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

|  |  |
| --- | --- |
| **Public Document** | |
| N34 | 2020/10/21 |
| Source | MPAI-1 |
| Title | MPAI Application Note #1 Rev. 1 - MPAI-CAE - Context-based Audio Enhancement |
| Target | MPAI Members |

### MPAI Application Note #1 Rev. 1

### Context-based Audio Enhancement (MPAI-CAE)

**Proponents**: Michelangelo Guarise, Andrea Basso (VOLUMIO)

**Description**: The overall user experience quality is highly dependent on the context in which audio is used, e.g.

1. Entertainment audio can be consumed in the home, in the car, on public transport, on-the-go (e.g. while doing sports, running, biking) etc.
2. Voice communications: can take place office, car, home, on-the-go etc.
3. Audio and video conferencing can be done in the office, in the car, at home, on-the-go etc.
4. (Serious) gaming can be done in the office, at home, on-the-go etc.
5. Audio (post-)production is typically done in the studio
6. Audio restoration is typically done in the studio

By using context information to act on the content using AI, it is possible substantially to improve the user experience.

**Comments**: Currently, there are solutions that adapt the conditions in which the user experiences content or service for some of the contexts mentioned above. However, they tend to be vertical in nature, making it dif­ficult to re-use possibly valuable AI-based components of the solutions for differ­ent applications.

MPAI-CAE aims to create a horizontal market of re-usable and possibly context-depending components that expose standard interfaces. The market would become more receptive to innov­ation hence more compet­itive. Industry and consumers alike will benefit from the MPAI-CAE stan­dard.

**Examples**

The following examples describe how MPAI-CAE can make the difference.

1. *Enhanced audio experience in a conference call*

Often, the user experience of a video/audio conference can be marginal. Too much background noise or undesired sounds can lead to participants not understanding what participants are saying. By using AI-based adaptive noise-cancellation and sound enhancement, MPAI-CAE can virtually eliminate those kinds of noise without using complex microphone systems to capture environment characteristics.

1. *Pleasant and safe music listening while biking*

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.

1. *Emotion enhanced synthesized voice*

Speech synthesis is constantly improving and finding several applications that are part of our daily life (e.g. intelligent assistants). In addition to improving the ‘natural sounding’ of the voice, MPAI-CAE can implement expressive models of primary emotions such as fear, happiness, sad­ness, and anger.

1. *Efficient 3D sound*

MPAI-CAE can reduce the number of channels (i.e. MPEG-H 3D Audio can support up to 64 loudspeaker channels and 128 codec core channels) in an automatic (unsupervised) way, e.g. by mapping a 9.1 to a 5.1 or stereo (radio broadcasting or DVD), maintaining the musical touch of the composer.

1. *Speech/audio restoration*

Audio restoration is often a time-consuming process that requires skilled audio engineers with specific experience in music and recording techniques to go over manually old audio tapes. MPAI-CAE can automatically remove anomalies from recordings through broadband denoising, declicking and decrackling, as well as removing buzzes and hums and performing spectrographic 'retouching' for removal of discrete unwanted sounds.

1. *Normalization of volume across channels/streams*

Eighty-five years after TV has been first introduced as a public service, TV viewers are still strug­gling to adapt to their needs the different average audio levels from different broadcasters and, within a program, to the different audio levels of the different scenes.

MPAI-CAE can learn from user’s reactions via remote control, e.g. to a loud spot, and control the sound level accordingly.

1. *Automotive*

Audio systems in cars have steadily improved in quality over the years and continue to be integrated into more critical applications. Toda, a buyer takes it for granted that a car has a good automotive sound system. In addition, in a car there is usually at least one and sometimes two microphones to handle the voice-response system and the hands-free cell-phone capability. If the vehicle uses any noise cancellation, several other microphones are involved. MPAI-CAE can be used to improve the user experience and enable the full quality of current audio systems by reduc­ing the effects of the noisy automotive environment on the signals.

1. *Audio mastering*

Audio mastering is still considered as an ‘art’ and the prerogative of pro audio engineers. Normal users can upload an example track of their liking (possibly obtained from similar musical content) and MPAI-CAE analyzes it, extracts key features and generate a master track that ‘sounds like’ the example track starting from the non-mastered track. It is also possible to specify the desired style without an example and the original track will be adjusted accordingly.

**Requirements:**

The following is an initial set of MPAI-CAE functional requirements to be further developed in the next few weeks. When the full set of requirements will be developed, the MPAI General Assembly will decide whether an MPAI-CAE standard should be developed.

1. The standard shall specify the following natural input signals
   1. Microphone signals
   2. Inertial measurement signals (Acceleration, Gyroscope, Compass, ...)
   3. Vibration signals
   4. Environmental signals (Proximity, temperature, pressure, light, ...)
   5. Environment properties (geometry, reverberation, reflectivity, ...)
2. The standard shall specify
   1. User settings (equalization, signal compression/expansion, volume, ...)
   2. User profile (auditory profile, hearing aids, ...)
3. The standard shall support the retrieval of pre-computed environment models (audio scene, home automation scene, ...)
4. The standard shall reference the user authentication standards/methods required by the specific MPAI-CAE context
5. The standard shall specify means to authenticate the components and pipelines of an MPAI-CAE instance
6. The standard shall reference the methods used to encrypt the streams processed by MPAI-CAE and service-related metadata
7. The standard shall specify the adaptation layer of MPAI-CAE streams to delivery protocols of common use (e.g. Bluetooth, Chromecast, DLNA, ...)

**Object of standard**: Currently, three areas of standardization are identified:

1. **Context type interfaces**: a first set of input and output signals, with corresponding syntax and semantics, for audio usage contexts considered of sufficient interest (e.g. audiocon­ferencing and audio consumption on-the-go). They have the following features
   1. Input and out signals are context specific, but with a significant degree of commonality across contexts
   2. The operation of the framework is implementation-dependent offering implementors the way to produce the set of output signals that best fit the usage context
2. **Processing component interfaces**: with the following features
   1. Interfaces of a set of updatable and extensible processing modules (both traditional and AI-based)
   2. Possibility to create processing pipelines and the associated control (including the needed side information) required to manage them
   3. The processing pipeline may be a combination of local and in-cloud processing
3. **Delivery protocol interfaces**
   1. Interfaces of the processed audio signal to a variety of delivery protocols

**Benefits:** MPAI-CAE will bring benefits positively affecting

1. Technology providers need not develop full applications to put to good use their technol­ogies. They can concentrate on improving the AI technologies that enhance the user exper­ience. Further, their technologies can find a much broader use in application domains beyond those they are accustomed to deal with.
2. Equipment manufacturers and application vendors can tap from the set of technologies made available according to the MPAI-CAE standard from different competing sources, integrate them and satisfy their specific needs
3. Service providers can deliver complex optimizations and thus superior user experience with minimal time to market as the MPAI-CAE framework enables easy combination of 3rd party components from both a technical and licensing perspective. Their services can deliver a high quality, consistent user audio experience with minimal dependency on the source by selecting the optimal delivery method
4. End users enjoy a competitive market that provides constantly improved user exper­iences and controlled cost of AI-based audio endpoints.

**Bottlenecks:** the full potential of AI in MPAI-CAE would be unleashed by a market of AI-friendly processing units and introducing the vast amount of AI technologies into products and services.

**Social aspects**: MPAI-CAE would free users from the dependency on the context in which they operate; make the content experience more personal; make the collective service experience less dependent on events affecting the individual participant and raise the level of past content to today’s expectations.

**Success criteria**: MPAI-CAE should create a competitive market of AI-based components expos­ing standard interfaces, processing units available to manufacturers, a variety of end user devices and trigger the implicit need felt by a user to have the best experience whatever the context.