

Conference report

42nd International Conference Ilmenau, Germany

AES 42nd International Conference

Semantic Audio

22 - 24 July 2011, Ilmenau, Germany





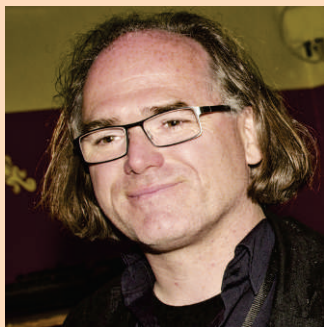
Conference cochairs: Mark Sandler (above) and Karlheinz Brandenburg (right)



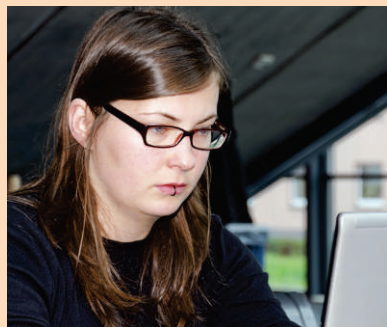
Papers cochairs: Christian Dittmar (left) and Anssi Klapuri (right)

The rapid evolution of digital audio technologies, audio compression, and streaming has significantly changed the way digital audio is created, processed, and consumed today. New audio content can be produced at a very low cost. Thus the sheer amount of available audio data grows more and more each day. 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





Holger Grossmann, facilities chair (left), and Peggy Walther, secretary and webmaster



Delegates enjoy getting to know each other over coffee.

becoming more important in object-oriented audio coding, editing, and processing. Recent product releases demonstrate this, but many more innovative functionalities are becoming available.

This was the motivation for the AES 42nd International Conference Semantic Audio, held in Ilmenau, Germany, July 22–24. The conference was concerned with research and development targeted toward content-based handling of digital audio recordings. Its aim was to give a broad overview of the 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 the Technical University of Ilmenau, it was a joint effort of the University, Fraunhofer IDMT, and Queen Mary University of London (QMUL). It was cochaired by Karlheinz Brandenburg (IDMT) and

Mark Sandler (QMUL). Papers cochairs were Christian Dittmar (IDMT) and Anssi Klapuri (QMUL). The local organization team was led by Peggy Walther of IDMT.

The technical program was full and varied. There were 21 technical talks and an additional 13 poster presentations. The schedule was arranged into sessions each dealing with particular aspects of semantic audio, such as music information retrieval, speech processing and analysis, automatic music transcription, audio source separation, and intelligent audio effects. It covered feature extraction techniques, ontologies, and classification and recommendation systems as well as evaluation methods.

Introductory remarks on the first day were given by Karlheinz Brandenburg. The technical sessions on Friday set the scene for the conference, with overview talks on advances in music information retrieval, interactive music applications, and semantic speech tagging. One inspiring talk was given by Masataka Goto,



who demonstrated several technologies that are changing the way we listen to and understand music. His presentation also produced universal laughter from the entire audience, when he pointed out that music is an environmental friendly, energy-saving resource. Unlike film, which is rarely watched by anyone many times over, music is a highly renewable resource that can be appreciated more with repeated listening. Of course, semantic audio is not just related to music, and the second paper session made this clear, with talks on semantic speech tagging, synchronization of speech and text, and automatic speech recognition. This session made clear just how challenging the tasks are and how advanced the technologies have become. For instance, it is possible to use machine learning in some cases to estimate the age, ethnic background, and even height of the speaker. The third paper session was dedicated to automatic music transcription, often called the holy grail or grand challenge in musical signal processing. A highlight in that session was Karin Dressler's talk on pitch estimation. As pointed out by session chair Anssi Klapuri, she has come first in community-based evaluation of pitch-estimation techniques several years in a row, and her talk gave everyone an opportunity to find out how she accomplished this. The first day concluded with a three-course dinner at the Hotel Lindenhof.

On Saturday, the talks began with an explanation of adaptive distance measures for exploring and structuring music collections by Sebastian Stober, bringing together several years' work in this field. This was followed by a thought-provoking talk by Cynthia Liem on applying alignment techniques in order to compare expressivity across musical performances. The morning also saw a focused session on audio source separation, including a method to extract singing voice from stereo recordings. The afternoon began with the second poster session. Two interesting posters were on a psychoacoustic approach to wavefield synthesis and on the use of gaze tracking to observe uncertainty in music

tagging, both illustrating the interdisciplinary aspects of semantic audio. The final paper session of the day concerned informed source separation. In contrast to blind source separation, the informed approach exploits some additional (semantic) knowledge for improved performance. This session included a talk by Jürgen Herre on parametric coding of audio objects, overviewing the state of the art, commercial applications, and opportunities for the future.

Saturday evening began with a visit to Fraunhofer IDMT labs. The attendees were treated to tours of several spatial audio labs and an anechoic chamber, a demonstration of high-performance flat-panel loudspeakers, and a preview of their new semantic audio application, Songs2See, which features many of the technologies discussed during the conference. This was then



Karin Dressler discusses pitch estimation during the session on automatic music transcription, which is often called the holy grail of musical signal processing.



Pablo Cabañas Molero discusses the separation of singing voices from stereo recordings.





Invited speakers educate the conference delegates: top row, from left, Masataka Goto, Björn W. Schuller, Meinard Müller, and Jürgen Herre; bottom row, Sebastian Stober, left, and Christof Faller.

followed by a barbecue and concert, and for some the festivities continued late into the night.

The final day began with a talk on automated chord recognition, a topic that has seen much interest recently. The final session of the conference was on intelligent audio effects. Josh Reiss gave an overview of how multichannel intelligent effects can be used for live sound. This was followed by Christof Faller's discussion of advances in microphone designs that enable better directivity and improved surround recording. These two talks illustrated how semantics can be exploited not just for postprocessing and interaction tasks, but also for recording and live performance. The final talk of the session was given jointly by Gyorgy Fazekas and Thomas Wilmering. It dealt with the design and use of semantic web ontologies for multitrack music production. This exploratory work was a fitting end to the technical sessions. For those not travelling back home that day, there was a delightful excursion to Eisenach. This included lunch at the charming, western-themed Lasso restaurant, and a visit to the Bachhaus, the first museum in the world dedicated to Johann Sebastian Bach.

Delegates crowd around posters to hear authors such as Sang Ha Park (below) present their ideas about semantic audio.



Editor's note: A printed proceedings (or a PDF of the entire book) of the conference papers can be purchased online at www.aes.org/publications/conf.cfm. Individual papers are available in the AES E-Library at www.aes.org/e-lib/.



From left, Thippur Srivas, Laurent Girin, and Estefanía Cano explain the principles of source separation.



The band plays on during an enjoyable evening of entertainment.



Social events during the conference gave everyone an opportunity to relax and get acquainted. Clockwise from top left: delegates enjoy one of the excellent buffet lunches; a trip to the Bach museum at Eisenach; Verena Hart of Fraunhofer IDMT is surprised with a birthday cake; and dinner at the western-themed Lasso restaurant.

