

AES 25th INTERNATIONAL CONFERENCE



In the 135-meter-high London Eye you can see as far as 25 miles away, and you have a bird's eye view of such major London sights as St. Paul's Cathedral, Buckingham Palace, the Houses of Parliament, and Big Ben. Chair John Grant and his committee are planning a conference this June 17–19 that will give you a great view of the critically important subject of metadata. *Metadata for Audio* will be held at Church House, the conference center that is just a stone's throw from Westminster Abbey and the Houses of Parliament in central London.

As the means for production and distribution of digital audio proliferate, appropriate metadata tools are needed to facilitate, control, and extend these activities. There has been a great deal of activity in individual organizations to develop metadata tools. However, substantial issues remain to be addressed before the desired goal of global exchange and common understanding can be reached. International standardization, such as the work on MPEG-7 and MPEG-21 may hold some important answers.

This conference seeks to describe the state of the art, identify the issues, and indicate directions for the development of advanced metadata systems, both for consumer distribution and business-to-business. It will bring together media publishers and software designers, media librarians and archivists, database managers and streaming engineers whose operations are increasingly dependent on the success of sophisticated metadata systems.

TUTORIAL DAY

Gerhard Stoll and Russell Mason, papers cochairs, have targeted a number of papers for tutorial presentations on Thursday, June 17 as a good way to offer attendees a thorough introduction to the subject of metadata. Two invited papers in the first morning session, "Metadata, Identities, and Handling Strategies," by Chris Chambers, and "Before There Was Metadata," by Mark Yonge, are introductory papers to set the stage for everything that follows.

The next session, *File Basics*, has three invited papers: "Introduction to MXF and AAF," by Philip DeNier; "XML Primer," by Claude Seyrat; and "Keeping it Simple: BWF and AES31," by John Emmett.

The first session on Thursday afternoon, *Practical Schemes*, starts with an invited paper by Philippa Morrell, "The Role of Registries." The next paper, by researchers from Pompeu Fabra University of Barcelona, will look at a system for managing sound effects. And Richard Wright will present an invited paper on the Dublin Core. The final session on Thursday is a workshop on MPEG-7. This tutorial day is also available as a single-day registration option (see the registration form on page 411).

Photos courtesy British Tourist Office



Metadata for Audio

London, UK

June 17–19, 2004

CONFERENCE DAY 1

On Friday the papers sessions begin with *Frameworks*, which includes an invited paper by Wes Curtis, “P-META: Program Data Exchange in Practice.” This will be followed by the first posters session. And the final morning session will be *Toolkits*, which will include two invited papers: “Digital Media Project,” by R. Nicol, and “MPEG-21: What and Why,” by Jan Bormans and Kate Grant. After lunch there will be the two-part session *Feature Extraction*. The second part of the posters session will also be on Friday afternoon.

On Friday evening there will be an optional (not included in conference registration fee) banquet and guided tour at the historic Houses of Parliament (see photo below). The evening will start with a tour of the debating chambers of the Houses of Commons and Lords. There will be a brief technical talk about the sound-reinforcement system in the Lords Chamber, which uses 84 microphones on motorized winches and has over 400 individually controlled loudspeakers. Afterwards dinner will be served in a room overlooking the River Thames.

CONFERENCE DAY 2

The entire Saturday morning session will be *Broadcast Implementations*, which will include papers on the metadata processes of British, German, and Japanese broadcasters.

Metadata is the “bread and butter” of libraries and archives, so the first afternoon session on Saturday will include papers about projects at the U.S. Library of Congress, Spanish National Radio, and Swedish Radio. The final conference session will be on metadata needed for the online delivery of audio. The calendar, complete program with abstracts, and conference registration form follow on pages 404–411.

And, of course, you should try to schedule an extra day or two to visit one of the world’s great cities. London preserves its magnificent history and at the same time encourages new music, art, literature, and architecture; just what good metadata does for audio. Meet your colleagues there June 17–19 for the AES 25th International Conference, *Metadata for Audio*, it’s going to be absolutely fabulous. Go to www.aes.org now for more details and register online.



AES 25th INTERNATIONAL CONFERENCE

Metadata for Audio

London, UK • June 17–19, 2004

THURSDAY, JUNE 17
TUTORIAL DAY

Tutorial Registration

Session T-1
Introduction

Session T-2
File Basics

Lunch

Session T-3
Practical Schemes

Workshop
MPEG-7

FRIDAY, JUNE 18
CONFERENCE DAY 1

Conference Registration

Session 1
Frameworks

Session 2: **Posters, Part 1**

Session 3
Toolkits

Lunch

Session 4
Feature Extraction, Session A

Session 5: **Posters, Part 2**

Session 6
Feature Extraction, Session B

Banquet (optional)
Houses of Parliament

SATURDAY, JUNE 19
CONFERENCE DAY 2

Session 7
Broadcast Implementations, Session A

Session 8
Broadcast Implementations, Session B

Lunch

Session 9
Libraries and Archives

Session 10
Delivery of Audio

Schedule is subject to change. Check www.aes.org for updates.

AES 25th International Conference Program

Metadata for Audio

2004 June 17–19

London, UK

Technical Sessions

Thursday, June 17

TUTORIALS

SESSION T-1: INTRODUCTION

T1-1 Metadata, Identities, and Handling Strategies—
Chris Chambers, BBC R&D, Tadworth, Surrey, UK
(invited)

With all the potential media material and its associated metadata becoming accessible on IT-based systems, how are systems going to find and associate the elements of any single item? How are the users going to know they have the correct items when assembling audio, video, and information for use within a larger project? This short talk will explore the way areas of our industry are hoping to tackle the problem and some of the standards being introduced to ensure management of this material is possible.

T1-2 Before There Was Metadata—*Mark Yonge* (invited)

Audio has never existed in isolation. There has always been a mass of associated information, both explicit and implicit, to direct, inform, and enhance the use of the audio. In the blithe days before information theory we didn't know it was all metadata. This paper reviews the extent of traditional metadata covering a range of forms. Some of them may be surprising; all of them need to be re-appraised in the light of newer, more formal metadata schemes.

Thursday, June 17

SESSION T-2: FILE BASICS

T2-1 Introduction to MXF and AAF—*Philip DeNier*, BBC R&D, Tadworth, Surrey, UK (invited)

The AAF and MXF file formats provide a means to exchange digital media along with a rich (extendible) set of metadata. This presentation will be a basic introduction into the content of these file formats and will include a description of the metadata scheme used.

T2-2 XML Primer—*Claude Seyrat* (invited)

Most audio professionals have heard of the term "XML" but not many know for sure what it means or have yet had to work with it. This paper sets out what XML is,

what it can do for the user, and various ways that it can be employed in the area of metadata for audio. The contents of this paper form the basis for a number of the papers that appear later in the conference.

T2-3 Keeping it Simple: BWF and AES31—*John Emmett*, Broadcast Project Research Ltd., Teddington, Middlesex, UK (invited)

Digital audio is spreading outward to the furthest reaches of the broadcast chain. Making the best use of the opportunities presented by this demands a standardization procedure that is adaptable to a vast number of past, present, and future digital audio formats and scenarios. In addition, would it not be just great if it cost nothing? This paper will point out the benefits of what we already have and tell a tale of borrowing economical audio technology from many sources.

Thursday, June 17

SESSION T-3: PRACTICAL SCHEMES

T3-1 The Role of Registries—*Philippa Morrell*, Metadata Associates Ltd., London, UK (invited)

Some forms of metadata, especially those that identify objects or classes of objects, form classes of their own that need to be administered centrally in order to avoid the risk of duplication and consequent misidentification. The concept of such a registry is not new; for example, International Standard Book Numbers (ISBN) derive from a central registry that was originally set up in 1970. The registry that ensures that every ethernet-connected device in the world is uniquely identifiable is another example. Formal identifiers and other metadata for use in commercial transactions will increasingly use the services of one or more metadata registries, as this paper will discuss.

T3-2 Sound Effect Taxonomy Management in Production Environments—*Pedro Cano, Markus Koppenberger, Perfecto Herrera, Oscar Celma*, Universitat Pompeu Fabra, Barcelona, Spain

Categories or classification schemes offer ways of navigating and having higher control over the search and retrieval of audio content. The MPEG-7 standard provides description mechanisms and ontology man-

agement tools for multimedia documents. We have implemented a classification scheme for sound effects management inspired by the MPEG-7 standard on top of an existing lexical network, WordNet. WordNet is a semantic network that organizes over 100,000 concepts of the real world with links between them. We show how to extend WordNet with the concepts of the specific domain of sound effects. We review some of the taxonomies to acoustically describe sounds. Mining legacy metadata from sound effects libraries further supplies us with terms. The extended semantic network includes the semantic, perceptual, and sound effects specific terms in an unambiguous way. We show the usefulness of the approach, easing the task for the librarian and providing higher control on the search and retrieval for the user.

T3-3 Dublin Core—*R. Wright, BBC (invited)*

Dublin Core metadata provides card catalog-like definitions for defining the properties of objects for Web-based resource discovery systems. The importance of the Dublin Core is its adoption as a basis for many more elaborate schemes. When the view ahead is obscured by masses of local detail, a firm grasp of the Dublin Core will often reveal the real landscape.

Thursday, June 17

WORKSHOP—MPEG-7

Coordinator: **G. Peeters**, IRCAM, Paris, France
(in association with SAA TC)

Managing Large Sound Databases Using MPEG—*Max Jacob, IRCAM, Paris, France*

Sound databases are widely used for scientific, commercial, and artistic purposes. Nevertheless there is yet no standard way to manage them. This is due to the complexity of describing and indexing audio content and to the variety of purposes a sound database might address. Recently there appeared MPEG-7, a standard for audio/visual content metadata that could be a good starting point. MPEG-7 not only defines a set of description tools but is more generally an open framework hosting specific extensions for specific needs in a common environment. This is crucial since there would be no way to freeze in a monolithic definition all the possible needs of a sound database. This paper outlines how the MPEG-7 framework can be used, how it can be extended, and how all this can fit into an extensible database design, gathering three years of experience during the CUIDADO project at IRCAM.

Integrating Low-Level Metadata in Multimedia Database Management Systems—*Michael Casey, City University, London, UK*

[Abstract Not Available at Press Time]

Tools for Content-Based Retrieval and Transformation of Audio Using MPEG-7: The SPOff and the MDTools—*Emilia Gómez, Oscar Celma, Emilia Gómez, Fabien Gouyon, Perfecto Herrera, Jordi Janer, David García, University Pompeu Fabra, Barcelona, Spain*

In this workshop we will demonstrate three applications for content-based retrieval and transformations of audio recordings. They illustrate diverse aspects of a common framework for music content description and structuring implemented using the MPEG-7 standard. MPEG-7 descriptions can be generated either

manually or automatically and are stored in an XML database. Retrieval services are implemented in the database. A set of musical transformations are defined directly at the level of musically meaningful MPEG-7 descriptors and are automatically mapped onto low-level audio signal transformations. Topics included in the presentation are: (1) Description generation procedure, manual annotation of editorial description: the MDTools, automatic description of audio recordings, the SPOffline; (2) Retrieval functionalities, local retrieval: SPOffline, remote retrieval: Web-based retrieval; and (3) Transformation utilities: the SPOffline.

Using MPEG-7 Audio Low-Level Scalability: A Guided Tour—*Jürgen Herre, Eric Allamanche, Fraunhofer IIS, Ilmenau, Germany*

[Abstract Not Available at Press Time]

Friday, June 18

CONFERENCE DAY 1

SESSION CD-1: FRAMEWORKS

1-1 Data Model for Audio/Video Production—*A. Ebner, IRT, Munich, Germany*

When changing from traditional production systems to IT-based production systems the introduction and usage of metadata is unavoidable. Direct access of the information stored in IT-based systems is not possible. Descriptive and structural metadata are the enablers to have proper access of selected material. Metadata does not focus on descriptive information about the content only. It describes the usage of the material, the structure of a program, handling processes, relevant information, delivery information about properties, and storage of information. The basis to achieve a complete collection of metadata is a detailed analysis of a broadcaster's production processes and usage cases. A logical data model expresses the relationship between the information and is the foundation for implementations that enable a controlled exchange and storage of metadata.

1-2 P-META: Program Data Exchange in Practice—*Wes Curtis, BBC Television, London, UK (invited)*

[Abstract Not Available at Press Time]

Friday, June 18

SESSION CD-2: POSTERS, PART 1

2-1 Low-Complexity Musical Meter Estimation from Polyphonic Music—*Christian Uhle¹, Jan Rohden¹, Markus Cremer¹, Jürgen Herre²*

¹Fraunhofer AEMT, Erlangen, Germany
²Fraunhofer IIS, Ilmenau, Germany

This paper addresses the automated extraction of musical meter from audio signals on three hierarchical levels, namely tempo, tatum, and measure length. The presented approach analyzes consecutive segments of the audio signal equivalent to a few seconds length each, and detects periodicities in the temporal progression of the amplitude envelope in a range between 0.25 Hz and 10 Hz. The tatum period, beat period, and measure length are estimated in a probabilistic manner from the periodicity function. The special advantages of the presented method reside in the ability to track tempo also in music with strong syncopated rhythms, and its computational efficiency.

2-2 Percussion-Related Semantic Descriptors of Music Audio Files—*Perfecto Herrera¹, Vegard Sandvold², Fabien Gouyon¹*¹Universitat Pompeu Fabra, Barcelona, Spain²University of Oslo, Oslo, Sweden

Automatic extraction of semantic music content metadata from polyphonic audio files has traditionally focused on melodic, rhythmic, and harmonic aspects. In the present paper we will present several music content descriptors that are related to percussion instrumentation. The “percussion index” estimates the amount of percussion that can be found in a music audio file and yields a (numerical or categorical) value that represents the amount of percussion detected in the file. A further refinement is the “percussion profile,” which roughly indicates the existing balance between drums and cymbals. We finally present the “percussiveness” descriptor, which represents the overall impulsiveness or abruptness of the percussive events. Data from initial evaluations, both objective (i.e., errors, misses, false alarms) and subjective (usability, usefulness) will also be presented and discussed.

2-3 Tonal Description of Polyphonic Audio for Music Content Processing—*Emilia Gómez, Perfecto Herrera*, Universitat Pompeu Fabra, Barcelona, Spain

The purpose of this paper is to describe a system that automatically extracts metadata from polyphonic audio signals. This metadata describes the tonal aspects of music. We use a set of features to estimate the key of the piece and to represent its tonal structure, but they could also be used to measure the tonal similarity between two songs and to perform some key-based segmentation or establish a tonal structure of a piece.

2-4 Phone-Based Spoken Document Retrieval in Conformance with the MPEG-7 Standard—*Nicolas Moreau, Hyoung-Gook Kim, Thomas Sikora*, Technical University of Berlin, Berlin, Germany

This paper presents a phone-based approach of spoken document retrieval, developed in the framework of the emerging MPEG-7 standard. The audio part of MPEG-7 encloses a SpokenContent tool that provides a standardized description of the content of spoken documents. In the context of MPEG-7, we propose an indexing and retrieval method that uses phonetic information only and a vector space IR model. Experiments are conducted on a database of German spoken documents with ten city name queries. Two phone-based retrieval approaches are presented and combined. The first one is based on the combination of phone N-grams of different lengths used as indexing terms. The other consists of expanding the document representation thanks to the phone confusion probabilities.

2-5 Efficient Features for Musical Instrument Recognition on Solo Performances—*Slim Essid, Gaël Richard, Bertrand David*, GET-Télécom Paris (ENST), Paris, France

Musical instrument recognition is one of the important goals of musical signal indexing. While much effort has already been dedicated to such a task, most studies were based on limited amounts of data that often included only isolated musical notes. In this paper we address musical instrument recognition on real solo performance based on larger training and test sets. A highly efficient set of features is proposed that is obtained from signal cepstrum but also from spectrum low- and higher-order statistical moments describing

signal spectral shape. The use of principal component analysis in conjunction with support vector machine classification yields to nearly perfect recognition accuracy on varied musical solo phrases from ten instruments issued from different instrument families.

Friday, June 18**SESSION CD-3: TOOLKITS****3-1 Digital Media Project**—*R. Nicol*, BT, Ipswich, UK (invited)

[Abstract Not Available at Press Time]

3-2 MPEG-21: What and Why—*Jan Bormans¹, Kate Gran²* (invited)¹IMEC, Leuven, Belgium²Nine Tiles, Cambridge, UK

The MPEG-21 vision is to define a multimedia framework to enable transparent and augmented use of multimedia resources across a wide range of networks and devices used by different communities. The technical report “Vision, Technologies and Strategy” describes the two basic building blocks: the definition of a fundamental unit of distribution and transaction (the digital item) and the concept of users interacting with digital items. The digital items can be considered the “what” of the multimedia framework (e.g., a video collection, a music album), and the users can be considered the “who” of the multimedia framework. MPEG-21 is developing a number of specifications enabling the integration of components and standards to facilitate harmonisation of “technologies” for the creation, modification, management, transport, manipulation, distribution, and consumption of digital items. This paper will explain the relationship of the different MPEG-21 specifications by describing a detailed use-case scenario.

3-3 A 3-D Audio Scene Description Scheme Based on XML—*Guillaume Potard, Ian Burnett*, University of Wollongong, NSW, Australia

An object-oriented schema for describing time-varying 3-D audio scenes is proposed. The creation of this schema was motivated by the fact that current virtual reality description schemes (VRLM, X3D) have only basic 3-D audio description capabilities. In contrast, MPEG-4 AudioBIFs have advanced 3-D audio features but are not designed as a metadata language. MPEG-4 BIFs are particularly targeted as a binary scene description language for scene rendering purposes only. Our proposed 3-D audio scene description schema features state-of-the-art 3-D audio description capabilities while being usable both as a metadata scheme for describing 3-D audio content (for example, 5.1 or Ambisonics B-format) and as a format for scene rendering.

Friday, June 18**SESSION CD-4: FEATURE EXTRACTION, SESSION A****4-1 A System for Harmonic Analysis of Polyphonic Music**—*Claas Derboven, Markus Cremer*, Fraunhofer IIS AEMT, Ilmenau, Germany

A system for harmonic analysis of polyphonic musical signals is presented. The system uses a transform with a nonuniform frequency resolution for the extraction of prominent tonal components and determines the key and the contained chords of a musical input signal with high accuracy. A statistical approach based on the

frequency of occurrence of musical notes for determining the key is described. An algorithmic solution for chord determination is presented with a concise explanation. Finally, a qualitative evaluation of the system's performance is conducted to demonstrate the applicability to real-world audio signals.

4-2 Robust Identification of Time-Scaled Audio—Rolf Bardeli, Frank Kurth, University of Bonn, Bonn, Germany

Automatic identification of audio titles on radio broadcasts is a first step toward automatic annotation of radio programs. Systems designed for the purpose of identification have to deal with a variety of postprocessing potentially imposed on audio material at the radio stations. One of the more difficult techniques to be handled is time-scaling, i.e., the variation of playback speed. In this paper we propose a robust fingerprinting technique designed for the identification of time-scaled audio data. This technique has been applied as a feature extractor to an algebraic indexing technique that has already been successfully applied to the task of audio identification.

4-3 Computing Structural Descriptions of Music through the Identification of Representative Excerpts from Audio Files—Bee Suan Ong, Perfecto Herrera, Universitat Pompeu Fabra, Barcelona, Spain

With the rapid growth of audio databases, many music retrieval applications have employed metadata descriptions to facilitate better handling of huge databases. Music structure creates the uniqueness identity for each musical piece. Therefore, structural description is capable of providing a powerful way of interacting with audio content and serves as a link between low-level description and higher-level descriptions of audio (e.g., audio summarization, audio fingerprinting, etc.). Identification of representative musical excerpts is the primary step toward the goal of generating structural descriptions of audio signals. In this paper we discuss various approaches in identifying representative musical excerpts of music audio signals and propose to classify them into a few categories. Pros and cons of each approach will also be discussed.

Friday, June 18

SESSION CD-5: POSTERS, PART 2

5-1 Toward Describing Perceived Complexity of Songs: Computational Methods and Implementation—Sebastian Streich, Perfecto Herrera, Universitat Pompeu Fabra, Barcelona, Spain

Providing valuable semantic descriptors of multimedia content is a topic of high interest in current research. Such descriptors should merge the two predicates of being useful for retrieval and being automatically extractable from the source. In this paper the semantic descriptor concept of music complexity is introduced. Its benefit for music retrieval and automated music recommendation is addressed. The authors provide a critical review of existing methods and a detailed prospect of new methods for automated music complexity estimation.

5-2 How Efficient Is MPEG-7 for General Sound Recognition?—Hyoung-Gook Kim, Juan José Burred, Thomas Sikora, Technical University Berlin, Berlin, Germany

Our challenge is to analyze/classify video sound track content for indexing purposes. To this end we compare

the performance of MPEG-7 audio spectrum projection (ASP) features based on several basis decomposition algorithms vs. mel-scale frequency cepstrum coefficients (MFCC). For basis decomposition in the feature extraction we evaluate three approaches: principal component analysis (PCA), independent component analysis (ICA), and non-negative matrix factorization (NMF). Audio features are computed from these reduced vectors and are fed into a hidden Markov model (HMM) classifier. We found that established MFCC features yield better performance compared to MPEG-7 ASP in general sound recognition under practical constraints.

5-3 Automatic Optimization of a Musical Similarity Metric Using Similarity Pairs—Thorsten Kastner, Eric Allamanche, Oliver Hellmuth, Christian Ertel, Marion Schalek, Jürgen Herre, Fraunhofer IIS, Ilmenau, Germany

With the growing amount of multimedia data available everywhere and the necessity to provide efficient methods for browsing and indexing this plethora of audio content, automated musical similarity search and retrieval has gained considerable attention in recent years. We present a system which combines a set of perceptual low-level features with appropriate classification schemes for the task of retrieving similar sounding songs in a database. A methodology for analyzing the classification results to avoid time consuming subjective listening tests for an optimum feature selection and combination is shown. It is based on a calculated "similarity index" that reflects the similarity between specifically embedded similarity pairs. The system's performance as well as the usefulness of the analyzing methodology is evaluated through a subjective listening test.

5-4 Automatic Extraction of MPEG-7 Metadata for Audio Using the Media Asset Management System iFinder—Jobst Löffler, Joachim Köhler, Fraunhofer IMK, Sankt Augustin, Germany

This paper describes the MPEG-7 compliant media asset management system iFinder, which provides a set of automatic methods and software tools for media analysis, archiving, and retrieval. The core technology of iFinder comprises several modules for audio and video metadata extraction that are bundled in the iFinderSDK, a commercial product offered to the media industry. The workflow for audio content processing together with pattern recognition methods used will be presented. Of special note, a technique for precise audio-text alignment together with a browser application for synchronized display of retrieval results will be demonstrated. An insight to using MPEG-7 as a standardized metadata format for media asset management will be provided from a practical point of view.

5-5 An Opera Information System Based on MPEG-7—Oscar Celma Herrada, Enric Mieza, Universitat Pompeu Fabra, Barcelona, Spain

We present an implementation of the MPEG-7 standard for a multimedia content description of lyric opera in the context of the European IST project: OpenDrama. The project goals are the definition, development, and integration of a novel platform to author and deliver the rich cross-media digital objects of lyric opera. MPEG-7 has been used in OpenDrama as the base technology for a music information retrieval system. In addition to the MPEG-7 multimedia description scheme, different classification schemes have been proposed to deal with operatic concepts such as musical forms (acts, scenes, frames, introduction, etc.),

musical indications (piano, forte, ritardando, etc.), and genre and creator roles (singers, musicians, production staff, etc.). Moreover, this project has covered the development of an authoring tool for an MPEG-7 standard, namely MDTools, which includes segmentation, classification scheme generation, creation and production, and media information descriptors.

5-6 Morphological Sound Description: Computational Model and Usability Evaluation—*Julien Ricard, Perfecto Herrera, Universitat Pompeu Fabra, Barcelona, Spain*

Sound samples of metadata are usually limited to low-level descriptors and a source label. In the context of sound retrieval only the latter is used as a search criterion, which makes the retrieval of sounds having no identifiable source (abstract sounds) a difficult task. We propose a description framework focusing on intrinsic perceptual sound qualities, based on Schaeffer's research on sound objects that could be used to represent and retrieve abstract sounds and to refine a traditional search by source for nonabstract sounds. We show that some perceptual labels can be automatically extracted with good performance, avoiding the time-consuming manual labeling task, and that the resulting representation is evaluated as useful and usable by a pool of users.

Friday, June 18

SESSION CD-6: FEATURE EXTRACTION, SESSION B

6-1 Drum Pattern-Based Genre Classification from Popular Music—*Christian Uhle, Christian Dittmar, Fraunhofer AEMT, Ilmenau, Germany*

This paper addresses the identification of drum patterns and the classification of their musical genres. The drum patterns are estimated from audio data automatically. This process involves the transcription of percussive unpitched instruments with a method based on independent subspace analysis and a robust estimation of the tatum grid and the musical meter. The rhythmic patterns are identified from pattern histograms, describing the frequency of occurrence of the percussive events. The classification procedure evaluates the meter information, the pattern histogram as well as other high-level rhythmic features derived from the estimated drum pattern.

6-2 Assessing the Relevance of Rhythmic Descriptors in a Musical Genre Classification Task—*Fabien Gouyon¹, Simon Dixon², Elias Pampalk², Gerhard Widmer²*

¹Universitat Pompeu Fabra, Barcelona, Spain
²Austrian Research Institute for AI, Vienna, Austria

Organizing or browsing music collections in a musically meaningful way calls for tagging the data in terms of, e.g., rhythmic, melodic or harmonic aspects, among others. In some cases, such metadata can be extracted automatically from musical files; in others, a trained listener must extract it by hand. In this paper we consider a specific set of rhythmic descriptors for which we provide procedures of automatic extraction from audio signals. Evaluating the relevance of such descriptors is a difficult task that can easily become highly subjective. To avoid this pitfall, we assessed the relevance of these descriptors by measuring their rate of success in genre classification experiments.

6-3 Music Genre Estimation from Low-Level Audio Features—*Oliver Hellmuth, Eric Allamanche, Thorsten Kastner, Ralf Wistorf, Nicolas Lefebvre, Jürgen Herre, Fraunhofer IIS, Ilmenau, Germany*

Despite the subjective nature of associating a certain song or artist with a specific musical genre, this type of characterization is frequently used to provide a convenient way of expressing very coarse information on the basic stylistic and rhythmic elements and/or instrumentation of a song. An audio database that is structured according to different musical genres is a first important step to provide an easy/intuitive access to a large music collection. Thus, a convenient way for indexing large databases by musical genre is desired. This paper describes a system for an automatic genre classification into several musical genres. Different features as well as classification strategies will be evaluated and compared. The system's performance is assessed by means of a subjective listening test.

Saturday, June 19
Conference Day 2

SESSION CD-7: BROADCAST IMPLEMENTATIONS, SESSION A

7-1 Audio Metadata in Radio Broadcasting—*Shigeru Aoki¹, Masahito Kawamor²*
¹TokyoFM Broadcasting, Tokyo, Japan
²NTT, Tokyo, Japan

Generally an audio sequence or program is produced by DAW (Digital Audio Workstation) and delivered as a digital audio file. However, the descriptive data of the audio program, such as the cue sheet of the radio program, is transferred apart from the audio file. The content descriptive data is commonly known as metadata. The most effective method to transfer the audio data and the metadata is to embed those as one digital file that an audio player plays and offer the description of that audio sequence simultaneously. This paper describes the format and scheme of the audio file with metadata.

7-2 Integrated Metadata in the Broadcast Environment—*Joe Bull¹, Kai-Uwe Kaup²*
¹SADiE UK, Cambridgeshire, UK
²VCS Aktiengesellschaft, Bochum, Germany

In a modern broadcast environment, efficient and effective handling of metadata becomes more important every day. Much time and money can be wasted reentering data that is already present in the digital domain. This money could be better spent on program-making. The authors will describe practical examples of how this can be achieved in a real broadcast environment using real products in use or in development.

Saturday, June 19

SESSION CD-8: BROADCAST IMPLEMENTATIONS, SESSION B

8-1 Broadcast Wave and AES Audio in MXF—*Bruce Devlin, Snell & Wilcox*

The SMPTE has established MXF is the new open standard file format for interchange in the broadcast world. One important aspect of the standard is audio mapping. This paper will be a basic tutorial on how MXF and the audio mapping standard work. It will include issues of physically interleaving audio and video as well as adding rich metadata using the MXF data model.

8-2 The Advanced Authoring Format and its Relevance to the Exchange of Audio Editing Decisions—*David McLeish¹, Phil Tudor²*
¹SADiE, Cambridgeshire, UK
²BBC R&D, Tadworth, Surrey, UK

This paper explores how the Advanced Authoring Format (AAF) model, a vehicle for exchanging metadata and media-rich content, can be used to describe audio program compositions so that they can be more seamlessly exchanged between audio editors who work with tools designed by different manufacturers. In addition, the extensibility of the format is discussed as a means of looking at its future potential.

Saturday, June 19

SESSION CD-9: LIBRARIES AND ARCHIVES

- 9-1 Development of a Digital Preservation Program at the Library of Congress**—*Carl Fleischhauer, Samuel Brylawski*, Library of Congress, Washington, DC, USA

This paper will trace the development of a digital preservation program for sound recordings at the Library of Congress. It will outline the Library's use of METS (Metadata Encoding and Transmission Standard); survey the challenges faced in the Library's work to create digital objects for public use, comprised of sound files and images of packaging and accompanying materials; and review the tools and methods utilized to create metadata.

- 9-2 Audio Metadata Used in the Radio Nacional de España Sound Archive Project**—*Miguel Rodeno¹, Jesus Nicolas², Isabel Diaz²*
¹Alcala University, Madrid, Spain
²Radio Nacional de España, Madrid, Spain

The 20th century Spanish sound history has been preserved in digital format and may now be consulted online through the Internet. This a pioneer project in the broadcasting industry around the world finished in December 2002. The archive is considered the most important audio archive in the Spanish language in the world. This paper describes the metadata used in this project. Radio Nacional de España followed the 1997/98 European Broadcasting Union (EBU) standard for the interchange of audio files and their broadcasting: the Broadcast Wave Format (BWF). The voice/word and the different kinds of music (classical, light, international or Spanish) have different types of metadata. Some examples are shown with the detailed metadata.

- 9-3 Integration of Audio Computer Systems and Archives Via the SAM/EBU Dublin Core Standard, Tech.doc 3293**—*Lars Jonsson¹, Gunnar Dah²*
¹Swedish Radio
²KSAD, Norsk Rikskringkasting, Oslo, Sweden

Dublin Core is a well-known metadata initiative from W3C that has been widely spread and used for text and Web pages on the Internet. The Scandinavian SAM-group, with 25 archive specialists and engineers have defined semantic definitions and converted the commonly used Dublin Core initiative for general use within the audio industry. The 15 basic elements of Dublin Core and new subsets have proven to cover most of the tape protocols and database fields existing in broadcast production chain from early capturing over various types of production and all the way to distribution and archiving. This presentation covers some examples of the use of metadata transfer with Dublin Core expressed in XML in Sweden and Norway. It ends in a discussion of the future possibilities of Dublin Core in comparison with other existing metadata initia-

tives in an integrated world of interconnected databases coming into all audio-related companies.

Saturday, June 19

SESSION CD-10: DELIVERY OF AUDIO

- 10-1 Watermarking and Copy Protection by Information Hiding in Soundtracks**—*Tim Jackson, Keith Yates, Francis Li*, Manchester Metropolitan University, Manchester, UK

In this paper digital audio watermarking techniques are reviewed and categorized. Applications of watermarking schemes are discussed, and their capabilities and limitations clarified in the context of audio copyright management and copy protection. Traditional watermarking schemes embed hidden signatures in soundtracks and are found to be effective in ownership authentication and copyright management. Nevertheless, they do not prevent unauthorized copying unless dedicated watermark detectors are added to the recording devices. Purpose-chosen hidden signals are known to interfere with some recording devices, for example, magnetic tape recorders, offering a potential solution to copy protection. It is therefore reasonable to postulate that watermarking techniques could be extended to the general audio copy protection without the resort to dedicated detectors.

- 10-2 Metadata Requirements for Enabling the On-line Music Industry's New Business Models and Pricing Structures**—*Nicolas Sincaglia*, MusicNow Inc., Chicago, IL, USA

The music industry has begun selling and distributing its media assets online. Online music distribution is vastly different from the normal means of media distribution. These untested methods of music sales and distribution require experimentation in order to determine which business models and pricing tiers will most resonate with the consumer. This translates into the need for versatile and robust data models to enable these market trials. Copyright owners and media companies require well-designed data structures to enable them to transmit and receive these complicated sets of business rules. This metadata is an essential part of an overall digital rights management system to control and limit access to the associated media assets.

- 10-3 Audio Meta Data Generation for the Continuous Media Web**—*Claudia Schremmer¹, Steve Cassidy², Silvia Pfeiffer¹*
¹CSIRO, Epping, NSW, Australia
²Macquarie University, Sydney, Australia

The Continuous Media Web (CMWeb) integrates time-continuous media into the searching, linking, and browsing function of the World Wide Web. The file format underlying the CMWeb technology, Annodex, streams the media content multiplexed with metadata in CMML format that contains information relevant to the whole media file (e.g., title, author, language) as well as time-sensitive information (e.g., topics, speakers, time-sensitive hyperlinks). This paper discusses the problem of generating Annodex streams from complex linguistic annotations: annotated recordings collected for use in linguistic research. We are particularly interested in automatically annotated recordings of meetings and teleconferences and see automatically-generated CMML files as one way of viewing such recordings. The paper presents some experiments with generating Annodex files from hand annotated meeting recordings.

AES 25TH INTERNATIONAL CONFERENCE REGISTRATION

2004 June 17–19, London, UK

Registration: Fee includes attendance at the 25th conference including a copy of the *Proceedings of the AES 25th International Conference* as well as lunch and refreshments Thursday through Saturday. There is an optional dinner on Friday evening at the Houses of Parliament. Rooms have been reserved at St. Ermin's Hotel, which is a few

minutes walk from Church House. Delegates will have the opportunity to continue their discussions there in a relaxed atmosphere. There are three price options—£99, £119, and £149—with VAT and breakfast included; these are special rates for the conference. To book a room fill out item 2 below; payment is made to hotel at end of stay.

Please return by mail or fax to:
 Heather Lane, AES, P.O. Box 645, Slough SL1 8BJ, UK
 Fax: +44 1628 667002 Tel: +44 1628 663725 Email: uk@aes.org

Conference registration is also available online at www.aes.org/events/25/registration.html.
 No registration is final until the conference fee has been received.

1 Please print/type all information as you wish it to appear on your badge.

First Name _____ Last Name _____

Affiliation (Company/Institution) _____

Street Address _____

City _____ State _____ Postal Code _____ Country _____

Telephone INT+ _____ Fax INT+ _____

Email _____ AES Membership No. _____

2 Book me a room at St. Ermin's Hotel Number of people _____

£ 99 Standard £ 119 Superior £ 149 Deluxe (Prices are per room, not per person)

Arrival Date _____ Departure Date _____ Number of nights _____

3 **Metadata Conference only (Friday and Saturday, June 18–19)**

<input type="checkbox"/> AES MEMBERS	£ 298.00 plus £52.15 VAT	Total: £ 350.15
<input type="checkbox"/> NONMEMBERS	£ 348.00 plus £60.90 VAT	Total: £ 408.90
<input type="checkbox"/> AUTHORS, STUDENTS	£ 198.00 plus £34.65 VAT	Total: £ 232.65

Tutorial Day only (Thursday, June 17)

<input type="checkbox"/> AES MEMBERS	£ 98.00 plus £17.15 VAT	Total: £ 115.15
<input type="checkbox"/> NONMEMBERS	£ 128.00 plus £22.40 VAT	Total: £ 150.40
<input type="checkbox"/> AUTHORS, STUDENTS	£ 68.00 plus £11.90 VAT	Total: £ 79.90

Tutorial Day and Conference

<input type="checkbox"/> AES MEMBERS	£ 368.00 plus £64.40 VAT	Total: £ 432.40
<input type="checkbox"/> NONMEMBERS	£ 448.00 plus £78.40 VAT	Total: £ 526.40
<input type="checkbox"/> AUTHORS, STUDENTS	£ 248.00 plus £43.40 VAT	Total: £ 291.40
<input type="checkbox"/> Optional Dinner ___number of tickets	£ 38.00 plus £ 6.65 VAT	Total: £ 44.65

4 Payment Modes (check box) Total Amount £ _____

Check (payable to AES Ltd.) Mastercard/Eurocard Visa

Card Number

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Expiration Date

		/		
Month			Year	

Cardholder's Name (print) _____

Signature of Cardholder _____