AES 25th INTERNATIONAL CONFERENCE

Metadata for Audio

London, UK
June 17–19, 2004

John Grant, conference chair

Metadata is the key to managing audio information and making sure that audiences know what is available in the way of audio program material. This was the theme for the AES 25th International Conference held from June 17–19 at Church House in London, drawing an audience of over 100 delegates from broadcasting, universities, digital libraries, and audio program-management operations from around the world. Chaired by John Grant, with papers cochairs Russell Mason and Gerhard Stoll, the conference spanned topics such as feature extraction, frameworks, broadcast implementations, libraries and archives, and delivery.

I'M AN AUDIO ENGINEER, TEACH ME ABOUT METADATA!

Since metadata is not audio at all, but rather structures for information about audio, it lies squarely in the domain of information professionals. For this reason audio engineers can find themselves at something of a loss when it comes to understanding what is going on in metadata systems and protocols. The metadata world is full of confusing acronyms like SMEF, P-META, MPEG-7, XML, AAF, MXF and so forth, representing a minefield for the unwary. Addressing the need for education in this important area, the first day of the conference was designed as a tutorial program, containing papers from key presenters on some of the basic principles of metadata.

John Grant, conference chair, welcomed everyone and then introduced the BBC's Chris Chambers who discussed identities and handling strategies for metadata, helping the conference delegates to understand the ways in which systems can find and associate the elements of a single program item. He outlined some of the standards being introduced by the industry to deal with this issue, neatly setting the scene for the following speakers who would describe the various approaches in greater detail.

Philip de Nier from BBC R&D proceeded to describe the Advanced Authoring Format (AAF) and Material Exchange Format (MXF), which are both file formats for the storage and exchange of audio–visual material together with various forms of metadata describing content and composition. This was followed by an XML primer given by Adam Lindsay from Lancaster University, UK. XML is a very widely used means of encoding metadata within so-called schema. Continuing the theme of file formats, John Emmett proceeded to discuss ways of "keeping it simple," describing the broadcast wave format (BWF) and AES31, which is an increasingly important standard for the exchange of audio content and edit lists, using BWF as an audio file structure.

Russell Mason chaired *Practical Schemes*, the second session. Philippa Morrell, industry standards manager of the BOSS Federation in the UK, took delegates through the relatively obscure topic of registries.These are essentially ➥

Photos courtesy Paul Troughton (http://photos.troughton.org)




INVITED AUTHORS

Clockwise, from top left: Max Jacob, Jügren Herre, Emilia Gómez, Michael Casey, Wes Curtis, Oscar Celma, Richard Wright, and Alison McCarthy.




databases that are administered centrally so as to avoid the risk of duplication and consequent misidentification of content. The issue of classification and taxonomy of sound effects is also something that requires wide agreement, and the management of this was outlined by Pedro Cano and his colleagues from the Music Technology Group at the University of Pompeu Fabra in Spain. Cano spoke about a classification scheme inspired by the MPEG-7 standard and based on an existing lexical network called WordNet, in which an extended semantic network includes the semantic, perceptual, and sound-effects-specific terms in an unambiguous way.

Richard Wright of BBC Information and Archives was keen to point out that the term metadata is not used in the same way by audio and video engineers as it is in standard definitions. Metadata, correctly defined, is the structure of information, or to more strictly refine the definition, the organization, naming, and relationships of the descriptive elements. In other words it is everything but data itself, it is beyond data. Describing so-called core metadata, he showed that this could be either the lowest common denominator or a complete description that fits the most general case. The first has limitations in what it can do and the second does not work; however, the first lends itself to standardization because it fits the requirements of being essential, general, simple, and popular.

Ending the tutorial day, Geoffroy Peeters from IRCAM in France chaired a workshop on MPEG-7, a key international standard involving metadata. There were four presentations on different aspects of the standard, including the first on managing large sound databases, given by Max Jacob of IRCAM, Paris, and a second on integrating low-level metadata in multimedia database management systems, given by Michael Casey of City University, London. Emilia Gómez, Oscar Celma, and colleagues from the University of Pompeu Fabra described tools for content-based retrieval and transformation of audio using MPEG-7, concentrating on the so-called SPOffline and MDTools. SPOffline stands for Sound Palette Offline and is an application based on MPEG-7 for editing, processing, and mixing intended for sound designers and professional musicians. MDTools was designed to help content providers in the creation and production phases for metadata and annotation of multimedia content. Finishing this session, Jürgen Herre of the Fraunhofer Institute in Germany gave a tour of the issues surrounding the use of MPEG-7 low-level scalability.

METADATA FRAMEWORKS AND TOOLKITS

The first main conference day following the tutorial sessions began with a discussion of frameworks for metadata. First, Andreas Ebner of IRT introduced data models for audio and video production, and then Wes Curtis and Alison McCarthy of BBC TV outlined the EBU's P/Meta scheme for the exchange of program data between businesses.

The Digital Media Project is led by Leonardo Chariglione, mastermind of the MPEG standardization forums, and this exciting project was reviewed at the conference by Richard Nicol from Cambridge-MIT Institute. He spoke about the need to preserve end-to-end management of digital content








INVITED AUTHORS

Clockwise, from top left: Philippa Morrell, Philip de Nier, Richard Nicol, Kate Grant, Chris Chambers, Andreas Ebner, John Emmett, and Adam Lindsay.

in a world where such content is potentially widely distributed and easy access is needed to the information about it. Following this Kate Grant of Nine Tiles provided delegates with an interesting summary of the "what and why" of the MPEG-21 standard, which is essentially a digital rights management structure. "The 21st Century is a new Information Age, and the ability to create, access, and use multimedia content anywhere at any time will be the fulfillment of the MPEG-21 vision," she proclaimed.

Turning to the issue of spatial scene description, Guillaume Potard of the University of Wollongong in Australia introduced an XML-based 3-D audio-scene metadata scheme. He explained that existing 3-D audio standards such as VRML (X3D) have very basic sound-description capabilities, but that MPEG-4 had powerful 3-D audio features. However, the problem is that these are based on a so-called "scene graph" with connected nodes and objects, which can get really complex even for simple scenes. Such structures are not optimal for metadata and cannot easily be searched. Consequently he bases the proposed XML-based scheme on one from CSound, using "orchestra" and "score" concepts, whereby the orchestra describes the content or objects and the score describes the temporal behavior of the objects.

POSTERS

A substantial number of poster papers were presented during the 25th International Conference, almost all concerned with the topic of feature extraction and automatic content recognition. Many of the posters described methods by which different musical features can be analyzed and described, such as instrument recognition, song complexity, percussion information, and rhythmic meter.

Some unusual presentations in this series included a means of retrieving spoken documents in accordance with the MPEG-7 standard, given by Nicolas Moreau and colleagues from the Technical University of Berlin, and an opera information system based on MPEG-7, presented by Oscar Celma from Pompeu Fabra.

FEATURE EXTRACTION

The topic of feature extraction is closely related to that of metadata. It also has a lot in common with the field of semantic audio analysis—the automatic extraction of meaning from audio. Essentially, feature extraction is concerned with various methods for identifying important aspects of audio signals, either in the spectral, temporal, or spatial domains, searching for objects and their characteristics and describing them in some way.

A variety of work in this field was described in papers presented during the afternoon of the first conference day, many of which related to the analysis of music signals. For example, Claas Derboven of Fraunhofer presented a paper on a system for harmonic analysis of polyphonic music, based on a statistical approach that examines the frequency of occurrence of musical notes for determining the musical key. Chords are determined by a pattern-matching algorithm, and tonal components are determined by using a nonuniform frequency transform.

Excellent lecture facilities were provided by Church House; located in the heart of London, it also hosts the annual meeting of the Church of England.

Bee-Suan Ong was interested to find out how representative sections of music can be identified automatically. She used the example of the Beatles' "Hard Day's Night" to show how humans can quickly name the title after hearing very short but easily recognizable bits of the song, whereas machines have to be shown how to do this. Musical items have to be segmented so as to separate the parts of a song; this can be done either using a model involving a training phase or an approach that is training-model free.

Christian Uhle of Fraunhofer looked at the matter of drum-pattern-based genre classification from popular music. The method involves the transcription of percussive unpitched instruments using independent subspace analysis and extraction of amplitude envelopes from each spectral profile. Note onsets corresponding to each percussive instrument are detected using extraction of amplitude envelopes. Continuing on the same theme of musical rhythm, a paper from Fabien Gouyon and his colleagues dealt with the evaluation of rhythmic descriptors for musical genre classification. This involved an attempt to detect the rhythmic pattern and tempo from an audio signal, followed by training and matching using a database of typical ballroom dancing numbers with specific genres such as jive, quickstep, and tango. In the final paper on feature extraction Oliver Hellmuth and colleagues from Fraunhofer described a process of music-genre estimation from low-level audio features; in this case attempting to detect the genre from nine broad possibilities including blues, classical, country, and so forth, using MPEG-7 low-level features such as spectral flatness, spectral crest factor, sharpness, and audio spectrum envelope. They found that the correct genre was assigned for 91.4% of all items.

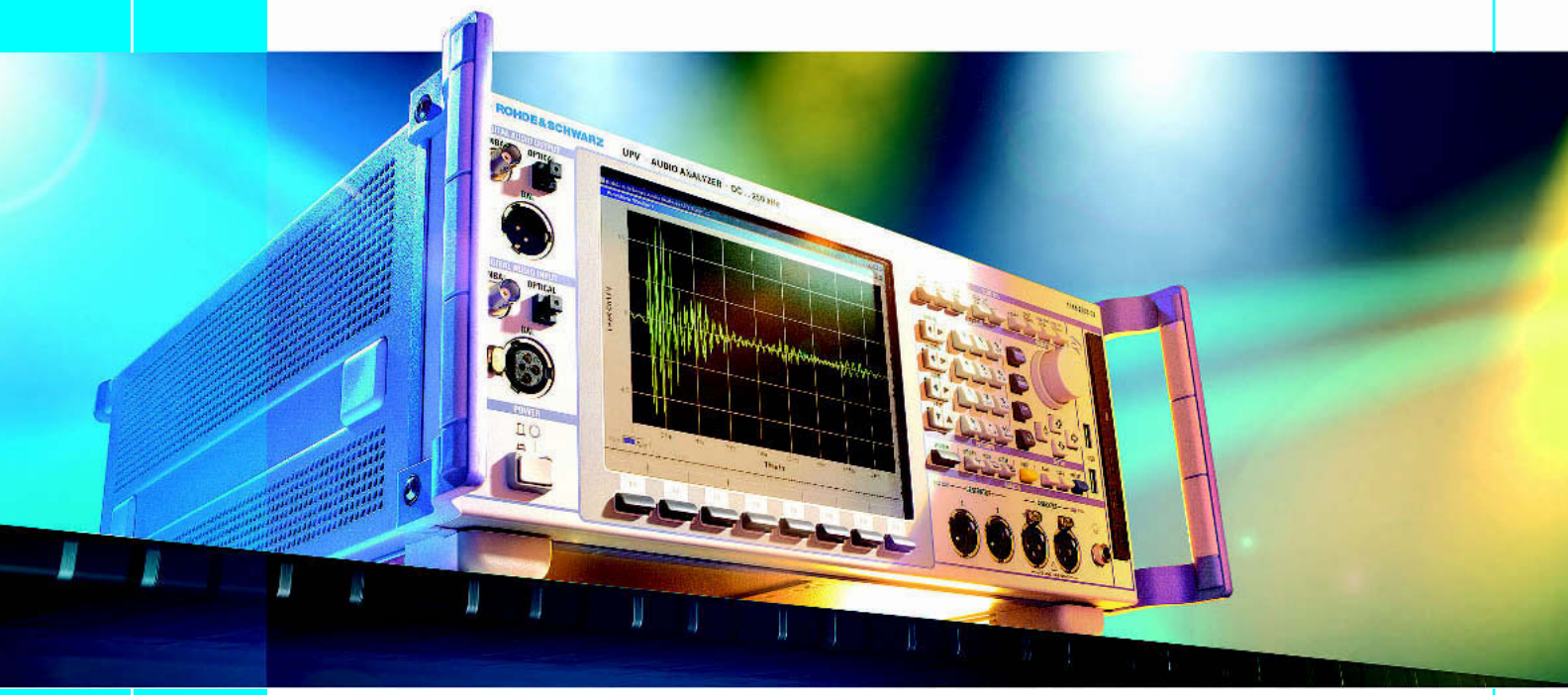
BROADCASTING IMPLEMENTATIONS

One of the primary application areas for metadata structures is in broadcasting, where multiple users need to be able to store and access program material in a way that can enhance the production and distribution process. Broadcasters have typically been among the first to develop and embrace the various standards for audio metadata. The second full conference day involved a morning devoted to this important topic.

Shigeru Aoki, from Tokyo FM Broadcasting, presented a paper on audio metadata in broadcasting that advocated a file format containing the audio and metadata within one file. ➥

Q&A sessions after each presentation at the conference afforded attendees the opportunity to get further clarification from authors and to offer their own perspective on the topic: Claudia Schremmer and Mark Plumbley are shown here.

Pure audio performance



pushing limits

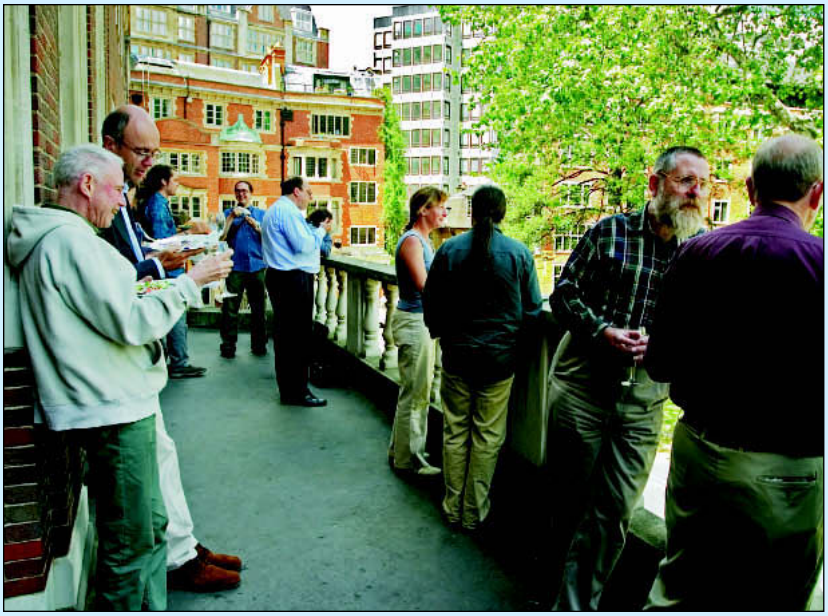
The R&S®UPV is the most advanced audio analyzer in the world

Rohde & Schwarz presents the new R&S®UPV Audio Analyzer. Its performance makes it the new reference standard – there is simply nothing else like it.
The R&S®UPV pushes the limits of technology for broadcasting and R&D, and is an extremely fast instrument for the production line. It handles high resolution digital media just as easily as analog measurements up to 250 kHz.
With all of its performance the Windows XP-based R&S®UPV is simple to operate. It's also expandable, and remarkably affordable. Take a look at its capabilities, and then contact us to find out more.

- For all interfaces – analog, digital and combined
- Real two-channel signal processing for maximum measurement performance
- Digital AES/EBU interface with sampling rate up to 192 kHz
- 250 kHz analysis bandwidth
- Recording and replaying of audio signals
- Overlapping FFT analysis
- Expandable, with easy addition of further audio interfaces like I^2S
- Built-in PC with Windows XP operating system
- Intuitive, ergonomically sophisticated user interface for efficient operation

www.upv.rohde-schwarz.com

ROHDE&SCHWARZ



Left, the balcony at Church House was a favorite spot for casual conversation during coffee breaks and lunches. Above, Mike Story (left) and John Richards, with the southern tower of the Houses of Parliament as a backdrop, share a toast on the balcony.

He concluded that the Broadcast WAVE file is suitable for this purpose and that cue-sheet metadata, for example, can be stored in an <adtl> chunk so that it remains with the audio.

Joe Bull of SADiE and Kai-Uwe Kaup of VCS spoke about integrated multimedia in the broadcast environment, looking at concepts of workflow, media assets, and genealogy. Genealogy, for example, is concerned with the tracking of the processes performed on the media assets and the different version numbers, so that the "lineage" of a media item can be traced. This is particularly important if royalties are to be allocated appropriately. These two company representatives were keen to emphasize the importance of developing an integrated approach to metadata so that material can be transferred between systems, but they felt that it is unlikely that the whole process can be automated, and they can also see that each broadcaster will end up with an approach specifically tailored to its operation.

The second session on broadcast implementations concerned interchange file formats. Broadcast WAVE is increasingly used as a universal file format for audio in broadcast environments. Phil Tudor, speaking on behalf of authors from Snell and Wilcox, showed how this can be encapsulated within the Material Exchange Format (MXF), a SMPTE standardized interchange format. A move to enable this encapsulation is currently under way, which if successfully standardized will enable both AES3 digital audio data and Broadcast WAVE contents to be transferred in MXF and reconstituted in their original form if required. David McLeish of SADiE then described ways in which editing information can be exchanged within the Advanced Authoring Format. He concluded that AAF provides the flexibility and extensibility to enable the necessary compositional decisions and effects parameters to be exchanged between users or systems. AAF and MXF fulfill different purposes within the broadcast production chain, but are intended to be complementary. A so-called zero-divergence doctrine exists between them to avoid incompatibility.

LIBRARIES AND ARCHIVES

Three papers on the afternoon of the second day concerned the use of audio metadata in libraries and archives. Sam Brylawski of the U.S. Library of Congress spoke about the development of a digital preservation program, and in particular the National Sound Conservation Archive. He said that they were considering using a 96-kHz, 24-bit PCM format for archive masters, one-bit formats being "ahead of the game" for them at the present time. The Library of Congress archive will be based on OAIS (Open Archival Information System, see http://www.rlg.org/longterm/oais.html) in which media objects will be described using MODS (the Metadata Object Description Schema, see http://www.loc.gov/standards/mods/); and metadata encoded using METS (Metadata Encoding and Transmission Standard, see http://www.loc.gov/standards/mets/). Long-term content management will be enhanced by additional metadata, but the danger is that with so much information needed to be collected the process will have to be automated in some way.

Miguel Rodeño, of the Computer Science Department at the Universidad de Alcalá in Spain and also associated with IBM Storage and Sales in Madrid, discussed the audio metadata used in the Radio Nacional de España Sound Archive ➥

Poster presentations at the 25th Conference were well received.

Your battery-operated processor didn't sound as good as it should'a,

BUT

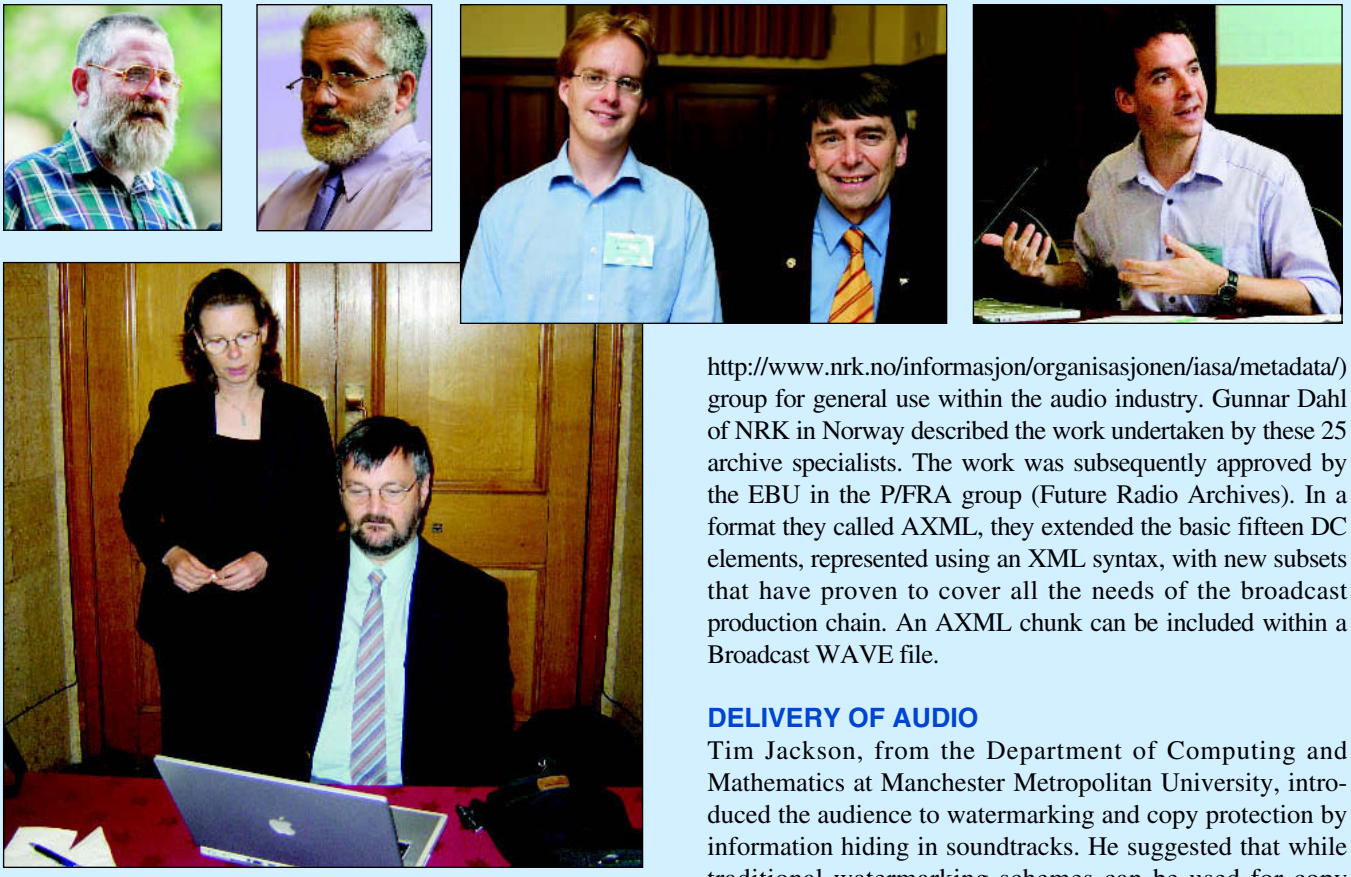
IT COULD'A!

THAT 4320 Analog Engine® lets you extend battery life without compromising sound quality in all your low-power audio signal processors. See how low voltage, and low power coexist with low distortion, wide dynamic range, and lots of processing power in a very tiny package.

THAT Corporation

Making Good Sound Better®

45 Sumner Street, Milford, MA 01757-1656, USA
Tel: +1 (508) 478-9200 Fax: +1 (508) 478-0990
Email: info@thatcorp.com Web: www.thatcorp.com



AES 25th Conference Committee: clockwise from top left, John Grant, Chris Chambers, Russell Mason, Gerhard Stoll, Paul Troughton, Mark Yonge, and Heather Lane.

Project. This was a large enterprise designed to store 20th Century Spanish sound history in a digital form. They used the Broadcast WAVE format but managed the use of the broadcast extension chunk of the file in a clever way to store description metadata, retaining information about the original physical media from which the file had been derived.

Dublin Core (DC) is a widely used metadata scheme that has been implemented in a particular way by the Scandinavian Audiovisual Metadata (SAM, see http://www.nrk.no/informasjon/organisasjonen/iasa/metadata/) group for general use within the audio industry. Gunnar Dahl of NRK in Norway described the work undertaken by these 25 archive specialists. The work was subsequently approved by the EBU in the P/FRA group (Future Radio Archives). In a format they called AXML, they extended the basic fifteen DC elements, represented using an XML syntax, with new subsets that have proven to cover all the needs of the broadcast production chain. An AXML chunk can be included within a Broadcast WAVE file.

DELIVERY OF AUDIO

Tim Jackson, from the Department of Computing and Mathematics at Manchester Metropolitan University, introduced the audience to watermarking and copy protection by information hiding in soundtracks. He suggested that while traditional watermarking schemes can be used for copy protection, information-hiding schemes might possibly be useful for copy prevention in the analog domain, perhaps by interfering with the subsequent audio coding devices such as sigma-delta modulators or subband coders.

Music is increasingly being distributed online, resulting in changing business models. Metadata enables business models to be developed that conform to specific rules for digital

Before the banquet (above) at the Houses of Parliament, the AES guests, including Francis and Sally Rumsey and Roger Furness (left), enjoyed a glass of wine on the terrace overlooking the River Thames.

The well preserved historic buildings of London, such as Westminster Abbey as seen from the balcony of Church House, served as reminders to conference attendees that preservation of historic audio is one of the important functions of metadata.

rights management. Nicolas Sincaglia of MusicNow covered a number of issues related to this topic, concluding that the technical side of the industry will continue to be challenged to build and operate the complex business systems to support the promotions and creative sales side of the industry.

Audio metadata transcription from meeting transcripts for the Continuous Media Web (CMWeb) was described by Claudia Schremmer from the CSIRO ICT Centre in Australia. The problem she described was that continuous media such as audio and video are not well suited to the current searching and browsing techniques used on the Web. A format known as Annodex can be used to stream the media context, multiplexed with XML markup in the Continuous Media Markup Language (CMML). The project she described dealt with the issue of automatically generating Annodex streams from complex annotated recordings collected for use in linguistics research.

DINNER AT THE HOUSES OF PARLIAMENT

Thanks to Heather Lane's superb organizational skills and auspicious connections, an opportunity was offered to delegates to visit the historic Houses of Parliament, just a short walk from Church House in Westminster. A guided tour enabled visitors to see the debating chambers and division lobbies, and even to play a part in acting out a typical voting procedure, which is done according to old customs. Statues of former prime ministers such as Winston Churchill, Lloyd George, and Harold Macmillan lined the walls. The House of Lords throne glistened with 24-carat gold worth over a million pounds. The audio systems used in the debating chambers were explained; the system uses numerous microphones suspended on small flat cables from the ceiling so that they retain their angles of orientation and can be raised and lowered for events requiring a clear line of view, such as television relays of the state opening.

After the tour the visitors had the rare privilege of enjoying a glass of wine on the famous terrace of the House overlooking the River Thames, followed by an excellent dinner around a long table in one of the dining rooms of the Houses of Parliament.

THANKS TO...

The conference committee: John Grant, chair, Gerhard Stoll and Russell Mason, papers cochairs, and the rest of the conference committee—Mark Yonge, Chris Chambers, Paul Troughton, and AES UK Secretary Heather Lane, with support from AES Executive Director Roger Furness—worked hard to make *Metadata for Audio* mega-enjoyable and to ensure that attendees experienced a rewarding three days in London. The conference proceedings and CD-ROM can be purchased online at http://www.aes.org/publications/conf.cfm.