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 22 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

AUDIO FOR TELECOMMUNICATIONS

Bob Zurek, Chair
Antti Kelloniemi, Vice Chair

The trend in mobile telecommunications across the globe has been toward 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.

From a voice quality standpoint, the AMR Wide Band codec is in widespread use globally, and the 3GPP EVS (Enhanced Voice Services) codec has been finalized and is being developed for commercialization.

The audio recording capabilities of the devices have improved to offer high bit rate 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 compliment the video recording on the devices.

Voice controlled user interface in the mobile devices has become commonplace, with new advances being introduced each product generation such as natural language control, always-on voice recognition that does not require the user to touch the device, and user defined triggers to wake up the recognition engine.

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 benefited from algorithms that adapt uplink noise reduction or algorithm parameters based on noise type and level. Likewise downlink audio and content playback have benefited 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 in the 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 are reflected in the standards world where the scope of standards as well as their names have had to change to keep up.

AUDIO FORENSICS

Jeff M. Smith, Chair
Daniel Rappaport, Vice Chair
Eddy Bøgh Brixen, Vice Chair

Forensic audio has important applications in law and investigations. These continue to grow as the ability to record and share digital media proliferates. Therefore, it is important for the practitioner working with forensic audio to be trained in and apply processes related to proper evidence handling and laboratory procedures. Analog audio evidence, once commonly presented on compact cassette and micro-cassette tapes, make up an increasingly small percentage of cases. As such, digital audio evidence, which in practice is closely related to and often generated by computer forensics investigations, requires handling practices such as imaging physical digital media, hashing file duplicates, authenticating recordings, and recovering and/or repairing corrupt or carved files.

Modern challenges for the forensic audio examiner have expanded to include dealing with evidence from social networks and sharing websites such as YouTube and Facebook. These pose unique challenges in attributing provenance due to processing applied automatically by the hosting site. Additionally, growing concern for privacy rights has led citizens and providers of Internet services to question the authority that law enforcement entities may (or may not) have in requesting and accessing digital media hosted on these sites. Similarly, mobile phone manufacturers have adopted policies and operating system designs resistant to the collection of probative material and the use of "backdoor" access to password-protected devices. As social media awareness and use continue to evolve, so will public and commercial policies related to privacy rights.

The use of probabilistic methods in the analysis and presentation of scientific evidence is of increasing interest to the forensic sciences as a whole. In audio forensics it is of particular interest in voice comparison where an unknown speaker can be compared to a known suspect or set of suspects within a reference population in order to derive a Bayesian likelihood of similarity. We anticipate further developments in this field and the possible integration of methods based upon such techniques in other areas of forensic audio.

Difficulties in obtaining successful audio enhancement can be exacerbated by the lossy data compression algorithms commonly used within small digital recorders, data compression and bandwidth limited signals in telecommunications, and the non-ideal recording environments common to surveillance and security. In the presence of heavy data compression, classical algorithm designs may be challenged, and this may lead to the development of new processing technologies.

Numerous papers on audio forensics appear in the *Journal of the AES* and are presented at AES Conventions each year. Additionally, there have been five AES Conferences on audio forensics (AES 26th in 2005, 33rd in 2008, 39th in 2010, 46th in 2012, and 54th in 2014) with the next slated for 2016. Additionally, regular workshops and tutorials organized by the TC-AF exploring these and many other emerging trends appear at AES Conventions in the US and Europe each year.

CODING OF AUDIO SIGNALS

Jürgen Herre and Schuyler Quackenbush, Chairs

Introduction

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. 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 tremendous evolution since its commercial beginnings in the early nineties. Today, an audio codec provides many more functionalities beyond just bitrate reduction while preserving sound quality. A number of examples are given below.

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 bit rates.

First, 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 load on the encoder core is reduced significantly. Conversely, 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. This allows transmitting spatial sound even at low bit rates. Examples for such spatial pre/postprocessors are MPEG Parametric Stereo (PS), MPEG Surround, MPEG Spatial Audio Object Coding (SAOC), and, as a more complex system, MPEG-H 3D Audio.

Second, to alleviate the burden of the core codec, various pre/post processors were devised that 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 bit rates. Well-known examples for such technologies are Spectral Bandwidth Replication (SBR), Harmonic Bandwidth Extension (HBE), and Intelligent Gap Filling (IGF).

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 audio coding technology and state-of-the-art 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).

Coding of 3D or immersive audio

After stereo and surround sound, the next generation of formats that further increased spatial realism and listener immersion involve 3D or immersive audio, i.e., reproduction including height (e.g., 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. While the time is ready for embracing such formats, two key challenges need to be solved.

First, the loudspeaker setup / "format" compatibility challenge: immersive audio content is produced today in many different formats, i.e., for different loudspeaker setups, thus impeding the introduction of immersive audio because of the incompatibility between them. Furthermore, reproduction in consumer homes will not happen with a large number of (e.g., 22) loudspeakers. Thus, a codec for immersive audio needs to be able to reproduce immersive content on any available loudspeaker setup, adapting the content and providing best possible listening experience for this setup.

Second, the input format compatibility challenge: immersive 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). While the latter two are independent of particular loudspeaker setups, these are different approaches. An audio codec for immersive audio needs to reconcile these paradigms, preferably by being equipped to support all of them.

As a recent example, the MPEG-H 3D Audio codec addresses the challenges by integrating many core technologies, such as MPEG-D Unified Speech and Audio Coding (USAC), and Spatial Audio Object Coding (SAOC), and additionally supports channels, objects and HOA as input signal formats.

Binaural rendering on headphones

Some codecs provide binaural rendering to headphones as an additional feature that is especially important for mobile devices. Headphone listening is another presentation format beyond that of any of a variety of possible loudspeaker layouts, and binaural rendering to headphones provides an immersive experience for what might be the most pervasive mode of audio consumption. The rendering comes at very low computational cost and supports use of personalized HRTFs/BRIRs. Examples include MPEG-D MPEG Surround and Spatial Audio Object Coding (SAOC) as well as 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 more and more attention. Most prominently, dialog enhancement of coded sound provides the possibility of boosting (or attenuating) dialog for hearing impaired listeners. Interactive sound modification can also be used for 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 bit rate applications, MPEG-D Spatial Audio Object Coding (SAOC).

Rate adaptive audio representation

With an ever-greater consumption of broadcast or streaming audio on mobile devices, the reality of variable transmission channel capacity needs to be addressed. Recent systems support point-to-point transmission in which a mobile device can signal to the transmitting device the instantaneous channel capacity (e.g., IP-based systems). 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, as operating over IP channels.

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 US 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 electro-acoustical 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.

An additional technology that has been refined over the years is the use of induction loop systems in public spaces. The technique is very simple yet very attractive to users, since a hearing aid user needs no additional equipment to receive the signals (as long as the hearing aid has a telecoil), which is socially much more acceptable than having to wear a conspicuous additional device. In its simplest form, a single conductor is installed in the form of a loop at floor or ceiling level, surrounding the area to be covered by the transmission. An audio-frequency current, generated by an amplifier from one or more microphones or a recording, flows in the loop conductor and produces a magnetic field with and outside the loop. It is practicable to generate a sufficiently strong and uniform field pattern with amplifiers of reasonable cost and complexity. In addition to coupling sounds in public spaces, induction transmission coils are commonly available in cellular and cordless phone systems, so as to enhance the signal to noise ratio associated with telephone communications.

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 electroacoustic 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 is planning its second conference devoted to technological solutions to hearing loss. Pending confirmation, this conference will most likely occur in late 2015 or early 2016.

HIGH RESOLUTION AUDIO

Vicki Melchior and Josh Reiss, Chairs

Growth in high resolution audio (HRA) over the last several years has been robust. HRA is an established mainstay in the professional and audiophile markets. However the arrival of new formats and refinement of the associated processing, along with the growth of internet delivery, and now a significant industry effort to make HRA a mainstream format, all herald an interesting and promising next few years. The HRA Technical Committee sponsors workshops, tutorials, and discussions highlighting significant aspects of these developments for the AES community.

New HRA formats

Especially notable in the last two years has been the rapid 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 f_s (64 × 44.1 kHz, or 2.8224 MHz) has now been expanded to include both 128 f_s and 256 f_s. The main advantage of the higher rates is that the rise in shaped noise that 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 f_s. 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 / 24 bit championed by Merging Technologies as an intermediate stage in DSD processing. Single bit streams cannot easily be filtered or processed and so are typically converted to PCM at high sample rates to facilitate production. More than just an intermediate stage, DXD's uses are evolving with some recording engineers employing it as a direct recording format for release in DSD, as an intermediate between DSD record and release, and possibly in the future as a 352.8 kHz PCM release format.

This trend to higher sampling rates in both PCM and DSD is supported by consumer and professional hardware. Many current DACs and ADCs support both PCM and DSD. New converters, software, and even portables increasingly support PCM from Red Book (44.1 kHz / 16 bit) up to 384 kHz / 32 bit and DSD to 256 f_s as the industry continues to explore both the merits and the degree of consumer interest in these formats. An open standard for packing DSD into PCM frames known as DoP has been adopted by major manufacturers to facilitate transfer of DSD across USB interfaces as well as AES and SPDIF.

Improved converters, filters, and signal processing

While high quality audio has always sought to define the sources of sonic deterioration associated with processing and filtering music data, high resolution and now 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 upsampling chips and multi-bit sigma delta modulator chips used nearly universally in PCM DAC processing. Techniques include substitution of FPGA or computer-based upsampling for that found on chips, custom filter design including minimum phase designs, increase of processing bit depth to double precision floating point (64 bit) or above, and custom sigma delta modulation and decimation. Several chip makers however have developed improved chips incorporating similar processing upgrades, plus improved noise shaping, jitter control, clocking, and isolation. Such chips are increasingly appearing in new HRA-capable hardware.

The theoretical and practical influence of filters on sound has long been debated, and a new test program initiated by Meridian Audio seeks to explore some of the audibility questions. In an important first paper given at the AES 137th convention, Meridian authors H. M. Jackson et al. measured the audibility in double blind tests of downsampling filters typical of those used in CD preparation when such filters were applied to a higher resolution stream without decimation and played through a high quality audio system. Their result disputes that from an earlier paper by E..B. Meyer and D..R. Moran (*JAES*, 55: 775–779, 2007) and provides evidence and a likely mechanism for an audible distinction between CD and higher resolutions.

Distribution, storage, and replay

Distribution of HRA files is now primarily internet-based. Download websites ranging from large aggregations down to individual labels and orchestras now exist and offer both new work and remastered back catalog. PCM resolutions ranging from 192 kHz / 24 bit to 44.1 kHz / 16 bit are available, and DSD at 64 f_s and 128 f_s is increasingly available. PCM above 192 kHz and DSD at 256 f_s are not yet significant factors but DAC manufacturers are including support for them anyway due to the rapid upward trend in bandwidths. FLAC, WAV, and AIFF are the dominant PCM transmission formats. Streaming is likely to supplement or replace downloading in the future as it is doing now with compressed music and lower resolution video. Although streaming bandwidths currently limit music resolution to losslessly compressed CD, a new codec designated MQA was recently introduced by Meridian Audio that is said to losslessly encode higher resolutions at bit rates slightly below those of CD. If successful, MQA may greatly influence the streaming of high res audio.

The emphasis on downloads correlates with the continuing strong trend toward adoption of computers, file servers and portables into all areas of music including the traditional two-channel audiophile music marketplace. Blu-ray movies incorporating HD audio also continue to sell well despite the continued decline of physical media, and there is a small dedicated market in high quality audio-only Blu-Ray discs.

New markets

Notable are two new initiatives to bring high resolution audio into the mainstream mass market. Behind these efforts are several factors: first, the considerable business success of the larger HRA download websites in the audiophile market, and then, a broader quest for higher quality as a result of the many complaints about the ubiquity of compressed low bit rate audio. Pono, an HRA music download service coupled with a well-engineered portable music player, is the product of multiple years of effort by the artist Neil Young. Pono is set to begin sales in late 2014. The second is a significant combined initiative from the Digital Entertainment Group (DEG), Consumer Electronics Association (CEA), the Recording Academy, and the major labels. They have developed an HRA definition and a set of provenance designators for future releases, and are currently sponsoring talks and demonstrations of HRA at trade events, including the AES 2014 Los Angeles convention. Provenance has been a major source of consumer complaint in the past because many DVD-A, SACD, and downloaded files labeled high resolution were merely upsampled redbook. Thus the optional use of designators is an attempt to redress the issue.

Also of note are several initiatives to make multitrack audio available to the research and education communities. These include the Open Multitrack Testbed, MedleyDB, the Free Multitrack Download Library, and the Structural Segmentation Multitrack Dataset. Many of these include better than redbook quality tracks, stems, and mixes and typically contain content available under Creative Commons licensing, allowing some degree of reuse or redistribution.

MICROPHONES AND APPLICATIONS

Eddy B. Brixen, Chair
David Josephson, Vice Chair

The microphone is an amazing device. What piece of audio equipment 20–50 years of age would be considered a hot pick for modern recording? The microphone it is.

Oldies but goodies(?)

In the marketplace of today, we find many old designs still manufactured. We find many new brands offering products that 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 partly explained by the opportunity of making good business on the general assumption that exotic looking microphones provide exotic audio.

Microphones as an industrial product

Microphones are designed and manufactured for many purposes other than music production. Communication is the area that implement the highest number of units. However, in microphones for industrial use like providing feed back signals for control systems is still getting more important.

Transducer technology

Even though traditional transducer technologies dominate the market, there has been some interest in development of new ideas.

MEMS

The micro-electronic mechanical systems (MEMS) are improving. Regarding the specifications, they still do not provide any advantages that on a commercial basis could not be achieved more efficiently by traditional transducer technologies. However, in array systems, which consist of a larger number of units, MEMS technology is finding its way due to the low cost and the fact that some of the inferior performance is to a reasonable degree overcome by technical work arounds, primarily provided by DSP.

Laser

Laser technology has been in use for years. So far state of the art was presented by NASA in 2007 in tech briefs: "A New Kind of Laser Microphone Using High Sensitivity," (Chen-Chia Wang, Sudhir Trivedi, and Feng Jin). They demonstrated experimentally a new kind of laser microphone using a highly sensitive pulsed laser vibrometer. By using the photo-electromotive-force (photo-EMF) sensors, they presented data indicating the real-time detection of surface displacements as small as 4 pm.

Ultrasonics

In 2012 Merkel, Lühmann, and Ritter presented their experiment on creating a virtual microphone by the aid of ultrasonics. A highly focused ultrasound beam was sent through the room. At a distance of several meters, the ultrasonic wave was received with an ultrasonic microphone. The wave field of a common audio source was overlaid with the ultrasonic beam. It was found that the phase shift of the received ultrasonic signal obtains the audio information of the overlaid field. So the ultrasonic beam itself acts as sound receiver, there is no technical device like membranes necessary at direct vicinity of sound reception. (See AES convention paper 8587).

Ionic microphones

In 2012 Akino, Shimokawa, Kikutani, and Green presented a study of ionic microphones: Diaphragm-less ionic loudspeakers using both low-temperature and high-temperature plasma methods have already been studied and developed for practical use. This study examined using similar methods to create a diaphragm-less ionic microphone. Although the low-temperature method was not practical due to high noise levels in the discharges, the high-temperature method exhibited a useful shifting of the oscillation frequency. By performing FM detection on this oscillation frequency shift, audio signals were obtained. Study results showed that the stability of the discharge corresponded to the non-uniform electric field that was dependent on the formation shape of the high-temperature plasma, the shape of the discharge electrode, and the use of inert gas that protected the needle electrode. (See AES convention paper 8745).

Optical wave microphone

In 2013 Toshiyuki Nakamiya and Yoshito Sonoda presented a solution for an optical wave microphone. An optical wave microphone with no diaphragm, which uses wave optics and a laser beam to detect sounds, can measure sounds without disturbing the sound field. The theoretical equation for this measurement can be derived from the optical diffraction integration equation coupled to the optical phase modulation theory, but the physical interpretation or meaning of this phenomenon is not clear from the mathematical calculation process alone. In this paper the physical meaning in relation to wave-optical processes is considered. This property can be used to detect complex tones composed of different frequencies with a single photo-detector. (See AES convention paper 8924).

Membrane material

Graphene: in 2014 Todorovic et al. presented recent trends in graphene research and applications in acoustics and audio-technology. The possibilities of application of single or multi-layer graphene as membranes in transducers were the scope of the research of the graphene group. FEM and experimental analysis of single and multi-layer graphene, as well as realization of the first samples of acoustic transducers, reported in progress. (See AES convention paper 9063).

Microphones and directivity

There is a still increasing interest on microphones with high—and frequency independent—directivity. Array designs have been around for years. Today array technologies are applied in laptops and handheld devices. Newer designs try to obtain constant directivity by applying sophisticated algorithms for beam forming.

Multichannel microphones

Single unit multichannel microphone technology is to a large extent building on Higher Order Ambisonics. In connection with the applied DSP for the control of beam forming, it is obvious to split signals into several channels. NH acoustics' Eigenmike is an example.

Microphone for 22.2 multichannel audio

In 2013 Ono, Nishiguchi, Matsui, and Hamasaki presented a portable spherical microphone for Super Hi-Vision 22.2 multichannel audio.

NHK has been developing a portable microphone for the simultaneous recording of 22.2 ch. multichannel audio. The microphone is 45 cm in diameter and has acoustic baffles that partition the sphere into angular segments, in each of which an omnidirectional microphone element is mounted. Owing to the effect of the baffles, each segment works as a narrow angle directivity and a constant beam width in higher frequencies above 6 kHz. The directivity becomes wider as frequency decreases and that it becomes almost omnidirectional below 500 Hz. The authors also developed a signal processing method that improves the directivity below 800 Hz. (See AES convention paper 8922).

Digital adaptation

Innovation in the field of modern microphone technology is to some degree concentrated around the 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 the microphones with integrated interfaces are available. However, different widely implemented "device-to-computer" standards like the USB and the FireWire standards—which are not specifically reserved for audio—also have been applied in this field. These standards have become de facto standards for the interfacing of various types of digital equipment in the area of home recording, including microphones.

Microphones and IP

Audio networking has for a long time been implemented in all corners of audio systems design using, for example, Ethernet based networks like DANTE, Ravenna, Q-LAN, etc. However, not until recently has this technology been implemented in microphones.

In 2014 Audio-Technica introduced a Network Microphone taking advantage of the DANTE network protocol control features and the low latency.

Standards

AES Standards Committee works on standardizing control functions from AES42 for the implementation in IP microphone solutions.

Materials

It should be noted, that 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 fewer dynamic microphones or rising prices.

NETWORK AUDIO SYSTEMS

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

AES67

AES67 is an interoperability standard for high performance audio over IP. It can be implemented as an interoperability mode or it can be used as a system's native low level audio networking protocol. The standard is tightly focused on getting audio across a network using standard protocols. Since publication in September 2013 it has been downloaded nearly 1000 times and several manufacturers have announced AES67 implementations and plans for implementation. An interoperability "plugfest" test organized by the AES occurred in October 2014 in Munich; another is being planned for North America in 2015.

MNA

The Media Networking Alliance is a new trade association whose purpose is to provide industry support for the proliferation of AES67. This is done by providing information and resources to the industry as well as promoting the benefits of AES67. Its focus is on three areas of activity, technical, marketing, and education. The MNA held its first meetings and presentations at the recent AES convention in Los Angeles. The presentations were well attended and included a lot of dialog and discussion with audience members. Since then, the MNA has created a web site (http://www.medianetworkingalliance.com/) and has received a lot of interest from manufacturers and others in becoming members.

AVB

AVnu Alliance announced the first certified switch at the end of 2013 and first certified pro audio endpoint in mid 2014. New certification programs to be announced in 2015 include automotive and pro video. The IEEE has expanded the scope of AVB—now denoted as Time Sensitive Networking (TSN)—with new features such as ultra-low latency and seamless redundancy. The AVnu Alliance has announced that it will certify TSN features in addition to AVB product features.

ACIP2

ACIP2 is an EBU working group addressing the contribution of audio material over IP. Participants are members of the EBU and manufacturers from all over the world. ACIP2 was set up as a follow-up group to ACIP. Basing on the formerly produced Tech. Doc. 3326, a standard for Audio Contribution over IP (known as N/ACIP), ACIP2 answers new rising questions and harmonizes the new world of audio over IP for broadcasters. During last IBC a new version of Tech. Doc. 3326 has been agreed on, and also a Tech. Doc. on profiles (a set of parameters describing how to transmit and receive audio streams and for the decoder to successfully decode the audio, based on the sent parameters) will be published by end of this year. Ongoing work covers SIP infrastructures, control, and management. Recently, the North American Broadcasters Association (NABA) has shown great interest in this work.

JT-NM

A Joint Task Force on Professional Networked Streamed Media (JTNM) was formed in 2013 by The European Broadcasting Union (EBU), The Society of Motion Picture and Television Engineers (SMPTE) and The Video Services Forum (VSF). The task force collected and categorized use cases for broadcast applications of media networking and then invited manufacturers and technology providers to indicate which use cases are addressed by existing and emerging technology. The task force completed phase 1 work with the publication of a Gap Analysis Report that compared use cases with technology capabilities. The report is available for download at https://tech.ebu.ch/docs/groups/jtnm/GapAnalysisReport_231213.pdf. The task force has since embarked on phase 2 projects including definition of a reference model for media networking.

Internet performance

Audio networking relies on realtime performance from networks. Due to a variety of factors, notably buffer bloat, delays across the internet can be measured in seconds. Active queue management (AQM) and fair queuing are being developed and deployed to reduce these delays. The IETF AQM and RMCAT working groups are preparing proposed standards in these areas (https://datatracker.ietf.org/doc/draft-nichols-tsvwg-codel/, https://tools.ietf.org/html/draft-pan-aqm-pie-02 and https://tools.ietf.org/html/draft-hoeiland-joergensen-aqmfq-codel-00).

Lip sync standard

Synchronizing separate audio and video streams in a broadcast plant and other facilities has been a detailed and error-prone engineering problem. With the inclusion of networked AV in these applications, the situation appeared to only be getting more difficult. However, in June 2014, the IETF published RFC 7272 that continues work started by ETSI and establishes an architecture and communication protocols for time synchronizing multiple media streams with sub-nanosecond accuracy. Any media networking based on RTP may use the techniques described in this standard. Other use cases addressed include social TV, video walls, and phase-coherent multichannel audio (e.g., stereophonic sound, line arrays, mic arrays and wave field synthesis).

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 microprocessor devices and programmable gate arrays, solid-state memory, and disk drives. 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. Also the number of DSP-related papers in other specific application fields increases yearly. DSP is also a key field for other professional conferences, including those sponsored by IEEE, ASA, and CMA.

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 in word length and sampling frequency), sophistication of algorithms, as well as quality and availability of consumer-ready hardware building blocks, such as special signal processors (DSPs), sampling rate converters (SRCs), digital audio network chips, ADCs, and DACs.

Third, 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 upconverting 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 the home AV environment there is a trend to use closely placed smaller loudspeaker and soundbars using algorithms that allow adequate spatial effects mimicking multi-speaker systems. The increasing number of people listening to music over headphones and mobile devices requires better algorithms for translation of music content to binaural format. In connection with head trackers similar algorithms are used in virtual reality systems.

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, as well as for ADCs, DACs, and SRCs with incorporated linear-phase FIR filters and thus introduce more latency.

Fourth, 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, de-reverberation, echo cancellation, acoustical feedback suppression, intelligent automixing, and steerable microphone arrays are expected in the audio teleconferencing field. Many of them are also suitable for modern hearing aids. In music production and mastering so called plug-ins enjoy great popularity and almost completely replaced hardware studio processors like PEQs, dynamics, reverb, and other effects. However there is still space for psychoacoustically related improvements. A new trend here is digital modeling of legendary analog gear.

And still there is a need for development of basic algorithms and transformation, better suitable for representing features of audio signals. There are complete analytically derived filters (not using analog reference), constant-Q frequency-domain and mixed time-frequency-domain transformations. Also in latency-minimized convolution algorithms the last word might be not still pronounced. Statistical and genetic algorithms, as well as fuzzy logic and neural network algorithms are more and more involved helping to create intelligent audio algorithms.

Audio coding algorithms as well as speech recognition, synthesis, and speaker identification are also based on audio signal processing technology such as described above, but are not currently the main focus of this committee, as they are now handled by specialised committees or organizations of their own.

Sixth, 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.

Seventh, 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 portion of the standard studio processing algorithms.

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

SPATIAL AUDIO

Sascha Spors and Nils Peters, Chairs

Overview

The AES Technical Committee on Spatial Audio addresses fundamental issues in production, transmission, and reproduction of spatial audio content employing loudspeaker and headphone-based techniques for creating immersive auditory illusions.

An increasing number of conferences, standards activities, and product developments in the last few years indicates a growing interest from industry and academia in spatial audio.

Object-based and scene-based audio

Audio content has been conventionally delivered as a channel-based mix. The audio mix is intended to be reproduced in a predefined standardized loudspeaker layout, such as 2.0 stereo or 5.1 surround. Consequently, a consumer with a non-standard loudspeaker layout may be unable to reproduce the audio content as intended.

Object-based audio is an approach to deliver content without being constrained to a standardized loudspeaker layout. The individual audio objects are transmitted separately, together with metadata describing their spatial properties. On the consumer side the audio objects are panned according to the consumer's loudspeaker layout. Further the consumer could adjust the audio mix in real time. For instance, she could change the loudness or language of the commentator independently from other audio elements. In the last few years object-based audio has been successfully deployed in the cinema context. This trend will likely continue and trickle down to home cinema, television, and mobile entertainment applications.

Higher-Order Ambisonics (HOA), a scene-based audio technique, is another way to circumvent the aforementioned constraints of channel-based content. Independent from the reproduction layout, HOA describes the sound field based on spherical harmonics. For the audio reproduction, the HOA data are rendered according to the desired loudspeaker layout or into binaural headphone feeds. HOA can either be created from single-channel audio tracks within a digital audio workstation or it can be derived from microphone-array recordings, a compelling use case for live-recording scenarios. Because HOA received significant attention in recent audio standardization and commercialization efforts and is supported by an active research community, it will presumably become a popular alternative to channel-based audio.

In general, hybrid content delivery is likely. For instance, object-based audio elements are transmitted in combination with conventional channel-based content or in combination with scene-based HOA recordings. Hybrid content delivery will account for the individual preferences of content providers and for technical limitations, such as bandwidth capacities.

Headphone-based reproduction

In many situations spatial audio reproduction via headphones is desired. This so-called binauralization can be achieved by filtering the audio content with head-related impulse responses (HRIRs) or binaural room impulse responses (BRIRs).

HRIRs are to some extent unique to an individual. The process of capturing individual HRIRs is time-consuming, requires specific equipment and precise measurement conditions. Yet, the optimal listening experience relies on binauralizing the audio content using a listener's own HRIRs. Current compromises involve approximations of HRIRs, through, for example, the database selection of well-matching HRIRs, or novel image analysis techniques.

With image analysis techniques, HRIRs can be estimated with the help of photos from the listener's pinnae. Also, high-resolution 3D head scans can lead the simulation of the sound field inside the ear canals and at the pinnae.

In the database selection approach, many different methods exist for rapidly selecting well-matching HRIRs from large HRIR databases, for example, selection based on geometric head and ear features. An increasing number of HRIR databases have become available, necessitating a standardized way to store and exchange HRIRs or their parameterized versions. AES experts developed the AES-X212 HRIR file format, which provides a flexible and scalable solution to store and exchange HRIR-related data. If the AES-X212 file format is largely deployed, we can expect an increase in the use of future binaural audio technologies.

Compared to static HRTFs, research demonstrates that binauralization based on dynamic head movement enhances the immersive audio experience. To take advantage of these dynamic cues, top tier consumer products already feature head-tracking. Dynamically binauralization may become available in more consumer products as long as head-tracking technology decreases in cost.

Spatial audio coding

Currently there is a lot of activity in the area of spatial audio coding and delivery of immersive audio content. Compared to previous spatial audio coding standards, the new generation of audio codecs has an increased coding efficiency, supports different content formats (e.g., channel-based, object-based, scene-based Higher Order Ambisonics), and is equipped with features, such as dynamic range control, interactivity for audio objects, and adaptive rendering to the consumer loudspeaker layout as well as to headphones. For the best possible adaptation of the spatial audio content to different reproduction scenarios, the embedding of content-related metadata into the bitstream becomes increasingly important.

Other ongoing activities strive to formalize ways to describe (uncompressed) spatial audio content for exchange and archival purposes. Those formats may be fed into spatial audio codecs for the content delivery.

Loudspeaker layouts

The most common consumer loudspeaker layouts for spatial audio are 5.1 and 7.1 surround. While these setups are horizontal-only, the next generation of loudspeaker setups incorporate elevated loudspeakers to create immersive audio experiences. Several consumer loudspeaker layouts range from 7.1 plus 2 or 4 elevated loudspeakers up to NHK's 22.2.

The increase in the number of loudspeakers comes with an additional purchasing cost and installation effort for the consumer. To account for physical restrictions in the placement of loudspeakers, the trend is to accommodate irregularly placed (non-standardized) layouts. This is often accompanied by automatic calibration techniques basing on acoustic measurement of the loudspeaker positions. To further simplify the installation and reduction of cabling costs, wireless loudspeaker setups are promising.

An emerging alternative solution to the problem of loudspeaker placement is loudspeaker frames around a TV, or soundbars below.

Spatial audio on mobile devices

The success of smartphones, phablets, and tablets as media devices creates demand for spatial audio solutions optimized for mobile devices. Due to the power constraints of mobile devices, DSP algorithms that either consume low-power or can dynamically trade-off processing quality with power consumption are desired.

Spatial audio on mobile devices is often reproduced binaurally via headphones, or with transaural (cross-talk cancelling) methods via the built-in near-field stereo loudspeakers. For the latter, the quality, variability, and placement of the built-in loudspeakers affect the reproduction quality and should be considered. To guarantee a dropout-free delivery of audio content over the network, several techniques exist that can seamlessly accommodate changes in bandwidth and latency.

For the creation of spatial audio content, top tier mobile devices are now capable of recording 5.1 surround sound.

Research on sound field synthesis

The deployment of Higher Order Ambisonics (HOA) and Wave Field Synthesis (WFS) arrays in research labs continues. For example, current research in sound field synthesis investigates the perception of synthetic sound fields. Besides localization, a special interest lies on coloration since this has major impact on the quality. The practical realization of arrays and synthesis techniques including height is a further research focus these days.

Besides sound field synthesis, a related and emerging field of interest, which was also the focus of the 52nd AES conference, is the active control of sound fields. Sound field control deals with aspects such as the active compensation of reflections using multiple loudspeakers, or the creation of different acoustical zones in close proximity where different audio content can be reproduced.

Content production

New content production tools are necessary for creating professional audio content beyond 7.1, such as content including height, object-based content, and content in Higher Order Ambisonics.

A number of digital audio workstations, the prefered working environments for professional content creation, already feature a flexible bus structure to support novel content formats, and novel spatial audio effects have started to take advantage of this.

Market trends point to growing interest in audio immersiveness, and immersiveness relies on those in the field of spatial audio having access to more and better content that will test, exploit, and showcase emerging technologies. Consequently, there is a lot of room for new content production and authoring tools that simplify the content production workflow and enable new immersive audio experiences.

TRANSMISSION AND BROADCASTING

Kimio Hamasaki and Stephen Lyman, Chairs
Lars Jonsson, Vice Chair

The growth of digital broadcasting is the trend in this field. Digital terrestrial TV and radio broadcasting have been launched in several countries using the technology standards listed below. Loudness and True Peak measurements are replacing the conventional VU/PPM methods of controlling program levels, which has largely eliminated significant differences in the loudness of different programs at home. "Object-based audio" and "immersive audio" are currently the subject of great interest in the EBU, ITU, AES, and SMPTE.

Digital terrestrial TV broadcasting

In Europe DVB-T2 has been deployed in several countries for HD services. ATSC is used in USA, Canada, Mexico, and Korea, while ISDB-T is employed in Japan, Uruguay, 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 fulfils this role in the USA and Canada. 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 has been standardized and chip-sets are available, while ISDB-T is used in Japan. DMB is employed in Korea and there have been a few trials in Europe. In the USA, the Advanced Television Systems Committee (ATSC) has begun development of a non-backwards compatible system with next-generation video compression, transmission and Internet Protocol technologies via local broadcast TV stations. This effort is termed "ATSC 3.0" and is covered in some detail within the professional journals and the industry trade press. It is too early to fully characterize the eventual results.

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 Ultra-High Definition image (4k video), several broadcasters are experimenting with immersive audio (for instance, 3D-multichannel-audio, Ambisonics, wave-field synthesis, directional audio coding).

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.

The fundamental document upon which all other work was based is 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." Importantly, it not only documents loudness but a standardized method to measure true peak in digital audio streams. This is a vital, and often overlooked aspect, as the loudness measurement is a time averaged value that should not be changing rapidly. So the operator needs a reliable method of ensuring they are not clipping their signals. The true peak provides this method.

The next document released (in historical order) tat was based upon BS.1770, was ATSC: A/85 "Techniques for Establishing and Maintaining Audio Loudness for Digital Television." A/85 covers the related topics exhaustively: Making Loudness Measurements; Target Loudness and True Peak Levels for Content Delivery or Exchange; Metadata Management Considerations Impacting Audio Loudness; Methods to Effectively Control Program-to-Interstitial Loudness; Dynamic Range Management; and last (and perhaps least obvious) Audio Monitoring Setup.

Starting roughly a year after the ATSC drafting group, the EBU drafting group has produced a similar set of documents:

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

EBU has, beyond minor revisions of R128 and its belonging Tech docs, a major deliverable planned for 2015 concerning the revision of the loudness range algorithm (Tech 3342).

The seeming "disconnect" between A/85 and the R128-family can be best understood by the very different scopes of each document. A/85 is focused on DTV audio only, while R128 attempts to cover almost any audio delivery mechanism (seemingly including analog). Beyond the seeming differences between A/85 and R128, the US Congress adopted a seemingly simple, but with unintended consequences, law called CALM ("Commercial Advertisement Loudness Mitigation Act"). Free TV Australia produced Operational Practice OP–48 (which the Australian Government made mandatory also). ARIB in Japan has issued. TR-B32 1.2 "Operational Guidelines for Loudness of Digital Television Programs."

Object-based audio

"Object-based audio" is currently the subject of great interest in the EBU, ITU, AES, and SMPTE. At its heart is the metadata describing the audio content that will enable its correct handling along the broadcast chain and optimal rendering for the individual listener. As part of this, the EBU has produced the "Audio Definition Model," published in EBU Tech Doc 3364 (which is currently the subject of standardization activities in the ITU). In the ITU-R there are two relatively recent publications outlining the future: Report ITU-R BS.2266 "Framework of future audio broadcasting systems," and Recommendation ITU-R BS.2051 "Advanced sound system for programme production."

Immersive audio

EBU FAR strategic program group has founded a work group for 3D audio, renamed Immersive Audio. Object-based production embraces 3D audio production, 3D audio codecs, 3D audio broadcasting, 3D audio file formats, and 3D audio creativity

The MPEG Committee has begun work on "immersive sound" coding.

Audio/video ("lip") sync

The audio/video-sync issue remains unsolved but is being discussed in digital broadcasting groups. SMPTE is close to issuing new standards for measuring the time differences between audio and video.

Technology Trends in Audio Engineering

A report by the AES Technical Council

INTRODUCTION

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.

Last month we published a comprehensive report from many of the Technical Committees. This month we add a report from the TC on Semantic Audio Analysis.

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

SEMANTIC AUDIO ANALYSIS

Mark Sandler, Chair
Gautham Mysore and Christian Uhle,
Vice Chairs

INTRODUCTION

Semantic audio analysis is concerned with the extraction of meaning from audio signals and with the development of applications that use this information to support the consumer in identifying, organizing, and exploring audio signals and interacting with them. This includes music information retrieval and semantic web technologies, but also audio production, education, and gaming.

Driven by the increasing availability of digital audio signals and compared to other audio technologies, semantic audio analysis is a relatively new research area. Semantic refers to the study of meaning of linguistic expressions and to the interplay between descriptors and their meaning. Semantic technology involves some kind of understanding of the meaning of the information it deals with, for example a search engine that can match a query for "bird" with a document mentioning "eagle." In audio, it incorporates methods and techniques for machine learning, digital signal processing, speech processing, source separation, perceptual models of hearing, and musicological knowledge. It also uses several principles from computer science, including metadata, ontologies, first order logic and brings these together with more traditional signal-based technologies. By doing this, musical features extracted from audio can be represented symbolically, which opens up the full potential of artificial intelligence and natural language processing techniques to be smoothly integrated into future music informatics systems.

FROM MUSIC COLLECTIONS TO MUSIC STREAMING SERVICES

Semantic analysis supports managing and browsing large music collections. It enables the identification of artist and title of a recording, the classification of the musical genre, finding alternative recordings of the same piece, and identifying and removing duplicate songs. It also helps the users to take a journey through a musical world, for example by finding similar songs, but also by finding similarities between artists such as birth place or musical influences. New ways of browsing large collections have been developed, e.g., using audio thumbnails, i.e., a short and representative segment of a song, and the search of songs by humming.

One major trend is to use mobile devices and services for on-demand streaming instead of accessing a physical music collection. Music streaming services offer their users access to millions of tracks on demand and functionalities for discovering new music. The number of users and the share of revenues have constantly increased in recent years.

With the advent of music streaming service, the volume of accessible music has increased even more, and with it the demand for semantic audio analysis and tools for exploring new music. A recent trend is to improve the performance of these tools and the user experience by adapting their functioning to the musical preferences and the listening behavior of the user.

DECOMPOSITION AND INTERACTIVITY

Techniques for both guided and unguided source separation are applied in various audio editors, dialog enhancement applications and for upmixing. These techniques include the extraction of the center channel signal from two-channel stereo, the decomposition into direct and ambient signal components, and the manipulation of transient signal components.

As a recent trend in source separation, informed separation techniques incorporating additional information about the input signal are further developed, with score-informed source separation being a prominent example. New developments also include the application of non-negative tensor factorization (e.g., for multichannel input signals) and sparse representations in general.

SEMANTIC AUDIO PRODUCTION

Recently developed applications of semantic analysis use automated transcription of rhythmic, harmonic, and melodic information in digital audio workstations.

Newly developed audio effects apply tempo estimation for synchronous temporal effects (e.g., delays and modulation effects like tremolo), tonal and harmonic analysis (e.g., harmonizers) and perceptual models for estimating the perceived intensity of artificial reverberation.

Semantics have been used for organizing, searching, and retrieving presets in digital synthesizers in the past, and current research also investigates the automated programming of synthesizers using example recordings.

SEMANTICS IN AUDIO CODING

State-of-the-art audio codecs are capable of delivering high sound quality for both, music and speech signals, by using dedicated coding schemes and tools for these signal types. Semantic analysis of the input signal is applied for the control of the signal processing in the codec.

MUSIC EDUCATION AND ENTERTAINMENT

The use of semantic audio analysis in music education and learning software includes music transcription and source separation for analyzing a musical performance and creating content, e.g., play-alongs and automated accompaniment. A clear trend is the use of such software on mobile devices.

Semantic audio analysis also includes the automated synchronization of recordings and the recommendation of recordings based on rhythmic and harmonic similarity in DJ software.

TRENDS IN RESEARCH

Semantic analysis can be improved by combining audio analysis with textual and image analysis into multimodal analysis, e.g., music recommendation using lyrics when searching for Christmas carols. Furthermore, the performance of machine learning methods is improved by fusion of multiple concurrent methods. Alternative classification methods are investigated, e.g., deep neural networks and random forests. An ongoing activity in the past ten years is the improvement of the evaluation of new methods and technologies by using formal evaluation frameworks with common data sets and evaluation metrics.

Other applications of semantic audio analysis besides mentioned above are broadcast monitoring (e.g., determining the amount of speech versus music and the content broadcast and usage), loudness control (by using speech detection for estimating the loudness of the dialog only), and archiving.