Fraunhofer
DIGITAL MEDIA

FRAUNHOFER DIGITAL MEDIA ALLIANCE

INTO IMMERSION

TRENDS AND TECHNOLOGIES IN DIGITAL MEDIA

Issue 6

CONTENT

PREFACE

In a time when FullHD resolution has become the basis for TV consumption, and even smartphones have the processing power of a computer, the question arises as to how technology can continue to improve media consumption.

Broadly speaking, there are three aspects: the transition to "all-IP" technology, i.e. all digital data will be exchanged and transmitted via Internet protocols, allowing for more efficient production and timely rollout to the consumer; ubiquitous media consumption independent of the device and the location of consumption; and the increase in immersiveness due to new image and sound quality – entering virtual worlds promises new experiences.

This is reflected in the development of new codecs that not only allow for optimized production workflows, but can also transmit content with additional objects or information to the end user in top quality. Using a completely digital work process which opens up new computer-aided analysis to improve quality and content analysis.

3DOF, 3DOF+, Windowed 6DOF, and 6DOF – the variety of terms shows that immersiveness of both image and sound still has a long way to go. This trend brochure will show you some of the most teasing solutions for new immersive experiences.

Enjoy reading it!

Dr. Siegfried Foessel

REALISTIC ACOUSTICS IN VIRTUAL WORLDS

We cannot turn our ears off – they give us a constant, three-dimensional sound image of our environment. A sound image of this kind can now be generated in virtual worlds, too: for the first time, the noises are directly linked to the objects that create them.

A door creaks, footsteps approach, a bird flies away, squawking loudly – in the real world, we are used to three-dimensional auditory impressions. The noises are directly linked to a certain object; the chirping, for example, is inextricably linked to the bird whose beak makes the noise. Once the bird flutters away above us, its chirping follows it. Creating 3D auditory impressions in virtual worlds, and making them seem even more realistic, has shown itself to be a great challenge, however.

The only option at the moment is offered by headphones, but many users find these annoying, particularly as they are already wearing a head-mounted display. A 5.1 loudspeaker system only provides limited assistance: it can only create a reasonable spatial sound impression when the user is standing in a certain space. If the user turns around or moves a few steps, the realistic sound is lost. And if there are several people in the virtual space at the same time, the sound system has no chance.

Linking sounds to objects

The SpatialSound Wave technology from Fraunhofer IDMT opens up entirely new possibilities. "Intelligent algorithms allow us to link the sound to the objects," Christoph Sladeczek, group manager at Fraunhofer IDMT, explains. An example shows what this could mean for visitors to virtual worlds. Let us assume that a group of users is in a VR world at the

same time. If a virtual motorbike rattles by, each individual user hears it pass just as they would in the real world – regardless of where they are in the space; they also have no control over the position of the loudspeakers. "We can place an audio object at any position in the room and move it through the room – you can no longer think in terms such as loudspeakers," according to the acoustic engineer. Wave field synthesis makes it possible: Overlaying the waves creates the physically correct sound impression of a noise source.

Stages, theaters, and concerts

The technology also offers benefits to concerts, theaters, or at home in your own living room. At the moment, sound engineers – when recording a concert, say – save pre-mixed loudspeaker signals that are tailored precisely to the loudspeaker arrangement, such as stereo or 5.1. This means that the sounds can only be played back by exactly these loudspeakers to attain good quality.

"SpatialSound Wave no longer generates finished loudspeaker signals, it creates an audio object that comprises an input source such as a violin, the position, and the volume," explains Sladeczek.
An intelligent algorithm uses this to calculate the loudspeaker signals in real time, customized for the loudspeaker situation in question.
Ergo: the recording can be played on all imaginable loudspeaker systems – without any editing and in optimum quality. SpatialSound Wave thus opens up a wide range of possibilities, not only in the area of VR.

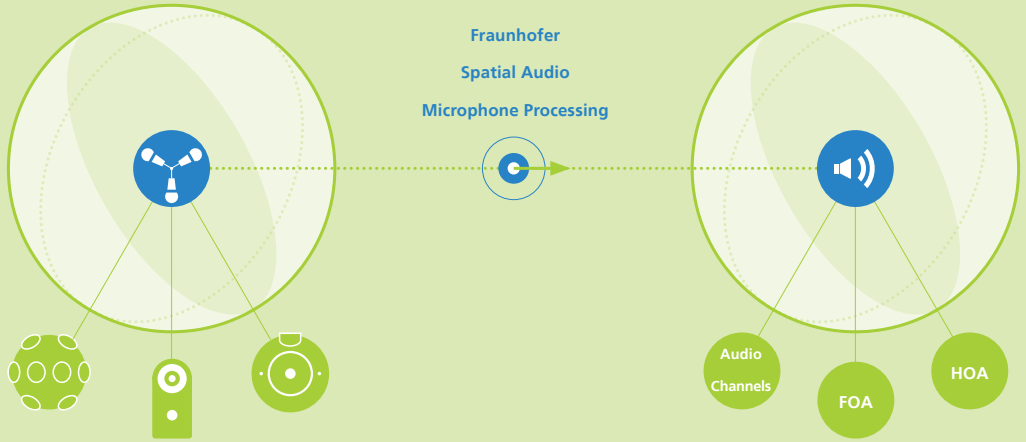
Jean-Pascal Beaudoin, Head of Audio at Felix & Paul Studios and Director of Sound at their subsidiary Headspace Studios commented: "Fraunhofer's technology allows us to faithfully and precisely capture scene-based audio that we can later combine with other sources for a fully immersive audio experience."

upHear® – SPATIAL AUDIO FOR VR

The first audio technology to be delivered under Fraunhofer's upHear® brand of immersive audio innovations is the Spatial Audio Microphone Processing SDK. The automated processing algorithm is designed to ease the high-quality capture of authentic spatial audio content on set.

.

The upHear Spatial Audio Microphone Processing solution will enable VR Studios to enhance their immersive VR content with spatial audio in a simple, yet robust way. The technology was designed for professional and consumer 360-degree cameras and mobile devices, improving sound capture using the built-in microphones of the device. The algorithm automatically transforms captured sound signals in real time to any popular surround or immersive reproduction format, including First-Order Ambisonics (FOA) or Higher-Order-Ambisonics (HOA), immersive 3D audio with 7.1+4 height channels, or 5.1 surround for legacy systems, while preserving the authenticity of the original audio scene. Its flexible semantic signal analysis approach allows for adoption into devices with three or more microphones in various configurations, as well as more sophisticated configurations.
Manufacturers, professional content creators, as well as consumers can benefit from the Fraunhofer Spatial Audio Microphone Processing technology.
Felix & Paul Studios, the Emmy® Award-winning studio behind the world's leading cinematic virtual reality experiences, will be the first to integrate Fraunhofer's Spatial Audio Microphone Processing solution into its next-generation proprietary 360-degree 3D camera system.



FILMS YOU CAN ENTER: 3D HBR MAY MAKE IT A REALITY

How about walking around within a film instead of just watching it passively from the sofa? What's more, you wouldn't be watching clumsy avatars, but realistic-looking actors. With 3D Human Body Reconstruction (3D HBR), this won't be a problem – realistic people can be added directly to virtual environments.

Watching a movie usually means sinking into the sofa or theater seat to watch the events unfold on the screen. And following what the director and cameraman have put together. Fraunhofer HHI has now made a completely new type of movie possible: here, the moviegoer will no longer sit idly and watch a finished product, but will be transported to the middle of the action, will be able to move around "within" the film, take a look at side scenes, view actors from all angles – just as the viewer chooses.

Realistic people in virtual movie environments

The first necessary step: The actors are not intended to move through the action like awkward avatars, as is so often the case in virtual realities, but ought to be as realistic as in "normal" movies or even in real life. This is made possible thanks to the technology 3D Human Body Reconstruction, 3D HBR for short, from Fraunhofer HHI.
"3D HBR is forming the basis for films you can enter: It allows producers to show realistic people directly within VR environments," says Dr. Oliver Schreer, group manager at Fraunhofer HHI. "To make this a reality, we record the person, currently with up to 32 cameras, from different spatial directions; the three-dimensional model thus calculated is output in the form of wire frames – it thus does not need to be post-processed any further. Nor is it necessary to animate the person any further."
The result is that the actor moves within the VR world as fluidly and naturally as in the real world, and the texture seems like real life to viewers.

Two large test productions

In two large test productions, the technology is already showing us what it can do: one project is Gateway to Infinity in collaboration with UFA, and the other is Tiger Tale, a joint production with Trotzkind. "While UFA and Trotzkind are responsible for the creative part, we provide the groundbreaking technology – an ideal combination. This allows us to unearth new ways of telling stories in virtual worlds; there are infinite possibilities," according to Schreer. More films and projects are in progress, including for example edutainment, gaming, walk-in movies, fashion shows, cultural heritage and historical documentation.

All-round recording

The Fraunhofer HHI researchers are currently working on a new lighting and recording system. Where, previously, 180-degree recordings around the actor were used, 360 degrees are now possible – from all sides, in other words. To this end, the researchers are developing a rotunda with illuminated walls in which up to 32 cameras are placed. The first test recordings are planned for September 2017.

HIGH-DEFINITION PANORAMIC VIDEOS

The resolution of 360-degree videos offers room for improvement – the picture quality is usually still quite low. A new technology known as tile-based streaming is now here to change that.

With the VR glasses on your head you can enter a panoramic video that seems as realistic as real life itself. That is the theory, at least. De facto, the image offered to viewers in these panoramic videos is usually quite sobering. The resolution leaves plenty of room for improvement. The reason for this is the resolution of the images transmitted. This is because it is not enough only to show the viewer's current virtual field of view – otherwise, when he turned his head suddenly, he would be staring into darkness. Instead, the video all the way around the viewer's head needs to be transmitted simultaneously.

At the moment, this takes place in two different ways. With the first method, all data for the panoramic video is transmitted in equal quality. In this case, the resolution of the virtual world is the same in all viewing directions, but is rather mediocre. With the second method, the focus is on the quality of the user's current viewing direction. Wherever the user turns his head, the virtual world will appear in a higher resolution than behind him. But this method, too, has its drawbacks: the server needs to calculate many different videos and provide them to the use at short notice; each new perspective needs a new video. The end device downloads the correct video depending on how the user moves. For the provider, however, this represents a significant cost factor.

High resolution, low storage requirements

Researchers at Fraunhofer HHI are now developing a third option. "We divide the entire 360-degree video into tiles that are encoded independently of one another – that is why we call our technology tile-based streaming," explains Robert Skupin, scientist at Fraunhofer HHI. "Within the user's field of vision, the image is high resolution; behind him, it is low." The major difference when compared to the previous method is that the end device decides which tiles to download in high resolution and which in low resolution. It puts the tiles it needs together to form the required image in the right resolution instead of downloading one of several prepared videos. This means that, unlike the existing videos, each tile only needs to be stored on the server once. "The resolution is high, the operating costs are low, and, what's more, the technology opens up several degrees of freedom in various areas such as dynamic adaptation of the transmission rates," says Skupin.

In order to create virtual worlds in high definition using the tile-based streaming technology, one needs – on the one hand – a suitable video codec and – on the other – an efficient transmission standard. As far as the video codec goes, it already exists in the form of HEVC. "HEVC is the first video standard where all end devices support a tile-based approach and are therefore already prepared for tile-based streaming," Skupin can advise. The necessary transmission standard, called "Omnidirectional Media Format" (OMAF), is currently being developed by the MPEG standardization committee; it should be ready by the end of 2017.
The Fraunhofer HHI researchers are also involved: They are expanding the standard so that it can allow tile-based streaming, in order to optimally supply current and future devices with high-definition data.

360-DEGREE VIDEO FOR EVERYONE

Interview with Dr. Stephan Steglich, Fraunhofer FOKUS

Videos that allow viewers to choose the view according to their wishes need more bandwidth than most methods of transmission can offer. A new streaming technology, however, makes it possible to offer 360-degree videos to a much larger public.

You hear a lion's loud roar behind you. Instinctively, you turn on your heel in order to react to the dangerous situation. Luckily enough, the predator isn't real; it's just inside a 360-degree video. All the same, the viewer can change their all-round viewpoint, just like in real life. For the time being, however, there are still some challenges that need to be overcome when it comes to 360-degree video. What are they, and how can we meet them? A conversation with Dr. Stephan Steglich, Director Business Unit Future Applications and Media (FAME) at Fraunhofer FOKUS.

Dr. Steglich, where are the challenges to 360-degree video to be found?

One major challenge is in the transmission and processing of the huge quantities of data involved. For a 360-degree video, you see, ten to twelve times as much data is needed as for a normal TV image, assuming that the image quality is comparable. A conventional DSL line cannot handle that, which means that the range of 360-degree videos is still very limited. It only makes sense to produce content, however, if enough people are going to watch it. The corollary of this is that we currently have to live with a data rate that is limited to 25 megabits, which – for 360-degree videos – is not sufficient to achieve TV quality (HD, UHD).

Might there be a way to make good-quality 360-degree videos available to a wider public?

At Fraunhofer FOKUS, we have developed a solution that addresses just these drawbacks. We are mostly focusing on flat screens (TV and HDMI sticks devices), which enables us to cover various end devices as well as possible. We prepare the video sequences in a cloud in such a way that they are able to show the various perspectives. If the TV viewer presses the right arrow button, for example, the TV sends this information to the cloud.

The pre-calculated video sequences that match the newly chosen perspective are then streamed. The end device doesn't even notice that it is receiving a 360-degree video. This approach allows a resolution of up to 4k for the Field-of-View..

Has the technology already been put through its paces?

We have already had various test runs with different TV stations – and we have been able to pass all tests with flying colors. Public broadcaster Rundfunk Berlin Brandenburg RBB, for example, produced a TV documentary series called Band Camp Berlin for children's channel KIKA. The additional material included a 360° video based on our technology. It was the world's first 360-degree video to be broadcast to TV sets.

Will 360-degree video change television?

The most significant changes, but also many creative possibilities, will probably occur in the area of storytelling. This is because, with 360-degree videos, the viewers decide what they want to see. The question is how can viewers avoid getting bogged down and missing the most important scenes? We could achieve that by means of sounds or graphic fade-ins – a little bird, say, that flies through the picture and attracts the viewers' attention back in the desired direction. Here, too, we are working on solutions.

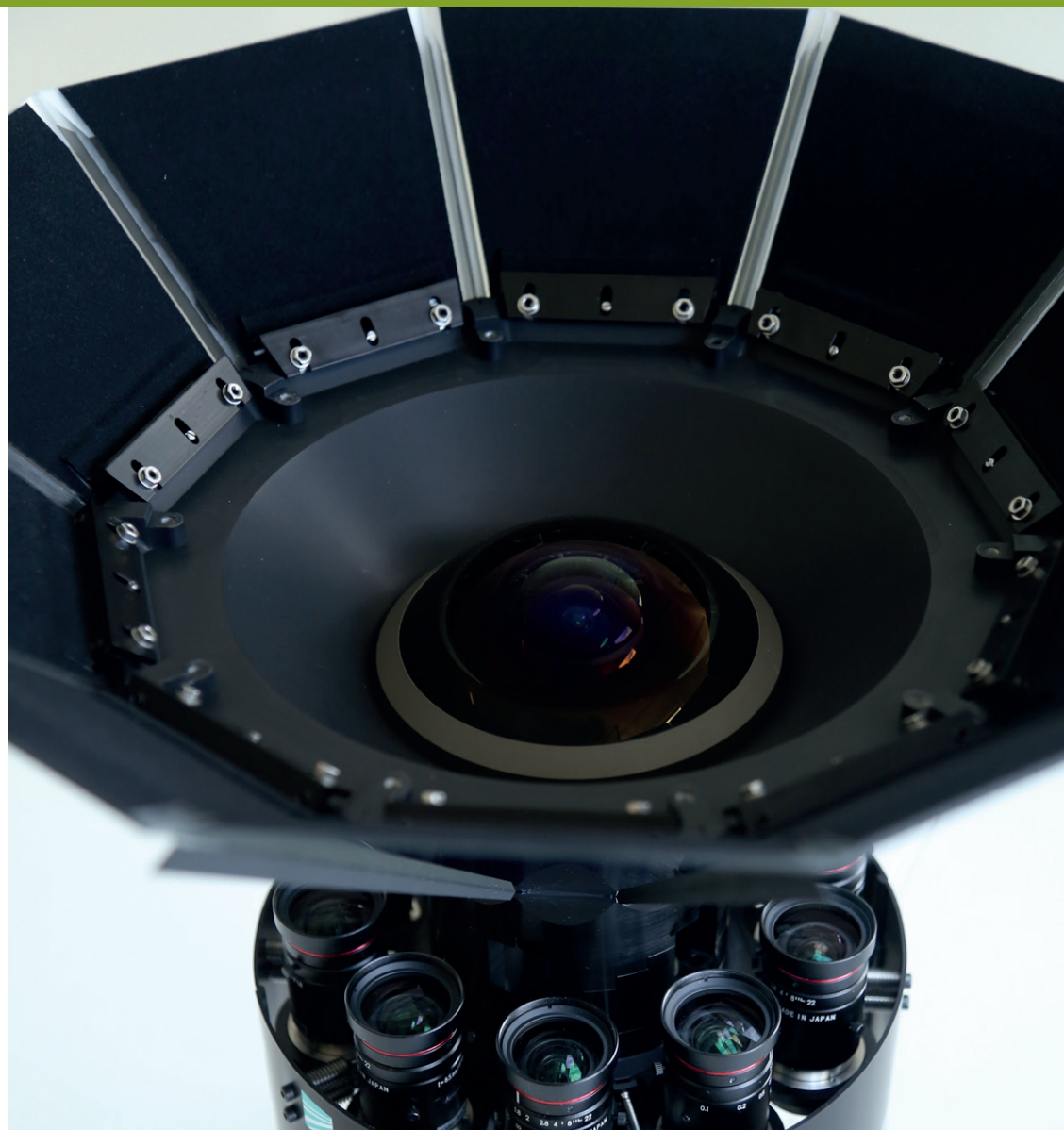
TRANSMITTING PANORAMA VIDEO LIVE

Interview with Christian Weissig, FraunhoferHHI, Capture and Display
Including panoramic views in films has previously been complicated and expensive. Furthermore, huge quantities of data are produced, and these are difficult to transmit. The Omnicam-360 from Fraunhofer HHI solves these deficits.

Reality offers a unique 360-degree panoramic view. If you attempt to capture it with a camera, however, you will only snap a small section. Panorama cameras are here to change that. There is one catch, however: the quantities of data involved are too large to be transmitted using conventional methods. The new Omnicam-360 from Fraunhofer HHI can overcome these deficits – the first 360-degree videos have already been transmitted live.

Mr. Weissig, what is so special about the new Omnicam model?
As with the previous version, the new model has ten cameras that can record the panoramic view without distortions or breaks – and in a simple-to-handle manner. To this, an additional camera is added. This one looks upwards from the optical center and films the sky. Furthermore, the new model, including its recording system, only weighs approximately 10 kilograms, allowing it to be attached to drones for aerial shots.

What has changed with regard to transmission?
Together with SES Astra, we streamed the Omnicam-360's panoramic recordings live for the first time – using encoding standard HEVC from Fraunhofer HHI. In order to be able to transmit the data via conventional means such as satellites, it needs to be compressed to a manageable bandwidth. We succeeded in getting it down to a bandwidth of only 22 Mbit/s. This means that we can broadcast 360-degree Omnicam videos straight into people's living rooms.



ORPHEUS – OBJECT-BASED AUDIO EXPERIENCE

ORPHEUS is a European research project dedicated to improving the management of object-based audio content. It will develop, implement and validate a new end-to-end object-based media chain for audio content. Orpheus started in December 2015 with a duration of 30 months. It receives funding from the European Commission under the Horizon 2020 program.

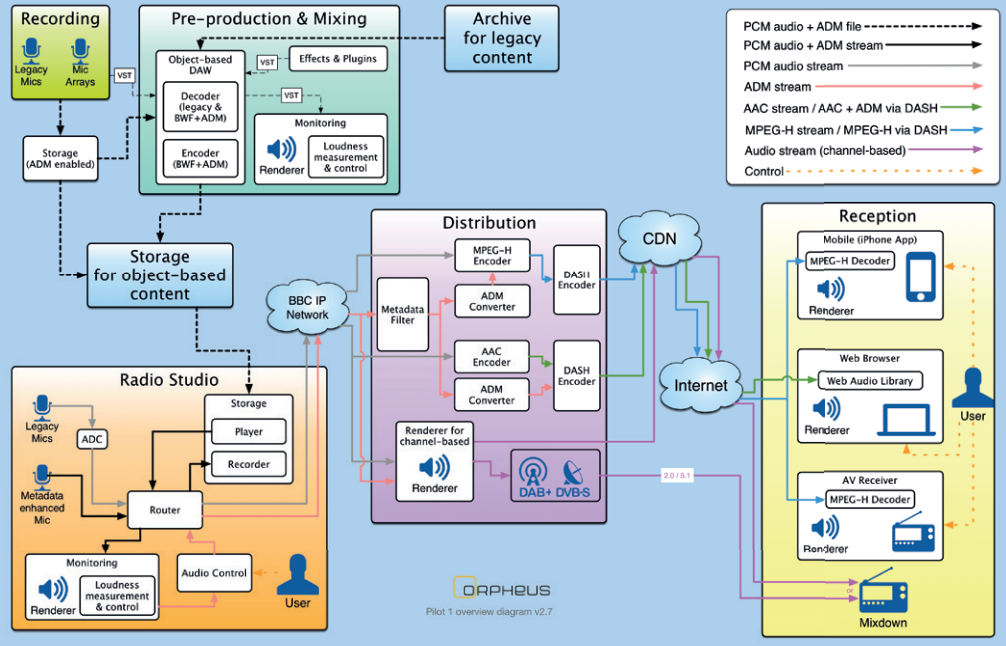
Object-based media is a revolutionary approach for creating and deploying interactive, personalized, scalable and immersive content, by representing it as a set of individual assets together with metadata describing their relationships and associations. This allows media objects to be assembled in groundbreaking ways to create new user experiences.

Orpheus is currently in the status of pilot phase 1 that focusses on live broadcast. This phase has been divided into three stages.

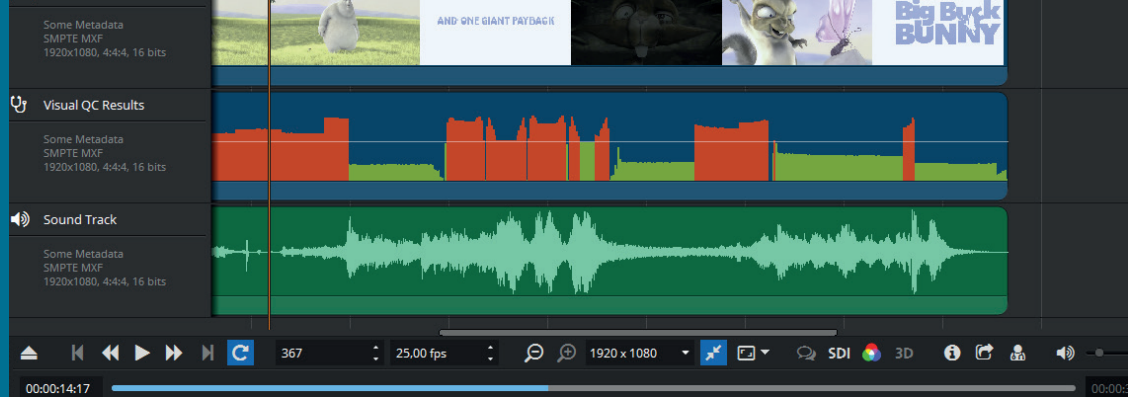
The first stage is a live production of an interactive object-based radio drama that can be experienced using a web browser.

The second is a selection of material encoded using MPEG-H and made available through an iPhone and AV receiver.

The third is an 'as-live' broadcast, live encoded using MPEG-H and made available over the Internet. The aim of the project pilot is to demonstrate a full chain of object-based audio production, distribution and reproduction, based upon the pre-defined pilot implementation architecture.



Pilot Phase 1 Infrastructure



IMF WORKFLOWS WITH INTEGRATED QC

New distribution channels for video-content arise almost daily. The Interoperable Master Format (IMF), the standardized and released universal format by SMPTE, has become a recognized exchange format in the professional film production for defining and automating transcoding steps. IMF is first choice to today´s challenges of exchanging content in the highest quality without supporting a lot of different formats.

But especially when the transcoding process is subject to automation, intensive quality control is indispensable to guarantee the required level of efficiency in the transcoding workflow. Usually today, both, workflow compositing and verification, require human interaction. Using the IMF Output Profile List (OPL) facilitates the process of defining the automation process in an already approved format. The OPL defines how an IMF package (IMP) is converted into a particular distribution format.

QC makes transcoding processes more efficient

For automated Quality Control (QC) of these formats, the experts of Fraunhofer IDMT and Fraunhofer IIS present a demo version of the technology that allows for quality checks of IMPs as well as for the derived distribution formats based on the OPL.

The big advantage of the Fraunhofer solution is that typical quality issues introduced by video transcoders, like audio and video errors due to incorrect operations, wrong parameters, interruptions within the processing chain, or inconsistent interpretation of standards can be detected right away. Furthermore, technical requirements in the transcoding process, like the necessity for low bitrates, to name one example, can significantly decrease the A/V quality. Therefore it is indispensable to detect, evaluate and, if necessary, change the parameters before and also after the transcoding.

For OPL composing, in addition to macros already defined by SMPTE, new custom macros have been developed. These involve saving a QC configuration and also choosing a specific output video or audio encoder and its parameters for the creation of a distribution format.

The developed solution shows a set of important quality checks and the way of detection and documentation of detected issues in the well-known and widely-used easyDCP software for IMF mastering.

THE POTENTIAL OF LIGHT-FIELD FOR PROFESSIONAL PRODUCTION

Interview with Dr. Joachim Keinert, Fraunhofer IIS,Computational Imaging

One of the most promising approaches to advance the production of immersive media content for VR applications, mixed reality content or compositing in post-production is currently light-field.

Dr. Keinert, why and when do you start to consider light-field technology as a very promising approach for future content production?

We started our research and development in computational imaging with the light-field technology in 2010. Based on the permanent increasing number of cameras that are used on production sets to generate different views of the same scene many partners in the industry asked for new efficient ways to work with these data. So we applied our recognized knowledge in camera electronics and post-production to allow special effects already known and established in computer graphics for mixed reality content. By these means, it helps to avoid complex and time-consuming photorealistic modelling and rendering of CGI-content.

What are the essential benefits of light-field data?

The prior advantage when thinking about light-field data sets used for further post-production is the possibility to generate different perspectives from any position of the scene. For example, reflections, lighting and textured surfaces can be done in 2D CGI but are a big challenge for mixed reality content. The different perspectives of the light-field data open the possibility to look and walk around objects or persons, which is essential for VR applications.

The immersive feeling for VR application is one of the most discussed topics. Which potential does light-field could bring in?

For VR applications light-field seem to be one of the possible key solutions to enable live-action content rendered for VR glasses with the perfect perspective. With the possibility to create an almost "complete" natural representation of the scene the user gains six degrees of freedom to move not only his head, but furthermore to step back and forward or from one to the other side and get always a realistic impression of the content.
Today, the dimension and amounts of cameras that were used for recording define the light-field similar to a window thru which all possible perspectives a person can see when he moves inside this frame. It is possible to calculate and render inside these dimensions billions of additional views to use them for further effects like focus or camera position shifts or camera rides in an almost unlimited way.

But looking at production sets the need to decrease the amount of cameras is obvious. Though, we are presently investigating which sort of camera array is suited for different applications like VR, games, advertisements and mixed reality movies.

What can light-field provide for professional production?

Light-field technology as mentioned records directly a lot of different perspectives of a scene. This is a key information to understand the depth and geometry of the scene similar to classic 3D reconstruction. By having a well-defined camera array, light-field capture allows to perform this 3D reconstruction even for videos. Moreover, the textural information of an object is captured to correspond to the reality as closely as possible. Combining these two methods could therefore be a promising option to save time and money in CGI.

Another option is to edit a scene without a 3D reconstruction step by so-called image based rendering. This simplifies the processing because you neglect if the computer understands which object points are essential for a surface or not. Such an operation can most easily be compared to vectorizing a pixel image. Given that this is a tedious task that easily introduces errors, it is only applied when absolutely necessary.

What is your Fraunhofer team working on?

Currently we offer a software plug-in called Realception® to test and work with multi-camera data in post-production for partners. It provides additional editing possibilities related to multi-camera capture, that with today`s tools are not available, because they do not follow the light-field concept.

Together with the Max-Planck-Institute for Informatics we have recorded sample light-field material which is suited to be used as test material to fully comprehend the performance and possibilities when working with these data in a post-production environment or workflow. Future partnerships and developments will enforce special application for light-field in the workflows and to develop practice-oriented tools to make this technology available for professionals in the industry.

What are your goals for the future?

The long term vision for our light-field developments is to enable a perfect 3D capture of objects or scenes that represent not only the geometric shape, but also their correct texture, the reflections and shadows to generate photorealistic representations. This opens up the door to various creative possibilities in post-production to edit, adopt, re-use and re-define the recorded content and will provide the potential for storytelling, game design and visits of remote scenes never known before.

DIGGING DEEPER – DEEP LEARNING MORE THAN A TREND IN MEDIA?

With the increasing amounts of image and moving picture data the challenge to retrieve, search and find the desired sequences is one of the hottest topics ever. And, in a situation where abundant multimedia data is generated each day it is not enough to search and classify the content but also to optimize recommendation systems or to find inappropriate content.

The Herculean task is to extract key information from a movie and summarize it in a few tags which best describe the movie. This requires an intelligent, highly efficient approach when it comes to automating and optimizing the analyzation process.

Fraunhofer IIS scientists use and evaluate the performance and capabilities of deep learning methods. Deep learning currently is the most promising way in the area of artificial intelligence methods when working with high volumes of data. The power of deep learning methods is similar to structures of the human brain creating new connections in only milliseconds whenever new incoming information can be added.
The task, however, is to continuously enhance the evaluation and classification of content or situations with this information.

The Fraunhofer IIS scientists trained e. g. a Convolution Neural Network (CNN) for the high-level classification of movies and movie scenes. The predictions of this classifier are filtered based on their confidence to generate a compact set of the most relevant tags. The results of this method are very promising to gain good accuracy.

The big advantage of this method is its flexibility in tagging vocabulary size without losing performance.
The Fraunhofer IIS experts clearly focus on achieving better quality of the tagging that is similar to the human way of describing/tagging the movie content. This approach leads to fewer false tags, more abstract tags and the possibility to search for scenes with a particular tag.

"The tests and evaluations we carried out in the last few months are very promising regarding the accuracy of the movie tags. The deep learning techniques are primarily focused on retrieving higher level semantics from a movie," explains Dr. Heiko Sparenberg. "We did not start at level zero with key-frame tags like ´car´, ´explosion´ or ´door closes´. We take these already existing classifications as an input to much more enhanced and deeper levels, so that abstract search features like ´science fiction´, ´action´ or ´accident´ can be identified."

The scientists use available state of the art techniques to extract key frames from a movie and pre-trained models to identify low-level features and objects in these key frames. This input was then used to train an additional Deep Learning network which draws much more abstract conclusions such as the theme or genre of the movie and scenes thereof. Based on this information it is then possible to run advanced search queries like "list all action movies in my repository" or "jump to the next scene in this movie that contains accidents".

SOUTH KOREA LAUNCHES UHD TV WITH MPEG-H AUDIO

MPEG-H TV Audio is the first next-generation audio system to be on air 24/7 as South Korea launched its 4K UHD TV Service in May 2017. This offers two never-heard-before novelties in sound to the TV audience: Immersive sound will pull viewers in the center of the action and interactive sound will enable them to personalize their audio.

The new TV experience will first be available in the Seoul capital area, and will be extended to the South Korean metropolitan areas and the venue cities of the Olympic Games in Pyeongchang in 2018.

By 2020, the service is scheduled to be available nationwide. South Korean consumers can already purchase a growing number of MPEG-H enabled TVs from Samsung and LG today.

Leading up to the MPEG-H TV Audio System going live, Fraunhofer IIS put in a lot of effort in the detail of acquainting Korean broadcasters with the new audio format. Fraunhofer IIS trained a team of Korean sound engineers in Germany who are now supporting the South Korean broadcasters on site during production and mixing of audio content in the MPEG-H format.

Additionally, a permanent demo room with a studio and training center was established at the Digital Media City in Seoul. At the official opening on May 15, more than 100 guests were able to tour the facilities and get a first-hand look of the new possibilities that the MPEG-H TV Audio System provides.

Fraunhofer IIS Demo Room in Seoul

FULL-HD – NOW AVAILABLE VIA ANTENNA

It was one of the biggest switchovers in the history of German broadcasting: At the end of March 2017, broadcasters switched from DVB-T to DVB-T2. Broadcaster ZDF put its trust in technology from Fraunhofer HHI. And that was a good call: the switchover, in just one night, went completely smoothly.

You can almost see the individual hairs in the coat of the gazelle on screen – that is how clear the TV image is in high definition. Previously, however, only viewers who had satellite or cable were able to enjoy this picture quality. Terrestrial – or antenna – reception was restricted to an inferior quality. Since March 29, 2017, this is no longer the case in many regions of Germany: broadcasters switched from DVB-T to DVB-T2 – which offers a lot more capacity. They are now able to transmit in Full HD quality wirelessly. The encoding standard High Efficiency Video Coding HEVC, which owes a great debt to researchers from Fraunhofer HHI, makes this possible.

ZDF's collaboration with Fraunhofer HHI went further: to encode the signals, they used broadcast software from Fraunhofer HHI running on a device from Rohde & Schwarz. "In a shoot-out, which is a type of technology competition, our solution won. We offered the best encoding quality," Mark Palkow, scientist at Fraunhofer HHI, is pleased to report. And, at the end of March, he was proved correct: the switchover ran very smoothly although it had to take place in just one night.
The broadcasting authorities often transmit not just one channel, but several – the various regional channels, for example. In principle, fast-moving images, such as the transmission of a soccer game, need considerably more capacity than a newsreader sitting in a studio.
"Statistical multiplexing, known as Statmux, allows us to divide up capacity flexibly so that every channel is transmitted optimally," Palkow explains. At the moment, the encoding system decides every few seconds how much capacity is required. The researchers are working on allowing this decision to be made in a more finely grained manner.
During the expert analysis at half-time, for example, capacity could be immediately shifted to the nature documentary running on another channel, giving the viewers an ever better view of the antelope's hair.

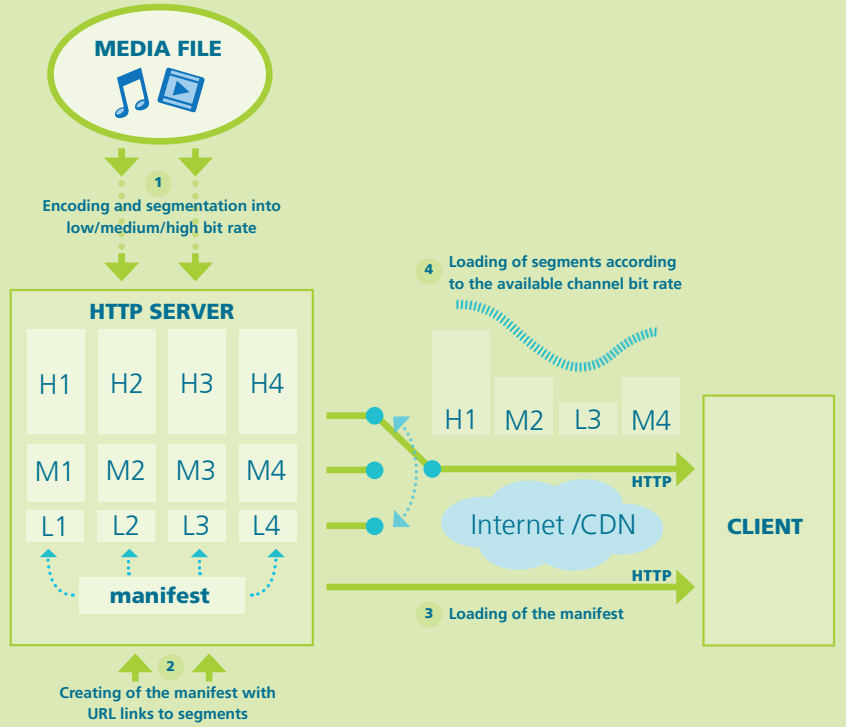
Fraunhofer solution successful for satellite transmission

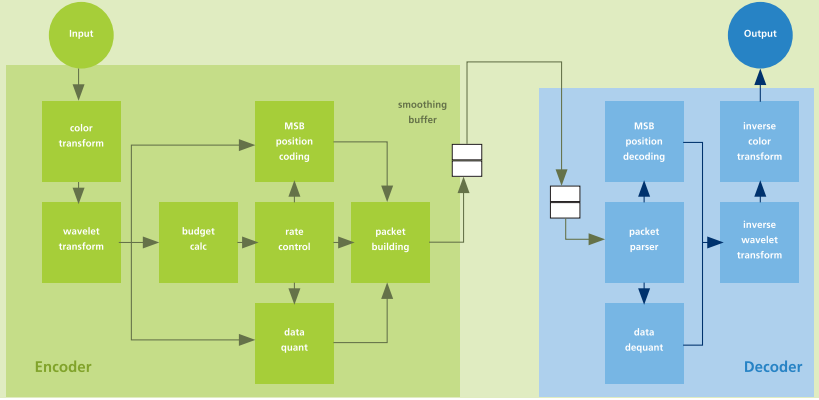
The Fraunhofer HHI system was not only a hit for terrestrial TV transmission – it was also the victor for satellite television. While the resolution for terrestrial television increased to Full HD thanks to HEVC, it went from HD to UHD (Ultra High Definition) for satellite television. "In the shoot-out, the aim was to be able to optimally transmit recordings of confetti flying around at the end of a soccer game. For encoding methods, a confetti "rain" like this is really chaotic," Palkow summarizes. But the Fraunhofer HHI software produced a very convincing result even in this difficult task. The result was that broadcaster Sky now also puts its trust in Fraunhofer HHI's solution when it comes to satellite transmission.

xHE-AAC – THE CODEC OF CHOICE FOR ADAPTIVE STREAMING

Streaming music wherever whenever has become normal for most of the smartphone users today. But when leaving the well-covered city centers with their 4G/LTE networks and falling back to 2G-coverage, or when joining mass events with plenty of mobile data usage, the network conditions decline, which can lead to dropouts in the music or video stream. This lowers the enjoyment significantly. With adaptive streaming such dropouts are a thing of the past.

There are two important adaptive bit rate streaming technologies available today: One is MPEG DASH (Dynamic Adaptive Streaming over HTTP), which was published as ISO standard in 2012, the other is the broadly adopted Apple HLS (HTTP Live Streaming). Both technologies work in a similar way: the content is broken into several small file segments; each containing a short sequence of the content (usually between 5-10 seconds long) which is encoded at different bit rates. The content segments are accompanied by a so-called manifest that describes the streams and their URL addresses. After loading the manifest to initiate the playback, the client automatically selects the segments with the highest possible bit rate that can be downloaded in time for playback without causing drop-outs. The MPEG codec xHE-AAC was designed right from the start for such adaptive streaming scenarios. It serves the whole spectrum from very low bit rates (12 to 16 kbit/s) as used in 2G networks up to transparent quality at 128 kbit/s for users in 4G/LTE networks. Additionally, it supports Dynamic Range Control and Loudness Control, which prevents peaks in loudness when the content changes (e. g. radio program and advertisement) or when switching between different stations. Professional streaming encoders featuring xHE-AAC with HLS are already available by StreamS.





JPEG XS – THE NEW LOW COMPLEXITY IMAGE CODEC FOR VIDEO PRODUCTION

Today two major trends can be observed in the video production industry. One trend is to use higher resolution displays like 4k TV (UHD-1) or even 8k TV (UHD-2) and higher dynamic ranges. This results in a tremendous increase of image data. The second trend is to get rid of specialized cabling and infrastructure like SDI and to use off-the-shelf infrastructure with Ethernet-based information technology. The question now is: In which way can these video data streams, which are uncompressed between 3-40 GBit/s, be transmitted effectively.
A well-established method in the industry is to use compression. However, as the data have to be processed during the delivery chain in a production facility, subsampled formats like '420' or high delays for compression and decompression are not suited. In many cases also the complexity is too high. For this reason the JPEG committee has started a new work item by collecting requirements and asking for relevant coding technologies. The new work item is called JPEG XS.

The most relevant requirements for the new codec in the first phase are (here as reduced set!):

- Image formats with RGB or YCrCb 444/422 with up to 12 Bit per color component
- Low latency (max. 32 lines end-to-end for compression and decompression)
- Low complexity (defined as max. per centage of some specific FPGA device families)
- No frame buffer required
- Multi-generation robustness
- Support of multiple platforms e. g. FPGA, ASIC, GPU, and CPU
- Real-time software implementation capability for 4k/60p formats on today's standard computers

Based on these requirements anchors were defined like JPEG, VC-2, JPEG2000 or even HEVC. As the main use case is 4k and 8k video production with an enormous demand for data throughput, it became clear that the compression ratio target should be between 2:1 and 6:1. Just as high that a lower Ethernet-based infrastructure generation can be used, for HD below 1GBit/s, for multiple 4k streams below 10GBit/s, and for 8k below 25GBit/s.

Development and Test

In September 2016 six technology proposals were submitted to the JPEG committee. These proposals were evaluated and the best technologies were selected to form a new codec. Various types of images were used as test set like natural scenes, synthetic or mixed content. Further technology enhancements were allowed to be added until July 2017. The final core experiments are in execution now, so that by the end of October 2017 the JPEG XS codec technology will be defined. First results showed that a quality like JPEG2000 is possible under the constraints above, but with much lower hardware and software resources. First implementations are expected for 2018.

Although the main focus of the codec was for usage in video production environments also other use cases are well suited for this technology. Examples are local video networks, keyboard-video-mouse (KVM) extender applications, connection of Virtual Reality (VR) glasses to PC or even internal connection of displays with a reduced set of wires.

Further work

Besides the core coding system additional standardization work was started like the definition of multiple profiles, the definition of a file format or methods for packaging of the coded bits into IP transport streams. Future extensions are targeting RAW Bayer compression, up to 16 Bit per color component

easyDCP SOFTWARE SUITE

In order to create a DCP (Digital Cinema Package) easily and compliant to all the requirements, or to test such packages subjected to delivery, Fraunhofer IIS developed the easyDCP software suite. More information about easyDCP, which is available as standalone tool-set – as well as plug-ins for various post-production solutions incl. BMDs Resolve – can be found at *www.iis.fraunhofer.de/easydcp*

REALCEPTION®

Plug-in for light-field post-production

Realception® tools from Fraunhofer allow for working with light-field or multi-camera data in post-production environment that is familiar to most of the professionals. The plug-in is available on a test user agreement basis.

easyDCP PUBLISHER

The easyDCP Publisher provides a cost-effective solution to generate DCPs with minimal effort and risk for approval and playback on a cinema server. To work with the software, professionals have to simply render, preview and fine-tune content and then the DCP is ready to be published. The project-based solution is available at *www.easydcp.com/publisher.php*

FRAUNHOFER CINGO®

Fraunhofer Cingo brings immersive 3D sound to mobile and virtual reality devices, creating the experience of "being there". For device manufacturers and service providers, Cingo is available as optimized software implementation for all major PC and mobile platforms, including Android and iOS. Equipment manufacturers using Cingo are LG (LG 360 VR) and Google (Nexus and Pixel family of devices).

3D SOUND WITH SPATIALSOUND WAVE

SpatialSound Wave is an object based system for producing and replaying true-to-life three-dimensional sound. It offers sound engineers and other users new options to easily and efficiently produce spatial sound. Application areas are full domes, live sound, entertainment, exhibitions, trade shows, and other events.

MPEG-H TV AUDIO SYSTEM

MPEG-H Audio provides interactive, immersive sound for TV and VR applications. It is available as software implementation to chip manufacturers, broadcasters and consumer electronics manufacturers. Professional broadcast encoders that support MPEG-H Audio software are DS Broadcast´s BGE9000 4K Ultra HD Encoder, the 4K UHD Live Broadcast Encoder KME-U4K from Kai Media, Media Excel´s latest UHD HEVC Encoder and the Pix HEVC-E4600HA Realtime UHD Encoder from PIXTREE. MPEG-H Audio enabled TV sets for the new UHD TV system in South Korea are available from LG and Samsung.

FRAUNHOFER DIGITAL MEDIA ALLIANCE

As an one-stop competence center for digital media we provide for our customers scientific know-how and the development of solutions that can be integrated in workflows and optimize process steps.

The members of the Digital Media Network are actively working in renowned organizations and bodies like International Standardization Organization ISO, ISDCF (Inter-Society Digital Cinema Forum), SMPTE (Society for Motion Picture and Television Engineers), FKTG (German Society for Broadcast and Motion Picture), and in the EDCF (European Digital Cinema Forum).

Fraunhofer Institutes in the Digital Media Alliance jointly offer innovative solutions and products for the transition to the digital movie and media world of tomorrow. The Institutes in the Alliance are available as renowned contacts and partners for all of the digital topics connected to digital media, digital movies, and standardization, as well as new cinematography, audio, and projection technologies, post-production, distribution, and archiving. The goal of the Fraunhofer Digital Media Alliance is to quickly and easily help find the right contacts, partners, and suitable technology.

The Fraunhofer Institute members are
- Digital Media Technologie IDMT, Ilmenau
- Integrated Circuits IIS, Erlangen
- Telecommunications, Heinrich-Hertz-Institut HHI, Berlin
- Open Communication Systems FOKUS, Berlin

Contact
Fraunhofer Digital Media Alliance
Angela Raguse M.A.
Phone +49 9131 776-5105
alliance-dc@iis.fraunhofer.de
www.digitalmedia.fraunhofer.de

Publication Information

Fraunhofer Digital Media Alliance
c/o Fraunhofer Institute for
Integrated Circuits IIS

Am Wolfsmantel 33
91058 Erlangen, Germany

Concept and Editor
Angela Raguse
Fraunhofer-Allianz Digital Media

Layout and production
Ariane Ritter

Photo acknowledgements
Cover picture: Fotolia.com
Page 3: Glosow Fotografie
Page 4: Istock.com
Page 11: Fraunhofer HHI
Page 13: Fotolia. com
Page 15 Fraunhofer Fokus
Page 17: Fraunhofer HHI
Page 27: Fotolia.com
Page 27: daviewkorea@gmail.com / Song In-hyuk