

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

issue 12

Preface



So happy to meet you in person!
IBC 2022 in Amsterdam

It feels like coming back home when I think that in only a few days the IBC 2022 will open its doors again and the industry gets back to one of its most favourite and important meeting places to discuss new ideas and innovations.

As Fraunhofer Business Area Digital Media we have plenty of news, updates and future-oriented projects in various fields and application of broadcast, streaming and moving pictures ready for IBC. So as for news about our codecs and implementation we will provide upgrades and updates for Versatile Video Coding VVC, for MPEG-H 3D Audio and for JPEG XS at our booth, showcasing the latest features, implementations and applications that help to guarantee even more quality and less streaming effort for excellent sound and video experiences.

We will demonstrate an interaction with dynamic 3D representations of real persons in a Mixed Reality context via eye contact. This truly immersive solution for the low-latency streaming of Volumetric Video is facilitated by the Low-Latency Low-Loss Scalable-Throughput (L4S) congestion control over 5G.

With Metaverse over 5G we demonstrate powerful multi-user Metaverse experiences on any platform, regardless of client device processing capabilities.

For speech-to-text conversion we show intelligent multimedia pattern recognition algorithms and with automatic subtitling for TV and video broadcasts with AI we facilitate multi-use of content in various contexts.

The latest version (R7) of our professional ContentServer technology for DAB+ digital radio and DRM, and on the receiver side, various technologies for decoding and displaying digital radio services as part of its SDR (Software Defined Radio) solutions are presented at our booth.

We are looking forward to demonstrating our new developments.

Enjoy reading our trends and topics.

Sincerely, yours



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

Topics

Preface 2

Topics 3

The JPEG XS codec for all IP workflows comes with new features 4

Video Data: Efficiently compressed, energy-saving playback 8

I was there: Experiencing stories up close and personal 10

The MPEG-H Authoring Suite 5.0 12

Understanding TV dialogue better – with MPEG-H Dialog+ Technology 14

Software Defined Radio 16

Music Analysis Technologies 18

Fraunhofer Business Area DIGITAL MEDIA 22

Impressum 23

The JPEG XS codec for all IP workflows comes with new features



JPEG XS is a special designed compression codec to enable transmission of high quality images up to 8k over Internet Protocol in production quality. The ISO standard for JPEG XS was issued in 2019 and the codec is now on its way to the market. A patent licensing pool program and several Software Development Kits SDKs were set-up to pave the way for industry partners to get access to and implement JPEG XS in various media workflows or devices.

The low-latency, low-complexity and standardized JPEG XS codec has now achieved good market acceptance. Many applications have already been realized, especially for broadcast content transfer to and from the cloud. The first edition of the standard ISO/IEC 21122 was originally designed to enable high-quality 422 and 444 broadcast content data transmission.

In parallel with the development of the coding standard in ISO, many companion standards have been developed that support the transport of JPEG XS over RTP (IETF RFC 9134) or MPEG-TS (ISO/IEC 13818-1) or store codestreams in MP4, HEIF or MXF (SMPTE ST2124) file formats. Associations like VSF and AMWA defined specifications for JPEG XS to enable higher interoperability in networked media. Due to market demand, further improvements resulted in a second edition recently published, and a third edition on its way.

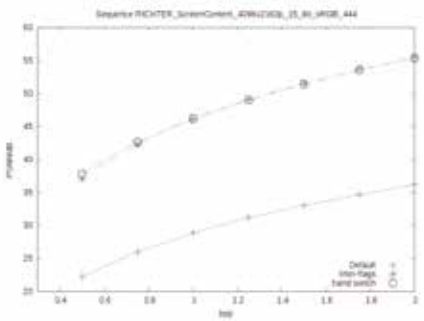
JPEG XS for RAW Bayer compression

The second edition added RAW Bayer data compression for CFA (Color Filter Array) image sensors to enable compression in digital cameras with highest quality. Image sensor data have specific characteristics in black level and noise behavior. While considering these characteristics an optimized coding with highest performance can be achieved. An improvement in quality up to 5dB could be demonstrated in comparison to data agnostic compression. Further features in the second edition are profiles for 420 image compression and mathematically lossless compression of image data.

Compression efficiency for screen and mixed signal content

A third edition of the standard has recently been launched. The focus is to achieve better compression efficiency for screen and mixed signal content. The build-in wavelet transform in the JPEG XS coding algorithm allows for a good energy compaction in natural images, resulting in good performance for this type of content. However, this alone is not ideal for screen content. Because of this, data from the previous frame in a sequence are now allowed to be used to predict the actual frame. This method does not increase the latency, but requires slightly more memory resources to store previous data in a buffer. We were able to improve the PSNR (Power Signal Noise Ratio) up to 10 dB for screen content. The third edition is still under development, but its publication is scheduled for begin of 2023.

Fraunhofer IIS offers JPEG XS SDKs for x86 CPUs, ARM CPUs and NVidia GPUs.



Contact

Siegfried Foessel
Fraunhofer IIS
siegfried.foessel@iis.fraunhofer.de



Video data: Efficiently compressed, energy-saving playback

If we want to continue coping with the flood of digital video data in all areas of life in the future, we need a way to compress it efficiently. This is possible with the VVC video compression standard. Where does it currently stand?

People used to get cozy in the evenings in front of the television, browsing whatever was on. Now the streaming services are becoming more and more popular.

The situation is similar for telephone calls: traditional audio connections are frequently being replaced by video conferences. As a result, compressed video data now accounts for 80 percent of the world's Internet traffic. To be able to handle this flood of data in the future as well calls for new compression standards.
One of these is Versatile Video Coding, or H.266/VVC – an international standard that Fraunhofer HHI played a major role in developing.

Compared to its predecessor, H.265/HEVC, VVC cuts the bit rate in half while maintaining the same quality. Benjamin Bross, Head of Video Coding Systems Group at Fraunhofer HHI, explains what has happened with the standard since last year. He also gives an update on the open-source VVC encoder and decoder implementations the Fraunhofer HHI researchers have developed.

Mr. Bross, what's the latest on the VVC standard?

For a great example of how VVC has been established with the help of our implementation, we can look to Brazil: right now the country is developing a new national broadcast standard for their next-generation digital television, TV 3.0. In the midst of these discussions, there was a call to propose a highly efficient codec for encoding videos with up to 8K resolution.

Of all the codecs submitted, the VVC standard we proposed was selected: it was the most efficient, and performed the best in all tests. The first task was to encode a whole series of test videos typical for TV, including in 8K resolution. Here the open-source VVenC encoder we developed at Fraunhofer HHI was used. At the second hurdle, the live demonstration, a live encoder from the world-leading company ATEME encoded the video data, while our VVdeC decoder handled the decoding for playback.



Benjamin Bross
Fraunhofer HHI

Turning closer to home, VVC was also included and standardized in Digital Video Broadcasting, or DVB. This means that broadcasters in Germany, too, can use VVC for future television transmission, whether terrestrial, cable, satellite or Internet.

What's been going on regarding the open-source VVenC encoder?

An important factor in encoders is to keep the energy consumption for software decoders as low as possible – all while ensuring high compression efficiency. In this area, we collaborate with the Friedrich-Alexander-Universität Erlangen-Nürnberg. The scientists there are investigating how different encoder parameters affect energy consumption. Together we have developed a parameterization for our VVC encoder, VVenC, which leads to low energy consumption in software decoders.

In another collaboration with Sony and Spin Digital – the latter is a spin-off of Fraunhofer HHI and TU Berlin that specializes in commercial codec development – we demonstrated that VVC is an ideal codec for high-resolution video content in 8K. This is where VVenC's excellent compression efficiency really shines.

In addition to the open-source VVenC encoder, Fraunhofer HHI has developed an open-source decoder, VVdeC, which decodes the transmitted files at the user's device. How do things stand at the moment?

A lot has happened regarding the VVdeC decoder as well. The supported computer architectures have been extended to include ARM processors, so that VVdeC now also runs on mobile devices and new Apple computers with an M1 processor. If the low decoding energy parametrization in VVenC that I mentioned before is used during encoding, VVdeC also consumes less power on mobile devices. For example, an Android tablet can play five hours of VVC-encoded video material without needing to be recharged. In addition, VVenC is going into the next round of testing at regional broadcaster Westdeutsche Rundfunk (WDR). Currently, the focus there is also on playback of VVCencoded videos on mobile devices.

Contact

Benjamin Bross
Fraunhofer HHI
benjamin.bross@hhi.fraunhofer.de



I was there: Experiencing stories up close and personal

Real experience is three-dimensional, while video is "flat" and 2D? Hardly! With interactive volumetric video, people captured on film appear to be standing live in the room right in front of the viewer, eye contact included. This is a particularly good idea in the case of reports by eyewitnesses to historical events.

If someone tells you their life story while standing right in front of you, looking you right in the eye, the narrative moves and touches you in a completely different way than if you were to hear it from an image on a flat screen. And when it comes to eyewitnesses from the time of the Holocaust, such personal accounts are naturally becoming increasingly rare. However, Fraunhofer HHI's interactive volumetric video technology can help preserve life stories – narrated by the witnesses themselves as they stand face to face with you. Not live, of course, even though it may feel that way to viewers. Using VR goggles, they see the eyewitness – such as Eva Umlauf, one of the youngest survivors of the Auschwitz concentration camp – right in front of them in 3D. They can even walk around her. What's special about this medium is that Eva looks at the listener and follows them with her gaze as they move.

"To do this, we recorded Eva Umlauf from every angle simultaneously using 32 cameras, all arranged in stereo pairs," explains Dr. Oliver Schreer, group manager at Fraunhofer HHI. The Fraunhofer researchers carried out the recordings as part of the VoviRex project, a collaboration with the Ludwig Maximilian University of Munich. The stereo pairs of cameras not only capture a "normal" image, but also estimate the distance to the person being filmed – in other words, they create a depth map. Having obtained 16 such depth maps, the researchers first combine them into a point cloud and then convert that into meshes. "This lets us subsequently animate the recorded person so they can, for example, maintain eye contact with the users," says Wieland Morgenstern, a scientist at Fraunhofer HHI.



Eva's life story can also be incorporated into school lessons, although equipping each of the students with a gaming PC and VR goggles is somewhat complicated. "For this and similar applications, we can render the three-dimensional scene in the edge cloud – basically, on a computer located close to the classroom," explains Dr. Cornelius Hellge, group manager at Fraunhofer HHI. From there, the user's current view of the scene is transmitted to their goggles as a 2D video. Each student has to have a headset, as the eyewitness's gaze is supposed to be adjusted to each viewer. In other words, every single student sees an individual volumetric video in their VR goggles. An alternative to this VR application is augmented reality, which eliminates the need for expensive VR headsets. Here, each student can look at their tablet to see both the classroom setting and the virtual eyewitness up at the front, sharing her moving life story.

Contact

Oliver Schreer
oliver.schreer@hhi.fraunhofer.de

Cornelius Hellge
cornelius.hellge@hhi.fraunhofer.de

Wieland Morgenstern
wieland.morgenstern@hhi.fraunhofer.de

Fraunhofer HHI

The MPEG-H Authoring Suite 5.0: Your complete tool chain for the production of customizable immersive sound



The MPEG-H Audio system delivers immersive sound on every kind of playback device – from home theaters to 3D soundbars to mobile devices. With its unparalleled range of customization options, it also helps meet modern audiences' demand for made-to-measure media content with a high degree of accessibility.

For the easy creation of such audio experiences, Fraunhofer IIS has developed a complete set of tools for the creation of audio and video content with object-based 3D Audio. The MPEG-H Authoring Suite (MAS) 5.0 now includes all tools required for authoring, encoding, and playback of MPEG-H Masters well as encoded MPEG-H. With the comprehensive range of instruments, producers and broadcasters deliver immersive soundscapes that users can adapt to their personal needs and preferences. This can happen within a predefined framework by including fully adjustable dialogue levels, customizable audio description, multiple languages, and even interactive object positioning.

The MAS 5.0 makes the production of MPEG-H easier, faster, and available for everybody. It can be used in all areas of the field from radio and TV broadcast to immersive music streaming and supports authoring and export in production environments running at 48kHz or 96kHz sample rates. It comprises:

- The MPEG-H Authoring Plug-in (MHAPi) that covers all steps of creating object- or channel-based MPEG-H Audio productions inside a VST3- or AAX-enabled Digital Audio Workstation (DAW) and its standalone version, the MPEG-H Authoring Tool (MHAT) that enables users to create MPEG-H metadata for existing audio material without the need of a DAW.
- The MPEG-H Conversion Tool (MCO) that converts Next Generation Audio (NGA) masters between different file formats. It also serves as an interface to the MPEG-H Audio ecosystem and supports the import and export of MPEG-H Master files.
- The MPEG-H Info Tool which analyzes NGA Master files and provides all relevant information in a convenient scene summary.
- The MPEG-H Production Format Player (MPF Player) that enables quality control before encoding with the new MPEG-H Encoding and Muxing Tool.
- The MPEG-H vvPlayer for playing back the mp4 delivery format.
- Learn more about MAS 5.0 and download it for free on https://www.iis.fraunhofer.de/mas

Contact

Mandy Garcia
Fraunhofer IIS
amm-info@iis.fraunhofer.de

Understanding TV dialogue better – with MPEG-H Dialog+ technology

Most TV stations are quite used to their audience complaining about hard-to-understand dialogue – be it in films, documentaries, sports coverage, and even the news. The matter is not an easy one to solve. Firstly, because the loudness difference between background sound and dialogue is a unique decision made by creators for every piece of content, and secondly, because the "perfect" dialogue loudness is a very personal issue.

The evolution of AI-based technologies and object-based audio (OBA), however, has enabled the creation of technologies such as MPEG-H Dialog+ by Fraunhofer IIS. The technology uses Deep Neural Networks (DNN) to automatically identify the dialogue of existing content, separate it from the background sounds, and remix it with a lowered background level. Using OBA, users can even adapt the dialogue level on their device to meet their personal requirements.

Recently, Fraunhofer IIS joined forces with German public broadcaster WDR and Telos Alliance to develop a professional workflow and bring MPEG-H Dialog+ into use. Fraunhofer IIS conducted field tests over DVB and the VoD platform "ARD Mediathek" to refine requirements and production workflows. The results were then fed into the product development of the Telos Alliance Minnetonka AudioTools Server Dialog+ module. The software has now been implemented as part of an automatic workflow – from archive to transcoding farm – in the WDR production infrastructure.

MPEG-H Dialog+ contains a deep neural network that performs dialogue separation. Most training data is real-world broadcast content, mostly provided by WDR and other ARD broadcasters. Dialog+ combines dialogue separation with a unique automatic remixing algorithm, where a global and a time-varying background attenuation can be combined. Global background attenuation lowers the relative level of the estimated background component by the same specified amount over the entire signal. This can be beneficial for users that prefer to always lower the background signal. For others, this might not be the optimal solution, as attenuating the background while the dialogue is not active does not improve speech intelligibility while potentially damaging



mood, atmosphere, and sounds of narrative importance. A solution is to lower the background level only when the dialogue signal is active and only as much as is necessary to reach the desired level.

Thanks to the implementation of MPEG-H Dialog+, which is called "Klare Sprache" in the ARD Mediathek, the VoD platform now provides a higher degree of accessibility. Additional benefits of the automated audio processing workflows include:

- Automated, cost saving, and scalable workflow approach
- State-of-the-art quality of the dialogue separation algorithm
- Dynamic remixing algorithm which only affects the background level when dialogue is present. This prevents unwanted changes to the mix and helps to preserve the artistic intent as much as possible.
- Set of presets customized for different use cases. This way, the content provider can apply processing optimized, for example, for documentaries, music films, and sports content.

MPEG-H Dialog+ is part of the MPEG-H Audio production software, providing all features of an OBA system like advanced user interactivity and personalization. This makes the use of MPEG-H Dialog+ a future-proof decision for broadcasters and content producers as MPEG-H Audio is one of the most advanced Next Generation Audio systems on the market. It has already been chosen as a TV audio standard by countries such as Brazil and South Korea.

Contact

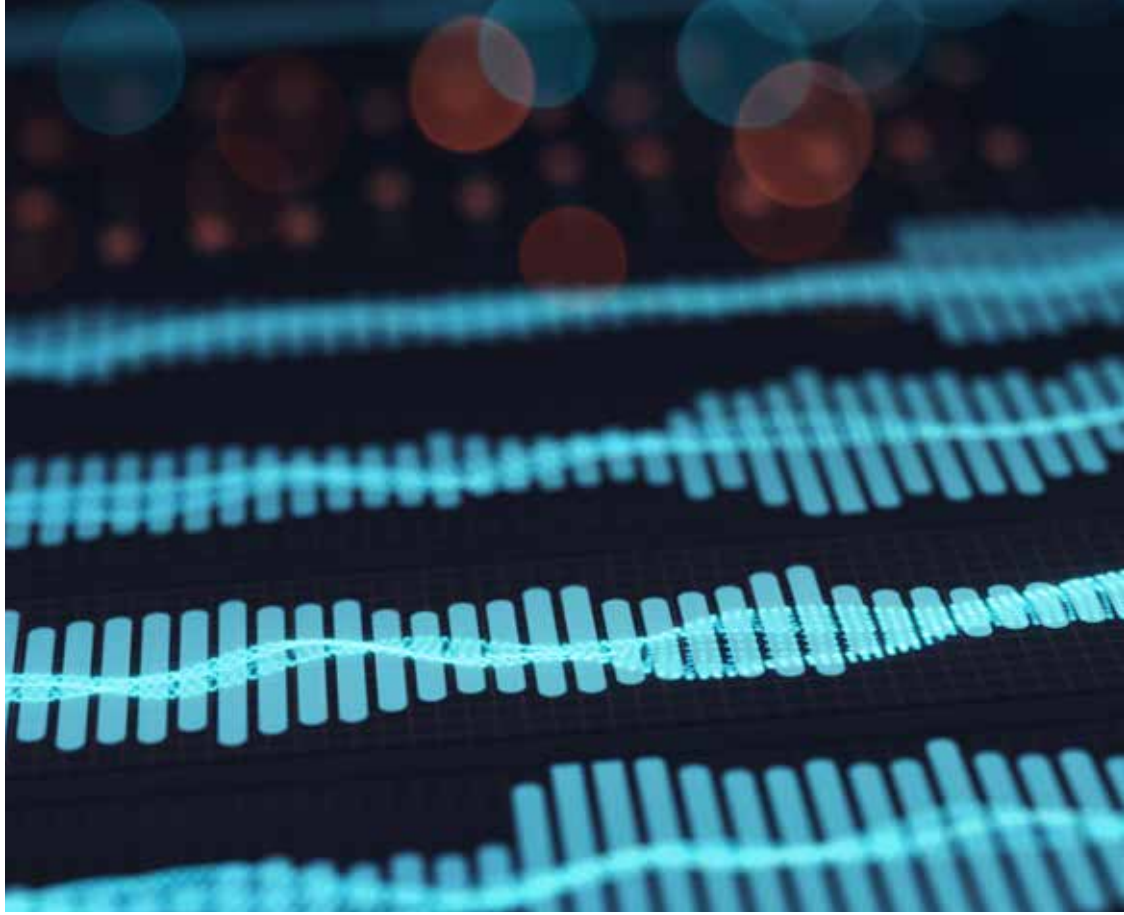
Mandy Garcia
Fraunhofer IIS
amm-info@iis.fraunhofer.de

Software Defined Radio

Bad reception is a thing of the past with the help of Fraunhofer software components: They allow radios to automatically switch between analog radio, digital radio, and Internet streaming for uninterrupted listening pleasure. New algorithms also guarantee improved reception quality everywhere.

The development of digital radio in recent years is remarkable, because it continues to gain listeners and can grow against the trend. At the same time, Internet radio primarily enhances the listening experience, without replacing its broadcasting counterpart. That can be seen for consumer, as well as car radios. In both cases, radio functionality will be just one component in the comprehensive infotainment system of the future. The radio receiver becomes an app for receiving broadcast content. This functionality is a software component that runs on modern computing platforms. Thus, it is flexible in using different platforms ranging from Android or any Linux operating system to even embedded systems without a powerful OS. Depending on the target market, the radio system can be configured for low-end audio devices to high-end infotainment systems with large displays, multiple antennas for best reception quality, and hybrid radio solutions that combine digital, analog and internet radio in one device.

For more than 20 years, Fraunhofer IIS is developing software libraries for DAB+ and DRM radio receivers for cars and consumer devices. The software development is following the demands of the automotive industry for high quality and reliable systems. New methods for hybrid radio systems ensure seamless listening experience by automatically switching between the best reception among analog, digital and internet radio. At the same time, the reception quality in the noisy environment of electric vehicles is being improved through advanced algorithms. The software radio architecture gives the highest flexibility for implementing the identical functionality on different hardware platforms to react on shortages in the semiconductor industry.



Contact

Martin Speitel
Fraunhofer IIS
martin.speitel@iis.fraunhofer.de

Music Analysis Technologies



How can I quickly locate a specific piece of music in my vast library of videos? Can I match a specific beat to a new production I'm working on? Which programs in my archive are the most successful? These are typical questions that a broadcaster might pose to Fraunhofer IDMT's Semantic Music Technologies Group, whose big picture is that when working with the content in an audiovisual archive, you need to be able to retrieve information by making it searchable.

Headed by Hanna Lukashevich, the group focuses on solving complex problems for music technology companies and professional media organizations involved in all kinds of audiovisual production. Use cases include locating a precise audio clip or specific type of sound, matching a particular mood, or enabling search across entire music libraries such as the broadcast archives of a TV or radio station, and analyzing programs statistically, based on acoustic analysis techniques.

"Our technologies are designed to give broadcasters an edge over their competition at a moment when competition for viewers is incredibly strong. When you can identify content and make informed business decisions quickly, this can drive new creative and business possibilities, says Hanna Lukashevich. "One cannot manually locate the content in a big archive with any degree of speed, and in today's business environment, that just doesn't scale."

The group uses classical signal processing algorithms together with AI/ML and recommendation technologies in a "multi-modal approach" that is very effective in bringing new solutions to common, but very challenging problems.

Its professional consulting services are used, for example, by big music archives that offer specialized search engines to professional producers. When enhancements are required, the group will first fully assess and then start to build the right components that improve performance or more fully automate the search processes.

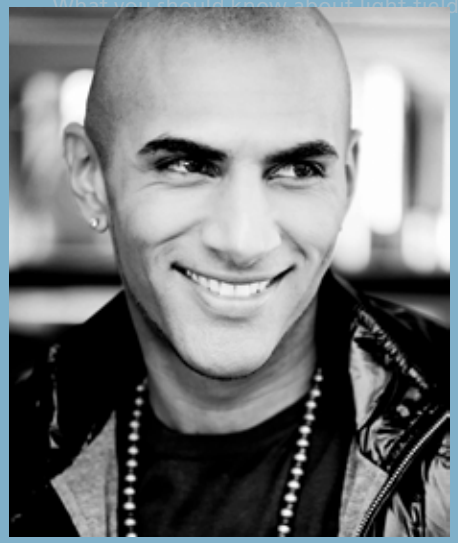
On delivery, the system is not only customized for a specific set of tasks, but future enhancements can be seamlessly added.

Professional producers always need fast access to many music options. They depend on such search engines to find specific sounds, or specific beats and loops that perfectly match and complement a given music production.

This is the business of Jamahook, a Swiss music technology company. Jamahook operates a vast database of loops and sounds and offers a plugin that allows music producers to find the most suitable audio loops for a production, based on examples extracted from an audio mix. The plugin is the result of a long-term technology cooperation with Fraunhofer IDMT.

Jamahook's latest feature is called "Pitch Shifted Matching". After analyzing an example track the software recommends a collection of loops from the Jamahook database and can even shift the loops into the pitch that matches the example. This provides tremendous flexibility enabling access to a larger selection of loops and sounds to find the perfect match for a creative work.

Another common use case is music replacement. Often in a production placeholder music is used until licensing and rights issues are resolved. "Our multi-modal approach enables us to match the overall feeling of a production by training our models with granular attributes like rhythms, tempo and instrumentation," says Hanna Lukashevich. This "learned"



© Michael Orlik

"After an initial meeting at Berlin Music Week, where we discussed the challenge of very quickly identifying the right beats that a producer might need, we decided to use "SoundsLike", Fraunhofer´s machine learning based software, for this task. With great results, Jamahook was able to release a unique new algorithm, "Soundmatch", that helped expand our business.
"This experience was the beginning of a solid, ongoing collaboration. Now we´re developing the world's first AI music mobile app based on Fraunhofer's music classification technologies, that will allow anyone to create unique, high-quality, royalty-freely music in the genre and instrumentation of their choice. We really look forward to continuing the collaboration to implement many more new and exciting features in the future.

"The combination of creative minds, experienced engineers and innovative team members has made it possible for Jamahook to glue "creativity" and "engineering" together to create unique products and get the best out of both worlds."

Karim Bhorania, CEO, Jamahook

knowledge can then be applied to a specific task, such as the need to identify the most important musical instruments in a score, or by making them searchable at the precise moment they are played.

The model keeps on learning. It can infer and make an interpretation of new, unseen music data on its own. Over time, the models get even better at their tasks. In the case of music detection, accuracy rates of 99% for foreground music and over 80% for the detection of low-volume background music can be achieved.

"Customers can win big in terms of time and efficiency savings", says Hanna Lukashevich. "There's just no time to manually sift through all of that material, so you need smarter tools when you are looking for something specific."

Like finding a needle in a haystack, Fraunhofer IDMT's Semantic Music Technologies Group addresses a class of work that simply cannot be done manually.

"We even help audience measurement companies acoustically analyze the performance of their programs and do similar work analyzing ads, to help them determine those that are the most successful," says Hanna Lukashevich.

Fraunhofer IDMT's Semantic Music Technologies Group helps its customers improve their archive management processes and optimize their businesses. Based on careful analysis and then, by delivering the components that are right for the job, customers are enabled to work in a highly accurate and automated way, at scale, using the multi-modal approach which is the strength of Fraunhofer IDMT.

Contact

Hanna Lukashevich
Fraunhofer IDMT
hanna.lukashevich@idmt.fraunhofer.de



Fraunhofer Business Area DIGITAL MEDIA

The cooperation of Fraunhofer Institutes within the business area Digital Media provides innovative solutions and products for digital media workflows.
We provide technological innovations for workflows, 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 of the Fraunhofer Digital Media Technologies, Fraunhofer USA, Inc. These contributions enable research and development activities worldwide 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

Impressum

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
Florian Meister
Janet Greco

Layout and Production
Ariane Ritter

Photo acknowledgements
Cover picture: © istock.com
Page 2: ©HHF München / Robert Pupeter
Page 5: ©istock.com
Page 6: ©istock.com
Page 7, 9: © Fraunhofer HHI
Page 11: © Tierney, Adobe Stock/Fraunhofer HHI
Page 13: © Fraunhofer IIS
Page 15: © Fraunhofer IIS/ David Hartfiel
Page 17: © istock.com
Page 19: © istock.com/cihatatceken
Page 22: © Michael Orlik
Page 22: © istock.com / Fraunhofer Digital Media

© Fraunhofer-Gesellschaft

Contact

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