Technology Trends in Audio Engineering

A report by the AES Technical Council

INTRODUCTION

Technical Committees are centers of technical expertise within the AES. Coordinated by the AES Technical Council, these committees track trends in audio in order to recommend to the Society papers, workshops, tutorials, master classes, standards, projects, publications, conferences, and awards in their fields. The Technical Council serves the role of the CTO for the society. Currently there are 23 such groups of specialists within the council. Each consists of members from diverse backgrounds, countries, companies, and interests. The committees strive to foster wide-ranging points of view and approaches to technology. Please go to: http://www.aes.org/technical/ to learn more about the activities of each committee and to inquire about membership. Membership is open to all AES members as well as those with a professional interest in each field.

Technical Committee meetings and informal discussions held during regular conventions serve to identify the most current and upcoming issues in the specific technical domains concerning our Society. The TC meetings are open to all convention registrants. With the addition of an internet-based Virtual Office, committee members can conduct business at any time and from any place in the world.

One of the functions of the Technical Council and its committees is to track new, important research and technology trends in audio and report them to the Board of Governors and the Society's membership. This information helps the governing bodies of the AES to focus on items of high priority. Supplying this information puts our technical expertise to a greater use for the Society. In the following pages you will find an edited compilation of the reports recently provided by many of the Technical Committees

Francis Rumsey
Chair, AES Technical Council
Bob Schulein, Jürgen Herre, Michael Kelly
Vice Chairs

ARCHIVING, RESTORATION, AND DIGITAL LIBRARIES

David Ackerman, Chair
Chris Lacinak, Vice Chair

Practical observations

Broadcast Wave File (BWF) format has become the de facto standard for preservation of audio content within the field, as has a digital audio resolution of 24 bit/96 kHz. Time-based metadata is also of particular interest, including time-stamped descriptive metadata and closed captions. Manufacturers have begun to enable preservation activities through additional metadata capabilities and support for open formats.

Sound for moving image is somewhat in limbo, currently being grouped with moving image preservation for the most part. Preservation of sound for moving image is a current focus for future attention of this committee. Moving image and sound preservation graduate programs are emerging throughout the world to support those who oversee and manage moving image and sound archives. This is acknowledgment of the differing skill set from traditional paper and still image archivists.

IT and programming skills are an ever-growing need in the fulfilment of preservation. This is an emerging required understanding / skill for audio engineers. Requirements and specifications for digital repositories serving preservation and access roles are currently in development.

Selected significant projects and initiatives

The 131st AES Convention featured an archiving track that was well attended. We believe archiving will continue to grow as an area of interest to AES members.

In addition, the Technical Committee on Audio Recording and Mastering Systems has completed a study on the persistence and interoperability of metadata in wav files, while Indiana University published "Meeting the challenge of media preservation; strategies and solutions."

The following standards activities recently took place:

AES60-2011 AES Standard for audio metadata—Core audio metadata was published September 22, 2011.

AES57-2011 AES standard for audio metadata—Audio object structures for preservation and restoration was published September 21, 2011

AES SC-07-01 Working Group on audio metadata was formed this October. This group continues the work to complete AES-

X98C, metadata for process history of audio objects.

AES SC-03 was retired this October.

The Federal Agencies Audio Visual Digitization Working Group (digitizationguidelines.gov) is investigating audio system evaluation tools for evaluating the performance of analog to digital converters and for detecting interstitial errors.

The Indiana University Archives of Traditional Music (ATM) and the Archive of World Music (AWM) at Harvard University have received a grant from the National Endowment for the Humanities to undertake a joint technical archiving project, a collaborative research and development initiative with tangible end results that will create best practices and test emerging standards for digital preservation of archival audio. This is known as Sound Directions.

The National Recording Preservation Board, mandated by the National Recording Preservation Act of 2000, is an advisory group bringing together a number of professional organizations and expert individuals concerned with the preservation of recorded sound. The group has published a report from the engineers' roundtable (CLIR).

The National Digital Information Infrastructure and Preservation Program (NDIIPP) has the mission to develop a national strategy to collect, archive, and preserve the burgeoning amounts of digital content, especially materials that are created only in digital formats, for current and future generations.

Presto Center is a European effort to push the limits of current technology beyond the state of the art, bringing together industry, research institutes, and stakeholders to provide products and services for bringing effective automated preservation and access to Europe's diverse audiovisual collections.

AUDIO FOR GAMES

Michael Kelly and Steve Martz, Chairs
Kazutaka Someya, Vice Chair

Emerging trends in audio for games are driven by continuing advances in game technology and the diversity of devices and operating systems that are now considered gaming devices. Trends are summarized under the headings below.

A general move from hardware to software processing

Audio DSP is now performed in software on CPUs or programmable DSP processors. Even on lower-power platforms there is a move away from dedicated audio chips and memory although exceptions still exist.

Game platforms are diversifying

Console platforms are very dominant in large budget titles and a lot of memory and DSP is leveraged for audio on these platforms. Consoles remain a major driver in game-audio trends and games often target high-end consumer playback environments. Portable platforms, particularly iOS and Android devices, now also account for a large portion of gameplay and present new constraints, development approaches, and creative styles. Production methodologies for console and mobile gaming will increasingly merge as handheld devices become more powerful. More recently, cloud gaming is demonstrating itself as a viable platform and offers new challenges including potential latency and network delivery issues.

Peripherals and interaction

Social gaming and new platforms offer new ways to interact with games using handheld devices (e.g., Wii Remote, PlayStation Move Controller), touch screens (e.g., iOS and Android devices) and non-contact technology (e.g., Microsoft Kinect, PlayStation Eye). These are able to track player position or gestures and are beginning to find useful applications in game-audio. 3-D video is yet to demonstrate a new counterpart in audio.

Spatial audio

Console games are largely geared around 5.1 and 7.1 playback or legacy formats. Some commercial games (e.g., Race Driver: Grid) are now making use of Ambisonics. Portable platforms are generally targeted at headphone playback or device speaker playback, although many tablets are equipped with other methods such as HDMI outputs. There is general trend toward scalability and adaptation to the consumer's configuration, particularly as the line between console and portable platforms becomes blurred. The driver for spatial audio formats largely comes from outside the games industry and future conventions include features such as height-channels to augment current multichannel setups.

Audio input

Speech input is now used in a number of games and devices for character control or player-to-player communication. Speech analysis and processing is a key research area in game-audio. Analysis of singing and research in this area has been applied in a number of leading console game titles. Rhythm based games (e.g., Rock Band, Guitar Hero) make use of varying degrees of instrument-style peripherals such as guitar controllers, piano keyboards, virtual drum kits; as well as motion controllers and touch screens. New technologies, such as those used in games like Rocksmith, permit the use of real instruments as game controllers.

DSP plugins and codecs

A move into software has made it possible for developers to write their own DSP plugins for use in games. There has been an increase in third-party companies providing DSP algorithms for licensing by game developers. Solutions often involve platform-specific optimized codecs and DSP for use in-game as well PC versions in the form of VST plugins or similar for authoring. There has been a growth in use of algorithms such as convolution reverb and efforts to further R&D in improving audio DSP for use in game. There has also a strong trend toward returning to synthesized sound in-game; this is partially driven by resource requirements of portable platforms, but also by the potential flexibility of synthesized sound as new R&D can provide improved quality for appropriate sounds. As well as low-level DSP, higher level systems such as intelligent or automatic mixing technologies are being used in games like the Battlefield series.

Tools and workflow

A number of studios now have extremely sophisticated tools for game audio content authoring, either developed in-house or licensed as middleware. Tools generally remain specific to the game domain. There are an increasing number of attempts though standards groups like the IASIG to increase interoperability between linear audio tools and game-tools.

Education and standards

Standards activity continues in the games industry and becomes more relevant as the industry matures. Current standards activity includes: interoperable file formats, digital audio workstation design, and loudness levels in games. There has been recent growth in the number of educational institutions that offer game audio courses and interest from academia continues to grow. This is an important step as informed game audio programmers are still in short supply. The IASIG recently introduced game audio curriculum guidelines for interested institutions. Research from academia is also directly impacting game development and many titles feature the results of collaboration between academia and industry.

AUDIO FOR TELECOMMUNICATIONS

Bob Zurek, Chair
Antti Kelloniemi, Vice Chair

The trend in mobile telecommunications has been toward moving advanced features down in price point to feature phones and using the more advanced mobile devices as a personal computing and multi-media capture and playback devices. The typical feature phone today exhibits all of the characteristics of a top of the line device of a few years ago with both private mode and hands free audio, multiple microphones with advanced noise reduction capabilities, and Bluetooth allowing a low end feature phone to serve as the center of a personal communications network.

Wideband audio communications has been rolled out in many countries over both cellular and wireless VOIP (voice over internet protocol) doubling the audio bandwidth used in speech communications. Multiple VOIP clients are available for download on the major mobile operating systems and many devices come with at least one VOIP client preinstalled.

The last few years have shown smartphones and tablet devices becoming a larger percentage of the total mobile telecommunications devices. They are no longer the niche devices of the mid to late part of the last decade. The move to common operating systems with thousands of applications allows the user to customize their device in ways not possible a few years ago. The downloadable application environments of the major mobile operating systems have allowed different users to take the same hardware and customize it into very diverse devices to suit their needs, from business oriented devices, to media and gaming devices, even as far as using the device as a configurable piece of test equipment.

The integration of sensing capabilities such as accelerometers, gyroscopes, light and infrared sensors into devices have allowed not only manufacturers but also creators of applications the ability to create more natural human interfaces to the device and have allowed the device to more accurately detect the environment that it is in. This allows the device to adapt its operation, to best function in any environment whether the device is being used for multimedia playback, communications, or computing. Voice control of communications devices has progressed to the point where networked voice recognition allows the use of natural language with larger vocabularies than previously possible on a standalone device.

Many people have replaced several individual pieces of mobile electronics with their portable communication device over the last few years. Integration of high quality optics has led to the replacement of still and video capture devices for some. Current devices are capable of both multi-megapixel still photography and HD video capture. Some of the devices feature multichannel audio capture capabilities. The combination of GPS and network connectivity has allowed the portable communications devices to become personal navigation devices with nearly continuous map updates and real time traffic information. The enhanced processing capabilities of separate application processors coupled with the over the air download of applications have led to the use of portable communications devices for office productivity, multimedia playback, and authoring, as well as gaming all in a single device.

Current 3G and 4G data rates allow the mobile devices to operate with bandwidths comparable to home-based high speed internet. This has led to the use of wireless devices as wi-fi hubs for a network of devices requiring internet access such as personal computers, gaming systems, automobiles, and televisions.

Many of the advances in handsets of a few years ago have migrated to the edge of the personal network allowing headphones, headsets, and car-kits to achieve handset levels of uplink voice quality. Consumers can upgrade the call quality of older devices by adding new Bluetooth headsets or car kits that contain many of the same noise adaptive algorithms found in much newer devices. This includes both noise adaptive downlink and advanced noise and echo suppression in the uplink signal.

Following the move toward using the mobile device as a user's main device for communication, computing, and media playback has led the creation of a number of multimedia docks, computing docks, and accessories for the devices. In many cases the portable communication device can serve as the hub for the home multimedia system, when paired to or placed in docking systems connected to the home audio video system. It is not uncommon for smartphones and tablets to have HDMI output for media playback on HDMI compatible monitors or sound systems. The creation of Bluetooth mice and keyboards, and laptop docks often in conjunction with HDMI video output has allowed the user to quickly and effortlessly transition from using the communications device as a portable phone to a home computer.

Software updating of not only the applications but also the operating systems allow for the devices to grow in capability after purchase much as personal computers have in the past. No longer is a customer forced to live with the limitations that a device is shipped with for the life of the device or service provider contract. As new features are developed and integrated into operating systems, as long as the hardware still supports the new functionality, a user of a year-old device can update to many of the features being released in the latest devices.

Over the next few years, the rapid growth in capabilities of portable communication devices tied with ever-expanding application environments will allow portable communications devices to evolve into tools unimaginable a few short years ago.

AUDIO FORENSICS

Jeff M. Smith, Chair
Christopher Peltier, Vice Chair
Eddy Bøgh Brixen, Vice Chair

Enhancement

The enhancement of forensic audio recordings remains the most common task for forensic audio practitioners. The goal of forensic audio enhancement is to increase intelligibility of voice information or improve the signal to noise ratio of a target signal by reducing the effects or interferences that mask it. Many tools are available through various software developers with the most common being noise reduction—either adaptive or linear. Difficulties in this area are caused by lossy data compression common to small digital recorders, data compression, and bandwidth limited signals in telecommunications, and non-ideal recording environments common to surveillance and security. One growing area of research is the assessment of speech intelligibility with multiple papers presented on the topic at the AES 39th Conference on Audio Forensics in 2010.

Authentication

The majority of audio media presented to the forensic examiner are digital recordings on optical disc, HDD, flash memory, and solid-state recorders. However, the analysis of analog tape cassettes and microcassettes is still required of examiners. In the area of forensic media authentication, digitally recorded audio files may be subject to various kinds of manipulation that are harder to detect than those in the analog domain. This leaves the forensic audio examiner with new challenges regarding the authentication of these recordings. Many new techniques have been developed in recent years for use in these analyses. These techniques continue to be published and presented through the AES *Journal* and proceedings of AES Conferences and Conventions. Among these techniques is the analysis of the Electric Network Frequency component (ENF) of a recording. If present, the remains of the ENF may be compared to a database of ENF from the same grid to authenticate the date and time the recording was made. In addition to automatic database comparison, it is possible to learn several other things from ENF analysis including whether portions of the recording were removed, if an audio recording was digitized multiple times, how the recorder was powered, and more.

Recent developments in digital audio authentication also include the Compression Level Analysis of an audio recording to determine if an uncompressed file had been previously subject to data compression or if the compression level present is consistent with an authentic recording. Also, a technique for determining the presence of butt-splice edits has been presented. In the digital domain, as in the analog, auditory, and spectral acoustic analysis continues to be necessary. However, it is also clear that analysis of the digital data that makes up a recorded audio file including its header and file structure must be exploited to ascertain a digital recording's authenticity.

Speech and speaker analysis

The analysis of speech and speakers present on audio recordings is a large domain that intersects many industries including forensics and security. The analysis of speakers present in recordings to ascertain identity continues to be a common request of forensic audio examiners. However, "identifying" persons in a 1:1 comparison is not supported within the scientific community that favors "recognition" of persons based on extracted features relative to a background model representing a population of speakers. Automatic systems based on cepstral coefficients, Gaussian Mixture Modeling, and likelihood ratios employ robust and validated techniques for speaker recognition. This quantitative approach better measures and takes into account intra- and inter-speaker variation. When used in a forensic environment where trained examiners base conclusions on likelihood ratios, this technique is valued greatly over other qualitative analyses.

The capability of a system to process multitudes of audio signals and sort them based on language, topic, speakers present, and acoustic environment continues to progress with many new advances. An interesting area of research and its application in audio forensics is Computational Auditory Scene Analysis (CASA). This field of audio processing is interested in developing machine systems that perform automatic signal separation using principles derived from the human auditory system's perceptual abilities to understand audio scenes. CASA systems have already proven very useful as preprocessors for automatic speech recognition systems and in hearing aids. New areas of study include their use in audio forensics. Also, automatic speaker segmentation based on extracted spectral features and statistical modeling can help automated systems tasked with speech and speaker recognition.

Other considerations

Since the fundamental aspect of forensic audio is its application to law with the litigation process benefitting from audio enhancement and analysis, it is important for the practitioner working with forensic audio to be aware of this process and the need for proper evidence handling and laboratory procedures. As digital audio proliferates so to have the identification of proper practices for imaging media, hashing file duplicates, and recovering and/or repairing corrupt or carved files.

Additionally, it is not only common for forensic audio to be played in a courtroom but for typed transcripts of recorded conversations to be prepared for the individuals involved in a case; the lawyers, judge(s), and/or jury. Specific to these needs, there are developments in addressing the inherent bias present in the human preparation of these transcripts. Also, the forensic audio practitioner must be aware of the audio samples being presented taking into consideration courtroom acoustics, psychoacoustics, and the hearing abilities of these individuals.

AES activities

Numerous papers on audio forensics appear in the *Journal* of AES and are presented at AES conventions each year. Additionally, there have been three AES conferences on audio forensics since 2005 (AES 26th, 33rd, and 39th) and the next will be in Denver, CO in 2012. Additionally, regular workshops and tutorials appear at AES conventions. At the AES 130th Convention in London there was a tutorial on forensic audio enhancement, and at the AES 131st Convention in New York there was a workshop on forensic audio authentication.

AUDIO RECORDING AND MASTERING SYSTEMS

Kimio Hamasaki, Chair
Toru Kamekawa and Andres Mayo, Vice Chairs

The growth of multichannel audio recording and production is the most remarkable trend in the audio recording and mastering systems. Recording and mastering using high resolution audio technology is also a notable trend in this area.

While 5.1 multichannel sound is widely applied in audio recording and mastering, audio recording and mastering using advanced multichannel sound formats such as 7.1, 9.1, and more channels have been increasing. Higher sampling frequencies such as 96 kHz and 192 kHz are also applied in audio recording and mastering. Most recording systems can now work at these higher sampling rates, including in some cases the very high rate used by DSD (Direct Stream Digital) systems. DXD (Digital eXtreme Definition), which samples multi-bit PCM at 352.8 kHz, is a new trend for digital recording. A-to-D converters and D-to-A converters for DXD are available, and some DAWs can record and edit DXD.

The digital audio workstation (DAW) is the principal tool for editing and mixing. Mixing consoles are becoming an interface for the DAW. Physical control surfaces for mixing are sometimes not used, but instead a virtual control surface on a PC display is often used for recording and mastering. DAWs use hard disks for storage, and music recording and mastering studios also intend to use server-based storage systems for recordings. Network attached storage (NAS) is widely used for audio recording. While removable hard disks had been widely used for audio recording and mastering, there is still no internationally standardized removable hard disk drive.

MADI (Multichannel Audio Digital Interface) has been gaining popularity in recording systems because multichannel sound recordings need many channels compared with 2-channel stereo recording. Stage boxes equipped with multichannel microphone preamps and A-to-D converters are now available with MADI output. Use of digital microphones according to the AES 42 standard is also gradually expanding in audio recording and mixing consoles and stage boxes equipped with AES 42 I/O are now available. IP networking is very often used in audio recording and mastering. Growth of IP networking, especially considering the increase of data transfer rates, is essential for the improvement of recording and mastering systems.

It is common to use DSP (digital signal processing) in recording and mastering systems. A new trend can be seen in the application of FPGAs (field programmable gate arrays) instead of DSP, and DAWs working on FPGA are already available. A remarkable trend in mastering systems is the development of new plug-in audio processing software for mixing and mastering. DAWs equipped with plug-in audio processing software are widely used for audio production and can be purchased quite inexpensively. The availability of such DAWs has been changing the nature of music productions.

AUTOMOTIVE AUDIO

Richard Stroud, Chair
Tim Nind, Vice Chair

Vehicles with built-in internet capability (via 3G, etc.) could present numerous music and talk selections at higher quality than most other data-reduced sources. At least one OEM is working on personal audio to allow people to have the same data and source material that they have at home in the car. Connectivity may be based on the user's mobile phone. Some OEM's are considering using a dedicated server to control quality.

There is an interest in providing sounds for very quiet cars such as electric vehicles. These include "engine start" and "engine running" sounds for inside the vehicle and pedestrian safety sounds for outside the vehicle.

Hard disk drives are now used in premium audio systems. These disks tend to be smaller than state-of-art home disk drives because of vibration requirements (40 to 80 Gbyte drives are becoming available, and larger drives are expected soon). Disk drive usages include navigation data and music. Systems allow storing of many CDs from on-board readers and music from available MP3 sources via the typically included USB connection. SSDs (solid state devices) will replace hard drives as preferred storage when cost permits. Premium receivers are beginning to appear that do not include CD players. Increasingly larger USB drives are becoming a primary music storage medium, along with Bluetooth-connected cell phones with their music libraries. Download of MP3 files into vehicles by home-based RF (radio frequency) links has been introduced.

Objective measurement is still battling subjective listening tests as a final authority for OEMs. SPL vs. distortion measurements are quite good now, and directionally correct frequency response measurements are improving. Spatial measurement capability is being developed and evaluated.

The trend toward higher performance audio systems is in direct conflict with recent trends of cost and weight reduction of components in automobiles. Increased application of neodymium magnets may help here. Neodymium magnet speakers, once attractive as an affordable means of mass reduction, have recently become much more expensive due to neodymium cost increases. Some reports indicate increases of as much as eight times their former prices. Vendors of smaller speakers were offering neodymium magnet speakers at prices similar to those of ferrite magnet speakers but are struggling to do so at present. Having a strong set of specifications will insure that sensitivity, X_{max} and other parameters are maintained in these speakers. Planar style speakers are now found in vehicles. These are not totally flat, but have profiles of 10 mm or less. Some examples have shown very low sensitivity.

HD radio components are now for sale. AM HD radio offers much higher fidelity and FM HD offers additional program sources. Because of the fidelity difference on AM, rapid switching in fringe areas must be carefully managed.

There are an increasing number of center speakers appearing in prestige class automotive system designs. Speakers have also appeared in the tops of front seats. "Surround sound" is becoming mandatory in high-end automotive systems even when the source is limited to two channels (so this is implemented using upmix algorithms). Some listeners sense that some surround systems provide limited envelopment on both stereo and much "surround" source material.

There is almost universal branding of audio systems in luxury cars, and newer brands are emerging. The maximum number of speakers used in luxury vehicle systems seems to be leveling out at 18±2. Aftermarket audio now represents a very small part of the automotive audio market. There are still parts of the world where 5.1 and high-level premium audio are not featured in most vehicles' audio line-ups. These systems can perhaps take advantage of inexpensive, powerful audio DSP systems to improve performance. Rear seat audio performance may be important in China and other countries, as some who can afford automobiles can also afford drivers.

Voice recognition systems for telephone and navigation functions are becoming more sophisticated and enjoy wider application. Automatic equalization is being offered for audio system tuning. Use of such automatic systems can significantly speed the tuning process but may not be ready to completely replace tuning for on-road performance by trained listeners. Active noise cancellation by the audio system is being used for exhaust drone under condition of cylinder deactivation. Active road noise bass and/or level compensation now enjoy a widespread market presence. Basic versions are available in many OEM head units while some high-end premium systems have more sophisticated implementations. Simple systems use the speedometer signal to apply predefined loudness curves. Others use microphones to measure the current cabin noise, after separating the music, allowing more targeted equalization or bass/level compression to be applied.

Switching audio is now commonly seen in automotive amplifiers. Switching audio costs are becoming comparable with older AB amplifiers, as the heat sink requirement is minimized. Important for electric vehicles is the low current draw under all audio power output conditions.

CODING OF AUDIO SIGNALS

Jürgen Herre and Schuyler Quackenbush, Chairs

Overview

Audio coding has emerged as a critical technology in numerous audio applications. In particular, it is a key component of mobile multimedia applications in the consumer market. Examples include wireless audio broadcast, internet radio and streaming music, music download, storage and playback, mobile audio recording, and Internet-based teleconferencing. Example platforms include digital audio broadcast radio receivers, portable music players, mobile phones, and personal computers. From this, a variety of implications and trends can be discerned.

Digital distribution of content is offered to the consumer in many formats with varying quality / bitrate trade-off, depending on application context. This ranges from very compact formats (e.g., MPEG HE-AACv2 and MPEG USAC) for wireless mobile distribution to perceptually transparent, scalable-to-lossless and lossless formats for regular IP-based distribution (e.g., MPEG AAC, HD-AAC and ALS).

The frontiers of compression have been pushed further, allowing carriage of full-bandwidth signals at very low bit rates to the point where recent coding systems are considered appropriate for some broadcasting applications, particularly relatively expensive wireless communication channels such as satellite or cellular channels. While such technology predominantly makes use of parametric approaches (at least in part) to achieve highest possible quality at lowest bit rates, they are typically not designed to deliver "transparent" audio quality (i.e., that original and encoded/decoded audio signal cannot be perceptually distinguished even under most rigorous circumstances). Nevertheless, "entertainment quality" services over wireless channels have been very successful. Examples of audio coding that facilitates these new markets include MPEG HE-AACv2 and MPEG USAC.

Transform-based audio coding schemes have been exploited to their full potential (quality vs. bitrate). As such, new paradigms will be exploited to gain further compression efficiency.

For broadcast-only applications where delay is not a constraint, there is the possibility to gain further compression efficiency by exploiting large algorithmic delays or even multi-pass algorithms in the case of "off-line" audio coding.

The role of higher-level psychoacoustics and perception is becoming increasingly important in audio coding. Detection of auditory objects in an audio stream, separation into auditory (as opposed to acoustic) objects, and storage and manipulation as auditory objects is beginning to play a role. This will be an important and ongoing area of research.

Hybrid and parametric coding

There is a consistent trend toward hybrid coding techniques that employ parametric modeling to represent aspects of a signal, where the parametric coding techniques are typically motivated by aspects of human perception. The core of most successful audio coders is still largely based on a classic filterbank based coding paradigm, in which the quantization noise is shaped in the time/frequency domain to exploit (primarily) simultaneous masking in the human auditory system. However, the recent success of parametric extensions to the core audio codec, in both market deployment and standardization, illustrates this tendency as follows.

Audio bandwidth extension technology substitutes the explicit transmission of the signal's high-frequency part (e.g., by sending quantized spectral coefficients) by a parametric synthesis of high-frequency spectrum at the decoder side based on the transmitted low frequency part and some parametric side information that captures the most relevant aspects of the original high frequency spectrum. This exploits the lower perceptual acuity in the high-frequency region of the human auditory system. An example is MPEG HE-AAC.

Parametric stereo techniques enable rendering of several output channels at very low bit rates. Instead of a full transmission of all channel signals, the stereo / multichannel sound image is re-synthesized at the decoder side based on a transmitted downmix signal and parametric side information that describes the perceptual properties (cues) of the original stereo / multichannel sound scene. Examples are MPEG Parametric Stereo (for coding of two channels) and MPEG Surround (for full surround representation).

Parametric coding of audio object signals provides, similarly to parametric coding of multichannel audio, a very compact representation of a scene consisting of several audio objects (e.g., music instruments, talkers, etc.). Rather than transmitting discrete object signals, the (downmixed) scene is transmitted, plus parametric side information describing the properties of the individual objects. At the decoder side, the scene can be modified by the user according to his/her preference, e.g., the level of a particular object can be attenuated or boosted. A recent example for such a technology is MPEG Spatial Audio Object Coding (SAOC).

There has been significant progress in the challenge of developing a truly universal coder that can deliver state of the art performance for all kinds of input signals, including music and speech, that has been achieved. Hybrid coders, such as MPEG USAC (Unified Speech and Audio Coding), have a structure combining elements from the speech and the audio coding architectures and, over a wide range of bit rates, perform better than coders designed for only speech or only audio.

Implications for technology and consumer applications

Solid-state and hard drive-based storage for audio has become extremely inexpensive and consumer internet connection speeds reach into the megabits per second range. When such resources are available, music streaming, download, and storage applications no longer require state of the art audio compression. Instead, what is occurring in the marketplace is that consumers are operating well-known perceptual coders at higher bit rates (lower compression) to achieve "perceptually transparent" compression of music, since the additional increment in resources required for such operating points is relatively inexpensive. For example, consumers are opting to use MPEG Layer III (MP3) or MPEG AAC at rates of 256 kb/s or higher to code their music libraries for their portable music players.

Processor speed has continued to increase at a tremendous pace. Even with the low-power restrictions imposed by battery powered portable devices, the quantity of CPU cycles potentially available for audio processing is large. Present audio coders work in a fraction of available CPU capacity, even for multichannel coding, and new research may be needed to discover how to use the additional CPU cycles and memory space. Some possibilities are improved psychoacoustic models and sophisticated acoustic scene analysis. Seen overall, the research in audio coding is moving to the extremes, both toward lowest bit rates (very lossy compression using parametric coding extensions) and highest bit rates (noiseless/lossless coding for high resolution audio at high sampling rates/resolutions), as well as the more complex high-level processing (scene analysis and sound field synthesis of various sorts).

Audio coding has successfully entered the world of telecommunication, providing low-delay high-quality codecs that enable natural sound for teleconferencing and video-conferencing. Such codecs deliver full bandwidth and high quality, not only for speech material but also for any type of music and environmental sound, enabling applications such as teleteaching for music. They support spatial reproduction of sound (stereo or even surround), which can greatly increase the ease of communication in conferences between several partners.

There is considerable research activity exploring audio presentation that is more immersive than the pervasive consumer 5.1 channel audio systems. One might apply the label of "3-D Audio" to such explorations, since their common thread is the use of many loudspeakers positioned around, above, and below the listener. This might range from proposed 22.2 channel systems for the consumer to tens or hundreds of loudspeakers for research in, e.g., wave field synthesis. Of great interest is exploring the impact of loudspeakers positioned above or below the horizontal plane of the typical 5.1 channel system. When systems with a large number of loudspeakers are considered, efficient coding of the audio speaker signals is of paramount importance. In addition, a flexible rendering method that permits high-quality playback on a wide range of conceivable consumer loudspeaker arrangements would be very desirable. It may be that audio coding and rendering to arbitrary loudspeaker setups can be realized in a unified algorithm. This will be an interesting trend to watch.

Finally, after quite some time, the "digital deadlock" regarding the legitimate commercial dissemination of authorized digital audio content has been successfully resolved, and the business models of the music industry have embraced the Internet. Besides a number of (mostly legal) sources of audio (and audio-visual) content with very limited audio quality and free access, several successful major distribution platforms exist now for the electronic distribution of audio. These download stores offer digital audio content in a variety of formats, quality levels and protection levels.

FIBER OPTICS FOR AUDIO

Ronald G. Ajemian, Chair (USA)
Werner Bachmann, Chair (Europe)

It is clear that there are new current and future trends in the area of fiber optics for audio. It has been hard to ignore that more and more companies are deploying fiber optics in their audio/video systems. One can witness this especially in the broadcast field of audio/video. In the current economy where jobs are diminishing, there is growth for expertise with using fiber optic-based audio/video systems. New start-up companies come to the AES Convention every year.

In the future, copper based systems will be inadequate to drive the demands for higher bit rates and bandwidth. It is clear from just the telecommunication and broadcast companies that everything is becoming more integrated. Optical fiber cables can carry multiple signals (audio, video, clock sync/time codes, control data, etc.) all over a single strand of fiber or two or more if necessary. The proof is in the application that has been proven to eliminate common noise, radio-frequency interference, electromagnetic interference, and mains hum.

Other trends include the use of fiber optic snakes, links, networks and switchers, cables and connectors, microphone preamplifiers, and feeds for stage/theater live sound. Fiber over Cat 5 or Cat 6 is an option, and fiber used in MADI. It is likely that fiber optics will affect every sector of audio/video and will eventually be ubiquitous.

HEARING AND HEARING LOSS PREVENTION

Robert Schulein, Chair
Michael Santucci and Jan Voetmann, Vice Chairs

Introduction

The AESTC on Hearing and Hearing Loss Prevention was established in 2005 with five initial goals focused on informing the membership as to important aspects of the hearing process and issues related to hearing loss, so as to promote engineering-based solutions to improve hearing and reduce hearing loss. Its aims include the following: raising AES member awareness of the normal and abnormal functions of the hearing process; raising AES member awareness of the risk and consequences of hearing loss resulting from excessive sound exposure; coordinating and providing technical guidance for the AES-supported hearing testing and consultation programs at U.S. and European conventions; facilitating the maintenance and refinement of a database of audiometric test results and exposure information on AES members; forging a cooperative union between AES members, audio equipment manufacturers, hearing instrument manufacturers, and the hearing conservation community for purposes of developing strategies, technologies, and tools to reduce and prevent hearing loss.

Measurement and diagnosis

Current technology in the field of audiology allows for the primary measurement of hearing loss by means of minimum sound pressure level audibility vs. frequency producing an audiogram record. Such a record is used to define hearing loss in dB vs. frequency. The industry also uses measurement of speech intelligibility masked by varying levels of speech noise. Such measurements allow individuals to compare their speech intelligibility signal-to-noise ratio performance to the normal population. Other tests are commonly used as well for diagnosis as to the cause of a given hearing loss and as a basis for treatment.

Within the past ten years, new tests have evolved for diagnosing the behavior of the cochlea by means of acoustical stimulation of hair cells and sensing their resulting motion. Minute sounds produced by such motions are referred to as otoacoustic emissions. Measurement systems developed to detect and record such emissions work by means of distortion product detection resulting from two-tone stimulations as well as hair cell transients produced from pulse-like stimulations. Test equipment designs for such measurements are now in common use for screening newborn children. Additional research is being conducted directed at using such test methods to detect early stages of hearing loss not yet detectable by hearing-threshold measurements. The committee is currently working to establish a cooperative relationship between researchers in this field and AES members, who will serve as evaluation subjects.

Emerging treatments and technology

Currently there is no known cure for what is referred to as sensorineural hearing loss, in that irreparable damage has been done to the hearing mechanism. Such loss is commonly associated with aging and prolonged exposure to loud sounds, although it is well established that all individuals are not affected to the same degree. Considerable research is ongoing with the purpose of devising therapies leading to the activation of cochlear stem cells in the inner ear to regenerate new hair cells. There are, however, drug therapies being introduced in oral form to prevent or reduce damage to the cilia portion of hair cells in cases where standard protection is not enough, such as in military situations. We are beginning to see the emergence of otoprotectant drug therapies, now in clinical trials that show signs of reducing temporary threshold shift and tinnitus from short term high sound pressure levels. New stem cell therapies are also being developed with goals of regenerating damaged hair cells.

Hearing instruments are the only proven method by which sensorineural hearing loss is treated. In general the task of a hearing instrument is to use signal processing and electroacoustical means to compress the dynamic range of sounds in the real world to the now limited audible dynamic range of an impaired person. This requires the implementation of level-dependent compression circuits to selectively amplify low-level sounds and power amplification and high-performance microphone and receiver transducers fitted into miniature packages. Such circuitry is commonly implemented using digital signal processing techniques powered by miniature 1-volt zinc-air batteries.

In addition to dynamic-range improvements, hearing aids serve to improve the signal-to-noise ratio of desired sounds in the real world primarily for better speech intelligibility in noise. Currently miniature directional microphone systems with port spacings in the 5-mm range are being used to provide improvements in speech intelligibility in noise of 4 to 6 dB. Such microphones have become rather sophisticated, in that many designs have directional adaptation circuits designed to modify polar patterns to optimize the intelligibility of desired sounds. In addition some designs are capable of providing different directional patterns in different frequency bands. Furthermore, some hearing aid manufacturers have introduced products using second-order directional microphones operating above 1 kHz with some success.

In many situations traditional hearing aid technology is not able to provide adequate improvements in speech intelligibility. Under such circumstances wireless transmission and reception technology is being employed to essentially place microphones closer to talkers' mouths and speakers closer to listeners' ears. This trend appears to offer promise enabled by the evolution of smaller transmitter and receiver devices and available operating-frequency allocations. Practical devices using such technology are now being offered for use with cellular telephones. This is expected to be an area of considerable technology and product growth.

Tinnitus

Another hearing disorder, tinnitus, is commonly experienced by individuals, often as a result of ear infections, foreign objects or wax in the ear, and injury from loud noises. Tinnitus can be perceived in one or both ears or in the head. It is usually described as a ringing, buzzing noise, or a pure tone perception. Certain treatments for tinnitus have been developed for excessive conditions in the form of audio masking, however most research is directed toward pharmaceutical solutions and prevention. We are also seeing the emergence of electro-acoustic techniques for treating what is commonly referred to as idiopathic tinnitus or tinnitus with no known medical cause. About 95% of all tinnitus is considered idiopathic. These treatments involve prescriptive sound stimuli protocols based on the spectral content and intensity of the tinnitus. In Europe, psychological assistance to help individuals live with their tinnitus is a well established procedure.

Hearing loss prevention

Hearing-loss prevention has become a major focus of this committee due to the fact that a majority of AES members come in contact with high level sounds as a part of the production, creation, and reproduction of sound. In addition, this subject has become a major issue of consumer concern due to the increased availability of fixed and portable audio equipment capable of producing damaging sound levels as well as live sound performance attendance. One approach to dealing with this issue is education in the form of communicating acceptable exposure levels and time guidelines. Such measures are however of limited value, as users have little practical means of gauging exposure and exposure times. This situation represents a major need and consequent opportunity for this committee, audio equipment manufacturers, and the hearing and hearing-conservation communities. In recognition of the importance of hearing health to audio professionals engaged in the production and reproduction of music, this committee has scheduled its first conference devoted to technological solutions to hearing loss. The 47th AES International Conference on Music Induced Hearing Disorders will take place in Chicago, IL, USA from June 20–22, 2012. This conference will focus on new technologies for measurement and prevention.

HIGH RESOLUTION AUDIO

Vicky Melchior and Josh Reiss, Chairs

Within the past decade, the types, distribution, and uses of audio have greatly diversified. Portables and internet sourcing have flourished and disc sales have fallen, although the balance between the two varies by country. High quality audio for formal listening has evolved simultaneously and mirrors many of the same influences. There is a notable broad trend toward increasing quality in many aspects of audio, and together with promised developments such as cloud storage and HD streaming, digital audio including high quality formal listening will continue to grow and evolve.

Music sources

High resolution remains a mainstay of professional recording and archiving due to its extended headroom, precision, and frequency capture. In the consumer marketplace, the principal current high resolution sources are discs, especially Blu-ray, and internet downloads. The music for these releases reflects a range of eras and recording techniques as well as resolutions, and may have been remastered, transcoded, or upsampled. Thus the frequency extension and dynamic range in some cases is less than that of newer recordings made directly at high resolution.

The original high resolution disc formats have not achieved wide success although SACDs continue to be released in small numbers, notably in classical music. SACD-capable players continue to be available and today's universal players may play Blu-ray Disc (BD), DVD, SACD, and CD. Some support for Direct Stream Digital, the single bit encoding technique behind SACD, can be found in professional recorders, players, and modern interfaces, but LPCM has largely supplanted single bit techniques as release and recording formats.

With the discontinuance of HD-DVD, BD is now the higher bandwidth successor to DVD and is well suited for high resolution multichannel audio, both alone and in combination with high definition (HD) video. The format provides an optional 8 channels of 96 kHz/24 bit audio or 2–6 channels of 192 kHz/24 bit. The great majority of current BDs include one or more of these optional formats. Audio-only discs are not yet common, but a nascent initiative exists on the part of several small companies to record audio-only high res multichannel on BD without the need for a TV monitor. Note that derivative HD discs also exist in some regions, for example China Blue HD in the Chinese market.

A rapid proliferation of BD-capable devices has resulted, encompassing players, laptops, external BD drives for PCs, PCI cards supporting 7.1 audio with BD decoding, recorders, and home theater processors. Many, though not all, support eight channels of high resolution audio. The retail industry in the U.S. also reports growing interest among ordinary consumers in BD and multichannel audio.

At least 40 websites ranging from large aggregators to individual orchestras and bands now exist and sell both new work and back catalog with resolutions from 192 kHz/24 bit to 44.1 kHz/16 bit. Tracks are principally stereo and favor classical music, although broader genre coverage is increasing. Websites currently sell without copy protection. Accordingly, few releases at the highest resolutions are available from the major labels.

The file formats of online downloads have coalesced around FLAC and WMA for lossless compression and WAV or AIFF for uncompressed LPCM. The popularity of FLAC relates to its free, open source nature and its compatibility with most computer operating systems. FLAC is not widely supported on mobile devices or in many lower priced home theater (HT) systems and can be difficult to route through an HT system without first transcoding.

Growth of computer and server-based audio

There is a strong trend toward adoption of computers and file servers into all areas of audio, especially evident in the U.S. and Far East. For high quality audio, there are excellent opportunities but a range of new technical and delivery issues. The term "computer audio" covers numerous configurations where the computer may act as front end disc player or file server; may output audio via a PCI sound card, external sound card, or motherboard ports; and may access downloads or streamed radio and AV from the internet. Files may be stored on hard drives, flash, network-attached storage (NAS), or redundant arrays with backup; and network file servers other than a computer may act as software players.

The traditional audiophile two-channel, music-only marketplace has embraced computers and file servers due to the convenience of file storage and downloads. In this market, which overlaps professional audio, the design ethos of low distortion, high quality engineering has spurred manufacturer research in identifying and eliminating technical problems associated with computers as front end devices. These include isolation of noisy computer power supplies, avoidance of jittered computer clocks, RFI shielding, special attention to computer layouts by makers of PCI sound cards, and design of digital interfaces to avoid contaminating an external DAC master clock with the jitter and noise from the PC. Examples of the latter include asynchronous USB, PLL chips in association with Firewire and SPDIF, and DAC-controlled data transmission. Much ongoing effort in computer related software aims to provide bit-accurate decoding, ripping, playback, and transcoding.

A trend to include computer audio in home theater is underway as well but with a greater mix of challenges for high quality audio. Home theater is above all a rapidly evolving and richly diverse area of wide price range and capability. HT components routinely support the lower resolution compressed formats streamed from the internet and cable, and variously the high resolution AV needed for DVD, BD, and HDTV. Support for the file types and resolutions typical of downloads, disc rips, and AV from other recording or non-movie sources may be absent. It continues to be challenging to transmit files without invoking unwanted sample rate conversion, unintended transcoding (e.g., FLAC to MP3), bit truncation, and loss of metadata.

Improving audio quality

Transmission of high quality, high bandwidth AV signals across networks and digital interfaces is a very active arena of work. In addition to advances in point-to-point interfaces discussed above, development continues on Ethernet and HDMI.

New Ethernet initiatives such as Audio Video Bridging (AVB) promise improved network attributes like bandwidth reservation, traffic shaping, phase synchronization across all channels, and low latency. AVB Ethernet is relevant to home and car systems, although the jitter performance of DAC clocks linked to the network will need to be assessed.

HDMI, the point-to-point connector required for BD and HD video, has excellent bandwidth and an Ethernet data link (HDMI 1.4), but lacks an audio clock. HDMI receivers must derive audio word clock from the video pixel clock, commonly resulting in very high jitter that affects quality and can be audible. Some high end receivers address the jitter and many companies are researching it but current solutions are expensive and uncommon.

Current wireless audio devices with few exceptions are limited to 48 kHz, but components and transmission protocols are underway that promise 96 kHz capability.

Convergence trends are strongly evident in AV design and will certainly continue in light of entertainment trends such as cloud storage and streamed HD live performance.

Research

High resolution formats in general are mature, although efforts to improve lossless compression continue. Inquiry continues into the perceptual characteristics and audibility of sample rates above 44.1 kHz/16 bit, and of the associated filtering and data conversion processes. Design research continues on loudspeakers, class D amps, and microphones in support of the wide bandwidth, low distortion, wide dynamic range requirements of high resolution. Also, surround algorithms emphasizing enhanced spatial coding are an especially active research area that should be mentioned in context of high resolution because of the improved spatial resolution they afford.

HUMAN FACTORS IN AUDIO SYSTEMS

Michael Hlatky, Chair
William Martens, Vice Chair
Jörn Loviscach

The Technical Committee on Human Factors in Audio Systems provides an industry forum for questions concerning the design of user interaction for audio applications, the integration of audio in man-machine interfaces (such as warning sounds, data sonification and auditory feedback), and the design of interfaces for musical instruments.

Touch screens and mobile devices

With the recent advent of ubiquitous touch-controlled computing devices, especially the first topic has gained considerable importance. Devices that provide touch-based on-screen manipulation such as smartphones and tablet PCs are heavily used to consume all things digital. Audio software on phones or tablets, however, is yet mostly targeted at the consumption end of the audio commercialization chain. The reason why we are not yet commonly seeing professional audio workstations running on a touch screens alone might be traced back to some of the obvious shortcomings of such devices when used to work with digital audio: Touch screens commonly lack pixel-precise navigation, parts of the screen will be visually obstructed by the user's hand and arm when manipulating an on-screen control, and there is relatively little to no tactile feedback during the interaction process.

These three reasons alone make the design of, for instance, a touch-controlled on-screen fader quite cumbersome. While the precision achievable by touch manipulation of an on-screen fader might be enough to set the playback volume when listening to MP3s on a phone, it can be by far not enough to set parameters when mixing music. Some manufacturers have therefore enabled swiping gestures on touch-controlled faders to increase precision; this does, however, take away direct controllability, as several micro-actions might be necessary to achieve a desired parameter value. Furthermore, the lack of pressure-sensitive touch screens on the mass market renders the expressive control of musical instruments with such devices nearly impossible. To enable an additional degree of freedom for expressive input, some software—for instance Apple's GarageBand on the iPad—incorporates data from the device's accelerometer sensor when the user plays virtual instruments with the on-screen keyboard.

The common smartphone's collection of sensors such as the touch screen, accelerometer, compass, GPS, microphone, and ambient light sensor also provides a whole new range of input capabilities that can be leveraged in conjunction with digital audio. There is a collection of new audio applications that enable users to influence the presented audio using these sensors. Software, such as for instance Smule's "I am T-Pain," RjDj's "Inception App" or the Black Eyed Peas' "BEP360" interactive music video, introduce a whole new level of interactivity into the formerly lean-back experience of listening to music. In addition, they raise the question whether music might in future not even be generally distributed as mere audio data, but as an application.

Interactive audio applications also pose a new set of problems to designers of the common digital audio workstation (DAW). How does a future digital audio workstation that is targeted at producing audio for interactive applications integrate itself well into the development environments for the iPhone and its siblings? Hints might be taken from software employed to design interactive music scores and dynamic sounds for computer games, such as Crytec's "CryEngine," or visual programming languages such as Cycling '74 Max, or Pure Data.

Novel game controllers

The experimental music scene has quickly picked up off-the-shelf devices for natural user interaction (NUI). Novel game controllers such as the Microsoft Kinect Sensor or the older Nintendo Wii Remote have, however, yet to arrive in the professional audio industry. In the gaming market especially the Kinect has had a huge impact; Microsoft reported selling more than eight million units within the first 60 days, making this the fastest selling consumer electronics device ever.

The Kinect controller enables data manipulation for multiple users by natural user interaction employing multiple users' whole bodies via skeleton tracking. This means that for instance the positions of the users' hands in three-dimensional space can be used to control parameters, or the software can react directly to full-body gestures. The hacking scene, such as the attendants of the industry-sponsored Music Hack Days, embraces these devices. In the case of the Kinect, this is fueled in particular by the SDKs provided by the open-source community and Microsoft itself. It took only few days after the Kinect came to market until the first software was published by the open-source community enabling control of software instruments.

The cloud

Another trend to be observed at Music Hackdays is the rise of web-based APIs (application programming interfaces). Whether it is finding new audio content, processing audio, or simply listening to music, companies such as SoundCloud, The Echo Nest or Spotify have an API for that. Music discovery and recommendation via interconnected web services are topics taken on now by Facebook and Google, and even Pro Tools got in its tenth incarnation equipped with a function to directly bounce a mix to SoundCloud. Even the DAW has moved into the cloud, with, for instance, PowerFX's "Soundation Studio" or OhmForce's "OhmStudio."

The key benefit of these new audio production platforms are the enhanced possibilities for remote collaboration in comparison to traditional DAWs. The move to the cloud does, however, also enable a whole new approach to designing user interfaces through so-called perpetual betas. As applications are running in the browser, update cycles are frictionless, because each time the user loads a session, a new version of the software can be delivered. Another fact to keep in mind is that the computing power in the cloud is decentralized. A limit to the number of plugins running in parallel might be a problem of the past as soon as audio processing has moved to the cloud.

With all this computational power available in the cloud, also completely new approaches to user experiences inside a DAW are possible. Already today, "UJAM" enables you to sing a few lines, from which it automatically generates a complete, professional-sounding song.

A drawback of the browser-based DAWs, however, might be that the long-learned, known, and expected standard user interfaces provided by the operating systems such as the default buttons or the behavior of menus are not easily replicable inside a web browser. With the advent of HTML5 as a "kind of operating system" of a cloud-driven audio experience, such standards might never exist again.

Modular hardware controllers

Tangible interfaces with knobs and faders are still a big topic, and it seems that extreme modularity is the new trend in hardware controllers. Steinberg's "CMC Series" or Euphonic's "Artist Series" controllers can be combined in any number, enabling the user to build a hardware controller setup for the bedroom studio or the scoring stage, all employing the same components.

Recent research in the HCI community has explored the combination of touch screen interfaces with superimposed physical controls in audio editing tasks, such as for instance "Slap Widgets" by Malte Weiss and his coworkers or "Dynamic Mapping of Physical Controls for Tabletop Groupware" by Rebecca Fiebrink and her coworkers. These approaches seem promising to unite the tactile controllability of physical input devices with the configurability of a touch screen.

MICROPHONES AND APPLICATIONS

Eddy B. Brixen, Chair
David Josephson, Vice Chair

The microphone is an amazing device. No other piece of audio equipment being 20 to 50 years old would be considered as a sensible choice for modern recording. However, that is to some degree the way microphones are regarded.

Oldies but goodies(?)

In the marketplace of today we find a lot of old designs still being produced. A high percentage of new products brought to market in reality are copies of aging technologies —ribbon microphones, tube microphones, and the like. The large number of these designs introduced to the market is better explained by the opportunity of making good business on the general assumption that exotic looking microphones provide exotic audio than it is by an increased level of research in understanding and improvements of these designs.

Transducer technology

There has been no major break-through in transducer technology during the last years. Microelectronic mechanical systems (MEMS) are not yet on the market for professional audio. However, in the near future the limited signal-to-noise ratios may not be a problem any longer.

Digital adaptation

Innovation in the field of modern microphone technology is to some degree concentrated around adaptation to the digital age. In particular the interfacing problems are addressed. The continued updating of the AES42 standard is essential in this respect. Now dedicated input/control stages for microphones with integrated interfaces are available. However, different widely implemented "device-to-computer" standards like

USB and Firewire—which are not specifically reserved for audio—have also been applied in this field. Regarding the data streams, USB3 is fully satisfactory for most audio purposes but USB microphones are outside standards. However they have reached a much higher level of popularity in semi-pro audio and home recording compared to AES42.

DSP-controlled microphones are still developing. This includes directional pattern control of multi-transducer units providing steering or multichannel output for surround recordings. These techniques are not necessarily applicable in professional audio. However, in the field of surveillance and security recordings the applications are obvious.

Other microphone developments

More attention has been paid to the reduction of EMC problems found in an environment of increasing high frequency electromagnetic fields that are being picked-up by microphones.

Higher order Ambisonics has taken a central position in the search for multi-format compatibility. Other dedicated formats for surround sound exist. However, it seems that the 9.1/13.1 formats are forcing many engineers to start reinventing arrays over again. This should not be necessary.

Some technologies earlier regarded as exotic are finding their way into practical applications. As an example NASA has published technical briefs on a laser microphone technology that must be regarded as a serious solution.

Battery technology—especially for wireless microphones—is an area of great attention. Surprisingly many engineers still prefer replaceable batteries from rechargeable. This will change.

The difficulties of getting some of the rare earth materials for magnets may affect the microphone selection available on the market. In the future the effect of this might be realized as fewer dynamic microphones or rising prices.

NETWORK AUDIO SYSTEMS

Kevin Gross, Chair
Umberto Zanghieri and Thomas Sporer, Vice Chairs
Tim Shuttleworth

This document is a compilation of contributions from numerous members of the Technical Committee on Networked Audio Systems. The committee has identified the following important topics related to emerging audio networking technologies. Technologies that have emerged since the last published Emerging Trends Report from the committee in 2007 are included. To provide structure to the report items are discussed in order of their maturity; commercialized technologies implemented in products available for purchase being discussed first and embryonic concepts in early development come up last. Other categorizations referred to in this document are consumer market orientation versus professional market focus, as well as media transport methods versus command and control protocols.

EBU N/ACIP

The European Broadcasting Union (EBU) together with many equipment manufacturers has defined a common framework for Audio Contribution over IP in order to achieve interoperability between products. The framework defines RTP as a common protocol and media payload type formats according to IETF definitions. SIP is used as signaling for call setup and control, along with SDP for the session description. The recommendation is currently published as document EBU Tech 3326-2008.

Audio video bridging

The Audio Video Bridging initiative is an effort by the IEEE 802.1 task group working within the IEEE standards organization that brings media-ready real-time performance to Ethernet networks. The IEEE is the organization that maintains Ethernet standards including wired and wireless Ethernet (principally 802.3 and 802.11 respectively). AVB adds several new services to Ethernet switches to bring this about. The new switches interoperate with existing Ethernet gear but AVB-compliant media equipment interconnected through these switches enjoy performance currently only available from proprietary network systems.

AVB consists of a number of interacting standards:

802.1AS – Timing and Synchronization
802.1Qat – Stream Reservation Protocol
802.1Qav – Forwarding and Queuing
802.1BA – AVB System
IEEE 1722 – Layer 2 Transport Protocol
IEEE P1722.1 – Discovery, enumeration, connection management and control
IEEE 1733 – Layer 3 Transport Protocol.

AVB standardization efforts began in earnest in late 2006. As of November 2011, all but the P1722.1 work have been ratified by the IEEE.

RAVENNA

A consortium of European audio companies has announced an initiative called RAVENNA for real-time distribution of audio and other media content in IP-based network environments. RAVENNA uses protocols from the IETF's RTP suite for media transport. IEEE 1588-2008 is used for clock distribution. Performance and capacity scale with the capabilities of the underlying network architecture. RAVENNA emphasizes data transparency, tight synchronization, low latency, and reliability. It is aimed at applications in professional environments, where networks are planned and managed.

All protocols and mechanisms used within RAVENNA are based on widely deployed and established methods from the IT and audio industry or comply with standards as defined and maintained by international standardization organizations like IEEE, IETF, AES, and others. RAVENNA can be viewed as a collection of recommendations on how to combine existing standards to build a media streaming system with the designated features.

RAVENNA is an open technology standard without a proprietary licensing policy. The technology is defined and specified within the RAVENNA partner community, which is led by ALC NetworX and supported by numerous well-known companies from the pro audio market.

AES X192

Audio Engineering Society Standards Committee Task Group SC-02-12-H is developing an interoperability standard for high-performance media networking. The project has been designated "X192."

High-performance media networks support professional quality audio (16 bit, 48 kHz and higher) with low latencies (less than 10 ms) compatible with live sound reinforcement. The level of network performance required to meet these requirements is achievable on enterprise-scale networks but generally not on wide-area networks or the public internet.

The most recent generation of these media networks use a diversity of proprietary and standard protocols (see Table 1). Despite

Technology	Purveyor	Date introduced	Synchronization	Transport
RAVENNA	ALC NetworX	In development	IEEE 1588-2008	RTP
AVB	IEEE, AVnu	In development	IEEE 1588-2008 advanced profile (IEEE 802.1AS)	Ethernet, RTP
Q-LAN	QSC Audio Products	2009	IEEE 1588-2002	UDP
Dante	Audinate	2006	IEEE 1588-2002	UDP
LiveWire	Telos/Axia	2004	Proprietary (native),	

Table 1 Media networks

a common basis in Internet Protocol, the systems do not interoperate. This latest crop of technologies has not yet reached a level of maturity that precludes changes to improve interoperability.

The X192 project endeavors to identify the region of intersection between these technologies and to define an interoperability standard within that region. The initiative will focus on defining how existing protocols are used to create an interoperable system. It is believed that no new protocols need be developed to achieve this. Developing interoperability is therefore a relatively small investment with potentially huge return for users, audio equipment manufacturers, and network equipment providers.

While the immediate X192 objective is to define a common interoperability mode the different technologies may use to communicate to one another, it is believed that the mode will have the potential to eventually become the default mode for all systems. It will be compatible with and receive performance benefits from an AVB infrastructure. Use of the standard will allow AVB implementations to reach beyond Ethernet into wider area applications.

While the initial X192 target application is audio distribution, it is assumed that the framework developed by X192 will be substantially applicable to video and other types of media data.

Dante

Dante is a media networking solution developed by Audinate. In addition to providing basic synchronization and transport protocols it provides simple plug and play operation, PC sound card interfacing via software or hardware, glitch free redundancy, support for AVB, and support for routed IP networks. The first Dante product arrived in 2008 via a firmware upgrade for the Dolby Lake Processor and since then many professional audio and broadcast manufacturers have adopted Dante.

From the beginning Dante implementations have been fully IP based, using the IEEE 1588-2002 standard for synchronization, UDP/IP for audio transport and are designed to exploit standard gigabit Ethernet switches and VoIP-style QoS (quality of service) technology (e.g., Diffserv). Dante is evolving with new networking standards. Audinate has produced versions of Dante that use the new Ethernet Audio Video Bridging (AVB) protocols, including IEEE 802.1AS for synchronization and RTP transport protocols. It is committed to supporting both IEEE 1733 and IEEE 1722. Existing Dante hardware devices can be firmware upgraded as Dante evolves, providing a migration path from existing equipment to new AVB capable Ethernet equipment.

Recent developments include announced support for routing audio signals between IP subnets and the demonstration of low latency video. Audinate is a member of the AVnu Alliance and the AES X192 working group.

Q-LAN

Q-LAN is a third-generation networked media distribution technology providing high quality, low latency, and ample scalability aimed primarily at commercial and professional audio systems. Q-LAN operates over gigabit and higher rate IP networks. Q-LAN is a central component of QSC's Q-Sys integrated system platform. Q-Sys was introduced by QSC Audio Products in June 2009. Q-LAN carries up to 512 channels of uncompressed digital audio in floating point format with a latency of 1 millisecond.

WAN based telematic/distributed performance and postproduction

Telematic or distributed performances are events in which musicians perform together synchronously over wide area networks, often separated by thousands of miles. The main technical challenge associated with these events is maintaining sufficiently low latencies for the musicians to be able to play together, given the distances involved. Emerging enabling technologies such as the low latency codecs CELT, which stands for "Constrained Energy Lapped Transform," Opus, a merging of CELT and SILK (a Skype codec) as well as ULD, which refers to "Ultra-Low-Delay" allow streaming over DSL or cable end point connections rather than high-bandwidth managed networks, such as Internet2, which are recently more commonly used.

Another wide area networked emerging use case is streaming audio for cinema postproduction, in which studios and postproduction facilities are connected with one another via high-bandwidth managed fiber networks. This allows studios to see and hear the latest version of a film in postproduction without having to physically move the assets to the studio or use a file-transfer system. Real-time streaming of uncompressed audio and video also allows greater collaboration between directors and postproduction facilities and between different departments in the postproduction process.

Networked postproduction uses two methods (at present) for streaming audio: when audio is streamed independently of video, hardware Layer 3 uncompressed audio-over-IP devices are used. When audio is streamed along with video, it is embedded in an HD-SDI video stream, and the stream is networked using a video codec. The former case is primarily used for audio postproduction, in which the audio engineers are mixing to a poor-quality version of the video; the video is then sourced locally at all locations, and the audio synced to it. Control information is streamed between all nodes using high-definition KVM-over-IP devices, along with MIDI-based control surfaces connected via Apple's MIDI Network Setup. KVM over IP is a server management technology. (Streaming of Ethernet-based control surfaces is forthcoming.) Video-conferencing to allow collaboration uses either H.323 devices or the same codec used to stream content video. Clock synchronization between nodes can be accomplished either with the hardware audio-over-IP devices, which usually stream clock information, or with GPS-based sync generators at each node.

XFN command and control protocol

XFN is an IP-based peer to peer audio network control protocol, in which any device on the network can send or receive connection management, control, and monitoring messages. The size and capability of devices on the network will vary. Some devices will be large, and will incorporate extensive functionality, while other devices will be small with limited functionality. The XFN protocol is undergoing standardization within the AES, and AES project X170 has been assigned to structure the standardization process. A draft standards document has been written and presented to the SC-02-12 working group for approval.

Home broadband audio-over-IP and home wireless LAN

Home broadband connections are increasing in speed, up to a typical rate, worldwide, of about 2 Mbps. This is sufficient for streaming audio services to produce a good performance, mostly using 256 kbps WMA or AAC, which yields pretty good quality at a low bit rate.

Use of wireless LANs in the homes, mostly WiFi, with some proprietary systems is increasing. IEEE802.11g routers and devices are realizing faster throughput rates, while IEEE802.11n achieves improved range, improved QoS, and speeds that exceed the needs of low bit rate compressed audio streaming. Two eco-systems co-exist at the moment. The first is the Digital Living Network Alliance (DLNA), which focusses on interoperability between devices using UPnP (Universal Plug and Play) as the underlying connectivity technology. DLNA is becoming available in more and more devices, such as PC servers and players, digital televisions with network connectivity, network attached storage (NAS) drives, and other consumer devices. The second eco-system is Apple AirPlay, which allows iTunes running on a PC or MAC to stream audio to multiple playback devices. AirPlay also supports streaming directly from an iOS device (iPhone, iTouch, iPad) over WiFi to a networked audio playback device. Both ecosystems are driving the rapid acceptance of audio networking in the home.

Cloud computing, in particular cloud storage of audio content, is another emerging trend. The increasing popularity of premium audio services, for example Rhapsody, Pandora, Last.fm, and Napster, are driving a trend away from users the need to keep a copy of their favorite music in the home or on a portable device. Connection to the internet allows real time access to a large variety of content. Apple is also driving this trend with iCloud, released with iOS5. Consumer devices are becoming more complicated and connecting devices to the network has been difficult for users, resulting in many calls to tech support. The good news is that devices are becoming easier to set up. The WiFi Alliance has created an easy setup method call WiFi Protected Setup (WPS). This makes attaching a new device onto the home network as easy as pressing a button or entering a simple numeric code.

Another trend driven by the adoption of home wireless LAN technologies is in the user interface (UI) of networked audio devices. More and more audio products are using the iPhone or iPad as the primary method of device control, via the home WiFi network. Some commentators are even announcing the death of the infrared remote control. Consumer Audio/Video Receiver manufacturers such as Denon and Pioneer offer free iPhone/iPad apps which allow complete, and intuitive control of their devices. This leads to another emerging trend, that of the display-less networked audio player. Once the player can be conveniently controlled from your smartphone, it may not be necessary for the device to continue to include an expensive display and user controls. Display-less high end audio players are already selling well (for example B&W Zeppelin Air). Such display-less networked audio players will become ubiquitous and be available for under $100.

Open Control Architecture Alliance (OCA)

The Open Control Architecture Alliance has been formed by a group of professional audio companies who are working in different product markets and represent a diverse cross section of vertical market positions and application use-cases. Each of the companies realized that relying solely on proprietary solutions for media networking system controls made interoperability with other manufacturers' equipment or across application domains difficult.

The member companies agreed that an open standardized control architecture was not only possible, but should be created and made available as an open, public standard that could be available to any participant in the audio market in order to facilitate an improved environment for the entire AV industry. It is the stated mission of the OCA Alliance to secure the standardization of the Open Control Architecture (OCA), as a media networking system control standard for professional applications. OCA in its current form is a Layer 3 protocol that has been created by Bosch Communications based around the earlier (abandoned) command and control protocol AES-24. The Alliance has been formed to complete the technical definition of OCA, then to transfer its development to an accredited public standards organization.

The founding group of OCA members is proceeding to complete the OCA specification and prepare it for transfer to a public standards organization without inviting new active members but welcomes any interested parties to join as an Observer Member.

International Telecommunications Union: Future Networks

ITU-T Q21/13, Study Group SG13 is looking at "Future Networks," which are expected to be deployed during 2015–2020. So far an objectives and design goals document has been published (Y.3001), and the study group is working on virtualization and energy saving (soon to be published as Y.3011 and Y.3021 respectively) and on identifiers. These deliberations are at a very early stage and a clear direction is not yet apparent. The underlying technology could be a "clean slate" design, or it could be a small increment to NGN (Next Generation Network, which is based on IPv6).

IEC/ISO: Future Network

ISO/IEC JTC1/SC6/WG7 is also working on Future Network, and also expects deployment during 2015–2020. Their system will be a "clean slate" design with a control protocol that is separate from the packet forwarding. It will support multiple networking technologies, both legacy technologies such as IPv4 and IPv6 and also new technologies able to provide a service suitable for the most demanding live audio applications.

It will carry two kinds of data, "synchronous" and "asynchronous." For synchronous data there is a set-up process (part of the control protocol) during which resources can be reserved. The application requests QoS parameters (delay, throughput, etc.) appropriate to the data to be sent, and the network reports the service the underlying technology is able to provide.

Asynchronous data can use a similar set-up process, or can be routed in a similar way to Internet Protocol. Thus it will also be efficient at carrying protocols such as TCP and will interoperate with IP networks. This provides a migration path from current systems.

SEMANTIC AUDIO ANALYSIS

Mark Sandler, Chair
Dan Ellis and Jay LeBoeuf, Vice Chairs
Gyorgy Fazekas

The scope of Semantic Audio Analysis has undergone a dramatic expansion over the past few years. As seen by many researchers and practitioners, the area is now best defined as the confluence of a multitude of technologies. These include digital signal processing tools that enable the extraction of characteristic features from audio, machine learning tools that connect raw feature data with high-level semantic representations of musical content, information management tools that allow us to effectively link and organize this information, and knowledge representation tools facilitating the use of automatic data aggregation and high-level inferences, thus the formulation of complex queries involving unique features of content, as well as social metadata about musical recordings. Web-based applications that allow us to pose queries like "find me upbeat and catchy songs between 130–140 bpm, performed by artists collaborating in the London-Shoreditch area, and sort them by musical key" are now imminent. The TC is concerned with overseeing and coordinating developments, disseminating knowledge, and promoting novel interdisciplinary tools and applications in the light of emerging trends in Semantic Audio. The most important of these novel trends and applications include the following.

Multi-modality and the use of contextual information

The process by which human beings understand music, and assign high-level semantic descriptions to musical events depends on a variety of information sources; precepts from different senses, memory, and expectations. As recently demonstrated during the first and highly successful AES conference on Semantic Audio Analysis in Ilmenau, Germany, researchers have started to recognize that semantic labelling of music relies on a number of different inputs, and started to develop techniques that take contextual information into account. This may be defined as a piece of complementary data that improves the results, but not in itself sufficient in a particular information extraction or audio processing task. Examples of new methods include informed source separation, which works by encoding information about the mixing process into the stereo signal, and enhance signal separation by using this data in the decoder, and informed music transcription, which takes prior information about the instruments into account. We can also observe the increasing use of studio stems, taking advantage of the multitrack format and the use of multiple modalities, the simultaneous analysis of audio, video, text, and other sources of information.

Ontologies and linked data

The heterogeneity and open-ended nature of musical data is often the culprit of developing complex systems that use many sources of information. Recent developments in other disciplines, namely Web-Science and the Semantic Web help us in developing methods for associating musical data with explicit meaning in a machine-processable way. Technologies such as the Resource Description Framework (RDF) and Semantic Web ontologies enable us to represent information and knowledge in a uniform interoperable framework, and lead to intelligent music processing tools of the future. Semantic Web ontologies such as the Music Ontology also provide the back-bone of the Linked Data, which eases linking and aggregation over disparate resources, containing increasing amounts of editorial and social data about music.

Educational games

New advances in semantic audio technologies enable the creation of interactive educational games for music learners. It is now possible to analyze the sound played on real instruments and thus avoid the need for using MIDI controllers, extract symbolic information such as chords or note names, and align this information with musical scores in real-time. Applications like Song2See demonstrate how semantic audio technologies may help to create content for music learners by using automated transcription, keep the user in the loop by allowing the correction of transcription errors, use the content to ease the learning process with fingering suggestions for each instrument, and provide real-time feedback about the quality of playing by means of sound analysis. The appearance of web-based platforms for content and metadata sharing, and advances in semantic analysis and recommendation technologies also provide for creating novel applications for music education. There is a growing trend in using community created web content, including lead sheets and chord charts, and to analyze YouTube videos to enhance machine analyses, or to create interactive games that are not limited to expert generated content. The use of the web thus provides an advantage over games like Rock Band or Guitar Hero.

Intelligent music production tools

Finally, there is a recent increase in adapting semantic audio technologies in music production. Examples of these applications include navigating sound effect libraries by using similarity defined by proximity in a characteristic feature space, using automatic audio-to-score alignment in audio editing, and developing intelligent audio effects and automatic mixing techniques that rely on semantic audio analysis.

SIGNAL PROCESSING FOR AUDIO

Christoph Musialik, Chair
James Johnston, Vice Chair

Signal processing applications in audio engineering have grown enormously in recent years. This trend is particularly evident in digital signal processing (DSP) systems due to performance improvements in solid-state memory, disk drives, and microprocessor devices. The growth in audio signal processing applications leads to several observations.

Observations

First, DSP has emerged as a technical mainstay in the audio engineering field. Paper submissions on DSP are now among the most popular topics at AES Conventions, while just some years ago DSP sessions were rare at AES. DSP is also a key field for other professional conferences, including those sponsored by IEEE and ASA.

Second, the consumer and professional marketplaces continue to show growth in signal processing applications, such as increasing number of discrete audio channels, increasing audio quality per channel (both word width and sampling frequency, and increasing quality of building block electronics, such as sampling rate converters, ADCs and DACs, due to continuously growing availability of consumer-ready DSP hardware

Third, there is growing interest in intelligent signal processing for music information retrieval (MIR), like tune query by humming, automatically generating playlists to mimic user preferences, or searching large databases with semantic queries such as style, genre, and aesthetic similarity.

Fourth, there are emerging algorithmic methods designed to deliver an optimal listening experience for the particular audio reproduction system chosen by the listener. These methods include transcoding and up-converting of audio material to take advantage of the available playback channels, numerical precision, frequency range, and spatial distribution of the playback system. Other user benefits may include level matching for programs with differing loudness, frequency filtering to match loudspeaker capabilities, room correction, and delay methods to synchronize wavefront arrival times at a particular listening position.

In professional sound reinforcement, loudspeakers with steerable radiation patterns can provide a practical solution for difficult rooms. Professional live audio applications often demand low-latency systems, which remain challenging for DSP because many algorithms and processors are optimized for block processing instead of sample-by-sample, and thus introduce more latency.

Algorithmic developments will continue to occur in many other areas of audio engineering, including music synthesis, processing and effect algorithms, intelligent noise reduction in cars, as well as enhancement and restoration for archiving and audio forensic purposes. Also, improved algorithms for intelligent ambient noise reduction echo cancellation, acoustical feedback suppression, and steerable microphone arrays are expected in the audio teleconferencing field.

Fifth, switching amplifiers (like Class D) continue to replace traditional analog amplifiers in both low power and high power applications. Yet even with Class D systems, the design trends for load independence and lowest distortion often include significant analog signal processing elements and negative feedback features. Due to advances in AD/DA converter technology, future quality improvements will require the increasingly scarce pool of skilled analog engineers to design input stages like microphone preamps, analog clock recovery circuits, and output amplifiers that match the specifications of the digital components.

Implications for technology

All of the trends show a demand for ever greater computational power, memory capacity, word length, and more sophisticated signal processing algorithms. Nevertheless, the demand for greater processing capability will be constrained by the need to minimize power consumption, since a continuously growing part of audio signal processing will be done in small, portable, wireless, battery-powered devices. On the other hand, due to the increasing capabilities of standard microprocessors, contemporary personal computers are now fast enough to handle a large proportion of the standard studio processing algorithms. Advanced algorithms still exceed the capabilities of traditional processors, so we see a trend in the design of future processors to incorporate highly parallel architectures and the compiler tools necessary to exploit these capabilities using high-level programming schemes. Due to the relentless price pressure in the consumer industry, processors with limited resolution will still challenge algorithm developers to look for innovative solutions in order to achieve the best price-performance ratio.

The Committee is considering forging contacts with digital signal processor manufacturers to convey to them the needs, experiences, and recommendations from the audio community.

SPATIAL AUDIO

James Johnston and Sascha Spors, Chairs

Loudspeaker layouts

Nowadays surround sound is available in many households, where the 5.1 layout is the most deployed loudspeaker configuration. The production chain from recording, coding, transmission to reproduction of surround sound for cinema is also well established. So far, the consumer market for surround sound has mainly been driven by movie titles; audio-only content is still quite rare. As successor of the 5.1 layout, various layouts with more loudspeakers arranged in a plane have been proposed, for instance the 7.1 layout. None of them had the commercial success of the 5.1 layout. Layouts that allow for the reproduction of height seem to be the next natural step in the evolution of surround sound. A number of proposed layouts that include elevated loudspeakers, for instance 9.1, 10.2 or 22.2, are becoming ready for the market. It remains to be seen whether the market accepts the increased number of loudspeakers that have to be installed at home. From a perceptual point of view, including height cues into the reproduction has a clear benefit. However, the optimal speaker layout is subject to a vital discussion within the community.

Novel recording, production, and coding techniques have to be developed and established for the new layouts including height. Upmixing algorithms for content that has been produced for systems not featuring height, for instance from stereo or 5.1 surround, to the novel formats including height are being developed. A number of proposals exist, however there is still a lot of room for new developments that can be foreseen to show up in the future. In addition, new delivery methods that provide specific information related to height and distance, as well as horizontal angle are being reported on at AES conventions.

3-D

With the increased spread of 3-D video in cinema and home cinema, new requirements must be met by spatial audio reproduction. While 3-D video adds depth to the image, this is not a straightforward task with stereophonic techniques. This holds especially for sound sources closer to the listener than the loudspeakers. Future spatial audio techniques have to provide solutions to the challenges imposed by 3-D video. First concepts have been presented.

Sound field synthesis

As alternatives to the traditional stereophonic systems, sound field synthesis techniques such as Wave Field Synthesis (WFS) and higher-order Ambisonics (HOA) are being deployed more and more. Sound field synthesis approaches are based on the principle of physically synthesizing a sound field. In the past two decades around 100 WFS systems have been installed worldwide with each up to 832 channels. Large scale Ambisonics systems are currently not so widespread, but it seems that such systems will show up in the near future. A vital research community exists for both WFS and HOA that investigates various aspects and combinations of both approaches.

Psychoacoustic motivation

Upcoming trends in spatial audio reproduction besides traditional stereophony are multichannel reproduction systems that are psychoacoustically motivated. Several techniques have been developed on the basis of WFS that aim at spatial reproduction with almost arbitrary layouts using a decreased number of loudspeakers compared to traditional WFS. Such approaches are already commercially available. Multichannel time-frequency approaches use techniques from short-term signal analysis to analyze and synthesize sound fields. Directional Audio Coding (DirAC) and Binaural Cue Coding (BCC) are representatives of the latter techniques. Time-frequency processing seems to be a promising concept since its basic idea is related to the analysis of sound fields performed by the human auditory system.

The psychoacoustic mechanisms underlying the perception of synthetic sound fields have been investigated in quite some detail. However there are still plenty of open issues in the field that should be researched in the future. Peer-reviewed, published perceptual research based on established psychoacoustic methodologies will definitely bring the community forward in this aspect.

Production and mixing techniques

So far, the traditional production and mixing techniques used in stereophony are channel-based. The different tracks are mixed in the studio for a particular target layout and then transmitted/stored. This process relies on the assumption that the setup at the receiver matches the setup used during production. Object-oriented audio, as an alternative approach, is based on the concept that each track or group of tracks forms an audio object. The signal(s) of the object together with side information (position, room, effects) is then transmitted to the receiver. Here the loudspeakers signals are generated by a suitable rendering algorithm, on the basis of the transmitted side information. A major benefit of the object-oriented approach lies in the fact that it is almost independent from the setup used by the receiver. It seems that in the future a combination of both approaches might be promising to cope for the needs of the producers on the one side and to allow setup independent mixing/production on the other side. Currently several different formats for the transmission of the content/side information have been proposed, however, none have yet been commercially adopted in any significant fashion.

Headphone listening

Although spatial audio is routinely used by the gaming industry, advanced techniques with better quality and realism can be expected with further increases in processing power. This holds especially for mobile devices, where spatial audio is currently rarely deployed. Due to the general shift toward mobile devices spatial audio will also be finding its way into the mobile world. As a consequence, an increasing number of individuals use headphones when listening to audio. In such scenarios the use of reproduction techniques based on head-related transfer functions provides truly three-dimensional spatial audio by relatively simple technical means. As a consequence binaural audio is expected to play a more prominent role in the future. As an alternative to headphone listening, near-field loudspeaker playback with cross talk cancellation may be used.

Diverse applications

Besides its traditional application fields, cinema and home cinema, spatial audio is increasingly being deployed in other areas. For instance, in teleconferencing systems, cars, and as auditory system for advanced human-machine interfaces. Here the use of spatial audio is expected to provide a clear benefit in terms of naturalness and transport of additional information. Another important area of application is virtual concert hall and stage acoustics using active architecture systems where spatial audio enhances the environment with which musicians and audiences interact during performance. Modern multichannel systems offer adjustability of acoustics and high sound quality suitable for live performance and recording.

Network standards

With respect to cabling, coping with the ever increasing number of loudspeakers, the new IEEE standards for Audio Video Bridging (AVB) seems promising. The standards are designed for the fully synchronized transmission of a high number of output/input channels via Ethernet. The standards are developed and currently supported by all major players in the field and devices are expected to be available in the near future. Such standards that work with intelligent processing to detect the listening setup are expected to be proposed soon.

TRANSMISSION AND BROADCASTING

Kimio Hamasaki and Stephen Lyman, Chairs
Lars Jonsson and Neville Thiele, Vice Chairs

The growth of digital broadcasting is the most remarkable trend in this field. Digital terrestrial TV and radio broadcasting have been launched in several countries using the technology standards listed below. Analog broadcasting has ceased in some countries. The World Wide Web has become a more common, alternate source of streamed or downloadable programming.

Digital terrestrial TV broadcasting

In Europe DVB-T2 has been deployed in several countries for HD services. ATSC is used in USA, Canada, and Korea, while ISDB-T is employed in Japan and Brazil.

Digital terrestrial radio broadcasting

In Europe and Asia DAB+ is state of the art in the DAB Eureka 147 family. HD-Radio or IBOC fulfills this role in the USA and Canada, with ISDB-SB in Japan. Large broadcasting organizations in Europe and Asia, and major countries like India and Russia with large potential audiences, are committed to the introduction of DRM (Digital Radio Mondiale) services and it is to be expected that this will open the market for low-cost receivers.

Digital terrestrial TV broadcasting for mobile receivers

DVB-T2 Lite (Europe) is still under development, while ISDB-T is used in Japan. DMB is employed in Korea and there have been a few trials in Europe.

In the U.S., the Advanced Television Systems Committee (ATSC) has published final reports of two critical industry planning committees that have been investigating likely methods of enhancing broadcast TV with next-generation video compression, transmission and Internet Protocol technologies and developing scenarios for the transmission of three-dimensional (3-D) programs via local broadcast TV stations. The final reports of the ATSC Planning Teams on 3-D-TV (PT-1) and on ATSC 3.0 Next Generation Broadcast Television (PT-2) are available now for free download from the ATSC web site.

Benefits of digital broadcasting

The introduction of digital broadcasting has introduced such benefits as High Definition TV (1080i, 720p). Due to the current availability of 5.1 surround sound in digital broadcasting, surround sound is an important trend in TV broadcasting. 5.1 surround sound is evolving with future extensions involving additional channels. Along with 3-D TV, several broadcasters are experimenting with 3-D audio (for instance 22.2, Ambisonic, wave-field synthesis, directional audio coding). Data broadcasting now includes additional information related to a program.

Digital broadcasting themes at AES conventions

Recent AES conventions have discussed the following digital broadcasting issues: strategies for the expansion of digital broadcasting; audio coding quality issues for digital broadcasting and transmission; the role and importance of audio in an era of digital multimedia broadcasting; new program-production schemes for audio in digital multimedia broadcasting; the future of radio including multicasting over the web and surround.

Internet streaming

The use of new methods for the distribution of signals to the home via the Internet with streaming services is an increasing trend. Web radio and IPTV are now getting audience figures that in a number of years from now will be closing in on the traditional systems. Distribution technologies with rapid growth in many countries are: ADSL/VDSL over copper or fiber, combined with WiFi in homes; WIMAX and 3G/UMTS; 4G and wi-fi hot spots for distribution to handheld devices.

Loudness

Loudness and True Peak measurements are replacing the conventional VU/PPM methods of controlling program levels. This has largely eliminated significant differences in the loudness of different programs (and advertisements) and the need for listeners to keep adjusting their volume controls. Supporting international standards and operating practices have been published by several organizations such as ITU-R, EBU and ATSC listed below. More and more broadcasters now apply these standards in their program production and transmission chains.

ITU-R: BS.1770: "Algorithms to measure audio programme loudness and true-peak audio level"; BS.1771: "Requirements for loudness and true-peak indicating meters."

The following five documents provide the core of the EBU Loudness work:

EBU R128 Loudness Recommendation
EBU Tech 3341 Metering specification
EBU Tech 3342 Loudness Range descriptor
EBU Tech 3343 Production Guidelines
EBU Tech 3344 Distribution Guidelines

ATSC: A/85 "Techniques for Establishing and Maintaining Audio Loudness for Digital Television".

Lip sync

The lip-sync issue remains unsolved, but is being discussed in digital broadcasting groups. Some international standards development organizations such as IEC and SMPTE are discussing new standards for measuring the time differences between audio and video.