|  |  |
| --- | --- |
|  | Moving Picture, Audio and Data Coding by Artificial Intelligence  www.mpai.community |

|  |  |
| --- | --- |
| **Public Document** | |
|  | 2020/12/18 |
| Source | Leonardo Chiariglione |
| Title | Use Case, Requirements and candidate technologies for MPAI-CAE CfT |
| Target | MPAI Members |

# Summary

This document reports the MPAI-CAE resources collected so far

1. Use Cases and Requirements of
   1. Emotion-Enhanced Speech
   2. Audio Recording Preservation
   3. Enhanced Audioconference Experience
   4. Audio-on-the-go
2. Candidate technologies to appear in the expected MPAI-CAE Call for technologies
3. Initial requirements of the said candidate technologies.

# Introduction

Moving Picture, Audio and Data Coding by Artificial Intelligence (MPAI) is an [international association](http://mpai.community/) with the mission to develop *AI-enabled data coding standards*. Research has shown that data coding with AI-based technologies is *more efficient* than with existing technologies.

The MPAI approach to developing AI data coding standards is based on the definition of *standard interfaces* of *AI Modules (AIM)* that can be combined and *executed* in an MPAI-specified *AI-Framework* that MPAI calls MPAI-AIF.

While AIMs must expose standard interfaces to be able operate in an MPAI AI Frame­work, their performance may differ depending on the technologies used by implementors. MPAI believes that *competing* devel­opers striving to provide more performing *proprietary* but still *inter­operable* AIMs will promote *horiz­ontal markets* of *AI solutions* that tap from and further promote AI *innov­ation*.

The content of this document is

|  |  |
| --- | --- |
| Chapter 2 | introduces the The MPAI AI Framework. |
| Chapter 3 | presents the 4 MPAI-CAE Use Cases with the following structure   1. Use Case description 2. Implementation architecture 3. AI Modules and External data 4. Workflow 5. Functions and potential technologies for CfT 6. An initial analysis of the requirements associated to each identified technology. |
| Chapter 4 | presents the technologies likely to be common across MPAI-CAE and MPAI-MMC. |
| Chapter 5 | gives information on the next steps of the process that will lead to the development of the MPAI-CAE standard. |

# The MPAI AI Framework (MPAI-AIF)

Most MPAI applications considered so far can be implemented as a set of AIMs – AI/ML and even traditional data processing-based units with standard interfaces assembled in suitable topol­ogies to achieve the specific goal of an application and executed in an MPAI-defined AI Framework. MPAI is making all efforts to iden­tify processing modules that are re-usable and upgradable without necessarily changing the inside logic.

MPAI plans on completing the development of a 1st generation AI Framework called MPAI-AIF in July 2021.

The MPAI-AIF Architecture is given by *Figure 1*



*Figure 1 – The MPAI-AIF Architecture*

Where

1. *Management and Control* manages and controls the AIMs, so that they execute in the correct order and at the time when they are needed.
2. *Execution* is the environment in which combinations of AIMs operate. It receives external inputs and produces the requested outputs both of which are application specific interfacing with Management and Control and with Communication, Storage and Access.
3. *AI Modules* (AIM) are the basic processing elements receiving processing specific inputs and producing processing specific
4. *Communication* is required in several cases and can be implemented, e.g. by means of a service bus and may be used to connect with remote parts of the framework
5. *Storage* encompasses traditional storage and is used to e.g. store the inputs and outputs of the individual AIMs, data from the AIM’s state and intermediary results, shared data among AIMs.
6. *Access* represents the access to static or slowly changing data that are required by the application such as domain knowledge data, data models, etc.

# Initial requirements of MPAI-CAE candidate technologies

## Introduction

So far, MPAI has indentified the following Use Cases in the “Context-based Audio Enhancement” Application Area benefiting from MPAI standardisation:

1. Emotion enhanced synthesised voice: Expressive speech model based on the primary emotions (fear, happiness, sadness, and anger)
2. Audio recording preservation: Automatic techniques to extract information from analog audio and video tapes: automatic analysis (preprocessing step); and second step (classification), in which a classifier is used to determine the content of each image saved during pre-processing.
3. Enhanced audio experience in a conference call: Adaptive audio processing pipeline to improve conference call experience.
4. Audio-on-the-go: Adaptive audio processing pipeline to improve Sound Quality on the go without loosing contact with the acoustic surroundings.

Other Use Cases in the pipeline are:

1. Efficient 3D sound
2. Audio mastering
3. (Serious) gaming
4. Normalization of TV volume
5. Automotive
6. Voice communication
7. Audio (post-)production

## Emotion-Enhanced Speech

### Use Case description

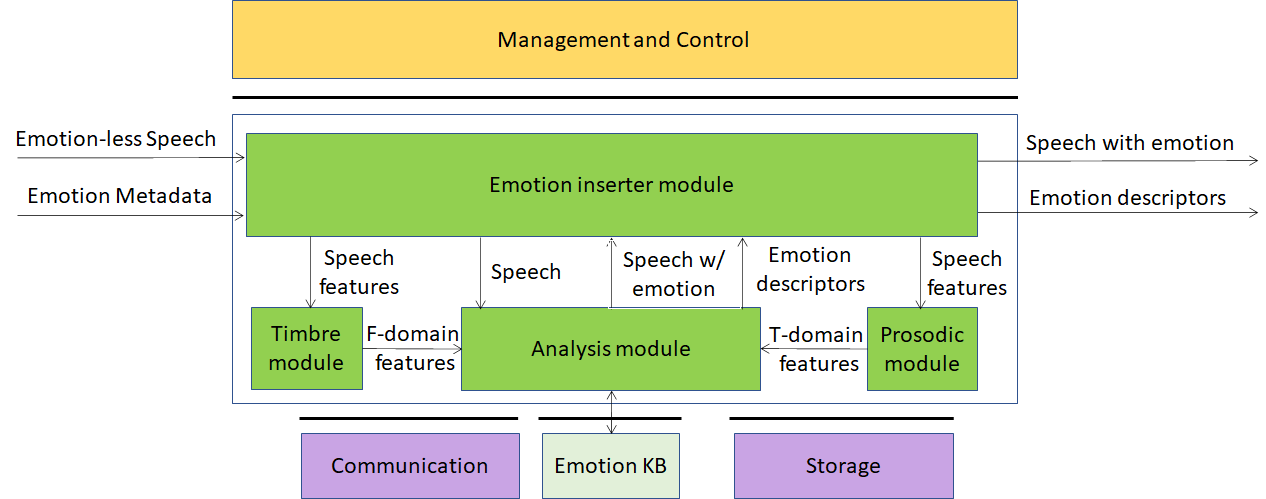
There are many cases where a speech without emotion need to be converted to a speech carrying an emotion, possibly with grades of a particular emotion. This is the case, for instance, of a human-machine dialogue where the message conveyed by the machine is more effective if it carries an emotion properly related to the emotion detected in the human speaker.

Emotion can be basic and with a universal applicability (e.g. fear, happiness, sadness, anger), or culture-specific.

### Implementation architecture

This Use Case can be implemented as in *Figure 2*.

The following AIM can be implemented either as AI or legacy modules: Analysis module. If this AIMs is implemented as a neural network, access to Emotion KB may not be needed.



*Figure 2 –* *Emotion-enhanced speech*

### AI Modules and External data

|  |  |
| --- | --- |
| **AIM** | **Function** |
| **Emotion inserter** | Inserts a particular emotional vocal timbre (fear, happiness, sadness, or anger) into a neutral (emotion-less) synthesised voice. It also changes the strength of an emotion (from neutral speech) in a gradual fashion |
| **Speech analysis** | Produces Emotion descriptors by querying the Emotion KB. Alternatively, an embedded neural network produces the Emotion descriptors. |
| **Prosodic module** | Produces Time-domain features by processing the prosodic parameters of emotions in speech (rhythm, speed of speech, intonation and intensity) |
| **Timbre module** | Produces Frequency-domain features by processing the vocal timbre-related parameters of emotions in speech (position of the formants and distribution of the spectral energy). |
| **Emotion KB** | Allows Emotion recognition to access features extracted from speech rec­ordings of different speak­ers reading/reciting the same corpus of texts, with the standard set of em­otions and without emotion, for different languages and genders. |

### Workflow

The workflow corresponding to Figure 2 is given below

1. Emotion-less Digital speech enters Emotion inserter module
2. Emotion metadata sync’ed with Digital speech enters Emotion inserter module
3. Emotion inserter module
   1. Extracts Speech features from Digital speech
   2. Sends Speech features to Timbre module and Prosodic module
   3. Sends Digital speech to Analysis module
4. Prosodic module
   1. Receives Speech features from Analysis module
   2. Computes Time-domain features
   3. Sends Time-domain features to Analysis module
5. Timbre module
   1. Receives Speech features from Analysis module
   2. Computes Frequency-domain features
   3. Sends Frequency-domain features to Analysis module
6. Analysis module
   1. Receives Time-domain features from Prosodic module
   2. Receives Frequency-domain features from Timbre module
   3. Produces Speech with emotion
   4. Produces Emotion Descriptors
   5. Sends Speech with emotion to Emotion inserter module
   6. Sends Emotion Descriptors to Emotion inserter module

### Functions and potential technologies for CfT

|  |  |
| --- | --- |
| **Function** | **Potential CfT items** |
| Digital Speech | 16-24 bit/s, 22.05-96 kHz |
| Emotion metadata | 1. Coded representation of Basic Emotion metadata 2. Coded representation of cultures 3. Coded representation of Specific Emotion metadata |
| Metadata attached to speech | Streamable file format of speech with sync’ed metadata |
| Speech features | Coded speech features |
| Emotion KB query format | Format of query to Emotion KB  Format of response from Emotion KB |
| Emotion descriptors | Coded representation of Emotion descriptors |
| Time-domain features | Coded representation of Time-domain features |
| Frequency-domain features | Coded representation of Frequency-domain features |

### Digital Speech

MPAI should not be too prescriptive. It should allow use of speech sampled in the 22.05-96 kHz with 16-24 bit/sample.

The CfT should not call for Digital speech technologies. However, it might be useful to collect comments on the choice made by MPAI.

### Emotion metadata

By Emotion metadata we mean attributes that classify emotion.

The most basic emotions are fear, happiness, sadness, anger. These can be taken as “universal” in the sense that they are common to all cultures. Emotions can have different grades.

However, in the literature other universal emotions are proposed. Therefore, the CfT should call for universal emotion metadata and their grades, their semantics and digital representation.

We need the requirements derived from the intended application: how is emotion added to em-otion-less speech.

It may be too early to ask for metadata that represent culture-dependent emotions. If we do, we should probably call for metadata that describe cultures and emotions within a culture. Even if we do not call for them, however, we should make it clear in the CfT that the basic emotion metadata are expected to be extensible, e.g., capable to represent culture-dependent emotions by adding appropriate metadata.

We should pay particular attention to this Emotion metadata technology because its use is not restricted to speech. A machine could produce images conveying a particular emotion. Conver-sation with emotion needs Emotion metadata for text and images (videos?).

### File format for speech and emotion

The format should allow the linking of emotion metadata to a particular speech segment. The speech can be contained in a file that includes the emotion metadata and the information about which segment of the speech a particular emotion metadata item refers to. The file can then be streamed. We need requirements.

### Speech features

Speech features are extracted by the Emotion inserter module using the input emotion-less speech and sent to the Prosodic and Timbre modules. The Prosodic module produces Time-domain feat­ures (see 3.2.11) and the Timbre module produces Frequency-domain features (see 3.2.12).

### Emotion KB query format

Emotion KB contains features extracted from the speech recordings of different speakers reading/ reciting the same corpus of texts with an agreed set of emotions and without emotion, for a set of languages and for different genders.

The Analysis module queries the Emotion KB providing emotions with the selected grade. The Emotion KB responds providing frequency-domain and time-domain features correlated to the emotions of the Speech.

### Time-domain features

Features to detect the arousal level of emotions: sequences of short-time prosody acoustic features (features estimated on a frame basis), e.g., short-term speech energy.

### Frequency-domain features

The following features might be considered because they have information about emotion.

1. Features related to the **pitch signal** (i.e., the glottal waveform) that depends on the tension of the vocal folds and the subglottal air pressure. Two parameters related to the pitch signal can be considered: pitch frequency and glottal air velocity. E.g., high velocity indicates a speech emotion like happiness. Low velocity is in harsher styles such as anger.
2. The shape of the **vocal tract** is modified by the emotional states. The formants (characterized by a center frequency and a bandwidth) could be a representation of the vocal tract resonances. Features related to the **number of harmonics** due to the non-linear airflow in the vocal tract. E.g., in the emotional state of anger, the fast air flow causes additional excitation signals other than the pitch. Teager Energy Operator-based (TEO) features, could be an example of measure of the harmonics and cross-harmonics in the spectrum.

An example solution of the features could be the **Mel-frequency cepstrum** (**MFC**).

### Emotion descriptors

A superset of parameters to be used by Emotion inserter to add the specific emotion to the input speech.

## Audio Recording Preservation

### Use Case description

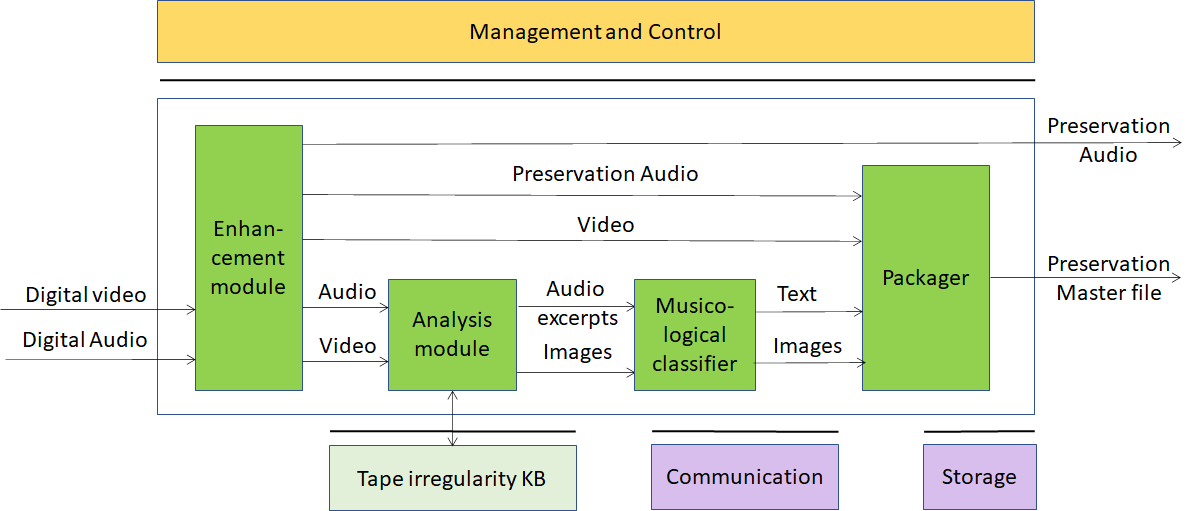
Preservation of audio assets recorded on a variety of media (vinyl, tapes, cassettes etc.) is important in an important activity for a variety of application domains, in particular cultural heritage. In the case of magnetic tapes, the carries may hold important information and the multiples splices of the tape can be annotated and display several types of irregularities.

In the use case considered here the audio is digitised and fed into the preservation system. The audio information is supplemented by the information coming from a video camera that is pointed to the head that reads the magnetic tape. The output of the restoration process is the preservation digital audio and a preservation master file that contains, next tp the preservation audio file, several other information types created by the preservation process.

### Implementation architecture

This Use Case can be implemented as in *Figure 3*.

The following AIM can be implemented either as AI or legacy modules: Analysis module. If this AIMs is implemented as a neural network, access to Tape irregularity KB may not be needed.



*Figure 3 – Tape Audio preservation*

### AI Modules and External data

|  |  |
| --- | --- |
| **AIM** | **Function** |
| **Enhancement module** | Produces Preservation audio using internal denoiser |
| **Analysis module** | Produces images and audio excerpts querying the Tape irregularity KB. Alternatively, an embedded neural network produces images and audio excerpts. |
| **Musicological classifier** | Produces images and text describing images |
| **Packager** | Produces file containing   1. Digital audio 2. Input video 3. Audio sync’ed images and text |
| **Tape irregularity KB** | Knowledge Base of data regarding visual and audio irregularities |

### Workflow

The workflow corresponding to *Figure 3* is given below

1. Digitised audio from tape (Digital audio) enters Enhancement module
2. Digital video from camera pointed to tape passing the head sync’ed with Digital Audio enters Enhancement module
3. Enhancement module
   1. Produces Preservation audio using internal denoiser
   2. Sends Preservation audio to output and to Packager
   3. Sends Preservation audio to Analysis module
   4. Receives Digital video
   5. Sends Digital video to Analysis module and to Packager
4. Analysis module
   1. Computes Audio features
   2. Extracts Images from Digital video
   3. Computes Image features
   4. Queries Knowledge base of tape irregularities
   5. Produces Audio excerpts from Preservation audio
   6. Selects Images
   7. Sends Musicological classifier
      1. Images with time information
      2. Audio excerpts with time information
5. Musicological classifier
   * 1. Selects and classifies Audio excerpts with time information
     2. Selects and classifies Images with time information
     3. Extracts Text description with time information from Audio excerpts
     4. Sends Images and Text to Packager
6. Packager
   * 1. Creates file with Preservation Audio, Video, Audio excerpts, Images and Text

### Functions and potential technologies for CfT

|  |  |
| --- | --- |
| **Function** | **Potential CfT items** |
| **Digital Audio** | 24 bits, 48-96 kHz |
| **Digital Video** | Digital video format |
| **Tape irregularity KB query** | Format of query to Tape irregularities KB  Format of response from Tape irregularities KB |
| **Audio excerpts** | Digital Audio excerpts |
| **Digital Image** | Digital Image format |
| **Text** | Character set(s) |
| **Packager** | Format of file containing   1. Preservation Audio 2. Video 3. Images 4. Text |

### Digital Audio

MPAI should not be too prescriptive. It should allow use of audio sampled in the 48-96 kHz with 16-24 bit/sample.

The CfT should not call for Digital audio technologies. However, it might be useful to collect comments on the choice made by MPAI.

### Digital Video

MPAI could identify fixed or a range of parameter values characterising acceptable Digital video formats.

1. Pixel shape: square
2. Bit depth: 8-12 bits/pixel
3. Aspect ratio: 4/3 and 16/9
4. Frame frequency 50-100 Hz
5. Scanning: progressive
6. Colorimetry
7. Colour format: RGB and YUV
8. Compression: uncompressed, which compression format?

### Digital Image

A Digital image is an uncompressed or compressed video frame with time information.

CfT should call for comments.

### Audio features

Audio features are used to describe the pre-equalization and the dynamic range compression (e.g., Dolby A, B, C and dbx), e.g., the first 13 Mel-frequency cepstral coefficients (MFCCs).

### Image features

Image features are used to describe

1. splices of
   1. leader tape to magnetic tape
   2. magnetic tape to magnetic tape
2. other irregularities such as brands on tape, ends of tape, ripples, damaged tape, markings, dirt, shadows

### Tape irregularity KB query format

Tape irregularity KB contains features extracted from images of different tape irregularities.

Analysis module queries the Irregularity KB by giving an image and obtain the type of irregularity detected. If an irregularity is found the Image is sent to Musicological classifier.

The CfT should call for digital representation of irregulaties.

### Text

Text is produced by Musicological classifier to describe an Audio excerpt.

Text should be encoded according to ISO/IEC 10646, Information technology – Universal Coded Character Set (UCS) to support most languages in use.

CfT should call for comments.

### Packager

Packager produces the Master Preservation Audio file which contains Preservation Audio, Video, Images and Text.

We should check whether ISO Based Media File Format can be a solution to the problem or a this item should be part of the CfT.

## Enhanced Audioconference Experience

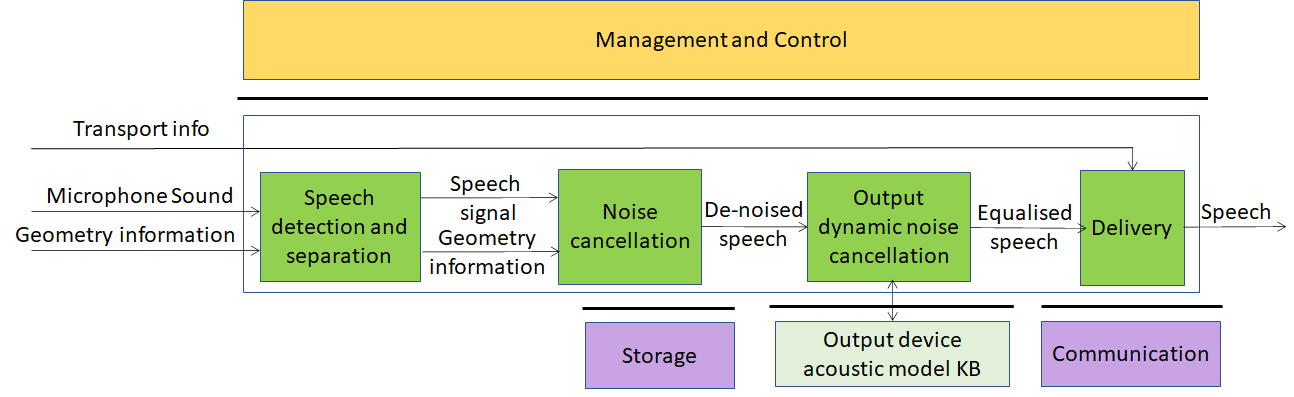
### Use Case description

Often, the user experience of a video/audio conference can be marginal. Too much background noise or undesired sounds can lead to participants not or mis-understanding what participants are saying.

By using AI-based adaptive noise-cancellation and sound enhancement, those kinds of noise can be virtually eliminated without using complex microphone systems that capture environment characteristics.

### Implementation architecture

This Use Case can be implemented as in *Figure 4*.



*Figure 4 – Enhanced Audioconference Experience*

### AI Modules and External data

|  |  |
| --- | --- |
| **AIM** | **Function** |
| **Speech detection and separation** | Separates relevant speech vs non-speech signals |
| **Noise cancellation** | Removes noise in Speech signal |
| **Output dynamic noise cancellation** | Reduces noise level based on Output Device Acoustic Model |
| **Delivery** | Wraps De-noised voice signal for Transport |
| **Output Device Acoustic Model KB** | Contains calibration test results for all output devices of a given manufacturer identified by their ID |

### Workflow

1. Digital sound from available microphone enters Speech detection and separation
2. Microphone Geometry information enters Speech detection and separation
3. Speech detection and separation
   1. Queries Microphone Physical characteristics KB
   2. Separates relevant Speech vs non-Speech signals
   3. Sends separated Speech signal to Noise cancellation
   4. Sends Geometry information information represented in standard metadata to Noise cancellation
4. Noise cancellation
   1. Removes noise in Speech signal
   2. Sends De-noised Speech signal to Output dynamic noise cancellation
5. Output dynamic noise cancellation
   1. Queries Output Device Acoustic Model KB
   2. Reduces noise level in De-noised Speech signal
   3. Sends equalised De-noised Speech signal to Delivery
6. Transport mechanism information enters Delivery
7. Transport info enters Delivery
8. Delivery
   1. Wraps equalised De-noised Speech for transport

### Functions and potential technologies for CfT

|  |  |
| --- | --- |
| **Function** | **Potential CfT items** |
| Digital Speech | 16-24 bit/sample, 22.05-96 kHz |
| Microphone Geometry information | Format of Geometry information |
| Output device acoustic model metadata KB query | Format of query to Output device acoustic model metadata KB  Format of response from Output device acoustic model metadata KB |
| Delivery | Transport mechanism information |

### Digital Speech

MPAI should not be overly prescriptive. It should allow use of speech sampled in the 22.05-96 kHz with 16-24 bit/sample.

The CfT should not call for Digital speech technologies. However, it might be useful to collect comments on the choice made by MPAI.

### Microphone geometry information

Formats to represent microphone geometry information are: MPEG-H 3D Audio and platform specific (Android, Windows, Linux) JSON Descriptors API.

### Output device acoustic model metadata KB query

Output dynamic noise cancellation queries the Output device acoustic model KB to obtain infor­mation about output device characteristics.

We need requirements for what is queried and for what type of information is expected from the Output device acoustic model KB to write the corresponding text in the CfT.

### Delivery

The CfT should call for a standard and extensible way to signal which transport mechanism is intended to be used.

## Audio-on-the-go

### Use case description

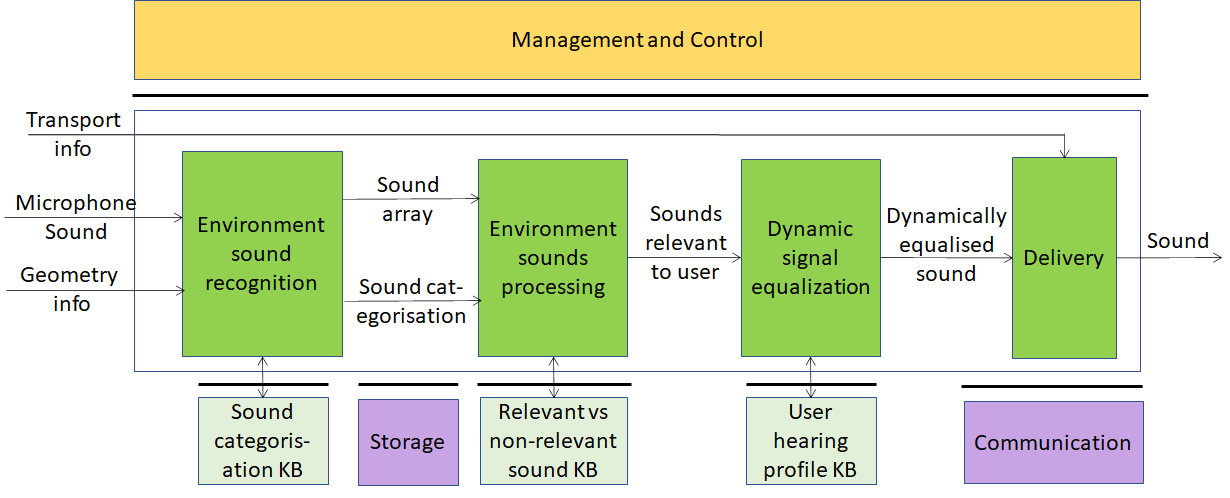
While biking in the middle of city traffic, AI can process the signals from the environment captured by the microphones available in many earphones and earbuds (for active noise cancellation), adapt the sound rendition to the acoustic environment, provide an enhanced audio experience (e.g., performing dynamic signal equalization), improve battery life and selectively recognize and allow relevant environment sounds (i.e., the horn of a car).

The user enjoys a satisfactory listening experience without losing contact with the acoustic surroundings.

### Implementation architecture

This Use Case can be implemented as in *Figure 5*.

The following AIMs can be implemented either as AI or legacy modules: Environment sound recognition and Environment sound processing. If any of these AIMs are implemented as a neural network, access to the corresponding KB may not be needed.



*Figure 5 – Audio-on-the-go*

### AI Modules and External data

|  |  |
| --- | --- |
| **AIM** | **Function** |
| **Environment Sounds Recognition** | Recognises and categorises surrounding environment sounds |
| **Environment Sound Processing** | Determines which sounds are relevant for the user vs sounds which are not |
| **Dynamic Signal Equalization** | Dynamically equalises the sound to produce the best possible quality output |
| **Delivery** | Wraps equalised sound for Transport |
| **Sound categorisation KB** | Contains audio features of the sounds in the KB |
| **Relevant vs non-relevant sound KB** | Contains audio features of relevant sounds |
| **User hearing profiles KB** | Contains user hearing profiles that a specific user can access at a given time. |

### Workflow

The workflow corresponding to Figure 5 is given below

1. Microphone sound enters Environment Sound Recognition
2. Environment Sound Recognition
   1. Produces array of sounds and related categorisation
   2. Queries Sound categorisation KB
   3. Sends array of sounds and related categorisation to Environment Sound Processing
3. Environment Sound Processing
   1. Produces Sounds relevant for the user
   2. Queries Relevant vs non-relevant sound KB
   3. Sends Sounds relevant to user to Dynamic Signal Equalisation
4. Dynamic Signal Equalisation
   1. Queries User hearing profile KB
   2. Produces Dynamically equalised sound
   3. Sends Dynamically equalised sound to Delivery
5. Transport info enters Delivery
6. Delivery
   1. Wraps Dynamically equalised sound for transport

### Functions and potential technologies for CfT

|  |  |
| --- | --- |
| **Function** | **Potential CfT items** |
| Digital Audio Sound | 24 bits, 48-96 kHz |
| Microphone Geometry information | Format of Geometry information |
| Sound array | Format of sound array |
| Sounds categorisation | Sound categories |
| Sound categorisation KB | Input query format of Sound categorisation KB  Input query format of Sound categorisation KB |
| Relevant vs non-relevant sound KB | Input query format of Relevant vs non-relevant sound KB  Input query format of Relevant vs non-relevant sound KB |
| User Hearing Profiles KB | Input query format of User Hearing Profiles KB  Output query format of User Hearing Profiles KB |
| Delivery | Transport mechanism information |

### Digital Audio Sound

MPAI should not be too prescriptive. It should allow use of audio sampled in the 48-96 kHz with 16-24 bit/sample.

The CfT should not call for Digital audio technologies. However, it might be useful to collect comments on the choice made by MPAI.

### Microphone geometry information

Exisitng formats to represent microphone geometry information are: MPEG-H 3D Audio and platform specific JSON Descriptors API for Android, Windows, Linux.

### Sound array

Format used by Environment sound recognition to package environment sounds.

### Sounds categorisation

The CfT should request a standard and extensible classification of all types of sound.

### Sound categorisation KB

Sound categorisation KB contains audio features of the sounds in the KB.

Environment sound recognition queries Sound categorisation KB giving sound as input. Sound categorisation KB reponds by giving the category of the sound.

### Relevant vs non-relevant sound KB

Relevant vs non-relevant sound KB contains audio features of the relevant sounds.

Environment sound processing queries Relevant vs non-relevant sound KB giving a sound as input. Relevant vs non-relevant sound KB responds by giving the relevant sound.

### User Hearing Profiles KB

User Hearing Profiles KB contains the user hearing profile for the specific user (identified via either a UUID or a third-party identity provider such as O-AUTH).

Dynamic signal equalisation queries User hearing profile KB giving the profiler ID as input. User hearing profile responds with the specific user hearing profile. The User hearing profile contains the hearing attenuation for a defined number of frequency spectrums. There are currently at least 2 SDKs on the matter: MIMI SDK, NURA SDK (both proprietary).

### Delivery

The CfT should call for a standard and extensible way to signal which transport mechanism is intended to be used.

# Potential common technologies

The following acronyms have been introduced

|  |  |  |
| --- | --- | --- |
| **Acronym** | **App. Area** | **Use Case** |
| EES | MPAI-CAE | Emotion-Enhanced Speech |
| ARP | MPAI-CAE | Audio Recording Preservation |
| EAE | MPAI-CAE | Enhanced Audioconference Experience |
| AOG | MPAI-CAE | Audio-on-the-go |
| CWE | MPAI-MMC | Conversation with emotion |
| MQA | MPAI-MMC | Multimodal Question Answering |
| PST | MPAI-MMC | Personalized Automatic Speech Translation |

The following technologies are potentially applicable to different Use Cases.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Function** | **EES** | **ARP** | **EAE** | **AOG** | **CWE** | **MQA** | **PST** |
| Digital speech | X |  | X |  |  |  |  |
| Digital Audio |  | X |  | X |  |  |  |
| Digital image |  | X |  |  | X | X |  |
| Emotion metadata | X |  |  |  | X | X | X |
| Speech features | X |  |  |  | X | X | X |
| Text |  | X |  |  | X | X | X |
| Image features |  |  |  |  |  |  |  |
| Image descriptors |  | X |  |  | X | X |  |

# Conclusions

The document in its current form is work in progress. MPAI intends to add more details to the Use Cases add more Use Cases to enable MPAI before issuing a Call for Technologies.

When the document will be considered sufficiently mature, MPAI will issue a Call for Technol-ogies requesting MPAI members and the industry members to submit proposals for:

1. Data formats suitable as inputs and outputs of the identified AIMs
2. Additions or removal of input/output signals to the identified AIMs with identification of data formats required by the new input/output signals
3. Possible alternative partitioning of the AIMs implementing the example cases providing
   1. Arguments in support of the proposed partitioning
   2. Detailed specifications of the inputs and outputs of the proposed AIMs
4. New Use Cases fully described as in the final version of this document.

Respondents will be asked to state in their submissions their intention to adhere to the Framework Licence developed for MPAI-MMC when licencing their technologies if included in the MPAI-MMC standard. Please note that “a Framework Licence is the set of conditions of use of a licence without the values, e.g., currency, percent, dates etc.”. The Framework Licence will give the MPAI-MMC standard a clear IPR licensing framework.

The MPAI-MMC Framework Licence will be developed, as for all other MPAI Framework Licences, in compliance with the generally accepted principles of competition law.