

20 Years

Fraunhofer
Research and Development
for Digital Media

Digital Media

In the Business Area Digital Media five Fraunhofer Institutes are cooperating to provide technologies and solutions for the media industry.

Fraunhofer IIS, HHI, FOKUS, IAIS, IDMT

www.digitalmedia.fraunhofer.de



Follow us on LinkedIn



Trends and Technologies in Digital Media

Pioneers in immersive digital media technologies

Topics

Preface	3
Together at the high table – a hybrid meeting via display.	4
JPEG XS for SMPTE Standard ST 2110 and the Industry Standard IPMX	6
Fraunhofer IIS unveils xHE-AAC with anchor loudnes features	8
Decoding Green Streaming: The energy footprint of video streaming devices.	10
Tracking down deep and shallow fakes reliably	14
Multi-lingual editing made easy!	16
MPEG-H Audio brings the television experience of the future to Brazil.	18
Enhancing Emergency Broadcasting thanks to EWFplus including WorldDAB EW.	20
Tailored Media Presence Measurement for the next level – InsightPersona	24
Large Language Model integration in Fraunhofer Audio Mining	26
Fraunhofer Business Area DIGITAL MEDIA	30
Imprint	31

Preface



It feels like a blink of the eye when we are looking back to 20 years of collaborative work in thrilling digital media projects and with cutting-edge technology developments in the Business Area Digital Media of Fraunhofer – formerly known as the Fraunhofer Digital Cinema Alliance. So, I am proud to coordinate this fantastic scientist crew who fights always for the best solution that can be transferred to our partners and customers.

From MP3 to Digital Cinema, from H.26x to VVC, from Hybrid Video Conferencing, 3D and Immersive Media to Metaverse Solutions, Green Streaming, high performing content verification, automatic and AI based localization and subtitling tools – the list of innovations, disruptive and vanguard technologies and the big number of fantastic pilot project together with industrial partners is quite too long to mentioned in a brief preface.

Therefore I like to invite you to get a sneak impression about our latest developments, activities and projects by reading this selection of highlight topics in our trend brochure.

Especially topics like AI assisted tools and solutions are driving our next-generation media solutions. If it's about content verification, fake news detection or comfortable custom localization tools for broadcast to digital cinema formats - all these new efficient and reliable technologies can be experienced in our demos at the Fraunhofer booth in Hall 8 B80.

One highlight project is the demonstration of a live hybrid conference where visitors will be able to talk to our personnel on-site as well to a specialist based in Berlin. A system that surpasses the border of a 'normal' videoconferencing session by far.

With new implementations of JPEG XS, MPEG-H Audio and several new features in VVC we will provide a whole spectre of new and novel developments and solutions for the media industry.

Get background information about our innovations and enjoy reading

A handwritten signature in black ink, appearing to read 'Siegfried Foessel'. The signature is fluid and cursive.

Prof. Dr. Siegfried Foessel
Spokesman Fraunhofer Business Area Digital Media

Together at the high table – a hybrid meeting via display

Video conferencing has its pitfalls: As the participants are usually looking at the display instead of the camera, there is no direct eye contact. It is also not possible to get a direct feedback from the gestures and direction of the other person's gaze who they are addressing. Dr. Oliver Schreer, Group Leader »Immersive Media & Communication« at the Fraunhofer Heinrich Hertz Institute HHI, explains »Live Conference 4.0«, which overcomes these problems.

Mr. Schreer, what is "Live Conference 4.0" all about?

While people try to hold eye contact with their counterparts in face-to-face meetings, the participants in video conferences usually look at the screen instead of the camera – making eye contact difficult. A further level of complexity is added when two people are taking part on one side of the conference.

Sample scenario: Two people are standing at a high table and are looking at a third participant on the display. With the technologies in Live Conference 4.0 we cannot only guarantee eye contact, but we can also enable that the person speaking on the display is also looking at the right person at the high table to whom her/his question is directed. While the person being addressed maintains eye contact, the person not being addressed sees the person from the side.

The display presents the two people at the high table with different views - just like in real life. A system that is currently unique in the world. The remote participant, on the other hand, has two displays on which he/she can see the other two persons separately.

How was this real-life high table scenario technically realized?

We record the distant person with several high-resolution cameras so that we obtain the most realistic representation possible and generate the appropriate 3D view from the images obtained using specially developed algorithms. In other words, we calculate a different view of the face shown so that it looks in the right direction. We use a standard 4k display for playback at the standing table. A lenticular is mounted in front of it, which directs different pixels to different regions in space and therefore



Live hybrid conference



generates different views depending on the viewing angle onto the display - similar to the well-known »wiggly postcards«. The manufacturer produced the lenticular according to our specifications, which were previously verified in simulations. The first prototype already exceeded our specifications, so it was even better than in the simulation!

What were the biggest challenges during development?

The main challenges were to generate novel views in very high resolution - and still meet the conditions required for video communication, i.e. real time at 15 frames per second and as little delay as possible. For video conferencing, the delay should be less than 270 milliseconds, otherwise people will start talking over each other. Our delay is only 160 milliseconds, so it is imperceptible.

What are you showing at IBC?

At IBC, we will be showing a live transmission from Berlin to Amsterdam. While the participant is sitting at a desk in Berlin, visitors to the IBC in Amsterdam can experience live at a high table in our booth how video communication can work in a class of its own. »Such a sort of hybrid video conference is unique in the world. You really get real-time reaction of the person by turning her/his head to the addressed person and direct eye-contact.«

Contact

Dr. Oliver Schreer
Fraunhofer HHI
oliver.schreer@fraunhofer.de

JPEG XS for SMPTE Standard ST 2110 and the Industry Standard IPMX

The video coding experts of Fraunhofer IIS, well known for their JPEG standardization activities in ISO, developed a JPEG XS SDK that perfectly complements the upcoming Digital Media over IP standards ST 2110 and IPMX (Internet Protocol Media Experience).

Whereas the SMPTE ST 2110 standard addresses absolutely frame-synchronous transmission of production quality video streams for broadcast, the industry standard IPMX from AIMS (Alliance for IP Media Solutions) is a set of opens standards and specifications for audio and video IP transmission in ProAV applications with less timing requirements. Although both Digital media over IP standards allow uncompressed transmission of audio and video data, compression with JPEG XS significantly reduces the required bandwidth while maintaining visual lossless quality.

JPEG XS is a very powerful codec for streaming of high-resolution image data over regular IP connections. It excels with its ultra-low-latency below one frame and its low-complexity so that the codec can be integrated in most of the existing hardware components without further need to extend them or even in standard computer equipment as software only solution.

Fraunhofer IIS JPEG XS SDK

The Fraunhofer IIS JPEG XS SDK enables a quick and easy integration in applications to profit from these advantages. The SDK is compliant to SMPTE Standards ST 2110-22 for professional workflows. It guarantees that each single stream is individually timed by the ST 2110 system. The audio-video-data synchronization using PTP (Precision Time Protocol) clocks ensures the accurate synchronization of all streams regardless of how the packets were routed.

The SDK is also compatible for all IMPX (Internet Protocol Media Experience) transmissions to enable the carriage of compressed and uncompressed video, audio and data over IP networks especially for the market.



Improving mixed signal content

The Fraunhofer JPEG XS SDK supports the newest edition 3 of JPEG XS with its "Temporal Differential Coding" TDC and adds application specific functions like Forward Error Correction FEC to stabilize the data transfer over IP. For an easy access to the JPEG XS functionality Fraunhofer IIS provides development kits for CPUs (Intel, AMD, ARM), GPUs (NVIDIA) and FPGAs (Intel, AMD).

Additionally, we set up a so called "Application Support Package" with sample codes alongside with an Implementation Guide for SDI to IP, IP to SDI, IP to Display, IP to MXF, MXF to IP, GStreamer and NVIDIA Jetson integration.

Application proof

The JPEG XS SDK is already in use by many renowned industry partners like broadcast equipment manufacturers, broadcast streaming providers, graphic card and component providers. If you want to get a trial or discuss licensing conditions – see you at our booth at IBC Hall 8 B80.



Contact

Prof. Dr. Siegfried Foessel
Fraunhofer IIS
jpeg-xs-business@iis.fraunhofer.de

Fraunhofer IIS unveils xHE-AAC with anchor loudness features

Fraunhofer IIS, the pioneers in audio coding, are set to showcase their groundbreaking xHE-AAC codec at IBC 2024. As the fourth-generation member of the AAC codec family, xHE-AAC is about to revolutionize audio streaming with its superior quality and adaptability. A new feature was added to even out loudness inconsistencies originating from ad insertion and highly dynamic content.

A highlight of this year's presentation is the new anchor loudness feature. The innovation focuses on speech-gated loudness measurement within the xHE-AAC encoder. By normalizing loudness based on speech segments, it ensures a consistent audio experience without volume jumps between different programs, which serves important standards like EBU R128 s4, and ATSC A/85.

xHE-AAC delivers exceptional audio experiences across various content types, including movies, music, audiobooks, and podcasts. It supports adaptive streaming from 12 to over 320 kbit/s stereo, ensuring continuous playback even in challenging network conditions. With MPEG-D DRC loudness normalization and dynamic range control, users can enjoy consistent volume and enhanced audio quality on any device.

Christian Simon, Business Developer at the Fraunhofer IIS Audio and Media Technologies Division, explains: "With the anchor loudness functionality, we address another common issue in audio streaming – the annoying volume changes between high-dynamic audio programs and ad breaks. This feature ensures an even loudness level of the dialog, making the listening experience more enjoyable for users."

xHE-AAC is natively supported across major operating systems such as Android, iOS, macOS, and Windows 11, making it a versatile solution for streaming services. It also plays a crucial role in Digital Radio Mondiale (DRM), where it is the mandatory audio codec since 2013.

Don't miss the opportunity to experience the future of audio streaming at the Fraunhofer IIS booth at IBC 2024.



xHE AAC
IIS

Contact

Christian Simon
Fraunhofer IIS
christian.simon@iis.fraunhofer.de

Decoding Green Streaming: The energy footprint of video streaming devices

In an era where streaming has evolved into an essential and omnipresent aspect of our everyday routines, the relevance of the ecological footprint of our multimedia entertainment is gaining prominence. Behind the scenes of the seemingly effortless transmission of our favorite movies, series, TV shows, and social media, complex interconnections and a multitude of components, parameters, and variables affect the energy consumption and, ultimately, the ecological footprint of our beloved video streaming.

The streaming supply chain, from content creation to display, is complex. Factors like scaling, concurrent viewers, and distribution significantly impact energy use. Measuring energy consumption across different scenarios, such as live streaming and video-on-demand, helps us analyze the overall demand. Key questions include the influence of bitrate, resolution, HDR, codecs, network interfaces, and display technologies (OLED, QLED, LED) on energy consumption.

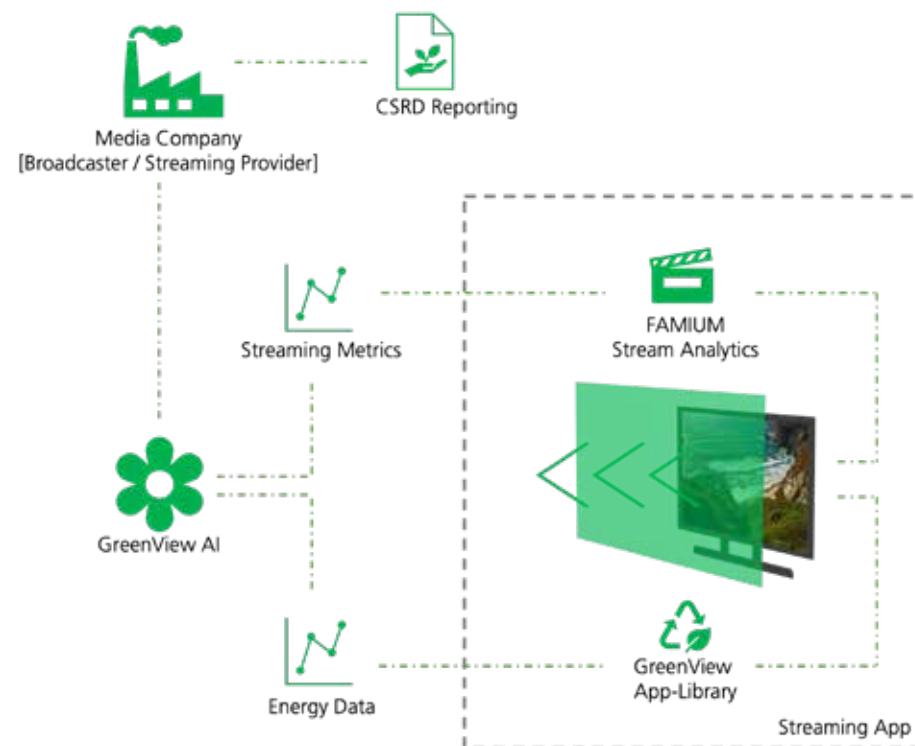
Green Streaming Measurement

Using our FOKUS FAMIUM Stream Analytics Solution, we combine energy consumption data and streaming metrics and do analyses based on it. For this purpose, we have developed a measurement framework that

enables the automation of measurements on streaming devices and transfers them, along with simultaneously recorded session metrics, into a common and reliable database through FAMIUM Stream Analytics. To evaluate the data, FAMIUM Stream Analytics provides individual dashboards. The measurement results are processed and used as a data basis for modeling a digital twin of the streaming chain and for training AI models to optimize its energy efficiency.

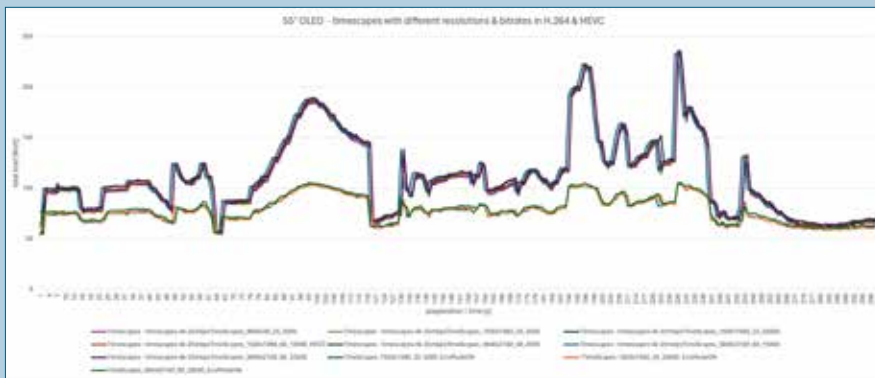
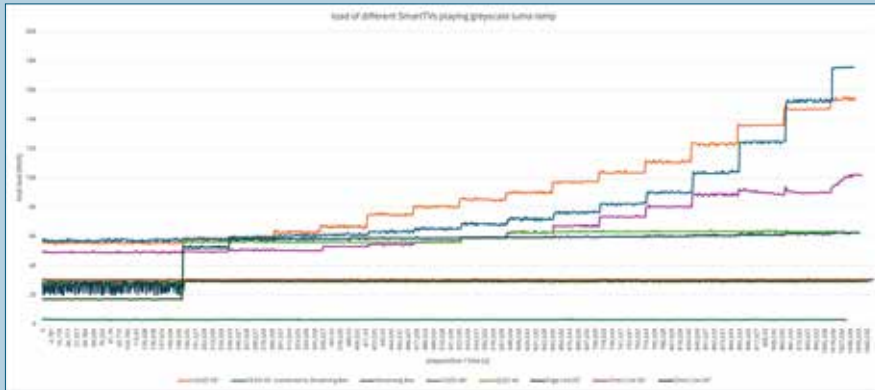
Setup and Methodology

In our initial measurements, we tested various streaming devices playing the same content to measure power consumption. We used Smart TVs with OLED, QLED, Full LED, and Edge LED displays to examine their energy signatures. An Android-based



We used Smart TVs with OLED, QLED, Full LED, and Edge LED displays to examine their energy signatures. An Android-based streaming box was also tested, serving as a player for the TVs via HDMI and measured separately.

This setup focused the TVs on their display function, excluding additional processing steps. The streaming box's measurements provided a benchmark for content reception and decoding energy use, which isn't separately measurable on Smart TVs.



Energy Signatures of different display types

In our test streaming session, the Smart TV's energy demand is dominated by display illumination and basic functions.

With a streaming box connected, potential energy savings from adjusting streaming parameters (bitrate, resolution, codec) are limited to the 3 watts used by the streaming box. In contrast, displaying bright content on the TV can consume up to 180 watts.

Stream settings and their affect on energy consumption

Evaluating measurement results shows that TV power consumption is significantly influenced by content brightness on OLED TVs. Various stream configurations were tested with energy-saving modes enabled and disabled.

Differences in power consumption between SD, HD, and UHD renditions with bitrates from 5Mbps to 25Mbps are minimal. However, enabling energy-saving mode can reduce power usage by over 50% in some scenes.

Our measurements at Fraunhofer FOKUS Smart TV-Lab reveal that Smart TV energy consumption is mainly affected by display brightness and TV settings, with limited impact from stream parameters like resolution or bitrate.

Tests on a 55" OLED TV showed that a 5% brightness reduction saved 9.0% energy, and a 10% reduction saved 14.6%, while maintaining acceptable visual quality. These findings confirm that display brightness significantly affects energy consumption, with substantial savings potential.

We can conclude, that bitrate and resolution in video streaming have a very low, practically almost negligible impact on the energy consumption of smart TVs and other streaming devices.

The real energy-saving potentials are hidden in the respective display and device settings, showing a characteristic dependence on the brightness of the content being displayed. FAMIUM GreenView is a solution that takes advantage of those findings and provide an impactful solution to reduce the energy footprint of stream devices.

Ongoing research results on sustainable and energy efficient video streaming can be found on our website www.green-streaming.de.



Contact

Robert Seeliger
 Fraunhofer FOKUS
 robert.seeliger@fokus.fraunhofer.de

Tracking down deep and shallow fakes reliably

Is an audio file genuine? Answering this question is becoming increasingly difficult, even for broadcasters and news agencies. Technologies from the Fraunhofer IDMT can expose deepfakes and manipulated media content.

For teenagers, it might be a fun tool to have AI read content using the individual voices of friends. For news agencies, however, forgeries are a serious issue - as fakes continuously improve, distinguishing between fake and original becomes more difficult. This applies to shallowfakes, such as manipulation of existing material or decontextualization, where statements are placed in a misleading context, as well as deepfakes, where content is entirely newly created.

In the future, broadcasters and other companies will receive support in detecting such fake content. »Our Content Verification Toolbox offers a wide range of audio analysis tools to detect manipulation and synthesis in media content,« says Patrick Aichroth, group leader at Fraunhofer IDMT. This toolbox can reveal fakes that are no longer discernible to humans, but it can also help to speed up and better document analysis even for audible forgeries.

Footprint analysis, synthesis detection and provenance analysis

One method applied by the Fraunhofer IDMT team to achieve this is based on the analysis of recording footprints, which are traces in the audio signals that result, for instance, from the specific recording device used, from the format and encoding of the file, or from the power grid frequency interference. When the material is subsequently altered, it leads to inconsistencies. »By analyzing such traces and comparing them with claims about the content, we can detect shallow-fakes. For example, if it is claimed that a specific smartphone was used for recording, we compare this claim against the traces in the recording to confirm or refute it,« explains Aichroth.

Knowledge of recording traces also helps the Fraunhofer IDMT team to develop advanced detectors for speech synthesis. These detectors focus on specific differences between natural and artificial recordings, achieving a high degree of explainability,



which is particularly important for deep-fake detection.

Additionally, the team is developing tools to detect content reuse and transformations: Were segments of one content item reused in other items? Where are they located, and in which order were they created? Such information can be extremely important in detecting fakes. While audio footprint analysis is already exotic – with only a few groups working on it - Fraunhofer IDMT's combination of all these approaches provides a unique approach to the challenge.

Tools from various projects

The tools that are integrated into the Content Verification Toolbox originate from various internal and collaborative projects that have been carried out at Fraunhofer IDMT over many years. »We have been working on audio forensics and audio manipulation detection since 2011 and have accordingly accumulated a wealth of experience,« says Aichroth.

In projects such as »vera.ai« or »news polygraph«, the researchers continue to advance the tools in collaboration with partners.



When we use artificial intelligence to develop technologies for content verification, explainability, and interpretability have top priority.«



Contact

Patrick Aichroth
Fraunhofer IDMT
patrick.aichroth@idmt.fraunhofer.de

Multi-lingual subtitling made easy!

In our globalized society, new movies and shows are available pretty much anywhere. But, the majority of these productions are natively created in one of a few languages. And while Mandarin is the native language of 900 million people in the world, and more than 1.3 billion humans speak English to some degree, the majority of movie-goers relies on translated or subtitled versions – at home as well as in movie theaters.

With our easyDCP Suite, we've been supporting filmmakers and distributors in bringing their productions to movie theaters for more than 15 years.

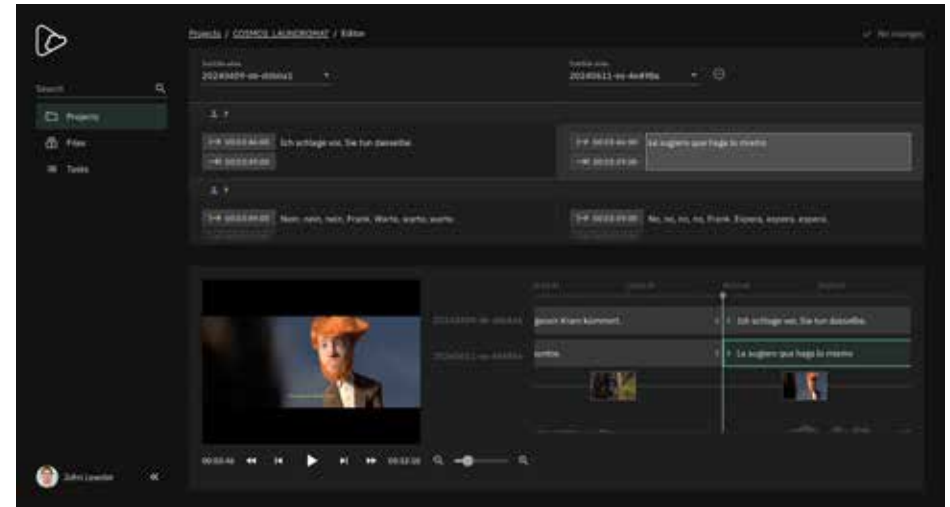
Continuous improvements to the software, frequent exchange with industry professionals, and our flexible licensing model have made our tools popular with film festivals and movie studios alike.

AI-based tools enable new workflows

The advancements in various fields of artificial intelligence let us re-think the workflows of movie localization. AI-assisted translations make it viable to create customized versions even for smaller markets or filmmakers on a budget.

Easier access to films for all audiences around the world thru localization

The ambitious plan for our Post-pro Cloud environment: Utilize the best AI-tools available and combine them into one easy-to-use localization workflow that does exactly what you need it to do. Starting with subtitle creation in the cloud, using your preferred transcription service, speaker detection and translation – available straight from easyDCP.



Translate with ease: Side-by-Side editing of different languages inside the easyDCP cosmos

Our flexible browser-based editor allows users to quickly compare and edit different language versions, using the side-by-side view. This makes subtitle adaptation for a wide range of language versions child's play.

Constantly expanding the feature set

As we are approaching the Post-pro Cloud next phase, we have already set our sights on the next technologies that will revolutionize what we expect from dubbed movies and shows.



Contact

Philipp Eibl
Fraunhofer IIS
philipp.eibl@iis.fraunhofer.de

MPEG-H Audio brings the television experience of the future to Brazil

The implementation of a next generation broadcast system in Brazil that includes enhanced audio experiences thanks to the Fraunhofer MPEG-H Audio system, recently celebrated some important milestones. The “TV 3.0 a TV do futuro!” campaign was unveiled by the Brazilian Ministry of Communications just before Globo presented a major showcase during an international sports event this summer and leading consumer electronics manufacturers released several new TV models on the Brazilian market.

Since Brazil started a major technological upgrade of their TV infrastructure, the eyes of the international broadcast industry have been on the leader of the Latin-American broadcasting ecosystem. To future-proof the new system, the most powerful technologies available today were chosen in a public and transparent evaluation process. They include HDR video as well as the latest in sound: personalized immersive audio. The Fraunhofer flagship technology MPEG-H Audio has been selected as the sole mandatory audio codec for the new TV 3.0 standard.

Leading broadcasters in Brazil, including Globo, Rede Amazonica, and TV Cultura worked together with Fraunhofer experts to implement MPEG-H Audio directly after its selection and have been using it in their regular ISDB-T broadcast services since 2021. Today, they already offer an

enhanced sound experience to their viewers and can ensure a seamless transition to the upcoming TV 3.0 next-generation broadcast system. As an accompanying activity, the Ministry of Communications recently presented the “TV 3.0 a TV do futuro!” campaign that introduces the new system to the public.

A strong tradition of sports enthusiasm means that the Brazilian audience is keen to enjoy high-profile sports events in highest quality and expects their broadcasters to deliver outstanding experiences. And this summer, Globo, the largest media group in Latin America, exceeded these expectations by presenting a major showcase for the occasion in an ambitious display of cutting-edge technologies, including MPEG-H Audio. To make sure that the showcase included the entire broadcast chain, the leading TV manufacturer Hisense released



MPEG-H Audio for Personalized Immersive Sound

TV sets supporting the new Brazilian broadcast technologies in the run-up to the games.

“We are proud to contribute to providing the best possible sound experience for broadcast and new media services to Brazilian viewers,” says Adrian Murtaza, Senior Manager Technology and Standards at Fraunhofer IIS, primary developer of the MPEG-H Audio system. “They will now be able to enjoy an entirely new experience in their own home and we believe that sports is a great way to showcase the advantages of our sophisticated audio system.”

Visit mpeg-h.com to discover all newly released devices and to learn more about MPEG-H Audio.



Contact

Elena Burdiel
Fraunhofer IIS
amm-info@iis.fraunhofer.de

Enhancing Emergency Broadcasting thanks to EWFplus including WorldDAB EWS

Fraunhofer IIS introduces the latest advancements in emergency broadcasting technology. With the integration of the new WorldDAB EWS standard in the DAB ContentServer technology as well as the innovative and comprehensive EWFplus solution, public safety and communication can be improved significantly. Stay informed and ahead in the ever-evolving world of digital radio.

In emergency situations, stable and efficient communication is essential.

Fraunhofer IIS played a key role in the standardization of the new and enhanced Emergency Warning System (EWS) signaling that was recently finalized by WorldDAB. Fraunhofer's DAB ContentServer head-end solution is the first to implement this new standard. EWS signaling wakes up receivers in standby mode so that authorities can transmit important information quickly and reliably at any time.

In Germany and other European countries, DAB+ is recognized as a robust and reliable technology for informing and warning the public in case of an emergency. EWS is now an integral part of EWFplus, the comprehensive solution for alert dissemination on the DAB platform that combines EWS signaling, audio announcements, and Journaline® advanced text information.

The full EWFplus functionality including EWS can be tested at the Fraunhofer IBC

booth. It has been further expanded by including the text-to-speech technology Allinga. In emergency situations, this enables maximum time savings from the arrival of an emergency message in the newsroom to its broadcast. In addition, information can be broadcast in multiple languages without having to wait for professional native speakers.

EWFplus: Symbiosis of audio and text for greater safety

In the event of a major disaster such as an incident at a chemical plant, every minute counts and communication between authorities and the public must be reliable and clear. EWFplus for the DAB platform not only wakes receivers from standby mode, but also takes care of the content. With the addition of the Journaline® advanced text application, it is possible to reach hearing-impaired people through text and provide multilingual information.



Thanks to the interactivity of the receivers, information and instructions can be transmitted: in case of a tsunami warning, listeners can look up the nearest shelter locations in a structured list with just two clicks.

The global Digital Radio Mondiale (DRM) standard already comprises the full EWF Emergency Warning Functionality – including alarm signaling, audio announcements and Journaline® text instructions – as a

default feature, allowing to reach listeners with urgent wake-up services and instructions in cases of natural disasters such as earthquakes or tsunamis approaching a shore. IBC visitors can test the DAB/DRM Multimedia Player App as an example implementation in a handset and find out which additional components make EWF and EWFplus a first-class solution.

New features of the MultimediaPlayer Radio app and the ContentServer Technology

The MultimediaPlayer radio app unleashes the full potential of digital radio on smartphones, tablets, and PCs for both consumers and professional users.

On Android devices, the audio output device can now be selected manually, and additional RF tuner dongle models are supported – including the popular RTL-TCP Blog V4 and HackRF.

The latest update to the DRM Content-Server technology saw the addition of a file-based DRM modulator feature, simply creating IQ files from an MDI multiplex recording. This is a particularly welcome addition for many companies operating DRM ContentServer-based systems for in-lab receiver development and testing.

In addition, both the DRM and DAB flavors of the ContentServer technology feature a whole new Journaline® encoder re-written from scratch, with greatly enhanced feature support, unmatched flexibility in XML and image file handling on upload, and a new XML format specification with lots of application examples and templates ready for direct copying.

Journaline® is the core of a range of innovative radio applications, including Distance Learning, Public Signage, and, most recently, EWF/EWFplus.

The new Journaline® encoder features offer broadcasters an even more straightforward way to create engaging content and tap into new revenue streams through features such as listener interactivity, audience measurement, and listener targeting.

Contact

Alexander Zink
Fraunhofer IIS
digitalradio@iis.fraunhofer.de



20 Years

Fraunhofer
Research and Development
for Digital Media

Fraunhofer Institute for Integrated Circuits IIS, Erlangen
Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institut HHI, Berlin
Fraunhofer Institute for Open Communication Systems FOKUS, Berlin
Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS, St. Augustin
Fraunhofer Institute for Digital Media Technologie IDMT, Ilmenau



Tailored Media Presence Measurement for the next level – InsightPersona

Media archives are a treasure as well as a nightmare for broadcasters. To unveil the potential and get the most out of these hidden treasures, it's necessary to fast, securely and reliably search and find the right information. How many female detectives are in the protagonist role of the popular German TATORT TV program in the last two years? Was Jennifer Aniston among the visitors of the Oscar Awards Ceremony? Media archives of broadcast and TV provide an almost endless amount of data, but it's hard to keep track of. InsightPersona provides an optimum tool to get these challenges answered.

InsightPersona, a solution designed by the Fraunhofer IDMT, enables you to use media data to answer all questions for a precise analysis or search of media presence. Which persons or actors are part of the afternoon series in my TV channels? Which key or buzz words were mainly used last week during the ballots and who were those persons who used them? Do we measure a special engagement or excitement in their speaking style?

When reporting counts

InsightPersona provides a lot of data for detailed analysis of media presence of individuals like politicians, celebrities, or groups on TV, in streams or archives. Using simultaneous face and speaker analysis as well as intelligent speech recognition you can either prompt images as well as spoken words or specific content to quickly

locate and identify the persons or voices you are looking for.

While reporting is becoming an essential part for an efficient exploitation of media archives, InsightPersona can help to optimize these tasks. Your content strategy can be optimized: Detailed statistics can support your ad marketing strategy or analyze and show whether guidelines for your content regarding gender, age classification or barrier-free access are fulfilled.

New Features at a Glance

- Combined face and speaker recognition
- Fast and reliable face search technology
- Analysis of vocal excitement level
- Word cloud functionality
- Listening effort analysis



Outstanding new features

"In the further development of our solution, we have focused on unique and outstanding new features", explains Christian Rollwage, Head of Audio Signal Enhancement at the Fraunhofer IDMT's Oldenburg Branch. Especially the analysis of the listening effort that helps to reduce complaints about hard-to-understand dialogues in films or shows is one of those features that are rarely offered on the market.

Addressing speaker training or text strategies, the analysis of the voice's excitement level together with the word cloud of the most used terms in audio and video content make a decisive difference for planning new formats or validating broadcast programs.

Many options tailored for your applications

"The big advantage for our customers are the tailored user interface that can be optimized for their specific analysis or reports", describes Rollwage the customer premise for InsightPersona. "If you use it either for analyzing talk shows or political debates or your media archive – we can flexibly adapt the software user interface to your needs".

More efficiency thanks to an optimized AI engine

InsightPersona is based on dedicated AI models optimized for huge media archives. The speaker engine was trained on 20,000 male and female speakers of every age and different pronunciations. Compared to common engines it uses a very small number of parameters to do its job – very fast and with low-energy consumption.

"The interplay of audio and video data is crucial. This allows for recognizing persons even when they are only seen without speaking or not being shown but can only be heard, such as in a talk show", explains Rollwage. The software runs completely on-site. This ensures that sensitive data does not have to be transferred via external networks.

An InsightPersona demo is available to get a closer look at its capabilities and to try it out.

Contact

Christian Rollwage
Fraunhofer IDMT
christian.rollwage@idmt.fraunhofer.de

Large Language Model integration in Fraunhofer Audio Mining

The Fraunhofer IAIS Audio Mining System is a media-content analysis platform that combines Automatic Speech Recognition (ASR) with speaker diarization and speaker recognition to make your content archive searchable. The platform has now been enhanced with cutting-edge technologies of exploring and finding content in your media content archives: at the IBC we will demonstrate how Large Language Models (LLMs) can help you achieve the best semantic search possible in your media archives.



LLM integrated in Fraunhofer's Audio Mining System

The new Audio Mining System uses LLMs, which are a sophisticated AI technology that understands and generates human-like text based on the input they receive, in addition to our already state-of-the-art ASR, which can analyze vast amounts of spoken content quickly and accurately, transforming speech into searchable text. LLMs are a technology which is the base of popular applications like ChatGPT, enabling interactive, conversational experiences.

The integration of this Generative AI (GenAI) into our Audio Mining introduces new, powerful features that enhance your ability to interact with your media archive. You can perform semantic search using natural language, meaning you can search your archive using specific questions rather than keywords. This makes it easier to find content related to broad topics or themes.



Moreover, this new way of searching also allows you to search for similar topics or statements within your archives, which is helpful for content creators and archivists in the media industry looking to streamline their workflows and enhance content discoverability.

With our single-channel speaker diarization and speaker recognition, our system can identify individual speakers in the audio and determine when they are speaking. Thus, you can ask if a specific person has made statements related to a particular subject anywhere in your entire archive. A final example is the ability to generate summaries of specific video files, such as news broadcasts.



Contact

Dr. Christoph Andreas Schmidt
Fraunhofer IAIS
christoph.andreas.schmidt@iais.fraunhofer.de

MP3
1995

SpatialSound Wave
2009

EVS
2014

MPEG-H Audio
2017

Content Verification Toolbox
2021

20 Years
Fraunhofer
Research and Development
for Digital Media

MPEG-4
1998

xHE-AAC
2012

Speaker Recognition
2017

Digital Radio Schooling
2020

2003
H.26x

2007
DCI Testplan

2010
HbbTV

2013
Omnicam 360

2014
CDMi

2018
JPEG XS

2020
VVC

2021
JPEG XS SDK

2005
ARRI D20/D21

2008
easyDCP

2010
Stereoscopic Analyzer

2012
Trifocal camera system

2016
Lightfield

2016
Volumetric Video

2019
360 Video

2022
Green Streaming

2023
Hybrid Video Conferencing

Fraunhofer Business Area DIGITAL MEDIA

The cooperation of Fraunhofer Institutes within the business area Digital Media provides innovative solutions and products for the digital age of motion picture.

We provide technological innovations for digital media workflows and for immersive viewing and sound experiences. Benefit from our expertise in research and development as well as in standardization. The institutes offer research and development in the areas of production, audio systems, data compression, post processing, transmission, projection, distribution and digital archiving.

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 Business Area 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). We are also a partner of the 3IT, the Innovation Center for Immersive Imaging Technologies and the Fraunhofer Digital Media Technologies, Fraunhofer USA, Inc. These contributions enable research and development activities based on international standards.

The Fraunhofer Institute members are

- Integrated Circuits IIS, Erlangen
- Telecommunications, Heinrich-Hertz-Institut HHI, Berlin
- Open Communication Systems FOKUS, Berlin
- Intelligent Analysis and Information Systems IAIS, St. Augustin
- Digital Media Technology IDMT, Ilmenau

Imprint

Publication Information

*Fraunhofer
Business Area Digital Media
c/o Fraunhofer Institute for
Integrated Circuits IIS
Am Wolfsmantel 33
91058 Erlangen, Germany*

*Concept and Publisher
Angela Raguse
Fraunhofer Business Area Digital Media*

*Editors
Dr. Janine van Ackeren
Angela Raguse
Jennifer Karbach
Marie Spies
Philipp Eibl
Hannah Pohl
Joran Poortstra
Robert Seeliger*

*Layout and production
Ariane Ritter*

*Photo acknowledgements
Cover picture: © Fraunhofer-Gesellschaft*

*Page 3: © HFF / Fraunhofer IIS
Page 5: © Fraunhofer HHI
Page 7: © adobe stock.com
Page 9: © istock.com
Page 15: © Fraunhofer IDMT/istock.com/
oatwa
Page 17: © Fraunhofer IIS
Page 19: © Adobe Stock/Thanaphon
Page 21: © istock.com
Page 25: © Fraunhofer IDMT/istock.com/vm
Page 27: © Adobe Stock/metamorworks*

Contact

*Fraunhofer
Business Area DIGITAL MEDIA
Angela Raguse M.A.
Phone +49 9131 776-5105
contact-digitalmedia@iis.fraunhofer.de
www.digitalmedia.fraunhofer.de*

Digital Media

In the Business Area Digital Media five Fraunhofer Institutes are cooperating to provide technologies and solutions for the media industry.

Fraunhofer IIS, HHI, FOKUS, IAIS, IDMT

www.digitalmedia.fraunhofer.de



Follow us on LinkedIn

