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 20 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 some of the Technical Committees.

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

AUDIO FOR TELECOMMUNICATIONS

Antti Kelloniemi
Kalle Koivuniemi, Co-chairs

The trend in mobile telecommunications across the globe has been towards devices that are the mobile window into a person's digital universe. In 2013 smartphones outsold feature phones globally for the first time on a unit basis. Several companies have focused efforts on creating inexpensive smartphones for the emerging markets, where, in many cases, these devices may be the only link those consumers have to the internet and the rest of the world. This expansion has resulted in devices that have many more communications bands and codecs than we have had in the past. It is not uncommon to find single devices that will communicate via UMTS, GSM, CDMA, LTE (VoLTE), and WiFi (VoIP). Some of these modes of communication have to support multiple voice codecs for each mode depending on the network or service provider. In all cases the switching between modes and performance of the system must be seamless, as past performance is no longer acceptable to the consumer. After 2013, mobile communication using VoIP and video conferencing apps has increased in accelerating pace.

From a voice quality standpoint, the AMR Wide Band codec is in widespread use globally, and the 3GPP EVS (Enhanced Voice Services) codec was finalized in 2014 and the first commercial deployment took place in South Korea next year. Demands for IP-based communication also in poor networks have caused a rising interest in low bitrate voice codecs.The audio recording capabilities of the devices have improved to offer high bitrate and 24 bit capture in some cases. There are also many products out on the market that allow multichannel audio recording via internal microphone arrays to complement the video recording on the devices. Use of stereo or spatial audio also for communication applications was increasingly studied in scientific research during the first decade of 2000. The first commercial applications have emerged, and we can see this as one trend that may change the way we communicate, together with the rise of technology and applications for augmented or virtual reality.

Voice controlled user interfaces in the mobile device have evolved to Virtual Personal

Assistants that extend to other platforms as well. Interaction with Natural Language is also coming to chatbots that are used for automating business processes like ordering a pizza by means of a phone call. Communication with these virtual assistants sets different requirements for smart phones and VoIP systems from what was common with human-to-human calls and new technologies are being developed to raise the quality of these features. We can estimate that in the near future using a mobile device for voice communication between a user and a digital assistant or a bot will be as typical as communication between humans. This communication will happen both through the same signal paths as normal voice communication and through dedicated paths, using signal processing and coding optimized for voice recognition engines.

Advancements in usability in all environments have come in the form of noise adaptive technology in voice call, content playback and voice control. Voice communication and voice recognition have benefitted from algorithms that adapt uplink noise reduction or algorithm parameters based on noise type and level. This becomes even more important when voice control and new types of endpoint devices emerge. Smart speakers or communication hubs for homes, conferencing phones for office environment are used from further away compared to personal mobile devices, so signal to noise ratio may be poor to start with. Likewise downlink audio and content playback have benefitted from noise adaptive algorithms. Another new feature that assists the usability in various environments has been the use of smart amplifier technology. The technology constantly monitors the content and the speaker to produce optimal loudness or bandwidth, while maintaining safe drive levels for the speaker. This technology paired with new high voltage amplifiers work together to produce impressive audio quality in many environments.

On the content side of the business, the trend has been from localized content to cloud based content. This has allowed the consumer to consume the same content across all devices, which has posed challenges for small power-conscious mobile devices.

The move from portable phones with a few extra features to the mobile hub of the consumer's digital world is reflected in the standards world, where the scope of standards as well as their names have had to change to keep up.

BROADCAST AND ONLINE DELIVERY

David Bialik, Kimio Hamasaki, Matthieu Parmentier, Co-chairs
Jim Starzynski, Vice Chair

Using LKFS and LUFS: Trends to Effectively Managing Video Content Audio Loudness

During the analog to digital television transition, professionals and consumers alike struggled with the unintentional loudness challenges that DTV's extreme audio dynamic range brought to audiences. Program-to-commercial loudness transition problems recognized by broadcasters, and shortly after by government, were effectively addressed and solved by the implementation of the recommended practices of industry-wide efforts (ATSC's A/85, EBU's R 128 and others) that documented new loudness measurement techniques.

A key ingredient to all solutions was the creation of the cornerstone loudness measurement recommendation—ITU-R BS.1770 that established a globally recognized means to assign a numeric value to perceptual loudness, based on a model of human hearing.

With these tools in hand, broadcasters around the world equipped their facilities with new loudness metering and trained their staff to mix using a new loudness measurement paradigm based on identically derived values (labeled LKFS by ATSC and LUFS by EBU) moving away from obsolete VU and PPM approaches. This development paved the way for file-based content transcoding devices to "measure and scale" content loudness automatically, with extreme accuracy and without audible artifacts. Smooth content transitions for over the air television programming and advertising were now possible, meeting the expectations of content creators, broadcasters, government and the home audience.

In January of 2016, on the heels of a very successful release of AES TD1004: Recommendation for Loudness of Audio Streaming and Network File Playback, the AES leadership and its members recognized the need to bring this new trend in TV loudness management to the rapidly expanding platforms of early Over-The-Top-Television (OTT) and Online Video Distribution (OVD) that were noticeably plagued by exactly the same loudness issues as early DTV. In winter of 2016, AES Audio Guidelines for Over the Top Television and Video Streaming (AGOTTVS—a subcommittee of this Technical Committee) group was formed and currently has over 50 members representing Amazon, Apple, Google, Hulu, Netflix, manufacturers, major US and European TV networks, content creators and many more.

AGOTTVS first charge was to develop "Preliminary Guidelines" to fulfill the group's urgent need to bring DTV's basic fundamental LKFS-LUFS loudness practices to the emerging online distribution community. Conceived during an informal face-to-face member meeting at NAB in April, work commenced on the paper and consensus was achieved in late summer. AES TD1005: Audio Guidelines for Over the Top Television and Video Streaming — Preliminary Loudness Guidelines was formally released in less than 6 months at the AES Convention in Los Angeles in late September.

Currently, AGOTTVS group members are working independently in a Coding, Metadata, Loudness and DRC Subgroup, an Architecture and Devices Subgroup, and a Content Creation Subgroup. All members meet every other week during a 90 minute teleconference. The group's goal is to leverage the successful loudness techniques and already in-place recommendations for over the air TV and bring them to OTT and OVD. Work being done on the behavior of current coding technology and the variety of devices reproducing on-line content will help to fine tune the new loudness guidelines.

AGOTTVS is well on its way to create an AES Comprehensive Loudness Recommendation to serve as a reference to the online content creation and distribution community. The work has potential to become an AES standard and is tentatively planned for release at AES 143 in New York this fall.

Interested individuals can contact **broadcast@aes.org** or **Jim.Starzynski@nbcuni.com**, Chairman AES AGOTTVS.

CODING OF AUDIO SIGNALS

Jürgen Herre and Schuyler Quackenbush, Chairs

The AES Technical Committee on Coding of Audio Signals is composed of experts in perceptual and lossless audio coding. The topics considered by the committee include signal processing for audio data reduction and associated rendering. This includes methods for reducing redundancy and irrelevancy. An understanding of auditory perception, models of the human auditory system, and subjective sound quality are key to achieving effective signal compression.

Audio Coding has undergone a continuous evolution since its commercial beginnings in the early nineties. Today, an audio codec provides many more functionalities beyond mere bitrate reduction while preserving sound quality. A number of examples are given in the following:

Evolution of the coder structure

In the nineties, a classic audio coder consisted of four building blocks: analysis filterbank, perceptual model, quantization and entropy coding of spectral coefficients, and bitstream multiplexer. Later, more and more tools were added to enhance the codec behavior for critical types of input signals, such as transient and tonal solo instruments, and to improve coding of stereo material. More recently, this structure was augmented by pre/post processors that provided significantly enhanced performance at low bitrates:

Joint coding of channels: firstly, spatial preprocessors perform a downmix of the codec input channels (two-channel stereo, 5.1 or more) into a reduced number of waveforms plus parametric side information describing the properties of the original spatial sound image. Thus, the bit rate can be significantly reduced. On the decoder side, the transmitted waveforms are upmixed again using the transmitted spatial side information to reproduce a perceptual replica of the original sound image. Examples for such spatial pre/postprocessors are Parametric Stereo (PS) and MPEG Surround.

Bandwidth extension: secondly, to alleviate the burden of the core codec, various pre/post processors were devised which exclude spectral regions (predominantly the high frequency range) from the regular coding process and send parametric information about them instead. On the decoder side, the transmitted waveform components are used to re-create the missing spectral regions using the parametric information. This allows full audio bandwidth even at low bitrates. Well-known examples of such technologies are Spectral Bandwidth Replication (SBR), Harmonic Bandwidth Extension (HBE) and Intelligent Gap Filling (IGF).

Both joint coding of channels and bandwidth extension have recently also been implemented without a dedicated filterbank, thus saving computational complexity and algorithmic delay.

Convergence between speech and audio coding

The latest generation of codecs often come as combined audio and speech codec, i.e. a synthesis of advanced perception-based audio coding technology and state-of-the-art speech-production-based speech coding technology into a truly universal codec with optimum performance for both music and speech signals. Examples include MPEG-D Unified Speech and Audio Coding (USAC) or 3GPP's codec for Enhanced Voice Services (EVS) where the latter is also a communication codec with suitably low delay.

Coding of immersive audio programs

After stereo and surround sound, the next generation of formats that further increased spatial realism and listener immersion is "3D Audio", i.e. reproduction including height (higher and possibly lower) loudspeakers. Examples are the 22.2 or 5.1+4 (i.e. four height loudspeakers added on top of the regular 5.1) configurations. For embracing such formats, two key challenges needed to be solved:

The loudspeaker setup / "format" compatibility challenge: 3D audio content may be produced in many different loudspeaker layouts. Furthermore, reproduction in consumer homes will not happen with a large number of (e.g. 22) loudspeakers that may have been used for content production. Thus, a codec for 3D Audio needs to be able to reproduce 3D content on any available loudspeaker setup, adapting the content and providing best possible listening experience for this setup.

The paradigm compatibility challenge: 3D audio content can be produced in different paradigms, i.e. either as channel signals intended for a specific loudspeaker setup, or as object signals intended for specific spatial playback coordinates and properties, or as Higher Order Ambisonics (HOA). Although the latter two are independent of particular loudspeaker setups, they are very different signal representations. These different signal representations require different coding approaches in order to provide high signal compression while preserving spatial fidelity.

As a recent example, the MPEG-H 3D Audio codec addresses these challenges by integrating a core signal compression technology (USAC) that can support channels and objects and HOA signals when equipped with preprocessing to achieve additional compression.

Integration of coding and rendering

Many modern codecs do not only perform efficient compression and decoding of audio signals but also include aspects of signal rendering to output devices, such as headphones and loudspeakers. Examples for such efficient combined processing include:

Binaural rendering for headphones: headphone listening is a presentation format that is especially important in view of the vast number of multimedia-enabled mobile devices. Binaural presentation provides an immersive experience via headphones. Integrated coding and binaural rendering at the decoder comes at very low computational cost and permits use of personalized HRTFs/BRIRs. Examples include MPEG-D MPEG Surround and Spatial Audio Object Coding (SAOC) as well as MPEG-H 3D Audio.

Rendering for multiple loudspeakers: for loudspeaker presentation, an immersive listener experience can be achieved by directly rendering the decoded audio output to the target multi-channel loudspeaker layout, which can be attractive from computational complexity perspective. Several recent audio codecs offer the ability to render to a user selectable output layout, i.e. either one of the standardized loudspeaker setups (ranging from 5.1 to 22.2) or even to arbitrary setups. Examples include MPEG-D Spatial Audio Object Coding (SAOC) and MPEG-H 3D Audio.

Interactive sound modification capability

The ability to modifying a particular sound within a complex sound mixture at the time of decoding is receiving increased attention. Most prominently, dialog enhancement of coded sound can boost dialog for hearing impaired listeners. Interactive sound modification also enables creating personalizing sound mixtures (e.g. commentator vs. stadium sound for sports events, personal music remixes). This can be achieved by using sound objects and, for low bitrate applications, MPEG-D Spatial Audio Object Coding (SAOC).

Rate adaptive audio representation

There is an ever-greater consumption of streaming audio on mobile devices. While mobile channels invariably suffer from variable transmission channel capacity, they also have the capability to signal to the transmitting device the instantaneous channel capacity. Compressed audio formats that support seamless switching to lower (or back to higher) bit rates permit the system to deliver uninterrupted audio under changing channel conditions. An example of such a system is the MPEG-3D Audio coding format in combination with the MPEG DASH delivery format.

HIGH RESOLUTION AUDIO

Vicki Melchior and Josh Reiss, Chairs

High resolution audio (HRA) is an established mainstay in the professional and audiophile markets. Growth in the last several years focused on a number of new formats and on quality considerations for the internet delivery of downloads in both the new and older HRA formats, typically at increasingly high bit rates. Rapid growth in streaming now suggests that formats appropriate to more stringent bit rates will be useful, and certain of those are in development. The ability to stream can also facilitate the introduction of HRA into the mainstream marketplace. The HRA Technical Committee sponsors workshops, tutorials, and discussions highlighting significant aspects of these developments for the AES community.

Recent HRA Formats

Notable in the last several years has been the emergence and uptake of DSD as an independent encoding and distribution format. DSD is the term used by Sony and Philips for the single bit stream output of a sigma delta converter that, together with related processing, is used in storage and transmission associated with the production of SACDs. The original DSD oversampling rate of 64 Fs (64 x 44.1 kHz = 2.8224 MHz) was expanded to include both 128 Fs and 256 Fs. The main advantage of the higher rates is that the rise in shaped noise which occurs as a consequence of dynamic range processing in sigma delta modulators can be pushed considerably further out beyond the audio band (> 60 kHz), and with less quantization noise remaining in the audio band, than is possible with 64 Fs. The DSD signal is said to sound cleaner and more transparent at the higher data rates.

Also related to DSD is DXD, a designation for PCM at 352.8 kHz/24b championed by Merging Technologies as an intermediate stage in DSD processing. Single bit streams cannot easily be filtered or processed and so typically are converted to PCM at high sample rates to enable production. DXD has been used as a direct recording format for release in DSD, as an intermediate when both recording and releases are in DSD, and less often as a 352.8 kHz LPCM release format.

"Master Quality Authenticated" (MQA) is a proprietary LPCM codec introduced in 2014 by Meridian Audio and MQA Ltd. It combines newly developed filtering and signal processing with a hierarchical fold-down and bit packing scheme, thus enabling the streaming of high resolution quality files at data rates comparable to CD while embedding a lower bitrate version for conventional CD playback. The high resolution data is not lossless in the conventional sense of exact bit retention but rather incorporates bandwidth and dynamic range processing reflective of the band limits of music and the dynamic range imposed by microphone noise.

Virtually all new professional and consumer hardware and software supports at least a subset of the LPCM and DSD high resolution formats. The MQA codec is gaining industry support and increasingly appears in newer portables, streamers and players, especially audiophile gear, although a significant backward-compatibility characteristic is the playability of the embedded standard resolution file in the absence of a decoder.

Improved Converters, Filters, and Signal Processing

While audio designers have always sought to identify the sources of sonic deterioration brought about by processing and filtering, higher resolutions are both the outcome of, and drivers of, this search. There is at present an effort by manufacturers of high quality converters to address shortcomings attributed to the up-sampling chips and multi-bit sigma delta modulator chips used nearly universally in PCM DAC processing. Techniques include substitution of FPGA or computer-based up-sampling for that found on chips, custom filter design including minimum phase designs, increase of processing bit depth to double precision floating point (64b) or above, and custom sigma delta modulation and decimation. Chip makers have developed improved chips incorporating similar processing upgrades, plus improved noise shaping, jitter control, clocking and isolation. Such chips increasingly appear in HRA-capable hardware.

The theoretical and practical influence of filters on sound has long been debated and continues to be. There has been little formal study, but a recent AES convention paper reported audibility in double blind tests of several down-sampling filters typical of those in common CD usage (AES 137, Los Angeles; Jackson, et.al.) Also worth mentioning in this regard is the filtering approach in the new MQA algorithm, which incorporates (proprietary) non-traditional sampling, filtering, and convolutional methods based on triangular sampling kernels. The designers claim, in AES and internet papers, that significant sonic enhancements occur over conventional sampling and filtering, and are readily audible to mastering engineers.

Distribution

Distribution of HRA files is now primarily internet-based, although discs including Blu-ray, Pure Audio Blu-ray, SACD, and

DVD-A continue in various regions. HRA downloads remain a growth sector. Specialty websites ranging from large aggregators to small labels and orchestras offer new work and remastered catalog at PCM resolutions from 192 kHz/24b to 44.1 kHz/16b, and DSD at 64 Fs and 128 Fs. Consumers and the industry generally have become increasingly conscious of release quality, leading some websites to require that older data be sourced from the best available masters while eschewing further transcoding or other media transfers.

Streaming is rapidly displacing downloads in the mass market and will certainly impact high end and pro audio, although the extent remains uncertain. Network and storage bandwidths have improved, allowing a number of services (e.g., Qobuz) to stream losslessly compressed HRA, but the high bitrates of these files mitigate against their general use. This has led to a search for viable data-packing formats for HRA. DSD files are one-fourth the size of comparable PCM files but are still large for general streaming. Two developing formats are MQA, mentioned earlier, and in some uses, MPEG-4 SLS. The main limitation to MQA adoption as yet is the lack of enough MQA-encoded music. Warner Music Group is currently converting their catalog to MQA, and if other major labels follow there may be significant content in the near future.

Initiative to Extend HRA to the Broad Market

A joint initiative from the Digital Entertainment Group (DEG), Consumer Electronics Association (CEA), the Recording Academy, and the major labels to define and promote HRA continues. Their definition and set of provenance designators for HRA were released in 2014 and, while not universally liked, are in current use. The group initially sponsored demonstrations of HRA at many trade events, including the AES, and are presently co-sponsoring extensive promotions with major consumer electronics resellers (Best Buy in the U.S.). Their labeling system has been extended to HRA streaming services in the industry-wide expectation that streaming will soon become important.

Academic Studies

A recent paper published by the Queen Mary University of London audio engineering group applied meta-analysis statistical techniques to a collection of 20 previously published perceptual studies of HRA. The result showed a small but significant audibility of HRA in comparison to CD that greatly increased when participants were initially trained. The paper is noteworthy both for the result, and for the use of meta-analysis in this context. (J. Reiss, "A Meta-Analysis of High Resolution Audio Perceptual Evaluation", JAES vol. 64, pp. 364–379, (June 2016))

LOUDSPEAKERS AND HEADPHONES

Steve Hutt, Chair; Juha Backman, Vice Chair

TC-LH is comprised of experts in loudspeaker and headphone systems with skills in transducer design, manufacturing, measurement and proficiencies in audio system integration, sound propagation and acoustics.

Consumer Electronics

Streaming Devices. This is a growing market segment but has recently become subject to disruption from "Smart Speakers". These are wireless systems with voice command control and interface to service platforms as part of the "Internet of Things" (IoT) connectivity network. By some accounts, smart speakers are already among the highest selling types of loudspeakers. So far, playback quality appears to be less important than interactive capabilities.

Home Theater. The trend for in-wall, invisible & low profile systems continues and drives development of transducers with minimal depth. Sound Bars continue their popularity, some with improved performance over earlier generations.

Headphones. While headphone performance was arguably less important to some users than fashion, efforts for sonic accuracy are progressing with research to quantify ideal performance goals, aligning objective measurements with subjective preferences. Wireless connectivity is increasing, driven in part by device design and convenience.

Headphone Transducers. These are dominated by "full-range" moving coil designs. Balanced armature devices and multi-way systems are becoming more prevalent. Alternative topologies are entering the market such as planar and in-ear electrostatics, also available in multi-way configurations.

Virtual Reality. Systems are garnering new opportunities requiring improved headphone performance, including requirements for rendering spatial attributes accurately.

Hi-Resolution. Hi-Res audio growth is robust. As loudspeakers are often referred to as the "weak link in the audio chain" and the Hi-Res effort includes extending audio bandwidth ever higher, loudspeakers capable of Hi-Res playback will require careful consideration and development. "Super-tweeters" and planar drivers have been available for years with >60 kHz bandwidth. Temporal requirements for Hi-Res audio are under investigation.

Technology

Micro Drivers. Portable devices drive a big demand for micro-drivers. Challenges lie in developing higher acoustic output with greater low frequency and quality, but with miniaturization of drivers, enclosures and cost. Micro-driver topologies are dominated by permanent magnet/moving-coil motors, though not necessarily with axisymmetric voice-coils. MEMS (microelectromechanical systems), with integrated electronics are finding their way into devices and show great promise.

Neo Magnets. The attractiveness of neo magnets (Neodymium Iron Boron, NdFeB) lie in the size and weight ratio to magnetic potential vs. ferrite or other magnets. Driven by weight reduction and though more expensive than ferrite, a trend to neo in automotive and portable pro sound started in the 1990s. In late 2008 neo prices began to rise dramatically and by 2011 were near 10 times their 2008 price. Automotive companies and suppliers absorbed the cost but pro sound has begun to revert back to ferrite in non-flying systems. Additional demand sustaining high neo prices comes from emerging industries such as electric vehicles and wind generation. Currently, neo prices have fallen to ~3.5 times their pre-2008 norm but dysprosium (used in neo for high heat applications) remains even higher. Neo continues to be favored for high performance compression drivers and tweeters.

Alternative Topology Transducers

Air Motion Transformer (AMT) & Planar. AMT transducers are finding increased use in studio monitors and pro systems. Planar magnetic drivers are found in some high end pro systems and vehicles.

MEMS. MEMS used in arrays are being touted for package, weight and potential for beam steering.

DML & BMR. The trend of Distributed Mode Loudspeakers (DML) systems popularized in the 2000s has waned with the exception of a few niche applications. Balanced Mode Radiator (BMR) device improvements continue.

Electro-Magnetic Drivers using a field-coil in place of permanent magnet, are subject to recent patent filings motivated by efforts to hedge neo price instability. Efficiency. Automotive and pro sound can benefit most, but improvements in loudspeaker efficiency lags potential advantages, partly because of the intellectual complexity of loudspeaker voltage sensitivity vs. true efficiency. Maximum efficiency solutions will trend with the integration of fully coupled DSP/amplifier/loudspeaker systems.

Moving Magnet topology has been researched for years and is now used in a potentially disruptive high output low-frequency system utilizing closely coupled amplifier and DSP control circuitry.

DSP Corrective Systems are evolving with the benefit of compensating for deficiencies in transducers, and have great merit with micro-drivers. Advancements in DSP solutions are indeed admirable and useful, but DSP on its own has not negated the effort to develop improved transducers.

Directivity Control

Pro Sound continues to trend directivity control with proliferation of line array systems and low frequency cardioid arrays. Purpose designed transducers and waveguides continue to evolve along with system modularity. Line array systems utilize passive or automated mechanical adjustments for vertical tilt and alignment. Horizontal coverage can be adjusted by manual or automated adjustment of horn boundaries. A trend is evolving with advanced adaptive electronics to manage beam steering. This allows reconfiguring a system's directivity for real time conformance to architectural and ambient conditions. An alternative approach loads multiple transducers onto a single horn that functions as a high output point source.

3D Sound in vehicles has evolved from spatial novelty to trend, incorporating overhead sound. The mechanical challenge is to develop transducers that fit shallow mounting depth in the headliner, yet have directivity appropriate for 3D sound fields.

Cinema. The trend to 3D sound continues though the specification for loudspeaker directivity and power capacity is still evolving.

Zoned-audio has not prevailed in consumer space but is gaining traction in automotive sound. The goal is to create active and null zones so that each passenger listens to discrete sound not audible to other passengers. DSP is used to manage the propagation and phase of purpose built transducers with waveguides that exhibit specific directivity and in some cases "flat" depth to fit into headliners.

Standards

AES2-2012 (Standard) [Methods of Measuring Drive Units] up for review in 2017. X223 (Information Document in progress) focuses on developing methods to reliably repeat loudspeaker driver measurements in different locations. AES-X241, "End-of-line testing for loudspeaker drivers" (standard in progress). Target release May 2018.

NETWORK AUDIO SYSTEMS

Kevin Gross, Chair
Richard Foss and Thomas Sporer, Vice Chairs

Professional AV synchronization

Professional AV media networking requires a synchronization component. Early audio network technologies such as CobraNet and Livewire used proprietary techniques for synchronization. In 2002, the IEEE 1588 precision synchronization standard was published. Initial deployment was in industrial networking applications but it soon found applications in audio networking in AVB, Dante, and Q-LAN.

Back in 2007 SMPTE recognized the need for network synchronization for professional video networking applications. A requirements study produced a Request for Standardization requirement document in August 2009. Standards development ensued and SMPTE ST 2059 was published in March 2015. AES67 was developed between October 2010 and September 2013. Coordination between the two standards development groups resulted in a synchronization component in AES67 based on IEEE 1588, which is compatible with ST 2059. The compatibility between the standards was perhaps not widely appreciated until the AES published AES Standards Report R16 PTP parameters for AES67 and SMPTE ST 2059-2 interoperability in May 2016. SMPTE has held interoperability testing events for manufacturers working with ST 2059 and has included synchronization interoperability testing with AES67 devices in recent events.

Video IP networking

Video infrastructure in broadcast facilities is primarily based on the synchronous digital interface (SDI) standardized in SMPTE ST 259M. In 2007 SMPTE published ST 2022-6, which describes carriage of SDI signals over an RTP/IP connection. Since SDI carries as many as 16 channels of digital audio multiplexed with video, ST 2022-2 also defines a type of audio networking.

During NAB 2014 several broadcast vendors demonstrated ST 2022-6 primarily in a point-to-point configuration. There was a clear need to define a more network-centric framework that could help realize the full potential of IP video. The Video Services Forum initiated such a development effort in April, 2014.

Two approaches were considered; the first was built on ST 2022-6 and thus moved the legacy SDI approach onto a network. This effort produced a technical recommendation called TR-04. The second approach removed the SDI legacy component and uses previous work done in the Internet Engineering Task Force (IETF), RFC 4175 for flexible uncompressed digital video over RTP/IP. This effort produced TR-03 in November 2015. The audio component under TR-03 is AES67 and synchronization is accomplished using SMPTE ST 2059 (see above). These media networking techniques have been adopted by many of the major players in the broadcast industry. A trade association, AIMS, was formed in April 2016 to promote the technology and SMPTE was charged with creating standards, expected to be called ST 2110, from the technical recommendations.

Audio Contribution over IP (ACIP II)

This EBU working group published EBU Tech. 3368 Profiles in November 2014. This document deals with a set of parameters that describe how to transmit and receive audio streams. These are to ensure the successful decoding of audio based on the parameters received and negotiated when each specific connection is established. The specification has been implemented by a number of manufacturers and is in use by some broadcasting organizations.

The specification defined a new Session Description Protocol (SDP) attribute named ebuacip, along with a number of attribute values. The attribute is now registered with IANA.

The ACIP II group is now officially closed but remains ready to be resurrected should the need arise for any particular topics that might be within its scope.

At the time of this writing there are two areas of new work under consideration. These are (i) dual streaming between devices; (ii) SIP peering between broadcasters, which would allow the transfer of audio between organizations that currently use closed SIP systems simply through the sender making an appropriate call to the destination. Any work in either of these two areas would be expected to make use of existing protocols and standards to achieve these goals.

AES67 includes a mode that allows interoperability with the ACIP standard EBU TECH 3326. The AES has held interoperability testing events for manufacturers working with AES67, and recent ones have included testing with ACIP devices. The next AES67 interoperability event is planned for February 2017.