which performers deviate from the musical score. An invited talk by Yves Raimond also generated a lot of

interest, especially from researchers interested in “big data,” when he described how a mixture of semantic web technologies, machine-generated annotations, and crowd-sourcing are exploited to enable the BBC to rapidly publish large archives of multimedia content.

Perhaps the highlight of this day was a live demonstration of the Complete Classical Music Companion by researchers from the Austrian Research Institute for Artificial Intelligence. They showcased the first integrated prototype of a system that identifies what piece is being performed and at what position within the piece. It then tracks the performance and will adjust to changes by the performer. This was demonstrated live for acoustic piano, with volunteer pianists from the audience. It was clear that the system could almost instantaneously adapt to whenever the performer switched to a new piece or left out segments within a piece. The prototype was still on after the demonstration had officially ended, and it provoked laughter from the audience when it tried to find the piano piece that matched the program chair's announcements about upcoming events.

## MAIN PROGRAM – DAY 2

The second day opened with a keynote talk by Gerhard Widmer, discussing some big-picture challenges in music semantics, and focusing on how a computer might identify, or even generate, expressivity in music. This was followed by a session on music transcription, another big challenge, which included technical talks on automatically transcribing bass guitar, detecting musical key (and using that to improve transcription), and unsupervised learning of the down beats on drum tracks. The poster session that followed included descriptions of several toolboxes, datasets and frameworks, as well as several presentations

related to intelligent audio effects and multitrack mixing.

The last session of the day had interesting talks on frontier research, one on modeling emotion in music, and the other on how



Much enthusiasm was expressed during animated previews of the various posters to be discussed in the poster session.



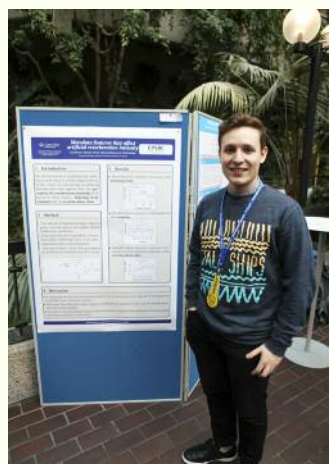
Yading Song, David Ronan, and Dawn Black of the conference committee during a break



Delegates meet to share ideas around the demos and posters.



Presenters at the podium to deliver their papers



Carl Bussey with his poster on metadata features for reverberation intensity



Tuomas Eerola and Simon Dixon enjoy a discussion during one of the poster sessions.



James Vegnuti and Mi Tian welcome delegates at the registration desk.

to design audio-related experiments as games to be played on social networks or websites.

This day ended with the highlight of the social program, a gala dinner on a boat cruise along the Thames River. The cruise went past many of London's most famous sites,

including Parliament, London Eye, St. Paul's Cathedral, and Tower Bridge. After the dinner awards were given for best paper (Tom Collins et al., "Bridging the Audio-Symbolic Gap: The Discovery of Repeated Note Content Directly From Polyphonic Music Audio"), best presentation (Katerina Kosta et al., "Practical Implications of Dynamic Markings in the Score: Is Piano Always Piano?"), best student papers (Christian Kirst et al., "On-Line NMF-Based Stereo Up-Mixing of Speech Improves Perceived Reduction of Non-Stationary Noise"; and Daniel Gärtner et al., "Unsupervised Learning of the Downbeat in Drum Patterns"), best poster (Chrisoula Alexandraki et al., "Using Computer Accompaniment to Assist Networked Music Performance"), and best demonstration (Andreas Arzt et al., "The Complete Classical Music Companion V0.9"). The awards included products raffled out by Native Instruments and iZotope. A Reproducible Research Prize was awarded to Dan Stowell.



Committee members Mi Tian and Jordan Smith



Jordan Smith and Brecht De Man look at one of the demos together.

## MAIN PROGRAM – DAY 3

The last day of the conference began with a keynote talk by Gaël Richard. He presented an excellent overview of the state of the art in audio source separation, focusing on informed techniques that offer superior performance if some additional information is known. This talk was invaluable for those wanting to understand the rapidly changing field of source separation, as well as to get a feel for some of the most important contributions in this large research area. This was followed by an invited talk by Jay LeBoeuf, focusing on the challenges and opportunities for commercializing semantic audio research. Next, Pedro Pestana gave a stimulating talk summarizing a vast body of work he has performed on determining best practices for music production and how this may be exploited by intelligent systems. Both Jay and Pedro were panellists in a special session immediately following this, where experts discussed the emerging field of intelligent audio production. The technical program ended with a session on speech processing and analysis, including a somewhat humorous, but still rigorous, technical talk on the influence of alcohol intoxication on speaker recognition, which used the "Alcohol Language Corpus" as a testbed. The conference ended on a high note, with a technical tour of BBC's new (and impressive) Broadcasting House.

Editor's note: a USB drive or downloadable PDF of the conference papers can be purchased online at [www.aes.org/publications/conf.cfm](http://www.aes.org/publications/conf.cfm)



The band plays on after the water-borne gala dinner on the third night



The conference chairs enjoy congratulating awardees.



London's Tower Bridge at night from the boat cruise



Committee with delegates: Brecht De Man, Katerina Kosta, Matthias Mauch, Shefki Kolozali, Panos Kudumakis, Carl Bussey, Yading Song



Meinard Müller and Olivier Lartillot



Katerina Kosta happy at having won the best presentation award



The assembled company gets together for a group photo at the end of the event.