AI-BASED MEDIA CODING AND BEYOND

L. Chiariglione, MPAI, IT; A. Basso, MPAI, IT; P. Ribeca, BioSS, UK; M. Bosi, Stanford University, US; N. Pretto, Audio Innova, IT; G. Chollet, IMT, FR; M. Guarise, Volumio, IT; M. Choi, ETRI, KR; F. Yassa, SpeechMorphing, US; R. Iacoviello, RAI, IT; A. Artusi, CYENS, CY; F. Banterle, CNR-ISTI IT; F.Gissi, Kebula, IT; A. Fiandrotti, Università di Torino, IT; G. Ballocca, Sisvel Technology, IT; M. Mazzaglia, Synesthesia, IT; M. Rosano, Politecnico di Torino, IT; S. Moskowitz, Wistaria Trading, US

ABSTRACT

MPAI – Moving Picture, Audio and Data Coding by Artificial Intelligence is the first body developing data coding standards that have Artificial Intelligence (AI) as its core technology. MPAI believes that universally accessible standards for AI-based data coding can have the same positive effects on AI as standards had on digital media. Elementary components of MPAI standards - AI Modules (AIM) - expose standard interfaces for operation in a standard AI Framework (AIF). As their performance may depend on the technologies used, MPAI expects that competing developers providing AIMs will promote horizontal markets of AI solutions that build on and further promote AI innovation. Finally, the MPAI Framework Licences provide guidelines to IPR holders facilitating the availability of compatible licences to standard users.

INTRODUCTION

MPAI - Moving Picture, Audio and Data Coding by Artificial Intelligence is an international, unaffiliated standard organisation. Its mission is to promote the efficient use of data by developing standards of coding any type of data, especially using new technologies such as Artificial Intelligence, and technologies that facilitate integration of such data coding components in complete systems.

Two-fold motivations drove the establishment of MPAI. The first is rooted in the belief that, while data processing technologies have enabled massive use of digital technologies benefitting industry players and consumers alike, the propulsive force of those technologies is reaching the limits. On the contrary, the scope, performance and applicability domain of Artificial Intelligence (AI) technologies are growing. The second motivation is the recognition that the system that has governed the transfer of innovation in the form of Intellectual Property (IP) to standards, and products/services is showing serious signs or wearing out. A revision of past practices to guarantee the irreplaceable role of IP and to iron out the path from innovation to standards and then to products/service is needed.

To perform its mission MPAI has worked out a process that allows it to reach out to those in need of a solution and those who have the technologies that can satisfy them. The process is designed 1) to live up to the promises implicit in its ".community" 2nd level domain name in the phase where needs and functional requirements are identified; 2) to allow all MPAI members to participate in the development of the standards that satisfy the identified needs, and 3) to delegate to those members who have elected to be allowed to do so (Principal Members) the task of developing the means to facilitate the development of IP guidelines.

The process goes through a first stage where needs are collected, followed by six stages: identification of one or more Use Cases (UC) for which Functional Requirements (FR) are

developed and made public. Anybody can participate in these three stages. When the FRs are consolidated, Principal Members who plan on contributing to the standard define the IP guidelines in the form of a Framework Licence (FWL). All members participate in the development of a Call for Technologies (CfT) which is published. Anybody can respond to a CfT. However, if some of the technologies proposed by a non-member are accepted on technical grounds, joining MPAI is a condition for having the technologies included in the standard. Principal Members adopt the standard for publication.

MPAI was established on 2020/09/30. In less than 3 months it has produced its first CfT, in less than 5 months, two more CfTs and in less than 6 months its fourth CfT. Responses to the first three CfTs have been received, new members have joined because their technologies have been accepted and the three standards are under development. This paper will introduce the work that is being carried out for the three standards and the fourth standard whose responses are due in a matter of days. MPAI has adopted an innovative standard development approach that relies on components called AI Modules (AIM) whose aggregation is executed in an AI Framework (AIF). An MPAI standard normatively defines the functionality and input/output data formats of an AIM, the topology of the AIMs in the AIF, and the functionality and input/output data of the AIF that implements a Use Case. MPAI is well aware of the impact that AI technologies will have on the use of some of its standards. Therefore, it is consolidating guidelines for conformance testing of AIMs and AIFs implementing Use Cases. The next steps will address the performance of AIMs and AIFs implementing Use Cases, performance being a name that includes reliability, robustness and fairness. This is work in progress and will not be reported in this paper.

The rest of the paper introduces standards work at different level of maturity: the first standard called AI Framework (MPAI-AI), the second and third standards called Context-based Audio Enhancement (MPAI-CAE) and Multimodal Conversation (MPAI-MMC), respectively, AI-Enhanced Video Coding (MPAI-EVC), Server-based Predictive Multiplayer Gaming (MPAI-SPG), Compression and Enhancement of Industrial Data (MPAI-CUI) and Integrative Genomic/Sensor Analysis (MPAI-GSA). Finally MPAI's Framework Licence approach to standardisation is presented.

AI FRAMEWORK

MPAI issued a Call for Technologies 'MPAI (1)' for MPAI-AIF on 16 December 2020. Responses were received on 14 February 2021. MPAI is currently reviewing the responses received that it will use to develop the standard, planned for July 2021. The reference model of MPAI AI Framework (MPAI-AIF) is depicted in has the reference model of Figure 1. The scope of the MPAI-AIF standard is to provide a framework that allows easy interconnection and interoperability of AI and data processing technologies implemented both in HW and SW, when encapsulated in modules with standard interfaces called AI modules (AIMs). MAPI-AIF adopts the philosophy of components-based development (CBD), enabling the reuse of independent components (AIMs) into systems.

In MPAI-AIF the AI Modules (AIMs) communicate with one another via standardised interfaces; they specify the services that other components can use, and how that can be done. In such a manner, there is no need to know about the specific implementation of the AIM in order to use it. A set of requirements have been defined as described in 'MPAI (2)' and include:

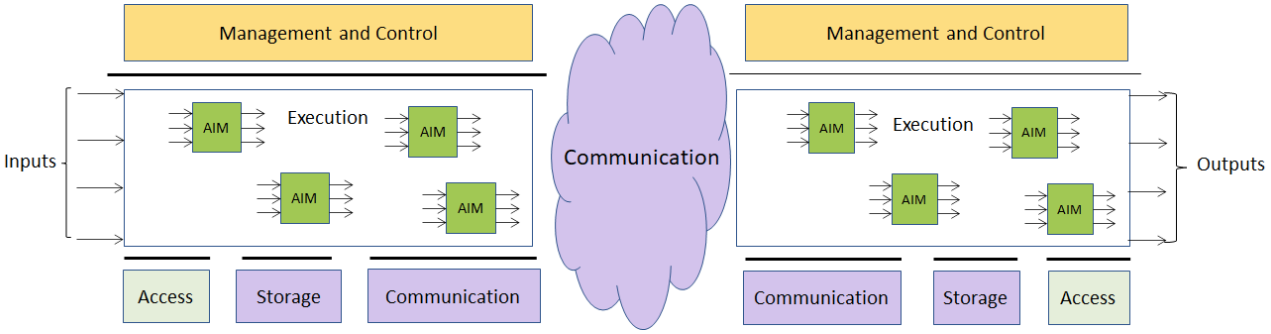

Figure 1: MPAI-AI Framework (AIF)

1. Support for single-AIM life cycle, including initialize, instantiate-configure-remove, start-suspend-stop, dumping/retrieval of internal AIM state, and enforcement of resource limits

2. Support for multiple-AIM life cycles including manual, automatic, dynamic-adaptive interface configuration of AIMs

3. Support AIM machine learning including training-retraining-update of AIMs — configure/reconfigure ML computational models, dynamical updates of the ML model, supervised/unsupervised/reinforcement-based learning paradigms

4. Support of workflows and in particular: (a) hierarchical execution of workflows represented as computational graphs, with Direct Acyclic Graphs (DAG) as a minimum; (b) AIM topologies can be synchronised according to at least one time base and to ML lifecycles; (c) One and two way signalling including initialisation; (d) Control, Communication functionalities; (e) Security policies between AIMs and at AIF level.

To reach HW-SW interoperability and maintain the framework general, MPAI plans to provide different profiles that cover scenarios SW-only and HW-SW. The extensive usage of metadata will minimize the issues related to differences in implementations. The difference in profiles is related to constraining the interfaces to subsets of elements to adapt to the limited flexibility of HW, i.e. in a signalling scenario the number of signals handled can easily programmed in SW to be very large while in HW, it will be constrained by the physical signals available.

A given logical graph of interconnected AIMs needs an execution model that determines the way in which the processing units are scheduled for execution. MPAI-AIF is currently evaluating different execution models, inheriting a unified approach to AI-based audio-visual data processing standardisation concepts from MPI (Message Passing Interfaces) and CWL (Common Workflow Language) and providing in addition a hierarchical level of workflow representation, that provides management and control of combinations of AI modules but also the possibility of AIMs to be interconnected and executed in constrained resources scenarios (MCUs). MPAI-AIF adopts secured, credential-based AIMs, static or dynamic registration together with their associated metadata. In dependence of the profiles MPAI-AIF will support event-based as well as signal-based signalling. MPAI-AIF provides mechanisms for resource management particularly useful in resource-constrained environments. Resource policies are enforced at both workflow level and at AIM level. One of the key characteristics of MPAI-AIF is its support for specific Machine Learning functionalities and in particular training and retraining and dynamic update of ML components. MPAI-AIF provides shared storage and communication mechanisms that are related to the execution models that will be adopted.

ENHANCED AUDIO APPLICATIONS

A specific MPAI area of work, the MPAI Context-based Audio Enhancement (MPAI-CAE) 'Bosi et al (3)', is showing tremendous potential for audio applications. MPAI-CAE applies context enhanced information to the input audio content to deliver the audio output via the most appropriate protocol. Four MPAI-CAE case studies have been currently identified: Audio recording preservation (ARP), a use case that covers the preservation/restoration process of audio documents, from historical audio documents to their digitized preservation/access copies; Audio-on-the-go (AOG), which aims to improve safety and listening quality for situations in which the users are in motion in different environments; and Emotion-enhanced speech (EES), a use case that implements a user-friendly system control interface that generates speech with various levels of emotions. In this section we focus on the MPAI submissions relevant to the ARP use case (see Figure 2): we describe a complete and detailed approach to the preservation of historical analog audio tapes, then we concentrate on a specific solution for the restoration of speech documents from different media.

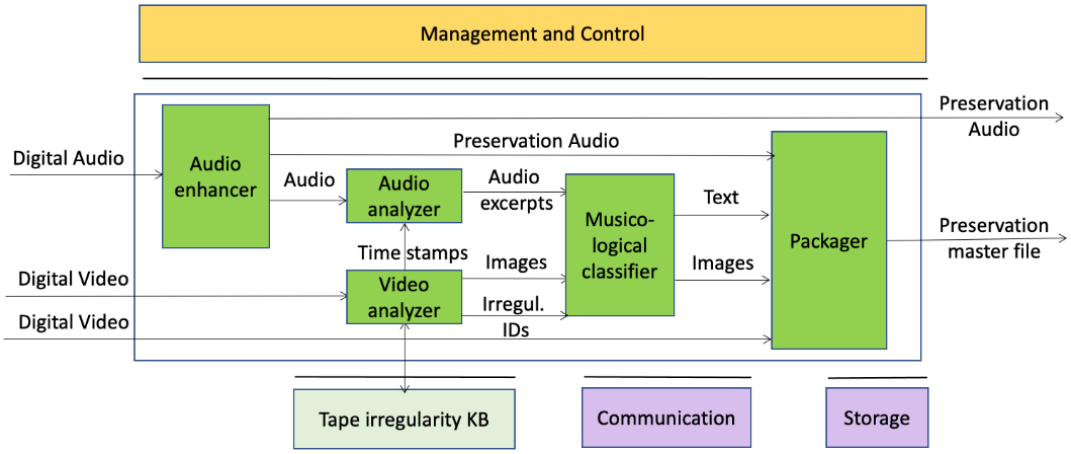

Figure 2: MPAI-CAE ARP workflow 'Bosi et al (3)'

MPAI is currently working on the responses to the Call for Technologies. One aspect concerns preservation of open-reel analog audio tapes, without the Audio Enhancer. An important contribution concerns a complete list of irregularities that can be found on a video of the tape (e.g., damages of the carrier, splices, marks), which extend the one proposed in 'Pretto et al (4)'. The Video Analyser detects the significant frames from the Digital video comparing consecutives frames. The detection is made on two areas: on the reading head and under the Pinch roller. The extracted frames are classified by the Musicological Classifier that select only the most relevant images of the irregularities that will be included in the Preservation master file. A second aspect is a novel way of restoring damaged or missing *audio* segments, and more specifically *voice* segments. Audio restoration has usually aimed to *repair* damaged vocal audio, for instance by filtering extraneous noise or by filling minor audio gaps by extrapolating from the surrounding audio, perhaps through exploitation of artificial intelligence techniques. However, if more substantial vocal gaps appeared, these methods might be blocked since in that case there would be nothing to repair, or can even distort the recordings due to extreme needs for cleaning of the voice track. In this proposal the damaged voice segments are completely *replaced*. A model of the relevant voice or voices can be learned from undamaged voice samples, and then – assuming that a usable script can be obtained – the damaged segments can be regenerated and, if necessary, fine-tuned. Thus, as a major advantage of this approach, even entirely

missing segments can be restored. The recordings could originate from tapes, disks or digital media. It is assumed that a text of the damaged segment is available. Good quality recordings of the target voice are necessary to adapt the Natural Language Speech Synthesiser (NLSS). The damaged segment is replaced by a synthetic segment adjusted in duration, intonation and rhythm according to the context of the document. While these technologies have myriad potential uses, the specific implementations presented in this section address important audio applications: the preservation/restoration of historical and possibly damaged audio documents in different media.

MULTIMODAL CONVERSATION APPLICATIONS

Multimodal conversation (MPAI-MMC) aims to enable human-machine conversation that emulates human-human conversation in completeness and intensity by using AI. Currently, MPAI-MMC standard includes 3 use cases: In Conversation with emotion (CWE; see the first picture in Figure 4, the human side of the dialogue includes speech, video, and possibly text, while the machine responds with a synthesized voice and an animated face.

MPAI is currently working on the responses to the Call for Technologies. One aspect concerns the CWE use case. It focuses on the description of output formats of each AIM. An important contribution concerns the list of main elements including emotion element of the Language understanding AIM which formally describes output of the module.

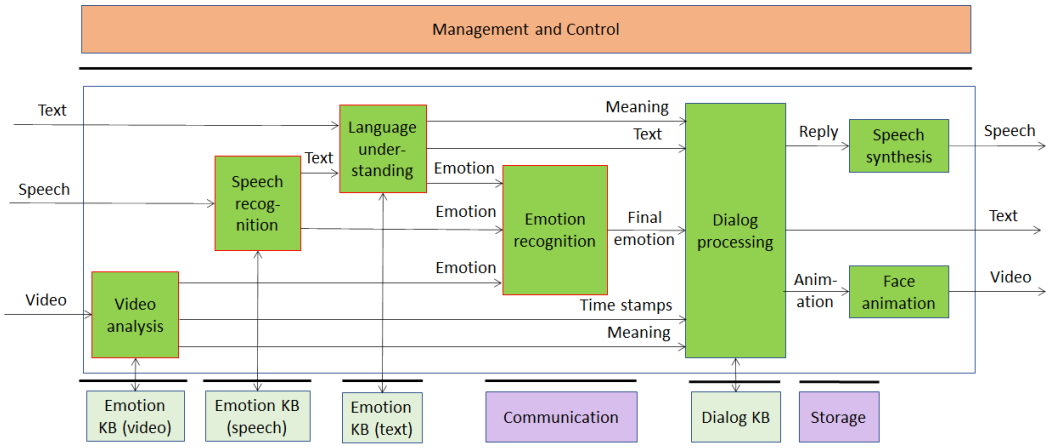

Figure 4: MPAI-MMC: Conversation With Emotions ‘MPAI (5)’

In *Multimodal question answering (MQA)*, a human requests information in natural language about a displayed object, and the machine responds with synthesised speech. MPAI is working on a formal description of output formats for each AIM in the workflow. Additionally, the Question analysis AIM produces the intention of the question of the user by several elements including topic, focus, answer types and domains of the question. The intention, together with the meaning of the question produced by the Language understanding AIM, are used by the Question Answering AIM as the input to generate a reply to the question.

In *Personalized automatic speech translation* (PST), a human-uttered sentence is translated by a machine using a synthesized voice that mimics the original speaker’s speech features. Two proposals related to the PAS use case were submitted to MPAI. The first proposal presents a formal description concerning the input and output of main AIMs. An important contribution concerns a list of speech features for the speech feature extraction AIM.

Utilizing an AI powerful speech analysis technology, in conjunction with multiple translation databases, personalized speech translation system aims to remove the language barrier

found in traditional instant messaging programs, so you can have friends all over the world and communicate with ease. This system implementation leverages cloud hosted services and handles all translation/voice morphing on the server side to increase speed and minimize local app size. This system features an intuitive interface that allows users to easily select the appropriate language for each contact in their international contacts list and automatically applies that language translation function to each chat window. Users will have access to a face-to-face conversational capability for local translations while traveling, or favourite contact chat to talk to friends they have connected to around the world.

Unlike other traditional speech translation products, which utilize computer-generated voices, this next generation proprietary technology allows users to talk to their contacts in their own voice. So now, you can actually hear what your friend would sound like if they were actually there in person and speaking your language. Also, unlike other standard translation products, this new technology implementation allows users to retain their original vocal inflection through the translation process, resulting in a more natural speech pattern.

Possible uses are: 1) Voice can be morphed into many languages; 2) Breaking down cross-language communication barriers; 3) Online dating, same interest group conversations around the globe; 4) Inviting friends from Facebook, phone or email address book or Twitter; 5) Chatting with foreign relatives; 6) Keeping in touch with foreign business associates; 7) Online gamers who don't want to use limited in-game preset phrase translators.

VIDEO CODING APPLICATIONS

To face the challenges of offering more efficient video compression solutions, research effort focused on radical changes to the classic block-based hybrid coding framework. AI approaches can play an important role in achieving this goal.

MPAI has recently carried out a literature survey on AI-based video coding 'Iacoviello (6)'. The result suggests that a performance enhancement of about 30% can be achieved. Therefore, MPAI is investigating whether it is possible to improve the performance of the Essential Video Coding (EVC) modified by enhancing/replacing existing video coding tools with AI tools keeping complexity increase to an acceptable level. Figure 5 describes the reference codec architecture. The red circles represent the data processing block candidate for enhancement/replacement with AI tools.

For each selected data processing tool three steps are needed: data extraction, coded block training and inference. The first step consists in extracting data from pre-selected video sequences to be used for training and evaluating each codec block. The second step selects an existing deep-learning algorithm that improves the performance of the coded block. To perform the third step a communication infrastructure is needed that abstracts the environment where the EVC codec is run from the specific AI Framework used for inference. For each codec block, the EVC codec communicates with the AI Framework using a socket and evaluates the hybrid EVC codec performance.

Currently, the group is working on two tools: Intra prediction and Super Resolution.

Intra prediction

We experimented enhancing the Intra predictor by means of an autoencoder neural network inspired by the Contex Encoder architecture 'Dumas et al (7)' as depicted in Figure 6. So far, we have considered only the 32x32 and 16x16 predictors, leaving 8x8 and 4x4 predictors to our future endeavours: therefore, in the following we refer to the 32x32 predictor case. The encoder receives in input a 64x64 patch representing the decoded context available at

the decoder (D0, D1, D2), whereas the bottom-right corner (P3) represents the EVC predictor as available also at the decoder. The network outputs an enhanced predictor (P3') and is trained to minimize the MSE between the enhanced intra predictor P3' and the original, uncompressed, 32x32 block O3 (Figure 6).

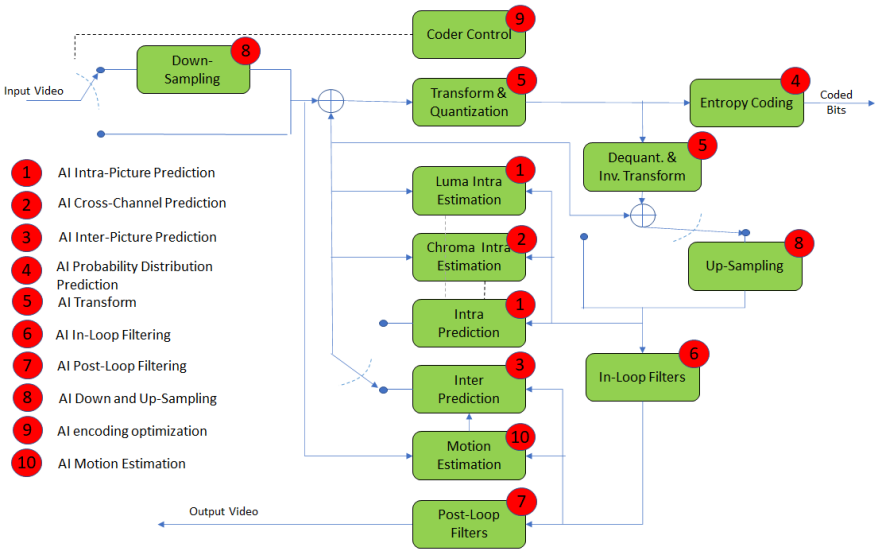

Figure 5: Traditional video coding with AI enhancement

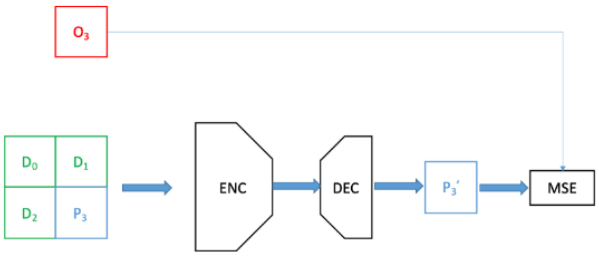

Figure 6: MPAI-EVC Convolutional Neural Network architecture 'Wang et al (8)'

Toward training our network, we have setup a dataset for training our autoencoder consisting in 1.5M patches of 32x32 predictors and 6M patches of 16x16 predictors extracted from the AROD dataset. The Original-Decoded-Predicted (ODP) data structure, has been used. These patches are arranged into 64x192 pixels images, constituted by three 64x64 pixels images, where the lower right hand side patch is the codec processed patch while the surrounding patches are the neighbourhood information (causal context).
The network trained over AROD patches is interfaced with the MPAI-EVC encoder via system sockets and the resulting AI-enhanced encoder is tested over the 1st frame of Class B sequences BasketballDrive, BQTerrace, Cactus, Kimono1, ParkScene and computed BD-rate over QPs = {22, 27, 32, 37, 42, 47}.
Table 1 shows the BD-rate gains for the case where the neural network unconditionally enhances all 5 Intra prediction modes (32x32 and 16x16 predictors only are enhanced). This experiment report gains above 1% especially at high QPs; it is to be noticed that no signalling overhead is implied in this case.

Sequence	BDRate [%]
BasketballDr	-0.4860
BQTerrace	0.0252
Cactus	0.0632
Kimono1	-0.4744
ParkScene	-0.2959
AVG	**-0.23358**

Sequence	BDRate [%]
BasketballDr	-1.5131
BQTerrace	-0.0835
Cactus	-0.2401
Kimono1	-1.0197
ParkScene	-0.6010
AVG	**-0.69148**

Sequence	BDRate [%]
BasketballDr	-2.3103
BQTerrace	-0.2734
Cactus	-0.5501
Kimono1	-1.3099
ParkScene	-0.9504
AVG	**-1.07882**

Table 1: BD-rate, unconditioned modes: QP 22-32 (left), 22-47 (centre), 32-47 (right).

Finally, we experiment selectively enabling the network for each predictor on the basis of the RD cost: this experiment provides an upper bound to the performance of our proposed method, and does not encompass the cost of signalling the network switch for each predictor. Table 2 shows BD-rate gains in excess of 5% for some sequences, prompting more investigations in this method that could encompass also 8x8 and 4x4 predictors and the signalling cost as well.

Sequence	BDRate [%]
BasketballDr	-2.2519
BQTerrace	-0.3196
Cactus	-1.4939
Kimono1	-1.6295
ParkScene	-1.5092
AVG	**-1.44082**

Sequence	BDRate [%]
BasketballDr	-5.5418
BQTerrace	-1.1324
Cactus	-2.8867
Kimono1	-3.2344
ParkScene	-2.5456
AVG	**-3.06818**

Sequence	BDRate [%]
BasketballDr	-8.0832
BQTerrace	-2.0522
Cactus	-4.3143
Kimono1	-4.0596
ParkScene	-3.7231
AVG	**-4.44648**

Table 2:BD-rate, Oracle modes: QP 22-32 (left), 22-47 (centre), 32-47 (right). Network switch signalling cost is not computed in the encoding rate.

Super Resolution

We added a super resolution step after the Post-loop filters (Figure 5, block 7) to improve the overall performance of the EVC decoding system. To achieve this, we have adopted a well-known deep learning algorithm, used in super resolution studies such as the Densely Residual Laplacian Network ‘Anwar et al (9)’. This architecture is employed as an up-sampler whenever the input sequence, into the decoding system, has been downsampled. Due to the complexity, in terms of memory and computational costs, of the used deep-learning approach we have designed an efficient training strategy. This has been solved by subdividing the input frame in crops and developing suitable training and validation sets based on the cropping strategy adopted. To achieve this, we have employed two strategies, both based on the entropy information of the input frame. This is calculated by estimating at each pixel position (i,j) the entropy of the pixel-values within a 2-dim region centred at (i,j). The first strategy uses a random crop if, and only if, its average entropy exceeds a given threshold. The second strategy selects n crops, of the same size, from the total crops available in each frame. This is based on the importance sampling technique applied on the entropy values distribution of all crops in each frame. A particular attention needs to be given to the right combination of the crop and batch sizes as in such a way a tradeoff with respect to GPU memory consumptions can be achieved. Table 3 shows the tested combination:

Input - Batch size	GPU memory usage (GiB)
48 - 16	9.7
72 - 16	18.0
96 - 10	19.7
128 - 6	20.0

Table 3: Tested combinations of crops and batch sizes.

In order to train the network, we have built three versions of the “Images 4k” Kaggle dataset with different resolutions, namely 4K-3840x2160, HD-1920x1080 and SD-960x540. Then, for each of them, the data have been encoded and decoded using three different values of Quantization Parameters (QP), i.e. 15, 30, 45 respectively, and two deblocking options, either enabled or disabled. The training procedure has been built on two different use cases, i.e. upscaling factor of 2 from SD to HD, and from HD to 4K resolution respectively.

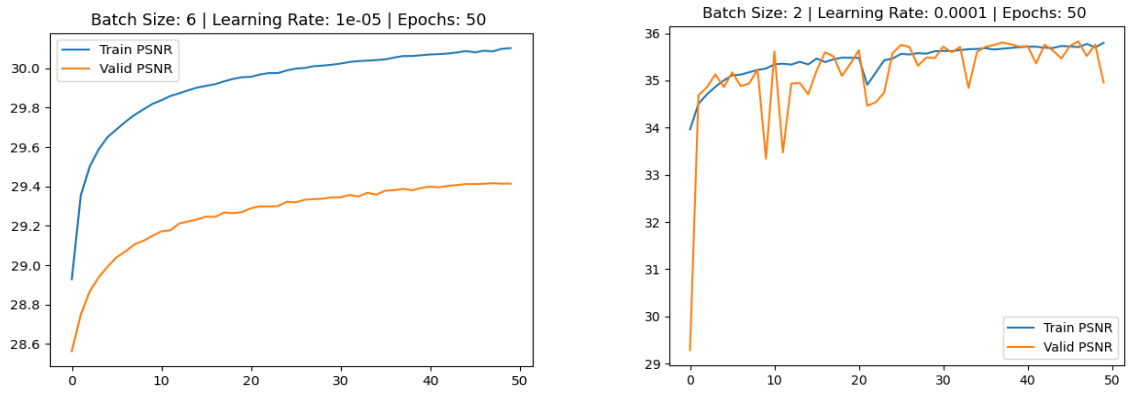

Figure 7: Training and validation results, as psnr metric, using the two cropping strategies: (left) random cropping and (right) importance sampling based cropping.

As described above, to efficiently perform the training tasks two cropping strategies have been employed. The hyperparameters and parameters used during the train phase were the followings: learning rate (lr) 10e-5, batch size 6 and 2 for the crop strategy based on importance sampling, epochs 50, the resolution of the crop input was 128x128, the crop output was 256x256, while the dataset used was the one with deblocking option activated. The Mean Square Error (MSE) metric was used as a loss function. The results using QP 15 from HD to 4K for both cropping strategies are shown in Figure 7.

SERVER-BASED PREDICTIVE MULTIPLAYER GAMING

Online Gaming is a large and growing industry with billions of users. However, two problems beset this industry: network packet loss and cheating systems. Server-based Predictive Multiplayer Gaming (MPAI-SPG) is providing a solution for both problems: minimising audio-visual and gaming discontinuities caused by high latency or packet loss during real-time online gaming sessions and designing a system that intercepts anomalous (cheating) situations. Currently, prediction is done exclusively on the ongoing data of the game and is used by the clients to achieve a smooth gaming experience that does not depend only on the arrival of packets from the game server.

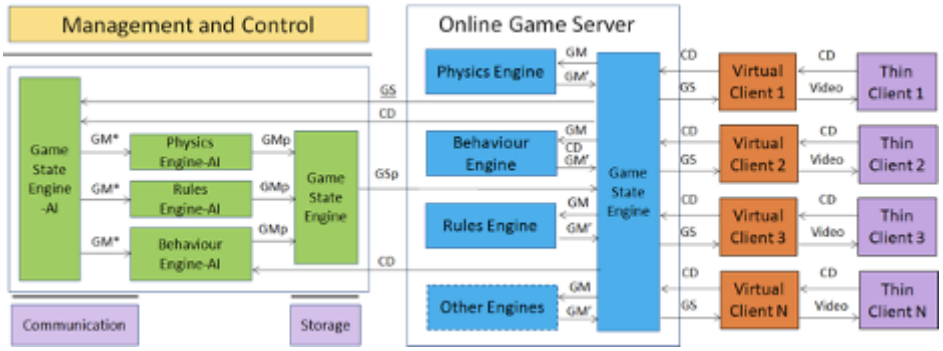

Figure 8: The MPAI-SPG reference model

With MPAI-SPG, if the information of a client is missing, the data collected from the clients involved in a particular game is fed to an AI-based system that predicts the moves of the client whose data is missing. The neural networks of MPAI-SPG are trained using all the games that have been played to arrive at the most accurate prediction of the missing parts. In the case of anti-cheating, the neural network suggests to the online game server what the current state of the system might be according to the input data and the previous game state. The online game server compares the information coming from MPAI-SPG with its own. In the event of a significant deviation, the online game server takes action against a particular player who is sending altered information. The system will appear as a plug-in to be integrated in game engines according to the prevailing philosophy available in the Game Industry. Additional information and updates can be found at 'MPAI (10)'.

OTHER MPAI STANDARD

Another MPAI challenge is to promote the efficient use of data by developing standards of filtering financial and organizational data to predict the performance of a company. It is the case of MPAI Compression and Understanding of Industrial Data (**MPAI-CUI**). The MPAI-CUI standard uses AI substantially to extract the most relevant information from company data, conduct forward-looking predictive analysis to identify the risk of bankruptcy long before it may happen, and take proper recovery actions. The interested reader may refer to 'MPAI (11)' for further information on MPAI-CUI.

Another area where AI-based analysis is becoming widespread is quantitative biology, in disciplines such as genomics, metabolomics, high-precision imaging and smart farming. The raw data sources these topics hinge upon, in particular sequencing and metabolomic data, are projected to be among the most relevant contributors to global data generation for the next few years. Integrative Genomic/Sensor Analysis (**MPAI-GSA)** will provide access to the file formats most commonly used in the field, allowing for existing and novel workflows to be transparently encapsulated and presented as a well-defined combination of AIMs performing well-defined tasks. More information on MPAI-GSA can be found in 'MPAI (12)'.

IMPLEMENTATIONS

In *Enhanced Audio Conference,* an intelligent microphone configures itself on the basis of the acoustic scene in which it has to operate, to provide the best user experience. The AIF comprises two AIMs. The first one performs Acoustic Scene Classification (ASC) while the second one can dynamically change the microphone configuration using beamforming and source localization algorithms on signals coming from a microphone array (4 microphones). On the basis of the classification of the acoustic environment performed by the fist AIM where 3 classes on scenes are considered: indoor outdoor in car, the second module can

adapt it processing ranging from signal pass-through to beamforming based on source localization.

The implementation relies of STM32 microcontrollers and ST AI software components freely available. The block diagram is shown in Figure 9. The AIF implementation includes a ST Bluecoin STEVAL-BCNKT01V1 which is equipped with a microphone array of 4 microphones and a STM32 micro controller and implements the AIM that can adaptively process the input microphone signals. The second sub-system composed by a ST Sensortile STEVAL-STLKT01V1 implements a second AIM perform Automatic Scene Classification (ASC) based on ML and the AIF management and control. Via the AIF control and management interface, the Bluecoin module changes it processing configuration adaptively from omnidirectional audio to strong beamforming based on ST algorithms of source localization. A mode detailed description of the prototype can be found in 'MPAI (13)'.

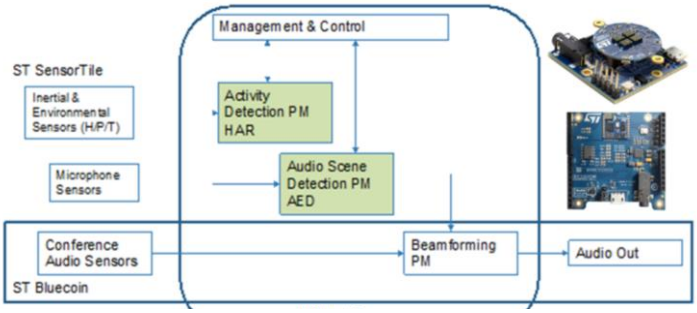

Figure 9: PM (Processing Module, AIF: Artificial Intelligence Framework, HAR: Human Activity Recognition, AED: Audio Event Detection)

We acknowledge Danilo Pau and Davide Ghezzi from ST Microelectronics for the precious support in developing the prototype.

FRAMEWORK LICENCES

Challenges in how to best assure competition for standards decisions was an underlying concern in developing MPAI "Framework License[s]" ("FWL"). The nature of AI differs from traditional licensing agreements in software and technology generally. One concept, the performance warranty, seeks to provide guidance between how the software performs given its accompanying documentation and specification. As discussed herein, AI and MPAI's standards efforts intend to provide greater flexibility given the changing nature of AI. Specifically, focusing on desired outcomes by the parties making a particular claim or attesting to same is of value to innovators and implementers alike.

In contrast with standards such as MPEG, MPAI places the business model part of a license into the process between defining requirements and calls for technologies. In other words, the MPAI approach is not a complete solution to providing timely licenses to data compression and representation standards, MPAI's FWLs facilitate at least one beneficial path forward. Significant differences in business models between adopters no longer dominate the standard discussions. Evaluation of functional and commercial requirements need not be undermined, FWLs instruct use cases and conditions, not specific cost.

Framework Licenses thus serve to overcome the uncertainties associated with FRAND (i.e., "fair, reasonable and non-discriminatory"), and even provide transparency over SEP (i.e., "standard essential patent") by establishing, to the extent possible, claims at the onset of the standardization process. In accordance with competition law and practice, MPAI replaced FRAND with FWLs developed by MPAI members with IP expertise as voluntary terms of use lacking any monetary consideration. In some cases, a range of values may provide a high and low estimate for understanding potential costs. FWL serves to provide guidance on business models for intellectual property rights ("IPRs") holders by eliminating specific

values such as percentages, royalty rates, cash values, dates, and simply proffering a cap for the cost of a given license based on comparable costs of similar standards and underlying technologies.

As MPAI develops standards and market adoption begins, FWLs will guide licensing of MPAI standard compliant technologies: FWLs will close the expectation gap between the innovators and licensors and the market adopters and implementers.

CONCLUSIONS

MPAI is a young organisation that leverages a decade-long data processing-based experience in digital media compression to attack the wider field of data coding. The approach is focused on the use of Artificial Intelligence, but is at the same time agnostic of the technology. The definition of data coding as the transformation of data from a given representation to an equivalent one more suited to an application, allows MPAI to address disparate areas of data coding standardisation with a unified approach. The definition of an AI system as a network of AI Modules executed in an AI Framework, as opposed to black boxes allows MPAI to advance the cause of explainability of AI Systems that conform to MPAI standards.

REFERENCES

1. MPAI Community (2021). Artificial Intelligence Framework. https://mpai.community/standards/mpai-aif/#Technologies/ Last accessed 2021/05/03
2. MPAI Community (2021). Artificial Intelligence Framework. https://mpai.community/standards/mpai-aif/#UCFR/ Last accessed 2021/05/03
3. Bosi et al. (2021) Sound and music computing using AI: Designing a standard. 18th Sound and Music Computing Conference
4. Pretto et al. (2019) Computing methodologies supporting the preservation of electro-acoustic music from analog magnetic tape, Computer Music Journal, 42, no. 4, 59–74.
5. MPAI Community (2021). Multimodal Conversation. https://mpai.community/standards/mpai-mmc/ Last access: April 26, 2021.
6. Iacoviello (2020). Analysis of performance of AI-based video codecs. MPAI-EVC.
7. Dumas et al. (2020). Context-Adaptive Neural Network-Based Prediction for Image Compression. IEEE Transactions on Image Processing, 29, 679–693.
8. Wang et al. (2019). Enhancing HEVC Spatial Prediction by Context-based Learning. ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - 2019-May, 4035–4039.
9. Anwar et al (2020). Densely Residual Laplacian Super-Resolution. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1–1.
10. MPAI Community (2021). Server-based Predictive Multiplayer Gaming. https://mpai.community/standards/mpai-spg/ Last access: April 26, 2021.
11. MPAI Community (2021). Compression and Understanding of Industrial Data. https://mpai.community/standards/mpai-cui/. Last access: April 22, 2021.
12. MPAI Community (2021). Integrative Genomic/Sensor Analysis. https://mpai.community/standards/mpai-gsa/ Last access: April 26, 2021.
13. MPAI Community (2021). https://mpai.community/wp-content/uploads/2020/12/N105-Validation-efforts-for-MPAI-AIF-model.docx