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

**Public document**

|  |  |
| --- | --- |
| **N878** | 2022/09/30 |
| **Source** | Requirements (CAV) |
| **Title** | MPAI-CAV Use Cases and Functional Requirements WD0.14 |
| **Target** | MPAI Community |

DISCLAIMER

This Use Cases and Functional Requirements document is a work in progress.

It defines the scope of the future Connected Autonomous Vehicle (CAV) standard identifying 4 subsystems and, for each subsystem, the workflow – AI Workflow (AIW) – whose function, interfaces, and I/O data formats will eventually be standardised and the basic components – AI Modules (AIM) – whose function, interfaces and I/O data formats will eventually be standardised.

The process envisages issuing a Call for Technologies open to all interested parties based on which the standard will be developed.

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

**MPAI Use Cases and Functional Requirements**

**Connected Autonomous Vehicles**

**MPAI-CAV**

|  |
| --- |
| **WD0.13.1** |

**WARNING**

Use of the technologies described in this Technical Specification may infringe patents, copyrights or intellectual property rights of MPAI Members or non-members.

MPAI and its Members accept no responsibility whatsoever for damages or liability, direct or consequential, which may result from use of this Technical Specification.

Readers are invited to review Annex 2 – Notices and Disclaimers.

© Copyright MPAI 2021-22. All rights reserved

**Connected Autonomous Vehicles**

**V1 (under development)**

[1 Introduction 4](#_Toc113452015)

[2 Scope of the MPAI-CAV Use Cases 5](#_Toc113452016)

[3 Terms and definitions 7](#_Toc113452017)

[4 References 10](#_Toc113452018)

[4.1 Normative References 10](#_Toc113452019)

[4.2 Informative References 11](#_Toc113452020)

[5 Subsystems 13](#_Toc113452021)

[5.1 Human-CAV Interaction (HCI) 15](#_Toc113452022)

[5.1.1 Use Case description 15](#_Toc113452023)

[5.1.2 Reference architecture 15](#_Toc113452024)

[5.1.3 Input and output data 16](#_Toc113452025)

[5.1.4 AI Modules 17](#_Toc113452026)

[5.2 Environment Sensing Subsystem (ESS) 18](#_Toc113452027)

[5.2.1 Use Case description 18](#_Toc113452028)

[5.2.2 Reference architecture 18](#_Toc113452029)

[5.2.3 Input and output data 19](#_Toc113452030)

[5.2.4 AI Modules 20](#_Toc113452031)

[5.3 CAV-to-Everything (V2X) 21](#_Toc113452032)

[5.3.1 Use Case description 21](#_Toc113452033)

[5.3.2 Input and output data 21](#_Toc113452034)

[5.4 Autonomous Motion Subsystem (AMS) 23](#_Toc113452035)

[5.4.1 Use Case description 23](#_Toc113452036)

[5.4.2 Reference architecture 23](#_Toc113452037)

[5.4.3 Input and output data 24](#_Toc113452038)

[5.4.4 AI Modules 24](#_Toc113452039)

[5.5 Motion Actuation Subsystem (MAS) 25](#_Toc113452040)

[5.5.1 Use Case description 25](#_Toc113452041)

[5.5.2 Reference architecture 25](#_Toc113452042)

[5.5.3 Input and output data 26](#_Toc113452043)

[5.5.4 AI Modules 26](#_Toc113452044)

[6 Functional Requirements 26](#_Toc113452045)

[6.1 Human-CAV Interaction 27](#_Toc113452046)

[6.1.1 I/O Data summary 27](#_Toc113452047)

[6.1.2 Digital representation of analogue signals 28](#_Toc113452048)

[6.1.3 Natively digital data formats 29](#_Toc113452049)

[6.1.4 Descriptors 33](#_Toc113452050)

[6.1.5 Interpretations 37](#_Toc113452051)

[6.2 Environment Sensing Subsystem 38](#_Toc113452052)

[6.2.1 I/O Data summary 38](#_Toc113452053)

[6.2.2 Sensor data 39](#_Toc113452054)

[6.2.3 Processed data 42](#_Toc113452055)

[6.2.4 Scene Descriptions 43](#_Toc113452056)

[6.2.5 Road Topology 45](#_Toc113452057)

[6.2.6 Basic World Representation 45](#_Toc113452058)

[6.2.7 Spatial Attitude 47](#_Toc113452059)

[6.3 Autonomous Motion Subsystem 47](#_Toc113452060)

[6.3.1 Summary of Autonomous Motion Subsystem data 47](#_Toc113452061)

[6.3.2 Basic World Representation 47](#_Toc113452062)

[6.3.3 CAV Identifier 47](#_Toc113452063)

[6.3.4 Command/Response 48](#_Toc113452064)

[6.3.5 Events 48](#_Toc113452065)

[6.3.6 Full World Representation 48](#_Toc113452066)

[6.3.7 FWR requirements 48](#_Toc113452067)

[6.3.8 Goal 49](#_Toc113452068)

[6.3.9 Offline map 49](#_Toc113452069)

[6.3.10 Path 49](#_Toc113452070)

[6.3.11 Pose 49](#_Toc113452071)

[6.3.12 Route 49](#_Toc113452072)

[6.3.13 State 49](#_Toc113452073)

[6.3.14 Traffic rules 50](#_Toc113452074)

[6.3.15 Traffic Signals 50](#_Toc113452075)

[6.3.16 Trajectory 51](#_Toc113452076)

[6.4 Motion Actuation Subsystem 51](#_Toc113452077)

[6.4.1 Summary of Motion Actuation Subsystem data 51](#_Toc113452078)

[6.4.2 Accelerometer data 51](#_Toc113452079)

[6.4.3 Brakes Command 51](#_Toc113452080)

[6.4.4 Brakes Feedback 51](#_Toc113452081)

[6.4.5 Command from AMS 52](#_Toc113452082)

[6.4.6 Feedback to AMS 52](#_Toc113452083)

[6.4.7 Motion Data 52](#_Toc113452084)

[6.4.8 Odometer Data 52](#_Toc113452085)

[6.4.9 Other Environment Data 52](#_Toc113452086)

[6.4.10 Road Wheel Direction Command 52](#_Toc113452087)

[6.4.11 Road Wheel Direction Feedback 52](#_Toc113452088)

[6.4.12 Road Wheel Motor Command 52](#_Toc113452089)

[6.4.13 Road Wheel Motor Feedback 53](#_Toc113452090)

[6.4.14 Speedometer 53](#_Toc113452091)

[7 Data privacy 53](#_Toc113452092)

[7.1 Human-CAV Interaction (HCI) 53](#_Toc113452093)

[7.2 Environment Sensing Subsystem (ESS) 53](#_Toc113452094)

[7.3 CAV to Everything (V2X) 53](#_Toc113452095)

[7.4 Autonomous motion subsystem (AMS) 53](#_Toc113452096)

[7.5 Motion Actuation Subsystem (MAS) 54](#_Toc113452097)

[Annex A – General MPAI Terminology (Normative) 55](#_Toc113452098)

[Annex B - Notices and Disclaimers Concerning MPAI Standards (Informative) 58](#_Toc113452099)

[Annex C – The Governance of the MPAI Ecosystem (Informative) 60](#_Toc113452100)

[Annex D – Datasets for CAV research 62](#_Toc113452101)

[Annex E – ETSI Technical Report 63](#_Toc113452102)

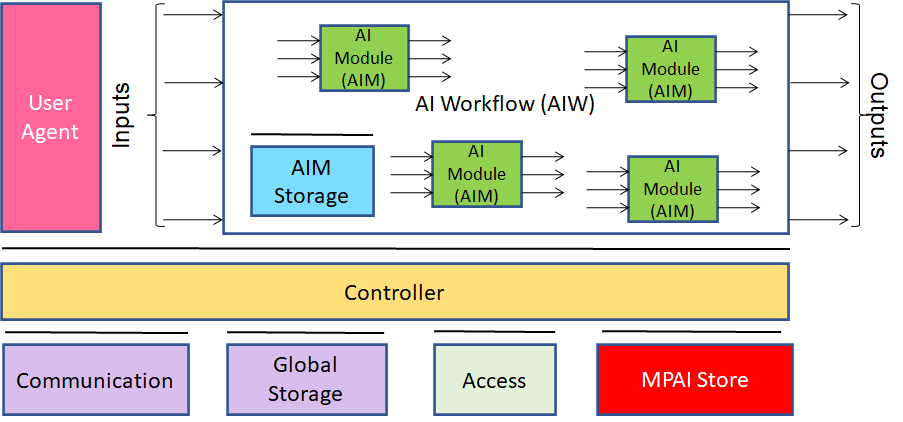
[Annex F – Some CAV Communication Technologies 65](#_Toc113452103)

# Introduction

Moving Picture, Audio and Data Coding by Artificial Intelligence (MPAI) is an [international Standards Developing Organisation](http://mpai.community/) with the mission to develop *AI-enabled data coding standards*. Research has shown that data coding with AI-based technologies is generally *more efficient* than with existing technologies. Compression and feature-based description are notable examples of coding. MPAI Application Standards enable the development of AI-based products, applications and services.

In the following, Terms beginning with a capital letter are defined in *Table 4* if they are specific to this Standard and in *Table 16* if they are common to all MPAI Standards.

MPAI Application Stan­dards Implementations operate in the AI Framework (AIF) specified by the MPAI-AIF Standard (MPAI-AIF). *Figure 1* depicts the MPAI-AIF Reference Model. This Introduction only describes the basic processing elements called AI Modules (AIM) which make up an AI Workflow (AIW) executed in an AI Framework (AIF).



*Figure 1 – The AI Framework (AIF) Reference Model and its Components*

MPAI Application Standards normatively specify:

1. *For the AIMs*: the Function, and the Semantics and Formats of the input and output data but not the internal architecture, which may be based on AI or data processing, and implemented in software, hardware or hybrid software and hardware technologies.
2. *For the AIWs*: the Function, the Semantics and Formats of the input and output data, and the Connections between and among the AIMs.

MPAI defines Interoperability as the ability to replace an AIW or an AIM Implementation with a functionally equivalent Implementation. MPAI also defines 3 Interoperability Levels of an AIW that executes an AIW. The AIW may have 3 Levels:

*Level 1 –* Implementer-specific and satisfying the MPAI-AIF Standard.

*Level 2 –* Specified by an MPAI Application Standard.

*Level 3 –* Specified by an MPAI Application Standard and certified by a Performance Assessor.

MPAI is the root of trust of the MPAI Ecosystem [1] offering Users access to the promised benefits of AI with a guarantee of increased transparency, trust and reliability as the Interoperability Level of an Implementation moves from 1 to 3. Additional information is provided by Annex 3.

# Scope of the MPAI-CAV Use Cases

There are several reasons why standards for the IT part of Connected Autonomous Vehicles (CAV) components should be developed:

1. the different nature of the interacting technologies making up a CAV.
2. the sheer size of the future CAV market [9].
3. the need for users and regulators alike to be assured of CAV safety, Reliability and Explainability.

At this point in time, a traditional approach to standardisation might consider CAV standards premature and some affected industries may not even be ready yet to consider them. CAVs, however, at best belong to an industry still being formed, that is expected to target the production of economic affordable units in the hundreds of millions p.a., with components to be produced by disparate sources. A competitive market of standard components can reduce costs and make CAV confirm their promise to have a major positive impact on environment and society.

A CAV Reference Model (RM) identifying components and their interfaces is required to accelerate the definition of standard components. Progression from research to standardisation can unfold as a series of proposals from research suggesting components and interfaces to standardisation, and standardisation either requesting more results, or refining the results, or adopting the proposals. Eventually, industry will receive a set of specifications for standard component functions and interfaces to be implemented as best available technology allows. Implementation in products will rely, as a minimum, on the know-how of those who have driven the development of the specific­ations.

*Connected Autonomous Vehicles* (MPAI-CAV) is an MPAI standard project seeking to define identified CAV standard components and their interfaces. MPAI-CAV comprises 4 Subsystems corresponding to Use Cases for each of which a Reference Model (RM) is defined. Each RM includes an AI Workflow (AIW) with a set of interconnected AI Modules (AIM).

*Table 1* identifies the names and the acronyms and characterises the 4 Subsystems.

*Table 1 – The 5 MPAI-CAV Subsystems*

|  |  |  |
| --- | --- | --- |
| **Subsystem name** | **Acr.** | **Function** |
| *Human-CAV Interaction* | HCI | Handles human-CAV interactions. |
| *Environment Sensing Subsystem* | ESS | Acquires and processes information from the Environment via a variety of sensors. |
| *Autonomous Motion Subsystem* | AMS | Issues commands to drive the CAV to the intended destination. |
| *Motion Actuation Subsystem* | MAS | Provides Environment information and receives/actuates motion commands in the physical world. |

Each of the 4 subsystems is defined for implementation as an instance of the MPAI-defined AI Framework (AIF) [2].

The Subsystems Reference Models identify and describe the requirements of the data types received or generated by the AIMs in each Subsystem. The Reference Models allows researchers to select data, define testing setups, propose update of interfaces, conduct contests, consider the influence of external components, and subdivide workload in a way that allows unambiguous comparison of results.

Unlike those of previously published papers (e.g., [10]), the Reference Models of this document have the following features:

1. They adopt a holistic approach that includes all IT components of a CAV.
2. Are based on AIF-AIW-AIM [2] as the unifying model to determine the Functions and the Data Formats of all CAV components.
3. Rely on AI Modules (AIM) having Functions and Data Formats that are being or already specified in other MPAI standards.
4. Focus on the Data Formats between AIMs rather than focus on the AIMs themselves because their internals are not part of a standard but left to proprietary implementations.
5. Envision a process where research is seamlessly integrated with a subsequent standardisation process.

The purpose of this document is:

1. To collect and describe the 4 identified Subsystems.
2. To identify the functions, and the input and output data of the AIWs that implement the Subsystems.
3. To identify the Topology of the AIMs making up the AIWs.
4. To identify the Functions, and the input and output Data Formats of the AIMs required to realise the AIWs.

Chapter 0 provides the functional requirements to be satisfied by the data formats identified in points 4. above and the connections identified in point 3 above.

# Terms and definitions

*Table 2* defines the terms used in this document. Terms are organised by the CAV Subsystems identified in *Figure 3*. The general MPAI Terms are defined in *Table 16*.

*Table 2 – Definition of Terms used in this document organised by CAV Subsystems*

|  |  |  |
| --- | --- | --- |
| Legend | AMS | Autonomous Motion Subsystem |
|  | CAV | Connected Autonomous Vehicle |
|  | ESS | Environment Sensing Subsystem |
|  | HCI | Human-CAV Interaction |
|  | MAS | Motion Actuation Subsystem |

|  |  |  |
| --- | --- | --- |
| **SubS** | **Term** | **Definition** |
| AMS | Command | High-level instructions whose execution allows a CAV to reach a Goal. |
| AMS | Full World Representation (FWR) | A description of Environment using the CAV’s and other CAVs’ Basic World Representations. |
| AMS | Goal | The planned Spatial Attitude at the end of the Decision Horizon. |
| AMS | Path | A sequence of Pose 𝑝𝑖 = (xi, yi, zi, αi, βi, γi) in the Offline Map. |
| AMS | Pose | Coordinates and orientation of the CAV in the Offline Map p = (𝑥, 𝑦, z, α, β, γ) |
| AMS | Route | A sequence of Way Points |
| AMS | Spatial Attitude | CAV’s Pose, Orientations and 1st and 2nd order time derivatives at a given time. |
| AMS | Traffic Rules | The digital representation of the traffic rules applying to a Waypoint. |
| AMS | Trajectory | A sequence of Spatial Attitudes (s1,s2,…si) and the expected time each Spatial Attitude will be reached. |
| AMS | Way Point | A point 𝑤𝑖 given as a coordinate pair (𝑥𝑖, 𝑦𝑖, z𝑖), in an Offline Map |
| AMS | Platooning | The joint moving of a group of vehicles along a shared trajectory |
| CAV | Accuracy | The degree of confidence of an estimated value (by computation, inference, etc.) |
| CAV | Connected Autonomous Vehicle | A vehicle capable to autonomously reach an assigned Pose by understanding human utterances, planning a route, sensing and interpreting the environment, exchanging information with other CAVs and acting on the CAV’s motion subsystem. |
| CAV | Decision Horizon | The time within which the decision is assumed to be implemented. |
| CAV | Health | The condition, e.g., mechanical, of an AIW (Subsystem) or an AIM. |
| CAV | Reference Model | The collection of the following resources:  1. AIW and their input/output data.  2. AIMs and their input/output data and connections. |
| CAV | Subsystem | One of the 4 components making up the CAV. |
| ESS | Accuracy | The precision with which a Data Type measures an entity. |
| ESS | Audio object | An aurally identifiable sound source. |
| ESS | Basic World Representation (BWR) | A digital representation of the Environment created with information available from the CAV’s ESS and an Offline Map or provided by another CAV in range. |
| ESS | Data Format | The standard digital representation of Data. |
| ESS | Data Type | An elementary component of a Data Format. |
| ESS | Deliberative Action |  |
| ESS | Environment | The portion of the world of current interest to the CAV. |
| ESS | Environment Sensing Technology (EST) | One of the technologies used by the Environment Sensing Subsystem, including Offline Map. to sense the Environment. |
| ESS | Format compatibility | The ability of a Data Format to be losslessly converted to another Data Format. |
| ESS | Global Navigation Satellite System (GNSS) | One of the systems providing global navigation informationsuch as GPS, Galileo, Glonass, BeiDou, Quasi Zenith Satellite System (QZSS) and Indian Regional Navigation Satellite System (IRNSS). |
| ESS | Inertial Measurement Unit | An inertial positioning device, e.g., odometer, accelerometer, speedometer, gyroscope etc. |
| ESS | Offline Map | An offline-created map of a location and associated metadata. |
| ESS | Scene Description | The organised collection of Descriptors enabling an object-based description of a scene. |
| ESS | Spatial Attitude | The set of 18 values (x,y,z,α,β,γ; their 1st order derivatives; and their 2nd order derivatives) corresponding to the Position (x,y,z) and Orientation (α,β,γ) representing pitch, yaw and roll and their derivatives of an Object. The position of the object corresponds to that of a representative point of the object, if the object is rigid. |
| ESS | Visual object | A visually identifiable set of pixels/voxels. |
| ESS | World Representation | A digital representation of the Environment produced by an Environment Sensing Technology in a CAV. |
| General | Compatibility | The ability of a data format to be converted to another data format without loss of information |
| General | Connected Autonomous Vehicle | A vehicle able to autonomously reach an assigned target by:   1. Understanding human utterances. 2. Planning a route. 3. Sensing and interpreting the Environment. 4. Exchanging information with other CAV. 5. Acting on the CAV’s motion actuation subsystem. |
| HCI | Attitude | An element of the internal status of a human related to the way s/he intends to position him/herself vis-à-vis the Environment or subsets of it, e.g., “Confrontational”, “Respectful”. |
| HCI | Audio | Digital representation of an analogue audio signal sampled at a frequency between 8-192 kHz with a number of bits/sample between 8 and 32, and non-linear and linear quantisation. |
| HCI | Audio Object | Coded representation of Audio information with its metadata. An Audio Object can be a combination of Audio Objects. |
| HCI | Audio Scene | The Audio Objects of an Environment with Object location metadata. |
| HCI | Audio-Visual Object | Coded representation of Audio-Visual information with its metadata. An Audio-Visual Object can be a combination of Audio-Visual Objects. |
| HCI | Audio-Visual Scene | (AV Scene) The Audio-Visual Objects of an Environment with Object location metadata. |
| HCI | Avatar | An animated 3D object representing a real or fictitious person in a Virtual Space. |
| HCI | Avatar Model | An inanimate avatar exposing handles for animation. |
| HCI | Cognitive State | An element of the internal status of a human reflecting his/her understanding of the Environment, such as “Confused” or “Dubious” or “Convinced”. |
| HCI | Colour (of Speech) | The timber of an identifiable voice independent of a current Personal Status and language. |
| HCI | Command | High-level instructions whose execution allows a CAV to reach a Goal. |
| HCI | Descriptor | Coded representation of audio, speech, or visual feature. |
| HCI | Emotion | An element of the internal status of a human resulting from the interaction with the Environment or subsets of it, such as “Angry”, and “Sad”. |
| HCI | Environment | A Physical or Virtual Space containing a Scene. |
| HCI | Face | The portion of a 2D or 3D digital representation corresponding to the face of a human. |
| HCI | Identity | The label uniquely associated with a human, an avatar or an object. |
| HCI | Manifestation | The manner of showing the Personal Status, or a subset of it, in any one of Speech, Face, and Physical Gesture. |
| HCI | Meaning (Gesture) | Information extracted from the head, arms, hands, and fingers conveying a coded message. |
| HCI | Meaning (Text) | Information extracted from an input text such as syntactic and semantic information. |
| HCI | Modality | One of Text, Speech, Face, or Gesture. |
| HCI | Personal Status | The ensemble of information internal to a person, including Emotion, Cognitive State, and Attitude. |
| HCI | Personal Status (Face) | The Personal Status, or a subset of it, conveyed by Face. |
| HCI | Personal Status (Gesture) | The Personal Status, or a subset of it, conveyed by Gesture. |
| HCI | Personal Status (Speech) | The Personal Status, or a subset of it, conveyed by Speech. |
| HCI | Physical Gesture | A movement of the body or part of it, such as the head, arm, hand, and finger, often a complement to a vocal utterance. |
| HCI | Response | Feedback autonomously generated by the CAV in response to a Command. |
| HCI | Scene | An Environment populated by humans and real objects or by avatars and virtual objects. |
| HCI | Scene Descriptors | The individual attributes of the coded representation of the objects in a Scene, including their location. |
| HCI | Speech | Digital representation of analogue speech sampled at a frequency between 8 kHz and 96 kHz with a number of bits/sample of 8, 16 and 24, and non-linear and linear quantisation. |
| HCI | Subword Lattice | A directed graph containing speech recognition sub-word candidates. |
| HCI | Text | A series of characters drawn from a finite alphabet. |
| HCI | Viewpoint | The point expressed as spatial coordinates and the direction expressed as (θ,φ) where a user is looking at the Environment. |
| HCI | Visual Object | Coded representation of Visual information with its metadata. A Video Object can be a combination of Video Objects. |
| HCI | Vocal Gesture | Utterance, such as cough, laugh, hesitation, etc. Lexical elements are excluded. |
| HCI | Word Lattice | A directed graph containing speech recognition word candidates. |
| MAS | Motion Data |  |

# References

## Normative References

This document references the following normative documents:

1. Technical Specification: The Governance of the MPAI Ecosystem (MPAI-GME) V1; https://bit.ly/30vp63g.
2. Technical Specification: AI Framework (MPAI-AIF) V1; https://bit.ly/30vp63g.
3. Technical Specification: Technical Specification: Multimodal Conversation (MPAI-MMC) V1; https://bit.ly/30vp63g.
4. Draft Technical Specification: Context-based Audio Enhancement (MPAI-CAE) VI, to be published at https://bit.ly/30vp63g.
5. Universal Coded Character Set (UCS): ISO/IEC 10646; December 2020
6. ISO/IEC 14496-10; Information technology – Coding of audio-visual objects – Part 10: Advanced Video Coding.
7. ISO/IEC 23008-2; Information technology – High efficiency coding and media delivery in heterogeneous environments – Part 2: High Efficiency Video Coding.
8. ISO/IEC 23094-1; Information technology – General video coding – Part 1: Essential Video Coding.

## Informative References

This document references the following informative documents:

1. Global Autonomous Cars (Semi & Fully) Market Report 2021, https://www.prnewswire.com/news-releases/global-autonomous-cars-semi--fully-market-report-2021-market-is-expected-to-reach-1-383-89-billion-in-2025---forecast-to-2030--301292821.html
2. Serban, Alexandru Constantin, Erik Poll, and Joost Visser. “A Standard Driven Software Architecture for Fully Autonomous Vehicles.” 2018 IEEE International Conference on Software Architecture Companion (ICSA-C). IEEE, 2018.
3. P. Avgeriou, "Describing, Instantiating and Evaluating a Reference Architecture: A Case Study," Enterprise Architecture Journal, June 2003.
4. S. Behere and M. Törngren, "A functional reference architecture for autonomous driving," Information and Software Technology 73, 2016, 136-150.
5. K. Jo, J. Kim, D. Kim, C. Jang, and M. Sunwoo, "Development of Autonomous Car - Part II: A Case Study on the Implementation of an Autonomous Driving System Based on Distributed Architecture," IEEE Transactions on Industrial Electronics, 62(8), 2015, 5119-5132.
6. K. Jo, M. Lee, D. Kim, J. Kim, C. Jang, E. Kim, and M. Sunwoo, "Overall Reviews of Autonomous Vehicle A1-System
7. Architecture and Algorithms," The International Federation of Automatic Control, 46(10), 2013, 114-119.
8. C. Maple, M. Bradbury, A. T. Le, and K. Ghirardello, "A Connected and Autonomous Vehicle Reference Architecture for Attack Surface Analysis," Applied Sciences, 9(23), 2019. doi:10.3390/app9235101
9. U. S. Department of Transportation, "Architecture Reference for Cooperative and Intelligent Transportation,", 2019, Retrieved March 03, 2020 from https://local.iteris.com/arc-it/html/architecture/architecture.html
10. SAE International Releases Updated Visual Chart for Its “Levels of Driving Automation” Standard for Self-Driving Vehicles, https://www.sae.org/news/press-room/2018/12/sae-international-releases-updated-visual-chart-for-its-%E2%80%9Clevels-of-driving-automation%E2%80%9D-standard-for-self-driving-vehicles
11. ISO 8855: "Road vehicles -- Vehicle dynamics and road-holding ability -- Vocabulary"
12. Rodolfo W. L. Coutinho and Azzedine Boukerche, Guidelines for the Design of Vehicular Cloud Infrastructures for Connected Autonomous Vehicles, IEEE Wireless Communications - August 2019
13. Claudine Badue, Rânik Guidolini, Raphael Vivacqua Carneiro, Pedro Azevedo, Vinicius B. Cardoso, Avelino Forechi, Luan Jesus, Rodrigo Berriel, Thiago M. Paixão, Filipe Mutz, Lucas de Paula Veronese, Thiago Oliveira-Santos, Alberto F. De Souza; Self-driving cars: A survey; Expert Systems With Applications 165 (2021) 113816
14. D. Cireşan, U. Meier, J. Masci, and J. Schmidhuber, “Multi-column deep neural network for traffic sign classification,” Neural Netw., vol.32, pp.333–338, Aug. 2012
15. ETSI TR 103 562 V2.1.1 (2019-12), Analysis of the Collective Perception Service (CPS); Release 2.
16. Gokulnath Thandavarayan, Miguel Sepulcre, and Javier Gozalvez; Generation of Cooperative Perception Messages for Connected and Automated Vehicles; IEEE Transactions on Vehicular Technology, Vol. 69, No. 12, December 2020
17. CAR 2 CAR Communication Consortium, https://www.car-2-car.org/
18. Usman Ali Khan and Sang Sun Lee; Distance-Based Resource Allocation for Vehicle-to-Pedestrian Safety Communication; https://www.mdpi.com/2079-9292/9/10/1640/pdf
19. Pranav Kumar Singhab, Sunit Kumar Nandiac, Sukumar Nandi; A tutorial survey on vehicular communication state of the art, and future research directions; Vehicular Communications Volume 18, August 2019, 100164
20. https://phantom.ai/assets/uploads/PAI%20Renesas%20Partnership%20Announcement%20(1).pdf
21. Charles R., Qi Li, Yi Hao Su, Leonidas J. Guibas; PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space; <https://arxiv.org/pdf/1706.02413.pdf>
22. Bernard, Olivier; Bradway, David; Hansen, Hendrik H.G.; Kruizinga, Pieter; Nair, Arun; Perdios, Dimitris; Ricci, Stefano; Rindal, Ole Marius Hoel; Rodriguez-Molares, Alfonso; Stuart, Matthias Bo; The Ultrasound File Format (UFF) - First draft, Proceedings of 2018 IEEE International Ultrasonics Symposium
23. LAS (LASer) File Format, Version 1.4 ,  
     https://www.loc.gov/preservation/digital/formats/fdd/fdd000418.shtml
24. https://www.lesliesikos.com/pcap/
25. https://pointclouds.org/documentation/tutorials/hdl\_grabber.html
26. https://www.mathworks.com/help/vision/ref/velodynefilereader.html
27. M. Heistermann, S. Jacobi and T. Pfaff,Technical Note: An open source library for processing weather radar data (wradlib), https://hess.copernicus.org/articles/17/863/2013/hess-17-863-2013.pdf
28. https://pro.arcgis.com/en/pro-app/latest/help/data/data-interoperability/supported-formats-with-the-data-interoperability-extension.htm
29. CDOT and Panasonic Take First Steps to Turn I-70 into Connected Roadway, <https://www.codot.gov/news/2018/july/cdot-and-panasonic-take-first-steps-to-turn-i-70-into-connected-roadway>
30. Navigation Data Standards, https://nds-association.org/
31. Matthias Schreier; Environment Representations for Automated On-Road Vehicles; https://www.researchgate.net/publication/323105152\_Environment\_Representations\_for\_Automated\_On-Road\_Vehicles
32. Toh Chai Keong, Juan Carlos Cano, Carlos-Javier Fernandez-Laguia, Pietro Manzoni; Wireless Digital Traffic Signs of the Future; IET Networks, September 2018
33. Vladimir Hahanov, Wajeb Gharibi, Eugenia Litvinova, Svitlana Chumachenko, Arthur Ziarmand, Irina Englesi, Igor Gritsuk, Vladimir Volkov, Anastasiia Khakhanova; Cloud-Driven Traffic Monitoring and Control Based on Smart Virtual Infrastructure; 2017-03-28
34. SharedStreets; <https://sharedstreets.io/>
35. Bojarski, Mariusz, et al. “End to end learning for self-driving cars.” arXiv preprint arXiv:1604.07316 (2016).
36. James S. Albus; The NIST Real-time Control System (RCS) An Approach to Intelligent Systems Research; <https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=820528>
37. Integrated Public Alert & Warning System; <https://www.fema.gov/gu/emergency-managers/practitioners/integrated-public-alert-warning-system/public>
38. OASIS Standard; Common Alerting Protocol Version 1.2; <https://docs.oasis-open.org/emergency/cap/v1.2/CAP-v1.2-os.pdf>
39. Dorais G., Kortenkamp D. (2001) Designing Human-Centered Autonomous Agents. Lecture Notes in Computer Science, vol 2112. Springer, Berlin, Heidelberg. <https://doi.org/10.1007/3-540-45408-X_32>
40. <https://youtu.be/nCaomvjPIok>
41. Stefano Mariani and Franco Zambonelli; Degrees of Autonomy in Coordinating Collectives of Self-Driving Vehicles; ISoLA 2021
42. Automatic Radar Performance Metrics and Testing Methods for Automated Driver Assistance Systems (ADAS) and Automated Driving Systems Applications (ADS); IEEE P3116; <https://sagroups.ieee.org/3116/>
43. Paulitsch, H., Fuchsberger, J., and Kock, K.: FM94-BUFR Encoding and Decoding Software, User Guidelines Version 1.6 For BUFR Software Version 3, http://www.knmi.nl/opera/bufr/doc/bufr\_sw\_desc.pdf, 2010.
44. Michelson, D. B., Lewandowski, R., Szewczykowski, M. and Beekhuis, H. (2011). EUMETNET OPERA weather radar information model for implementation with the HDF5 file format, version 2.1, Opera working document WD\_2008\_03, EUMETNET OPERA.
45. MPAI; MPAI-MMC V2 Use Cases and Functional Requirements; <https://mpai.community/standards/mpai-mmc/about-mpai-mmc/#V2>.
46. Intelligent Transport Systems (ITS); Facilities Layer function; Part 2: Position and Time management (PoTi); Release 2; ETSI EN 302 890-2.
47. Letizia Marchegiani, Xenofon Fafoutis; How Well Can Driverless Vehicles Hear? An Introduction to Auditory Perception for Autonomous and Smart Vehicles; IEEE Intelligent Transportation Systems Magazine; May/June 2022; <https://ieeexplore.ieee.org/document/9352026?mkt_tok=NzU2LUdQSC04OTkAAAGF_oHR8sRXCaOYj8X4r3C6PmWYuuhBVayVAyD0AGTR_-JhPa-VZS0LKHhpACcNGxFd7qmkZ83XnBW_DwGR8SQeXnmJOoIpkEqrxsjN1UzOSI3WMg>

# Subsystems

MPAI-CAV seeks to standardise all IT components that enable the implementation of a Connected Autonomous Vehicle (CAV), i.e., a system capable of executing the command to move its physical body autonomously – save for the exceptional intervention of a human – based on the analysis of

1. The data produced by a range of CAV sensors exploring the Environment.
2. The information transmitted by other sources in range, e.g., CAVs and roadside units (RSU).

*Figure 2* depicts the context and the actors where a CAV operates:

1. Other CAVs.
2. Roadside Units (RSU).
3. CAV-aware vehicles.
4. Pedestrians.
5. Traffic lights

A picture containing icon

Description automatically generated

*Figure 2 – The environment where a CAV operates*

MPAI-CAV includes 4 Subsystems:

**Human-CAV interaction (HCI)** recognises the human having right to the CAV, responds to humans’ commands and queries, provides extended environment representation (Full World Representation) for humans to use, senses human activities during the travel and may activate other Subsystems as required by humans.

**Environment Sensing Subsystem (ESS)** acquires information from the environment via a variety of sensors and produces a representation of the environment (Basic World Representation), i.e., its best guess given the sensed data.

**Autonomous Motion Subsystem (AMS)** computes the Route to destination, uses different sources of information – CAV sensors, other CAVs and transmitting units – to produce a Full World Representation and gives commands that drive the CAV to the intended destination.

**Motion Actuation Subsystem (MAS)** provides non-electromagnetic and non-acoustical environment information¸ receives and actuates motion commands in the physical world.

The interaction of the 4 subsystems in depicted in *Figure 3*:



*Figure 3 – The CAV subsystems*

The following high-level workflow illustrates the operation of the CAV envisaged by this docum­ent.

A *human* with appropriate credentials requests the CAV, via *Human-CAV Interaction*, to take the human to a given place.

*Human-CAV Interaction* authenticates the human, interprets the request, and passes a command to the *Autonomous Motion Subsystem*. Later, instructions can be integrated/ corrected.

*Autonomous Motion Subsystem*:

Requests *Environment Sensing Subsystem* to provide the current Pose.

Computes the Route.

Issues the start command.

*Environment Sensing Subsystem* computes and sends the Basic World Representation to *Aut­onomous Motion Subsystem* and CAVs in range*.*

*Autonomous Motion Subsystem*:

Receives and fuses the Basic World Representations to compute the Full World Representation.

Computes a Path.

Issues commands to *Motion Actuation Subsystem* to move the CAV accordingly.

While the CAV moves, humans

Interact and hold conversation with *Human-CAV Interaction* and/or other humans on board.

Issue commands.

Request views of the environment (Full World Representation) etc.

Interact with (humans in) other CAVs.

## Human-CAV Interaction (HCI)

### Use Case description

Human-CAV Interaction operates based on the principle that the CAV is impersonated by an avatar, selected/produced by the CAV rights-holder. The CAV avatar features are:

1. Visible: head, face, arms, hands, and fingers.
2. Audible: speech embedding as much as possible the sentiment, e.g., emotion, that would be displayed by a human driver.

The CAV’s avatar is reactive to:

1. The Environment, e.g., it can show an altered face if a human driver has done what it considers an improper action.
2. A human, e.g., it shows an appropriate face to a human in the cabins who has made a joke gazing at them.
3. Etc.

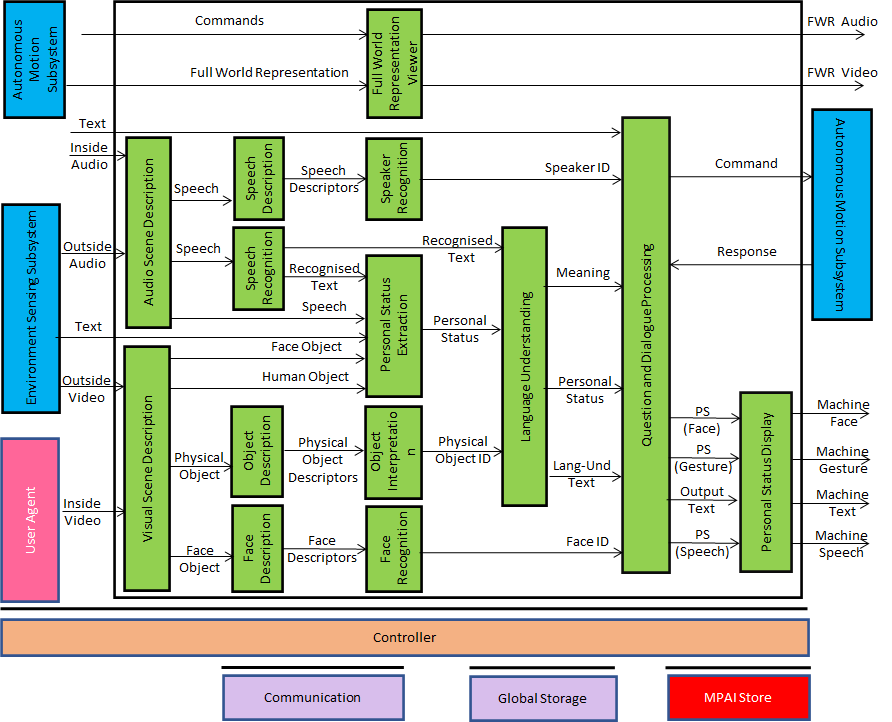
Other forms of interaction are:

1. CAV authenticates humans interacting with it.
2. A human issues commands to a CAV, e.g.:
   1. Commands to Autonomous Motion Subsystem, e.g.: go to a Waypoint or display Full World Representation (see 5.3), etc.
   2. Other commands, e.g.: turn off air conditioning, turn on radio, call a person, open window or door, search for information etc.
3. A human entertains a dialogue with a CAV, e.g.:
   1. CAV offers a selection of options to human (e.g., long but safe way, short but likely to have interruptions).
   2. Human requests information, e.g., time to destination, route conditions, weather at destination etc.
   3. Human entertains a casual conversation.
4. A CAV monitors the passenger cabin, e.g.:
   1. Physical conditions, e.g., temperature level, media being played, sound level, noise level, anomalous noise, etc.
   2. Passenger data, e.g., number of passengers, ID, estimated age, destination of passengers.
   3. Passenger activity, e.g., level of passenger activity, level of passenger-generated sound, level of passenger movement, emotion on face of passengers.
   4. Passenger-to-passenger dialogue, two passengers shake hands, or passengers hold everyday conversation.

It is important to point out that, although vehicles can exhibit different levels of autonomy, the exhibited autonomy should always be adjustable [1]. The system should recognise people as intelligent agents it should inform and be informed by. A CAV should be able to change its level of autonomy to one of several levels during its operation. Such an adjustment may be initiated by a human, another system, or the CAV itself. An important benefit of adjustable, user-centred autonomy is increased user acceptance of the system [47].

### Reference architecture

*Figure 4* represents the Human-CAV Interaction (HCI) Reference Model.



*Figure 4 – Human-CAV Interaction Reference Model*

HCI operates in two modes:

1. Outdoor: when humans are in the environment and approach the CAV:
   1. HCI separates and locates
      1. The human faces from the environment.
      2. The human speech sources from the Environment sound.
   2. HCI identifies the humans using face and speech.
   3. HCI converses with humans on “travel-by-car”-related matters.
2. In the cabin:
   1. HCI separates the speech sources from other sounds in the cabin and locates them.
   2. HCI locates the visual elements of humans in the cabin.
   3. HCI identifies the humans.
   4. HCI extracts Personal Status from Speech, Face, and Gesture.
   5. HCI converses with humans in the cabin.

When conversing with the humans in the cabin, the CAV responds by generating an avatar animated using the Personal Status Display [53].

Depending on the technology used (data processing or AI), the AIMs in *Figure 4* may need to access external information, such as Knowledge Bases, to perform their functions. While not represented in *Figure 4*, they will be identified, if required, in the AI Modules subsection.

### Input and output data

*Table 3* gives the input/output data of Human-CAV Interaction.

*Table 3 – I/O data of* *Human-CAV Interaction*

|  |  |  |
| --- | --- | --- |
| **Input data** | **From** | **Comment** |
| Audio (ESS) | Users in the Environment | User authentication  User command  User conversation |
| Audio | Cabin Passengers | User’s social life  Commands/interaction with CAV |
| Text | Cabin Passengers | User’s social life  Commands/interaction with CAV |
| Video (ESS) | Users in the Environment | Commands/interaction with CAV |
| Video | Cabin Passengers | User’s social life  Commands/interaction with CAV |
| Full World Representation | Autonomous Motion SS | For processing by FWR Viewer |
| **Output data** | **To** | **Comments** |
| Output Speech | Users in the Environment | CAV’s response to passengers |
|  | Cabin Passengers | CAV’s response to passengers |
| Output Face | Cabin Passengers | CAV’s face when speaking |
| Output Gesture | Cabin Passengers | CAV’s head when speaking |
| Output Text | Cabin Passengers | CAV’s response to passengers |
| Full World Representation Audio | Passenger Cabin | For passengers to hear Environment |
| Full World Representation Video | Passenger Cabin | For passengers to view Environment |

### AI Modules

*Table 4* gives the AI Modules of the Human-CAV Interaction depicted in *Figure 4*.

*Table 4* *– AI Modules of* *Human-CAV interaction*

|  |  |
| --- | --- |
| **AIM** | **Function** |
| **Audio Scene Description** | Receives Environment Audio (outside)  Produces Audio Scene Descriptors in the outdoor and indoor scenarios. |
| **Visual Scene Description** | Receives Environment Video (outside)  Produces Visual Scene Descriptors in the outdoor and indoor scenarios. |
| **Speech Description** | Receives Speech  Produces Speech Descriptors |
| **Speech Recognition** | Receives Speech  Produces Recognised Text. |
| **Physical Object Description** | Receives Physical Object  Produces Object Descriptors |
| **Face Description** | Receives Face Object  Produces Face Descriptors |
| **Speaker Recognition** | Receives Speech Descriptors  Produces Speaker ID |
| **Personal Status Extraction** | Receives Recognised Text, Speech, Text, Face Object, Human Object.  Produces Personal Status |
| **Object Interpretation** | Receives Physical Object Descriptors  Produces Object ID |
| **Face Recognition** | Receives Face Descriptors  Produces Face ID |
| **Language Understanding** | Receives Recognised Text, Personal Status, and Object ID  Produces Meaning, Personal Status, Lang-Und-Text |
| **Question and Dialogue Processing** | Receives Speaker ID, Personal Status, Meaning, Lang-Und-Text, and Face ID.  Produces PS (Face), PS (Gesture) Output Text, and PS (Speech) |
| **Personal Status Display** | Receives PS (Face), PS (Gesture) Output Text, and PS (Speech)  Produces Machine (Face), Machine (Gesture), Machine (Text), and Machine (Speech) |
| **Full World Representation Viewer** | 1. Receives Full World Representation (FWR) and user Commands 2. Presents a FWR view as instructed by human via FWR Com­mands. |

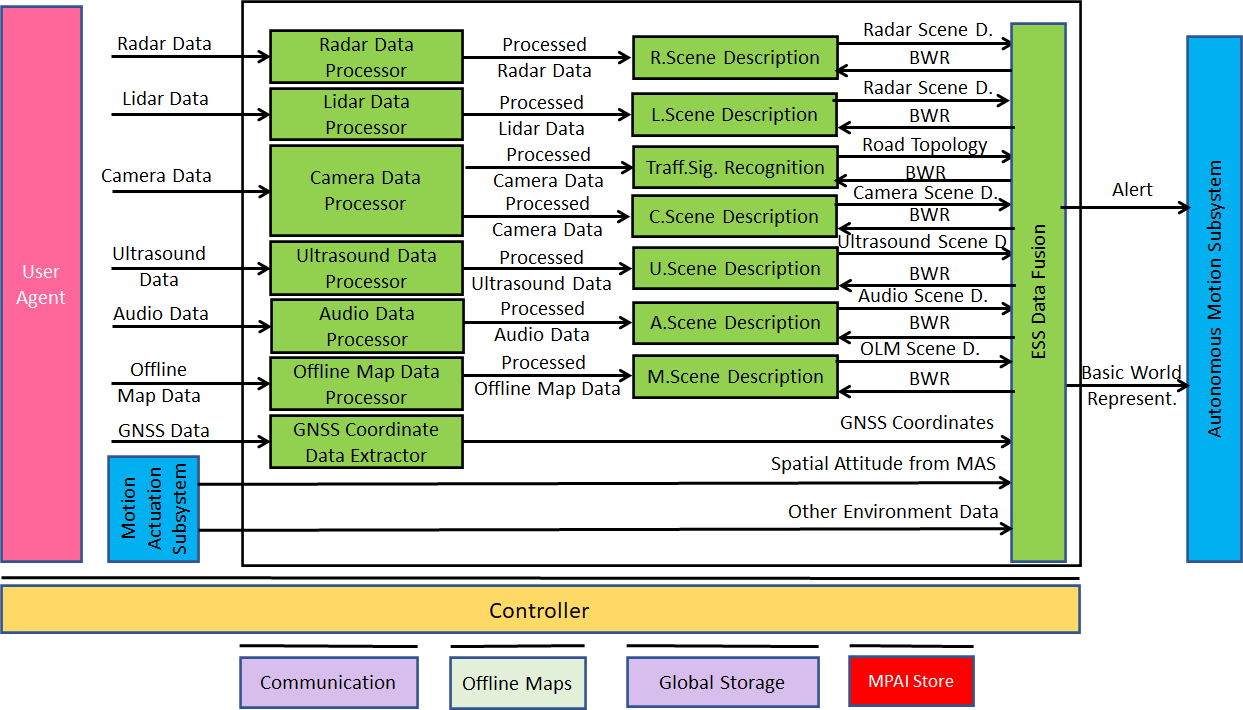
## Environment Sensing Subsystem (ESS)

### Use Case description

The purpose of the Environment Sensing Subsystem (ESS) is to acquire all information from the Environment – electromagnetic, acoustic, Spatial Attitude and time derivatives up to 2nd order, and other physical data (e.g., temperature, pressure, humidity etc.) – that can be sensed with the devices onboard with the goal of creating the Basic World Representation.

### Reference architecture

*Figure 5* gives the Environment Sensing Subsystem Reference Model.



*Figure 5 – Environment Sensing Subsystem Reference Model*

*Figure 5* represents the general case in which raw information provided by an Environment Sensing Technologies (EST), e.g., Radar Data, is acquired and processed by, e.g., the Radar Data Processor AIM, and then passed to the Radar Scene Description AIM. Of course, the Radar Data Processor AIM and the Radar Scene Description AIM can be integrated into a single AIM.

It is assumed that the Traffic Signalisation Recognition produces the Road Topology by processing the Processed Camera Data. The model can easily by extended to the case that another EST is processed to achieve produce the Road Topology.

Note that Online Map Data is considered as an EST. The data are restricted to an area around the current position of the CAV.

The ultimate purpose of the ESS is to produce the Basic World Representation (BWR) and the Spatial Attitude, an element of the BWR, using all Sensed Data from available ESTs.

### Input and output data

The Environment Sensing Technologies (EST) currently considered by MPAI-CAV are:

1. Global navigation satellite system or GNSS (~1 & 1.5 GHz Radio).
2. Geographical position and orientation, and their time derivatives up to 2nd order.
3. Data in the visible range, possibly supplemented by depth information (400 to 700 THz).
4. Lidar (~200 THz infrared).
5. Radar (~25 & 75 GHz).
6. Ultrasound (> 20 kHz).
7. Audio in the audible range (16 Hz to 16 kHz). [55]
8. Spatial Attitude (in addition to GNSS, from the Motion Actuation Subsystem).
9. Other environmental data (temperature, humidity, ...).

*Table 5* gives the input/output data of Environment Sensing Subsystem.

*Table 5 – I/O data of* *Environment Sensing Subsystem*

|  |  |  |
| --- | --- | --- |
| **Input data** | **From** | **Comment** |
| Spatial Attitude from AMS | Motion Actuation Subsystem | To be fused with GNSS data |
| Other Environment Data | Motion Actuation Subsystem | Temperature etc. to be added to Basic World Representation |
| Global Navigation Satellite System (GNSS) | ~1 & 1.5 GHz Radio | Get Pose from GNSS |
| Radar | ~25 & 75 GHz Radio | Capture Environment with Radar |
| Lidar | ~200 THz infrared | Capture Environment with Lidar |
| Ultrasound | Audio (>20 kHz) | Capture Environment with Ultrasound |
| Cameras (2/D and 3D) | Video (400-800 THz) | Capture Environment with Cameras |
| Microphones | Audio (16 Hz-16 kHz) | Capture Environment with Microphones |
| **Output data** | **To** | **Comment** |
| Spatial Attitude | Autonomous Motion Subsystem | For Route, Path and Trajectory |
| Basic World Representation | Autonomous Motion Subsystem | Locate CAV in the Environment |

### AI Modules

*Table 6* gives the AI Modules of Environment Sensing Subsystem.

*Table 6 – AI Modules of* *Environment Sensing Subsystem*

|  |  |
| --- | --- |
| **AIM** | **Function** |
| **Radar Data Processor** | Processes radar data. |
| **Lidar Data Processor** | Processes lidar data. |
| **Camera Data Processor** | Processes camera data. |
| **Ultrasound Data Processor** | Processes ultrasound data. |
| **Audio Data Processor** | Processes microphone array data. |
| **Online Map Data Processor** | Converts online map data to a format enabling the creation of the online map-based Scene Description. |
| **GNSS Data Coordinate Extractor** | Provides the CAV’s global coordinates as computed from GNNS Data. |
| **Radar Scene Description** | Produces Radar Scene Descriptors from Processed Radar Data |
| **Lidar Scene Description** | Produces Lidar Scene Descriptors from Processed LidarData |
| **Traffic Signalisation Recognition** | Produces Road Topology of the Environment. |
| **Ultrasound Scene Description** | Produces Ultrasound Scene Descriptors from Processed Ultrasound Data. |
| **Audio Scene Description** | Produces Audio Scene Descriptors from Processed Audio Data. |
| **Online Map Scene Description** | Produces Online Map Data Scene Descriptors from Processed Online Map Data. |
| **Environment Sensing Subsystem Data Fusion** | Creates the CAV’s Basic World Representation by fusing the Scene Descriptors of the different Environment Sensing Technologies, Road Topology, Spatial Attitude, and Other Environment Data. |

## CAV-to-Everything (V2X)

Earlier versions of this document envisaged a special subsystem dedicated to communication with entities external to the CAV. In the current version, communication external to the is handled directly by Subsystems in need for communication, e.g., the Autonomous Motion Subsystem.

### Use Case description

To improve its Environment perception capabilities, a CAV exchanges information via radio with other entities, e.g., CAVs in range and other CAV-aware communication devices such as Roadside Units and Traffic Lights in a secure fashion. The current version of this document does not consider the secure CA-to-CAV communication.

Multicast Communication adopted when a CAV broadcasts its identity and in case of heavy data types, e.g., Basic World Representation (BWR). Unicast mode may be used in other cases.

A Communication Device outside of the MPAI-AIF Trusted Zone of the Autonomous Motion Subsystem (AMS) in in charge of communication. The Device communicates with any AIF of the CAV which has communication needs or from which the Communication Device has received data.

The flow of operations of the Communication Device when handling communication with other CAVs or with devices having CAV functionality (e.g., a traffic light or a roadside unit) is:

1. Receives identities broadcasted by CAVs in range.
2. Establishes unicast sessions with CAVs in range. Issues to be considered are:
   1. Can a CAV’s Communication Device handle a large number of sessions?
   2. Can the AMS process data from a large number of different CAVs?
   3. Which application protocol is used by the Communication Device?
   4. Does the delay inherent in a unicast protocol have an impact?
3. Creates a list of CAVs in range with which it has established a session.
4. Sends the list with Basic World Representations (BWR) received via broadcast to the Autonomous Motion Subsystem (AMS).
5. ESS sends its own BWR to the Communication Device.
6. Communication Device broadcasts BWR in encrypted form with a key that is only known to CAVs in range that have an open unicast session with the Communication Device.

Note: no decision has been made on whether a CAV should send/receive Full World Representations (FWR).

The Communication Device is also made aware of the nature of the CAV and CAV-like device:

1. Traffic light.
2. Fire Truck.
3. Police.
4. Ambulance.
5. Flock Leader.

CAVs should communicate using a protocol that assigns a slice of the available transmission rate to each CAV based on the number of CAVs.

### Input and output data

#### CAVs within range

*Table 7* gives the data types a CAV broadcasts to CAVs in range via its Communication Device.

*Table 7 – I/O data of* *CAV’s Communication Device*

|  |  |  |
| --- | --- | --- |
| **Input Data** | **From** | **Comments** |
| Basic World Representation | Other CAVs | A digital representation of the Environment created by the ESS. |
| CAV Identity | Other CAVs | In principle, this should be the digital equivalent of today’s plate number including Manufacturer and Model information. The need to share this information is TBD. |
| CAV Intention | Other CAVs | The Path and other motion data relevant to other CAVs |
| Full World Representation | Other CAVs | A digital representation of the Environment created by fusing all available Basic World Representations. The need to share this information is TBD. |
| Information Messages | Other CAVs | Typical messages a CAV can broadcast. Potentially important messages for CAVs are given by [45, 46]   1. CAV is an ambulance. 2. CAV carries an authority. 3. CAV carries a passenger with critical health problem. 4. CAV has a mechanical problem of an identified level. 5. Works and traffic jams ahead 6. Environment must be evacuated 7. .... |
| **Output Data** | **To** | **Comments** |
| Basic World Representation | Other CAVs | Same as input for all other input data. |
| Full World Representation | Other CAVs | A digital representation of the Environment obtained from the fusion of all available Basic World Representations. The need to share this information is TBD. |

#### CAV-aware equipment

Examples of such equipment are traffic lights, roadside units, vehicles with CAV communication capabilities. The following data may be exchanged:

1. Identity and coordinates (exact coordinate reference).

Static Full World Representation regularly updated via download (may be part of the Offline Map).

Current objects in Environment.

State (Green-Yellow-Red) of traffic light and time to change state.

Lane markings.

Speed limits.

Pedestrian crosswalks

General information on the Environment (e.g., one way street etc.)

Etc.

Such equipment can:

1. Act as any other CAV in range.
2. Have the authority to organise motion of CAVs in range.

#### Other non-CAV vehicles

Other vehicles can be scooters, motorcycles, bicycles, other non-CAV vehicles, possibly transmitting their position as derived from GNSS. No response capability is expected. Vehicle may also have the capability to transmit additional information, e.g., identity, model, speed.

#### Pedestrians

Their smartphones can transmit their coordinates as available from GNSS. No response capability is expected.

## Autonomous Motion Subsystem (AMS)

### Use Case description

The typical series of operations carried out by the Autonomous Motion Subsystem (AMS) is described below. Note that the sequential description does not imply that an operations can only be carried out after the preceding one has been completed.

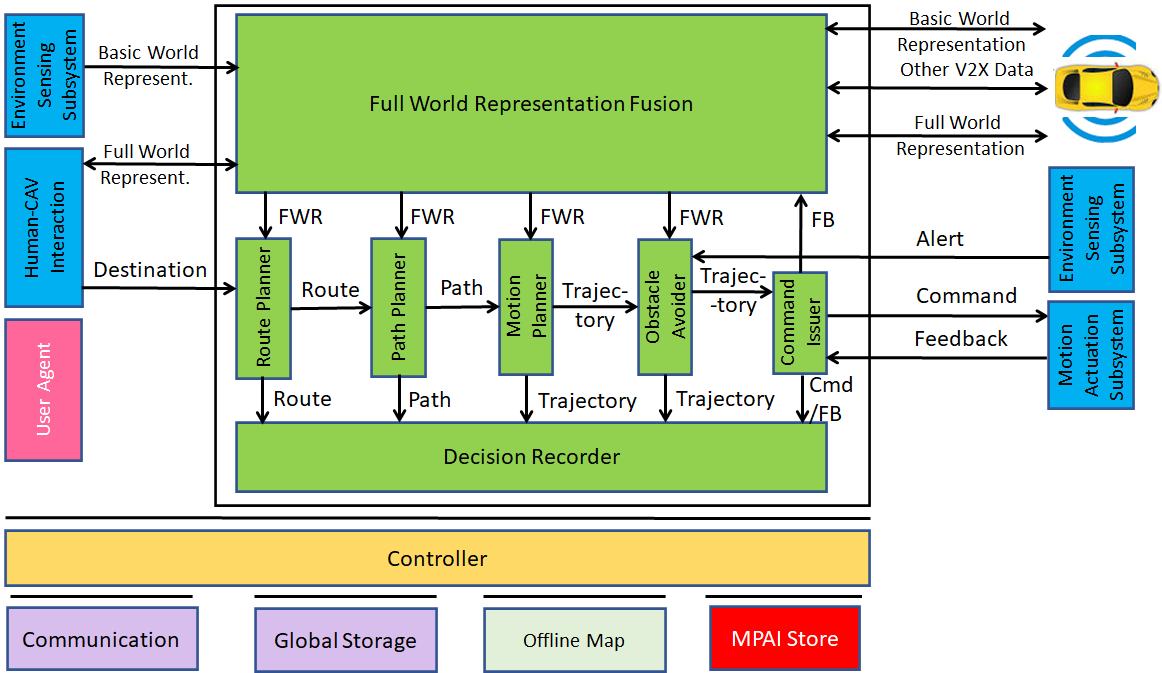
1. Human-CAV Interaction requests Autonomous Motion Subsystem to plan and move the CAV to the human-selected Pose. Dialogue may follow.
2. Computes the Route satisfying the human’s request.
3. Receives the current Basic World Represen­tation from Environment Sensing Subsystem.
4. While moving, CAV
   1. Transmits the Basic World Representation and other data to CAV-to-Everything.
   2. Receives Basic World Representations and other data from CAV-to-Everything.
   3. Produces the Full World Representation by fusing its own Basic World Representation with those from other CAVs in range.
   4. Plans a Path connecting Poses.
   5. Selects behaviour to reach intermediate Goals acting on information about the Goals other CAVs in range intend to reach.
   6. Defines a Trajectory that
      1. Complies with general traffic rules and local traffic regulations
      2. Preserves passengers’ comfort.
   7. Refines Trajectory to avoid obstacles.
   8. Sends the Motion Actuation Subsystem Commands to take the CAV to the next Goal.
5. Stores the data resulting from a decision (Route Planner, Path Planner etc.)

The AMS should be designed in such a way that different levels of autonomy, e.g., those indicated by SAE International [9], are possible depending on the amount and level of available func­tionalities.

### Reference architecture

*Figure 6* gives the Autonomous Motion Subsystem Reference Model.

A human activates the CAV requesting to be transported to a waypoint. This request activates the Route Planner and the Path Planner. The latter requests the Full World Representation from the Full World Representation Fusion AIM. The AIM has received the Basic World Representations from its ESS and from the other CAVs. The chain Behaviour Selection-Motion Planner eventually generates a command which is checked by the Obstacle Avoider AIM. If this has received a alert from the ESS containing an object captured from an EST, it checks the current FWR. If the check confirms that an action contradiction the decision of the Motion Planner AIM, the Obstacle Avoider AIM sends an appropriate command to the Motion Actuation Subsystem. Otherwise, the Motion Planner AIM’s decision is confirmed. The decision of each element of the said chain are recorded.



*Figure 6 – Autonomous Motion Subsystem Reference Model*

### Input and output data

*Table 8* gives the input/output data of Autonomous Motion Subsystem.

*Table 8 – I/O data of* *Autonomous Motion Subsystem*

|  |  |  |
| --- | --- | --- |
| **Input data** | **From** | **Comment** |
| Command from HCI | Human-CAV Interaction | Human commands, e.g., “take me home” |
| Basic World Representation | Environment Sensing Subsystem | CAV’s Environment representation. |
| Other Environment Data | Environment Sensing Subsystem | E.g., temperature, air pressure. |
| Other V2X Data | Other CAVs | Other CAVs and vehicles, and roadside units. |
| Feedback from MAS | Motion Actuation Subsystem | CAV’s response to Command. |
| **Output data** | **To** | **Comment** |
| Response to HCI | Human-CAV Interaction | MAS’s response to AMS Command |
| Command to MAS | Motion Actuation Subsystem | Macro-instructions, e.g., “in 5s assume a given State”. |
| Full World Representation | Other CAVs | For information to other CAVs |

### AI Modules

*Table 9* gives the AI Modules of the Autonomous Motion Subsystem.

*Table 9 – AI Modules of Autonomous Motion Subsystem*

|  |  |
| --- | --- |
| **AIM** | **Function** |
| **Route Planner** | Computes a Route, through a road network, from the current to the target Pose. |
| **Path Planner** | Generates a set of Paths, considering   1. Current Route. 2. State. 3. Full World-Representation. 4. Traffic Rules. |
| **Behaviour Selector** | Sets a Goal with a Driving Behaviour, to be reached within the Decision Horizon time frame. |
| **Motion Planner** | Defines a Trajectory, from the current State to the current Goal fol­lowing the Behaviour Selector’s Path to the extent possible, satisfying the CAV’s kinematic and dynamic constraints, and considering passengers’ comfort. |
| **Obstacle Avoider** | Defines a new Trajectory to avoid obstacles. |
| **Command** | Instructs the CAV to execute the Trajectory considering the Environment conditions. |
| **Full World-Representation Fusion** | Creates an internal representation of the Environment by fusing infor­mation from itself, CAVs in range and other transmitting units.. |

## Motion Actuation Subsystem (MAS)

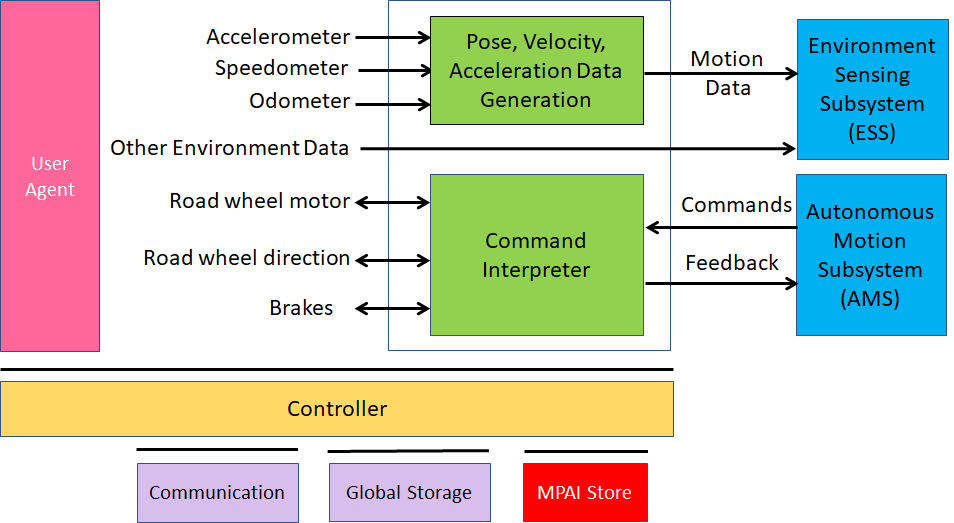
### Use Case description

The Motion Actuation Subsystem:

1. Transmits information gathered from its sensors and its mechanical subsystems to Environ­ment Sensing Subsystem.
2. Receives Command from Autonomous Motion Subsystem.
3. Translates such instructions into specific commands to its own mechanical subsystems, e.g., road wheels, wheel motors, brakes.
4. Receives feedbacks from its mechanical subsystems.
5. Packages feedbacks into high-level information.
6. Send Feedback (high-level information) to Autonomous Motion Subsystem.

### Reference architecture

*Figure 7* represents the Motion Actuation Subsystem Reference Model.



*Figure 7 – Motion Actuation Subsystem Reference Model*

### Input and output data

*Table 10* gives the input/output data of Motion Actuation Subsystem.

*Table 10 – I/O data of* *Motion Actuation Subsystem*

|  |  |
| --- | --- |
| **Input** | **Comments** |
| Odometer | Provides distance data. |
| Speedometer | Provides instantaneous velocity. |
| Accelerometer | Provides instantaneous acceleration. |
| Other Environment data | Provide other environment data, e.g., humidity, pressure, temperat­ure. |
| Road Wheel Motor | Forces road wheels rotation, gives feedback. |
| Road Wheel Direction | Moves road wheels by an angle, gives feedback. |
| Brakes | Acts on brakes, gives feedback. |
| Command from AMS | High-level motion command. |
| **Output** | **Comments** |
| Motion data | Position, velocity, acceleration. |
| Other data | Other environment data. |
| Feedback to AMS | Feedback from Command Converter during and after Command ex­ecution. |

### AI Modules

*Table 11* gives the AI Modules of Autonomous Motion Subsystem.

*Table 11 – AI Modules of Motion Actuation Subsystem*

|  |  |
| --- | --- |
| **AIM** | **Function** |
| **Pose-Velocity-Acceleration Data Generation** | Transforms odometer, speedometer, accelerometer data to standard data format. |
| **Command Interpreter** | Converts AMS Commands into specific actuation commands to Road wheel motor, Road wheel direction and Brakes. Forwards MAS feedbacks to AMS. |

# Functional Requirements

Functional Requirements developed in this document refer to individual technologies identified as necessary to implement MPAI-CAV Use Cases using AIMs operating in an MPAI AI Framework (AIF). They adhere to the following guidelines:

AIMs are defined to allow implementations by multiple technologies (AI, ML, DP).

DP-based AIMs may need interfaces, e.g., to a Know­ledge Base. AI-based AIM will typically require a learning process, however, support for this process is not included in the document. MPAI may develop further requirements covering that process in a future document.

AIMs can be aggregated in larger AIMs. Consequently, some data flows of aggregated AIMs may no longer be accessible.

MPAI has issued Calls for Technologies for the MPAI-MMC [3] and MPAI-CAE [4] standards and acquired a set of first-generation technologies related to some of the data types listed below. MPAI is ready to consider new technologies related to the data Formats requested in this Chapter if:

1. They support new requirements and/or they enhance capabilities.
2. The need to support such new enhanced capability requirements are documented.

## Human-CAV Interaction

### I/O Data summary

For each AIM (1st column), *Table 12* gives the input (2nd column) and the output data (3rd column).

*Table 12 – I/O data of Human-CAV Interaction AIMs*

|  |  |  |  |
| --- | --- | --- | --- |
| **AIM** | | **Function** | |
| **Audio Scene Description** | | Receives Environment Audio (outside inside)  Produces Audio Scene Descriptors in the outdoor and indoor scenarios. | |
| **Visual Scene Description** | | Receives Environment Video (outside and inside)  Produces Visual Scene Descriptors in the outdoor and indoor scenarios. | |
| **Speech Description** | | Receives Speech  Produces Speech Descriptors | |
| **Speech Recognition** | | Receives Speech  Produces Recognised Text. | |
| **Physical Object Description** | | Receives Physical Object  Produces Object Descriptors | |
| **Face Description** | | Receives Face Object  Produces Face Descriptors | |
| **Speaker Recognition** | | Receives Speech Descriptors  Produces Speaker ID | |
| **Personal Status Extraction** | | Receives Recognised Text, Speech, Text, Face Object, Human Object.  Produces Personal Status | |
| **Object Interpretation** | | Receives Physical Object Descriptors  Produces Object ID | |
| **Face Recognition** | | Receives Face Descriptors  Produces Face ID | |
| **Language Understanding** | | Receives Recognised Text, Personal Status, and Object ID  Produces Meaning, Personal Status, Lang-Und-Text | |
| **Question and Dialogue Processing** | | Receives Speaker ID, Personal Status, Meaning, Lang-Und-Text, and Face ID.  Produces PS (Face), PS (Gesture) Output Text, and PS (Speech) | |
| **Personal Status Display** | | Receives PS (Face), PS (Gesture) Output Text, and PS (Speech)  Produces Machine (Face), Machine (Gesture), Machine (Text), and Machine (Speech) | |
| **Full World Representation Viewer** | Full World Representation  Viewer Command | | FWRV Audio  FWRV Video |

### Digital representation of analogue signals

#### Microphone Array Audio

MPAI needs Microphone Array Audio to support the requirements of the following existing and new Use Cases:

1. Enhanced Audioconference Experience (EAE).
2. Conversation About a Scene (CAS).
3. Human-CAV Interaction (HCI).
4. Avatar-Based Videoconference (ABV).
5. Audio-On-the-Go (AOG).

In [4] MPAI has standardised Microphone Array Audio defined as Interleaved Multichannel Audio whose channels are sampled at a minimum of 5.33 ms (e.g., 256 samples at 48 kHz) to a maximum of 85.33 ms (e.g., 4096 samples at 48 kHz) and each sample is quantised with 16 or 24 or 32 bits.

**To respondents**

Respondents are requested to propose extensions to [4] or to propose alternative representation formats supporting the Use Cases described in this document:

1. HCI – Humans approach a CAV and converse with it.
2. HCI – Humans are in the cabin and converse with the CAV.
3. ABV – Human/s is/are in a room attending a virtual videconference

#### 2D Video

Video may be used to capture visual information to create a Visual Scene as identified in this document.

MPAI [3] specifies Video as:

1. Pixel shape: square
2. Bit depth: 8 or 10 bits/pixel
3. Aspect ratio: 4/3 or 16/9
4. 640 < # of horizontal pixels <1920
5. 480 < # of vertical pixels <1080
6. Frame frequency 50-120 Hz
7. Scanning: progressive
8. Colorimetry: ITU-R BT709 or BT2020
9. Colour format: RGB or YUV
10. Compression, either:
    1. Uncompressed.
    2. Compressed according to one of the following standards: MPEG-4 AVC [7], MPEG-H HEVC [8], MPEG-5 EVC [9]

**To respondents**

Respondents are invited to comment on MPAI’s parameter choice for 2D Video.

#### 3D Video

3D Video may be used to capture visual information to facilitate the creation of a Visual Scene. This can result from one camera an array of cameras having video + depth as the baseline format or other 3D Video data formats.

**To respondents**

Respondents are requested to provide a 3D Video format to create Visual Scene Descriptors as identified in this document:

1. CAS – A human in a room with objects.
2. HCI – Humans around a CAV.
3. HCI – Humans sitting in the cabin.
4. ABV – Humans sitting in a room attending a conference.

### Natively digital data formats

#### Text

For all instances of text, MPAI specifies ISO/IEC 10646, Information technology – Universal Coded Character Set (UCS) [5] as Text representation because of its ability to represent the characters used by most languages currently in use.

**To respondents**

Respondents are invited to comment on this choice.

#### Recognised Text

Recognised Text can be represented as simple **Text** or as a Word or Sub-word lattice.

**To respondents**

Respondents are requested to propose a format for Recognised Text:

1. Whose start is timestamped in relation to the selected unit, e.g., word, syllable, phoneme.
2. That provides:
   1. Time interval (duration).
   2. Strings of characters or phonemic symbols.
   3. Corresponding probability.

#### Language Understanding (Text)

Language Understanding (Text) is Text revised by the Language Understanding Module.

**To Respondents**

Respondents are requested to comment on the above.

#### Summary

Summary is data structure composed of Text possibly enhanced by characters expressing a Personal Status in a format suitable for human editing.

**To respondents**

Respondents are invited to propose a format satisfying the requirements.

#### Environment Model

The Environment Model is the data format describing a Virtual Environment.

An Environment Model can be:

1. Captured from a Real Environment.
2. Synthetically generated.

to produce a Virtual Environment:

1. Captured from a Real Environment, or
2. Synthetically generated.

The Environment Model shall have a Visual and Audio component and satisfy the following functional requirements:

1. The Visual component shall:
   1. Describe the elements bounding the Environment (e.g., walls, ceiling, floor, doors, windows).
   2. Describe the Visual Objects in the Environment (e.g., table, swivel chair, and furniture) at given coordinates.
   3. Support uniform lighting in the environment.
   4. Support integration of avatars with animated face and gesture.
2. The Audio component shall:
   1. Describe the acoustic characteristics of the environment, e.g., reverberation time.
   2. Describe the Audio Objects in the Environment.
   3. Support the integration of Audio Objects (e.g., the speech associated to an avatar) in the environment at given coordinates and directions.
3. It shall be possible to associate Audio Objects to Visual Objects, e.g., a Speech to an Avatar and vice-versa.
4. The party receiving an Environment Model shall be able to use it without additional information from the party who created it.

**To Respondents**

MPAI requests respondents:

1. To propose a format for the Environment Model that would allow the digital representation and rendering of a static Environment with the features described and with the ability to accommodate at runtime any other dynamic Audio and Visual Descriptors identified in this document.
2. To comments on the use of the Industry Foundation Classes (IFC)/BIM [] file for the visual component or to propose alternative formats.

#### Avatar Model

The Avatar Model is the data format describing a static avatar from the waist up displaying movements in face and gesture.

An Avatar Model can be:

1. Captured from an instance of a human.
2. Synthetically generated.

With the goal of producing an Avatar Model:

1. Captured from an instance of a human, or
2. Synthetically generated.

The Avatar Model shall be able to:

1. Represent the face, head, arms, hands, and fingers specific of a human.
2. Display a given Personal Status in face, head, arms, hands, and fingers.
3. Animate the lips based on Text, Speech, and Personal Status.

The Avatar Model shall have a standard format, i.e., a party shall be able to use an Avatar Model as intended by the party who created it without access to additional information.

**To Respondents**

MPAI requests respondents to propose an Avatar Model satisfying the stated requirements.

#### Human Object

Human Object is an object in a Visual Scene corresponding to a human satisfying the following requirements:

1. It represents the upper part of the body (from the waist up) with high accuracy.
2. It gives access to the following components of the human body: face, head, arms, hands, and fingers.

**To respondents**

Respondents are invited to propose a Human Object format that supports the requirements.

#### Face Object

Face Object is the 2D image captured from a camera or the 3D image of a face extracted from a Visual Scene Description that can be used for different purposes, such as:

1. To extract the identity from the Face Object.
2. To extract the Personal Status from the Face Object.
3. To extract the spatial coordinates of the Face Object.

The Face Object shall be organically part of a Human Object.

**To respondents**

Respondents are invited to propose a Face Object format satisfying the above requirements.

#### Head Object

Head Object is the 2D image captured from a camera or the 3D image of a head extracted from a Visual Scene Description that

1. Is organically part of the Human Object.
2. Can be used to extract Head Descriptors.

**To respondents**

Respondents are invited to propose a Head Object format satisfying the above requirements.

#### Arm Object

Arm Object is the 2D image captured from a camera or the 3D image of an arm extracted from a Visual Scene Description that

1. Is organically part of the Human Object.
2. Can be used to extract Arm Descriptors.

**To respondents**

Respondents are invited to propose an Arm Object format satisfying the above requirements.

#### Hand Object

Hand Object is the 2D image captured from a camera or the 3D image of an arm extracted from a Visual Scene Description that

1. Is organically part of the Arm Object.
2. Can be used to extract Hand Descriptors.

**To respondents**

Respondents are invited to propose a Hand Object format satisfying the above requirements.

#### Finger Object

Finger Object is the 2D image captured from a camera or the 3D image of an arm extracted from a Visual Scene Description that

1. Is organically part of the Hand Object.
2. Can be used to extract Finger Descriptors.

**To respondents**

Respondents are invited to propose a Finger Object format satisfying the above requirements.

#### Spatial coordinates

Use Cases in this document need to refer to a coordinate system to be able to express the spatial position of a real or virtual object. For example, in CAS machine needs to know the position and orientation of the human so that it can locate the object the human is pointing at.

Four different cases are considered:

1. The coordinates are absolute on the surface of the Earth, e.g., in the case of the objects in a World Representation used by a Connected Autonomous Vehicle.
2. The coordinates are defined with reference to a specific point on the Earth.
3. The coordinates are possibly arbitrarily defined local coordinates, e.g., in the case of Conversation About a Scene. The specific use case defines what is the point with coordinates (0,0,0).
4. The coordinates are in a virtual space. The specific use case defines what is the point with coordinates (0,0,0).

**To respondents**

MPAI requests respondents to provide a coordinate system that can represent a point on a known place on the surface of the Earth, represent a point with reference to an arbitarily defined point on the Earth, a physical point which does not have a defined correspondence with a point on the Earth and a point in a virtual space.

#### Orientation

The phi/theta angles of the perpendicular to the front face of an object with the φ=0 and θ=0 direction assumed for the scene.

#### Positions and Orientation

A set of coordinate points corresponding to the spatial positions and their orientations of a set of Visual Objects in an Environment.

**To Respondents**

MPAI requests respondents to propose a data format for Positions and Orientation.

#### Point of View

The point expressed as coordinates from where a user looks at the space around him/her or hears the sound field in a given direction (θ,φ).

**To Respondents**

MPAI requests respondents to propose a format to represent a Point of View.

#### Full World Representation

The Full World Representation requirements are developed in the context of Autonomous Motion Subsystem requirements.

**To respondents**

Respondents are invited to read the requirements there.

#### Full World Representation Viewer commands

The requirements of FWR interaction will be developed once the FWR requirements are defined.

**To respondents**

Respondents are invited to read the requirements there.

### Descriptors

#### Audio Scene Descriptors

Audio Scene Descriptors describe the structured composition of Audio Objects from two different perspectives:

1. The Audio Scene is *captured from the real world*, e.g.:
   1. A group of humans approach a CAV or are sitting in the CAV cabin. A microphone array in the CAV separates and spatially locates the different sound sources (speech and noise).
   2. A group of humans is sitting in a room. A microphone array in the room separates and spatially locates the different sound sources (speech and noise).
2. The Audio Scene is *used to create a sound field* in an Environment, e.g.:
   1. In Avatar-Based Videoconference, avatars are sitting around a table and the mouth of each avatar emits the sound of speech of the human it represents.

Audio Scene Descriptors shall have a standard format, i.e., a party shall be able to use the Audio Scene Descriptors as intended by the party who created it without access to additional information. MPAI-CAE V1 has standardised an audio scene description for 1.b [4].

**To Respondents**

MPAI requests respondents to propose Audio Scene Descriptors satisfying the needs of the Use Cases described above and other Use Cases of this document.

The Audio Scene Description should provide:

1. Access to the individual objects (Speech Objects).
2. The spatial coordinates and the directions of the individual objects in a specified coordinate system.

#### Visual Scene Descriptors

Visual Scene Descriptors describe the structured composition of Visual Objects in a Visual Scene serving two purposes:

1. *The Visual Scene is captured from an Environment* in:
   1. (HCI) A group of humans approach a CAV or are sitting in the cabin. The CAV
      1. Separates the different visual objects (humans and other objects),
      2. Identifies the faces of all humans.
      3. Spatially locates all objects as they move.
2. *The Visual scene is created in an Environment*, Virtual (VR) and Real (AR) in:
   1. (HCI) an avatar represents the virtual CAV driver displaying its Personal Status (face and head) congruent with the speech it utters.

A party shall be able to use the Visual Scene Descriptors as intended by the party who created it without access to additional information.

**To Respondents**

MPAI requests respondents to propose Visual Scene Descriptors representing the spatial and temporal arrangement of the objects that are relevant to the 3 Use Cases of this document, i.e.:

1. Human Objects (in CAS, HCI, and ABV).
2. Face Objects (in CAS, HCI, and ABV).
3. Physical Objects (in CAS and HCI).

All objects share the feature of being “static”, in the sense that they have a static coordinate set and a static direction (θ,φ), while the movements of face, head, arms, hands and fingers are handled outside the Visual Scene Descriptors.

The Visual Scene Description should provide:

1. Access the individual objects of a scene for specific processing such as the extraction of Face Object or the extraction Face and Gesture Descriptors from the Human Object.
2. The time-dependent spatial coordinates of the individual objects in a specified coordinate system.

#### Audio-Visual Scene Descriptors

Audio-Visual Scene Descriptors describe the structured composition of the Audio-Visual Objects in an audio-visual scene.

The Audio-Visual Scene Descriptors can be:

1. Captured from an instance of a human.
2. Synthetically generated.

with the goal of producing an Audio-Visual Scene:

1. Captured from an instance of a human, or
2. Synthetically generated.

or reproducing a Physical Environment.

The Audio-Visual Scene Descriptors shall:

1. Have a common time base.
2. Enable the association of co-located audio and visual objects if both are available.
3. Support the physical displacement and interrelation (e.g., occlusion) of audio and visual objects over time.

For example:

1. Description of the audio-visual objects in the Real Environment where a group of speaking humans with a background noise.
2. Description of the audio-visual objects in a Virtual Environment (a conference room) attended by a set of speaking avatars representing humans and a Virtual Secretary.

The Audio-Visual Scene Descriptors shall have a standard format, i.e., a party shall be able to use Audio-Visual Scene Descriptors as intended by the party who created them without access to additional information.

**To Respondents**

MPAI requests respondents to propose Audio-Visual Scene Descriptors that integrate Audio Scene Descriptors and Visual Scene Descriptors serving the needs of the Use Cases of this document.

#### Avatar Descriptors

Avatar Descriptors represent the instantaneous alterations of the face, head, arms, hands, and fingers of an Avatar Model.

Avatar Descriptors can be:

1. Captured from an instance of a human.
2. Synthetically generated.

with the goal of producing Avatar Descriptors of an avatar:

1. Captured from an instance of a human, or
2. Synthetically generated.

Avatar Descriptors shall have a standard format, i.e., a party shall be able to use Avatar Descriptors as intended by the party who created them without access to additional information.

**To Respondents**

MPAI requests respondents to propose a set of Avatar Descriptors supporting the stated requirements.

#### Face Descriptors

MPAI needs Face Descriptors for the following possibly overlapping purposes:

1. To recognise the identity of a limited number of humans, e.g.,
   1. Members of a family.
   2. Customers of a CAV-renting company.
   3. Participants in a meeting.
2. To describe the features of a face for the purpose of extracting:
   1. The physical features of a face.
   2. The movement of the mobile parts of a face, e.g., lips and eyes.
   3. The Personal Status conveyed by a Face Object.
3. To animate the face of an Avatar Model using any of the points in 2.
4. To specify the coordinates of the point representing the face assumed to be a rigid body.
5. To specify the trajectory of the point representing a moving face.

For instance, the x(t), y(t), z(t) coordinates of the point representing the face and the yaw, pitch, roll.

Key Points are examples of Descriptors. They should be technology-independent (i.e., image processing, or ML, or their combinations).

**To Respondents**

Respondents are requested to propose Facial Descriptions that can be used for the purposes identified above.

#### Gesture Descriptors

Gesture is the ensemble of head, arm, hand, and finger movement and Gesture Descriptors are the organised composition of Head Descriptors, Arm Descriptors, Hand Descriptors, and Finger Descriptors.

Gesture Descriptors used to:

1. Represent arbitrary movement of head, arms, hands, and fingers.
2. Recognise:
   1. Sign language.
   2. Culture-dependent signs (e.g., mudra sign).
   3. Coded hand signs, e.g., to indicate a particular object in a scene.
   4. A human’s Personal Status conveyed by Gesture.
3. Animate an avatar’s head, arms, hands, and fingers.

**To Respondents**

MPAI requests syntax and semantics of arm, hand, finger key points that are independent of the technology used to implement the Key Point Detection AIM – i.e., image processing, or ML, or combinations of the two.

#### Head Descriptors

Head Descriptors are used to describe:

1. The Head Object for the purpose of extracting:
   1. The physical features of a head.
   2. The Manifestation of Personal Status in the head.
2. The coordinates of the point representing the head assumed to be a rigid body.
3. The trajectory of the point representing a moving head.
4. The rotation of a moving head.

For instance, the x(t), y(t), z(t) coordinates of the point representing the head and the yaw, pitch, and roll rotation.

**To Respondents**

MPAI requests respondents to propose a set of Head Descriptors satisfying the requirements.

#### Arm Descriptors

Arm Descriptors are used to describe:

1. The Arm Object for the purpose of extracting:
   1. The physical features of an arm.
   2. The Manifestation of Personal Status in the arm.
2. The coordinates of the 2 components of an arm assumed to be a rigid body.
3. The trajectory of the points representing a moving arm.

**To Respondents**

MPAI requests respondents to propose a set of Arm Descriptors satisfying the requirements.

#### Hand Descriptors

Hand Descriptors are used to describe:

1. The Hand Object for the purpose of extracting:
   1. The physical features of a hand.
   2. The Manifestation of Personal Status in the hand.
2. The coordinates of the hand assumed to be a rigid body.
3. The trajectory of the points representing a moving hand.

**To Respondents**

MPAI requests respondents to propose a set of Hand Descriptors satisfying the requirements.

#### Finger Descriptors

Finger Descriptors are used to describe:

1. The Finger Object for the purpose of extracting:
   1. The physical features of a finger.
   2. The Manifestation of Personal Status in the finger.
2. The coordinates of the 3 components of an arm assumed to be a rigid body.
3. The trajectory of the points representing a moving finger.

**To Respondents**

MPAI requests respondents to propose a set of Finger Descriptors satisfying the requirements.

#### Speech Descriptors

MPAI-MMC V1.2 [3] and MPAI-CAE V1.4 [4] have standardised Speech Descriptors (Features). To support the Use Cases of this document, Speech Descriptors are needed satisfying the following requirements:

1. Ability to recognise the identity of a limited number of humans, e.g.,
   1. Members of a family.
   2. Customers of a CAV-renting company.
   3. Participants in a meeting.
2. Ability to describe the features of a speech segment for the purpose of extracting:
   1. The Personal Status conveyed by the speech segment.
3. Ability to use the identifiable speech features of 2.a to select the identifiable features of the speech synthesised by a third-party synthesiser.

**To Respondents**

MPAI requests respondents to propose Speech Descriptors satisfying the above requirements.

#### Text Descriptors

To support the Use Cases of this document, Text Descriptors are needed satisfying the following requirements: To describe elements of e.g., sentiment analysis of a text, such as positive /negative words, emojis.

**To Respondents**

MPAI requests respondents to propose Text Descriptors satisfying the above requirements.

### Interpretations

#### Emotion

MPAI has standardised a set of basic Emotions and their semantics to represent emotion conveyed by text, speech, face, and gesture. The same set is assumed to represent the emotion conveyed by any channel.

**To Respondents**

Respondents are requested to propose the following:

1. To comment on the use of [3] to represent emotion. If a respondent claims that Basic Emotion Set of [3] is unsuitable for the Use Cases of this document, respondents are requested to motivate their claims, and propose an extension of the MPAI Basic Emotion Set or a new solution.
2. To comment on the use of the same set of emotions for text, speech, face, and gesture.

#### Cognitive State

The internal status of a human reflecting his/her understanding of the Environment, such as “Confused” or “Dubious” or “Convinced”.

**To Respondents**

Respondents are requested to propose the following:

1. A basic set of Cognitive States suitable for the Text, Speech, Face, and Gesture channels preferably using the same structure adopted by MPAI for Emotion [3].
2. Motivate the use of different basic sets of Cognitive States for the different channels.

#### Attitude

The internal status of a human related to the way s/he intends to position him/herself vis-à-vis the Environment, especially social, or subsets of it, e.g., “Confrontational”, “Respectful”, “Soothing”.

**To Respondents**

Respondents are requested to propose the following:

1. A basic set of Attitudes suitable for the Text, Speech, Face, and Gesture channels preferably using the same structure adopted by MPAI for Emotion [3].
2. Motivate the use of different basic sets of Attitudes for the different channels.

#### Personal Status

Personal Status (PS) indicates a set of 3 factors {Emotion, Cognitive State, Attitude} conveyed by one or more of the Text, Speech, Face, and Gesture channels. Each factor may be present or absent, and may be time dependent, i.e., it could have a timestamp, say, T1. A factor could also have a duration D and represent a different Personal Status at time T2=T1+D.

Personal Status may be the result of fusing the Personal Statuses conveyed by at least two of the following channels: Text, Speech, Face, and Gesture.

**To Respondents**

Respondents are requested to propose the following:

1. A Personal Status format capable of describing the evolution of Personal Status over time.
2. A Fused Personal Status format supporting the requirements to:
   1. Include the Emotion, Cognitive Status, and Attitude making up a Personal Status.
   2. Retain information on the measured value of the different factors in a Personal Status conveyed by the different channels.
   3. Describe the evolution of Personal Status over time.

#### Meaning

Meaning is information extracted from an input text and Personal Status such as question, statement, exclamation, expression of doubt, request, invitation. MPAI-MMC [3] specifies a digital representation format for Meaning.

**To respondents**

Respondents are requested to comment on the suitability of the technology standardised in [3].

#### Object Identifier

MPAI-MMC [3] specifies an Object Identifier to be used to identify a visual object.

**To respondents**

Respondents are requested to comment on the suitability of the technology standardised in [3] for the Conversation About a Scene and Human-CAV Interaction Use Cases.

## Environment Sensing Subsystem

### I/O Data summary

For each AIM (1st column), *Table 13* gives the input (2nd column) and the output data (3rd column). The following 3-digit subsections give the requirements of the data formats in columns 2 and 3.

*Table 13 – Environment Sensing Subsystem AIMs and data*

|  |  |  |
| --- | --- | --- |
| **AIM** | **Input** | **Output** |
| **Radar Data Processor** | Radar Data. | Processed Radar Data. |
| **Lidar Data Processor** | Lidar data. | Processed Lidar Data. |
| **Camera Data Processor** | Camera data. | Processed Camera Data. |
| **Ultrasound Data Processor** | Ultrasound data. | Processed Ultrasound Data. |
| **Audio Data Processor** | Microphone Array Data. | Processed Audio Data. |
| **Online Map Data Processor** | Online Map Data. | Processed Online Map Data. |
| **GNSS Data Coordinate Extractor** | GNNS Data. | Global Coordinates. |
| **Radar Scene Description** | Processed Radar Data.  Basic World Representation. | Radar Scene Descriptors. |
| **Lidar Scene Description** | Processed Lidar Data.  Basic World Representation. | Lidar Scene Descriptors. |
| **Traffic Signalisation Recognition** | Processed Camera Data.  Basic World Representation. | Road Topology. |
| **Camera Scene Description** | Processed Camera Data.  Basic World Representation. | Lidar Scene Descriptors. |
| **Ultrasound Scene Description** | Processed Ultrasound Data.  Basic World Representation. | Ultrasound Scene Descriptors. |
| **Audio Scene Description** | Processed Audio Data  Basic World Representation. | Audio Scene Descriptors. |
| **Online Map Scene Description** | Processed Online Map Data | Online Map Scene Descriptors. |
| **Environment Sensing Subsystem Data Fusion** | Radar Scene Descriptors.  Lidar Scene Descriptors.  Road Topology.  Lidar Scene Descriptors.  Ultrasound Scene Descriptors.  Audio Scene Descriptors.  Online Map Scene Descriptors.  Spatial Attitude from MAS.  Other Environment Data, | Basic World Representation.  Alert. |

### Sensor data

#### Audio Data

Microphones are used to capture the external sound for the following purposes:

1. Create Audio Scene Description to
   1. Enable extraction of speech addressed to CAV by humans.
   2. Incorporate in Basic World Representation.
2. Suppress noise inside the passenger cabin.

MPAI-CAE [5] specifies Interleaved Multichannel Audio.

**To Respondents**

Respondents are requested to comment on the usability of the specified technology for MPAI-CAV and/or propose an Audio Array Format suitable to create a 3D sound field representation of the Environment for the stated purposes or to propose a new format highlighting its benefits.

#### Camera Data

Video camera data can be provided by one camera or by an array of cameras with or without depth measure.

**To Respondents**

Respondents are invited to propose a Camera format that facilitates identification, tracking and representation of visual objects with the goal to produce The Camera Scene Description.

#### Lidar Data

Radio Detection and Ranging (RADAR), LiDAR and Ultrasound are active sensors based on “time-of-flight” to measure the distance and speed of object based on the time it takes for a signal to hit an object and be reflected.

Unlike Radar, Lidar operates in the µm range – ultraviolet, visible, or near infrared light. The typical features of an eye-safe LiDAR are:

1. Has a frequency of ~200 THz and a wavelength ~1.5 µm (the visible range is 0.4 to 0.75 µm).
2. Measures the range of each voxel.
3. Measures pixel grayscale by the intensity variation of the reflected light.
4. Measures the colour of an object by using more than one wavelength.
5. Measures velocity by using the Doppler shift in frequency caused by motion, or by taking the position at different times.
6. Measures micro-motion by using the Doppler shift measured with a coherent LiDAR.
7. Produces 100 to 200 kpoints/frame.
8. Angular resolution is 0.1º and vertical field is 40º.
9. A Lidar scan captured at 25 fps generates 270 Mbit/s, i.e., 33.75 Mbytes/s.

The LAS (LASer) format is a binary file format for LiDAR point cloud data specified by the American Society for Photogrammetry and Remote Sensing (ASPRS) [31].

Pcap is a well-established data format for Lidar scans [32, 33, 34]. Other formats are listed in [36]. E57 is one of them.

**To Respondents**

Respondents are invited to provide a LiDAR data format that facilitates identification, tracking and digital representation of objects having the goal to produce Lidar Scene Descriptions.

#### Radar Data

Radar operates in the mm range. It can detect vehicles (CAVs and trucks) because they typically reflect radar signals while smaller and less reflecting objects, e.g., pedestrians and motorcycles have a poor reflectance. In a busy environment, the reflections of big vehicles can overcome a motorcycle’s causing missed detection of important objects (e.g., a child next to a vehicle), while a can may produce an image out of proportion to its size.

The main features of Radar are:

1. Measures distance.
2. Is independent of environment.
3. Provides low resolution images.
4. Provides distance (short range radar in the 25 GHz band).
5. Detects objects and measures speed @ ≤ 250 m (long range radar in the 76-77 GHz). Typical long-range radar systems have ranges of 80-200 m.
6. Only a small antenna is needed because wavelength is ~3.5-4 mm.
7. Atmospheric absorption limits interference with other systems.
8. A multitask 94-GHz pulse Doppler radar has 25-cm radial and 1.5 degrees angular resolution

Radar sensors build a representation of the environment based on the observation of complex, scattered radio waves, from which information of an object’s distance and velocity can be derived.

Some Radar data formats include:

Some Radar data formats include [35]:

1. OPERA BUFR format [51].
2. hdf5 formats [52].
3. NetCDF files generated by the commercial EDGE software.
4. hdf5 files generated by the commercial GAMIC software.
5. German Weather Services quantitative local scan format (DX).
6. Quantitative composite format (RADOLAN, see German Weather Service, 2004).

**To Respondents**

Respondents are invited to propose a format of Radar images that facilitates identification, tracking and representation of objects having the goal to produce the Radar Scene Description.

#### Ultrasound Data

These are the main features of Ultrasound:

1. Operates at frequencies above 20 kHz.
2. Is environment-independent of.
3. Yields low-resolution images.
4. Works on a limited range (≤ 10 m)

The Ultrasound File Format initiative has defined the Ultrasound File Format (UFF) format [30].

**To Respondents**

Respondents are invited to propose an Ultrasound Format that facilitates identification, tracking and representation of audio objects with the goal to produce the Ultrasound Scene Description.

#### GNSS Data

Global Navigation Satellite Systems (GNSS) provide spatial information with different accuracies. GNSS can only be relied on when reception conditions are above a certain level. This excludes GNSS in tunnels or urban canyons.

Some data formats are:

1. GPS Exchange Format (GPX) provides an XML schema providing a common GPS data format that can be used to describe waypoints, tracks, and routes.
2. World Geodetic System (WGS) includes the definition of the coordinate system's fundamental and derived constants, the ellipsoidal (normal) Earth Gravitational Model (EGM), a description of the associated World Magnetic Model (WMM), and a current list of local datum transfor­mations.
3. International GNSS Service (IGS) SSR is a format used to disseminate real-time products to support the IGS (igs.org) Real-Time Service. The messages support multi-GNSS and include corrections for orbits, clocks, DCBs, phase-biases and ionospheric delays. Extensions are planned to also cover satellite attitude, phase centre offsets and variations and group delay variations.

**To Respondents**

Respondents are requested to propose a single GNSS data format that is capable to represent the features of all GNSS types.

#### Offline Map Data

An Offline Map or HD maps or 3D maps is a roadmap with cm-level accuracy and a high environmental fidelity reporting the positions of pedestrian crossings, traffic lights/signs, barriers etc. at the time the Offline Map has been created.

A CAV should preferably use a single data format such as:

1. Navigation Data Standards [38] calls itself “The worldwide standard for map data in automotive eco-systems”. Their NDS specification covers data model, storage format, interfaces, and protocols.
2. SharedStreets [42] Referencing System calls itself a global non-proprietary system for describing streets.

or a minimum set of data formats that are recognised by the Offline Map Data Processor.

**To respondents**

Respondents are requested to provide one or, if unavoidable, more than one Offline Map Data Format as described.

#### Spatial Attitude from MAS

The CAV Spatial Attitude is received from MAS as the set of:

1. The spatial coordinates
2. Yaw, pitch, and roll
3. Their derivatives up to second order
4. Their accuracies.

A possible definition of the point the Spatial Attitude refers to has been provided by [54].

**To Respondents**

Respondents are requested to propose a format for the said set of data, and to comment on [54] or to propose a different definition.

### Processed data

#### Processed Camera Data

The Camera Data Processor converts Camera Data to a format that facilitates the creation of the corresponding Visual Scene Description.

**To Respondents**

Respondents are requested to propose a Processed Camera Data suitable for CAVs.

#### Processed Lidar Data

The Lidar Data Processor converts Lidar Data to a format that facilitates the creation of the corresponding Visual Scene Description.

**To Respondents**

Respondents are requested to propose a Processed Lidar Data suitable for CAVs.

#### Processed Radar Data

The Radar Data Processor converts Radar Data to a format that facilitates the creation of the corresponding Visual Scene Description.

**To Respondents**

Respondents are requested to propose a Processed Radar Data suitable for CAVs.

#### Processed Ultrasound Data

The Ultrasound Data Processor converts Ultrasound Data to a format that facilitates the creation of the corresponding Visual Scene Description.

**To Respondents**

Respondents are requested to propose a Processed Ultrasound Data suitable for CAVs.

#### Offline Map Data

The Offline Map Data Processor converts Offline Map Data to a format that facilitates the creation of the corresponding Visual Scene Description.

**To Respondents**

Respondents are requested to propose a Processed Offline Map Data suitable for CAVs.

#### Processed Audio Data

The Audio Data Processor converts Audio Data to a format that facilitates the creation of the Audio Scene Description.

**To Respondents**

Respondents are requested to propose a Processed Audio Data suitable for CAVs.

### Scene Descriptions

#### Introduction

The following points should be considered as requirements of the different referenced data formats:

1. Each Environment Sensing Technology (EST) produces snapshots of the Environment called sensed data at the frequency proper to the EST.
2. The Processed EST Data are passed to the specific EST Scene Description.
3. EST Scene Description produces Scene Descriptors. An SD may be a rather complex data structure that includes several elementary components called Data Types each having a Data Format.
4. Scene Descriptions (SD) are object-based, time-dependent, constantly updated, and contain objects that may have different resolutions, e.g., the background may be a single object with low resolution while a CAV that is 5 m away from the ego CAV is given with high resolution.
5. Scene Descriptors (SD) of different ESTs shall be compatible, i.e., an SD from an EST can be converted to the SD of another EST without loss of information
6. SD#1 of EST#1 may include Data Types not included in SD#2 of EST#2 and vice versa.
7. Some Data Types in SD#1 may be different than the Data Types in SD#2.
8. Data in SD#1 of EST#1 need not have the same values as the data of the same type as in SD#2 of of EST#2, i.e., their values may:
   1. Not have the same Accuracy.
   2. Conflict. even though their Data Types may be the same or compatible.
9. The formats of the Offline Maps shall allow for integration in a SD without loss of information.

The SD obtained by processing the snapshots of an EST shall have one of a limited number of formats standardised by MPAI, e.g.,

1. Object coordinates.
2. Bounding box.
3. 2D object description.
4. 3D, i.e., Volumetric object description.

The objects of a Scene Description of an EST shall contain the following attributes:

1. Identifier specific of a snapshot of the EST.
2. Timestamp.
3. Format (standardised by MPAI, see above).
4. State (Static, Dynamic, Unknown and their accuracies).
5. Spatial Attributes and estimated accuracy.
6. Identifiers of previously existing objects mapping to the current object in the given EST and estimated accuracy.
7. Any Semantics known at this stage.

The process whereby the Scene Description SD(t) of a particular EST is created from the actually sensed SDa(t) and the preceding SDs is:

1. Compute the predicted SDp(t) using SD(t-Δt), SD(t-2Δt) etc.
2. Finds and compares each predicted object in the predicted SDp(t) with each object in the SDa(t).
3. Creates SD(t) by:
   1. Updating the attributes of each object inherited from preceding SDs.
   2. Removing objects present in previous SDs and no longer present in SDa(t).
   3. Adding and assigning attributes to new objects sensed at time t (i.e., entirely new objects or the merge of two or more objects).

The EST SD keeps a memory of the past SDs. Recent objects may retain all attributes while objects in the far past may have coarser attributes or not be available.

The SD format of an EST is a list of SDs in the EST-specific time window. SD(t) is a list objects detected and confirmed at time t with their attributes.

The scene object format can be 2D, 2.5D or 3D.

**2D Scene Objects**

1. Static environment:
   1. Parametric free space representation represented as a single object.
   2. Alternative representations as individual static objects.
2. Dynamic environment: object-based representation.

**2.5D Scene Objects**

1. Static components of the scene
   1. Grid-based (elevation maps or stixel world), represented as a single object.
   2. Object-based for traffic poles and signals (e.g., Stixel world, Multi-level surface map).
2. Object-based for the dynamic parts (e.g., Stixel world, Multi-level surface map).

**3D Scene Objects**

1. Static components of the scene
   1. Voxel grids, meshes, possibly as a single.
   2. Object-based for traffic poles and signals (voxel grids, meshes).
2. Dynamic components of the scene (point clouds, voxel grids, meshes, …)

#### Camera Scene Descriptors

Camera Scene Description receives Processed Camera Data, a portion of a previous Basic World Representation and provides Camera Scene Descriptors.

**To Respondents**

Respondents are requested to propose Camera Scene Descriptors that provide the information identified in 6.2.3.1

#### Lidar Scene Description

Lidar Scene Description receives Processed Lidar Data, a portion of a previous Basic World Representation and provides Lidar Scene Descriptors.

**To Respondents**

Respondents are requested to propose Camera Scene Descriptors that provide the information identified in 6.2.3.2.

#### Radar Scene Description

Radar Scene Description receives Processed Radar Data, a portion of a previous Basic World Representation and provides Radar Scene Descriptors.

**To Respondents**

Respondents are requested to propose Radar Scene Descriptors that provide the information identified in 6.2.3.3.

#### Ultrasound Scene Description

Ultrasound Scene Description receives Processed Ultrasound Data, a portion of a previous Basic World Representation and provides Ultrasound Scene Descriptors.

**To Respondents**

Respondents are requested to propose Ultrasound Scene Descriptors that provide the information identified in 6.2.3.4.

#### Offline Maps Scene Description

Offline Map Scene Description receives Processed Offline Map Data and provides Offline Map Scene Descriptors.

**To Respondents**

Respondents are requested to propose Offline Map Scene Descriptors that provide the information identified in 6.2.3.5.

#### Audio Scene Description

Audio Scene Description receives Processed Audio Data, a portion of a previous Basic World Representation and provides Audio Scene Descriptors.

**To Respondents**

Respondents are requested to propose Audio Scene Descriptors that provide the information identified in 6.2.3.6.

### Road Topology

Traffic Signalisation Recognition receives Processed Camera Data containing:

1. Traffic signs.
2. Road signs.
3. Placement signs.
4. Traffic lights.

This AIM may also receive a portion of a previous Basic World Representation.

Possible classification as below (From: https://www.researchgate.net/figure/Traffic-Signals-ontology\_fig4\_271376130).

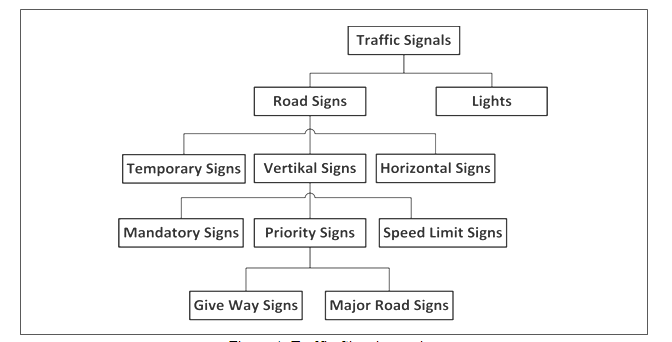


Figure 8 - Traffic Signal classification

**To Respondents**

Respondents are requested to propose a set of Traffic Signalisation Descriptors.

### Basic World Representation

The entire CAV’s SD is called Basic World Representation (BWR) resulting from the *integration* of the different SDs generated from different ESTs of a CAV [39].

The BWR shall result from the integration of all data sensed by a CAV:

1. Audio and Visual information.
2. Spatial information (e.g., GNSS, odometry).
3. Environmental data (e.g., weather, temperature, air pressure, ice and water on the road, wind, fog etc.)

The requirements of the BWR format to be standardised by MPAI are:

1. The BWR at time t shall include the BWRs of sufficient prior snapshots following at a frequency useful for synthesis of object tracking, inference of motion vectors, etc.
2. The BWR shall describe the traffic environment in terms of Road Topology (e.g., roads and lane geometry), and traffic rules and regulatory restrictions available at this stage applying to portions of the traffic environment[[1]](#footnote-1).
3. The BWR shall describe each Object with the following attributes:
   1. Timestamp.
   2. Type (Audio or Visual).
   3. ID.
   4. ID of corresponding object of different Type.
   5. Motion State (Static, Dynamic, Unknown).
   6. Spatial Attitude.
   7. Predicted Trajectory.
   8. Object dimensionality (2D, 2.5D and 3D). Applicable only to Visual Objects.
   9. Object shape
      1. Shape for Visual Objects
      2. Sound Field Shape for Audio Objects
   10. Semantics (to the extent that it is available at this stage)
       1. Another CAV or other object types
       2. Belonging to a group of objects (listed by their IDs).
   11. Accuracy of all the estimated values in the Object.
4. The BWR shall allow for easy check of the feasibility of a Trajectory (e.g., the AMS can easily check that the intended Trajectory of the ego CAV designed to reach the intended point does not collide with other Visual Objects in the Decision Horizon based on the current state of the BWR).
5. The BWR shall have a scalable representation, i.e., allowing for:
   1. Refinement of a BWR using new EST-specific Scene Descriptors.
   2. Extraction of part of the BWR based on a required Level of Detail (e.g., object bounding boxes and their Spatial Attitudes).
   3. Easy addition of new data (e.g., adding shape of an object when there was only the bounding box).
   4. Fast access to object metadata, e.g.:
      1. Spatial Attitude.
      2. Predicted Trajectory.
      3. Shape (e.g., bounding box for a Visual Object).
   5. Selected (read) access to data required by different AIMs, e.g.:
      1. Radar Scene Description accessing the current BWR to improve its description.
   6. Updates for Object and Scene from one snapshot to another.
   7. A CAV to communicate to another CAV a subset of its BWR containing objects with different degrees of details, e.g., starting from bounding boxes and their Position Attributes, depending on the available bandwidth.

**To Respondents**

Respondents are requested to propose formats for the following data types and their accuracy:

1. Timestamp.
2. Type (Audio or Visual).
3. Object IDs.
4. Motion State (Static, Dynamic, Unknown).
5. Spatial Attitude.
6. Trajectory (see AMS).
7. Object dimensionality (2D, 2.5D and 3D).
8. Object shape
   1. Shape for Visual Objects
   2. Sound Field Shape for Audio Objects
9. Semantics
   1. of individual objects
   2. of groups of objects (IDs, relative distances, etc.).

## Autonomous Motion Subsystem

### Summary of Autonomous Motion Subsystem data

*Table 14* gives, for each AIM (1st column), the input data (2nd column) and the output data (3rd column).

*Table 14 – CAV Autonomous Motion Subsystem data*

|  |  |  |
| --- | --- | --- |
| **CAV/AIM** | **Input** | **Output** |
| **Route Planner** | Pose | Route |
| Destination | Estimated time |
| Offline maps |  |
| **Full World Representation** | Basic World Representations | Full World Representation |
| **Path Planner** | Offline maps | Set of Paths |
| Route |
| Full World Representation |
| Traffic Rules |
| **Behaviour Selector** | Spatial Attitude | Path |
| Route |
| Full World Representation |
| **Motion planner** | Path | Trajectory |
| Full World Representation |
| **Obstacle Avoider** | Trajectory  Full World Representation | Trajectory |
| **Command to AMS** | Feedback | Command |
| **Autonomous Motion SS** | CAV identity and model | CAV identity and model |
| **Autonomous Motion SS** | Spatial Attitude-Path-Trajectory | Spatial Attitude-Path-Trajectory |
| **Autonomous Motion SS** | Basic World Representation | Basic World Representation |
| **Autonomous Motion SS** | Full World Representation | Full World Representation |
| **Autonomous Motion SS** | Messages | Messages |
| **Autonomous Motion SS** | Basic World Representation | Basic World Representation |

### Basic World Representation

Defined in Environment Sensing Subsystem.

**To Respondents**

Those intending to respond should check Basic World Representation in ESS.

### CAV Identifier

The CAV identification system should carry the following information

1. Country where the CAV has been registered
2. Registration number in the country
3. CAV manufacturer identifier
4. CAV model identifier

**To Respondents**

MPAI requests proposals for universal CAV identification system. Justified proposals for inclus­ion of additional data in the CAV Identifier are welcome.

### Command/Response

Defined in Human-to-CAV subsystem.

**To Respondents**

No response requested here. Comments welcome.

### Events

Events is used to provide CAV with information that is useful for its travel.

Examples are:

1. Road blocked at waypoint x,y,z
2. Traffic jam at waypoint x,y,z
3. ...

**To Respondents**

MPAI requests proposals for events, their semantics and coded representation.

### Full World Representation

The Autonomous Motion Subsystem of a CAV may produce its Full World Representation (FWR) by *integrating* the BWRs received from other CAVs in range.

### FWR requirements

The FWR requirements are:

1. The FWR shall be an extension of the BWR.
2. The FWR shall indicate where the BWR of the ego CAV has major discrepancies with a BWR another CAV and its identity.
3. The FWR shall include all available information from other CAVs, that enable a CAV to define a Path in the Decision Horizon Time.
4. The objects in the FWR may include additional semantics:
   1. Flags (e.g., warning)
   2. Platooning indicating which objects belong to a platoon
   3. Priority (e.g., police car, ambulance, vehicle carrying hazard material).
5. Fast access to:
   1. Priority.
6. Selected (read) access to data required by different AIMs, e.g., Obstacle Avoider may access the FWR to get information useful for avoiding obstacles.
7. The FWR shall allow for deliberative and reactive actions.
8. The BWRs of two CAVs with different positions and orientations may be different, have different levels of detail or have conflicting values.

**To Respondents**

Respondents are requested to:

1. Comment on the assumptions made and replaced, with justified new assumptions.
2. Propose a Basic World Representation format satisfying the requirements.

Proposals based on justified extended requirements will be considered.

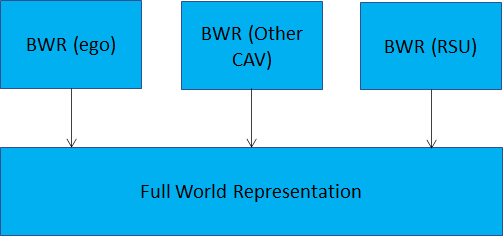


Figure 9 - Full World Representation

<https://www.mdpi.com/2079-9292/10/22/2825/pdf> propose an edge-fog-cloud computing-based

road dynamic object-mapping system. Our method processes the data from each connected

car using a standardized form and manages the dynamic information with grid-based

maps.

### Goal

A particular State.

**To Respondents**

No response requested. Comments welcome.

### Offline map

Defined in Environment Sensing Subsystem.

**To Respondents**

No response requested here. Comments welcome.

### Path

A sequence of Poses in the Offline Map

**To Respondents**

No response requested here. Comments welcome.

### Pose

A sequence of Poses in the Offline Map. The AMS issues micro-commands to MAS in case a Pose cannot be reached from the Pose in a straight line.

**To Respondents**

A format to represent Pose is requested.

### Route

A sequence of Waypoints. A Waypoint is at a sufficient high level in the Route-Path-Trajectory hierarchy.

**To Respondents**

A Route Format compatible with a proposed Offline Map Format is requested.

### State

Defined in Environment Sensing Subsystem.

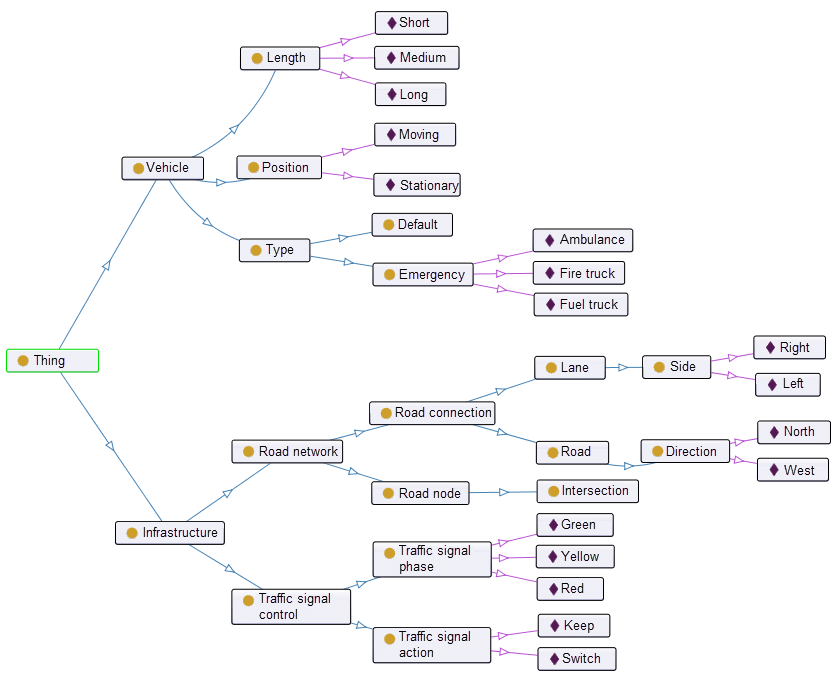
**To Respondents**

No response requested here. Comments welcome.

### Traffic rules

The traffic rules should be digitally represented to realise a route [41]. This could be based on an ontology, e.g., Control of vehicles and robots: creating of knowledge bases for mivar decision making systems robots and vehicles. See https://www.researchgate.net/figure/Ontology-for-traffic-signal-control\_fig1\_353182434

<https://www.researchgate.net/publication/339985273_Control_of_vehicles_and_robots_creating_of_knowledge_bases_for_mivar_decision_making_systems_robots_and_vehicles>.



**To Respondents**

MPAI requests a digital representation of traffic rules capable to:

1. Produce traffic rules from a given set of traffic signals.
2. Produce traffic signals from traffic rules.

A Traffic Ontology is a possible solution.

### Traffic Signals

Format to represent traffic signals on a road and around it, i.e., the semantics of the traffic signs. An ontology may be a solution.

**To Respondents**

MPAI requests a Traffic Signals Format capable to represent

1. All traffic signalisations required.
2. The specific local version of traffic signalisation.
3. The coordinates of the traffic signals.

### Trajectory

The Path and the Spatial Attitudes that allow a CAV to start from a Spatial Attitude and reach another Spatial Attitude in a given amount of time without violating Traffic Rules and affecting passengers’ comfort.

**To Respondents**

A digital representation of Trajectory is requested.

## Motion Actuation Subsystem

### Summary of Motion Actuation Subsystem data

*Table 15*gives, for each AIM (1st column), the input data (2nd column) from which AIM (column) and the output data (3rd column).

*Table 15 –Motion Actuation Subsystem data*

|  |  |  |
| --- | --- | --- |
| **CAV/AIM** | **Input** | **Output** |
| **Command Interpreter** | Command from AMS  Road Wheel Motor Feedback  Road Wheel Direction Feedback  Brakes Feedback | Feedback to AMS  Road Wheel Motor Command  Road Wheel Direction Command  Brakes Command |
| **Pose, Velocity, Acceleration Data Generation** | Accelerometer  Speedometer  Odometer | Motion Data |

### Accelerometer data

An accelerometer is an electronic sensor that measures the acceleration forces acting on a CAV. An accelerometer measures proper acceleration, i.e., the acceleration of a body in its own instantaneous rest frame, not to be confused with coordinate acceleration, i.e., acceleration in a fixed coordinate system. Therefore, an accelerometer at rest on the surface of the Earth measures an acceleration straight upwards of g ≈ 9.81 m/s2. In free fall (falling toward the centre of the Earth at ≈ 9.81 m/s2) measures zero.

**To Respondents**

Respondents are requested to propose a single Accelerometer data format.

### Brakes Command

The result of the interpretation of AMS Command to Brakes.

**To Respondents**

Respondents are requested to propose a set of command messages.

### Brakes Feedback

The feedback of Brakes to Command Interpreter.

**To Respondents**

Respondents are requested to propose a set of feedback messages.

### Command from AMS

The Command issued by the AMS

**To Respondents**

Respondents are requested to propose a set of high-level command messages.

### Feedback to AMS

The Feedback of Command Interpreter summarising the Feedbacks.

**To Respondents**

Respondents are requested to propose a set of high-level feedback messages

### Motion Data

**To Respondents**

Respondents are requested to propose a Motion Data Format bearing in mind that Motion Data will be used to create the CAV State by adding GNSS information.

### Odometer Data

An odometer converts as the distance travelled the number of wheel rotations times the tire circumference (π x tire diameter) from the start up to the point being considered.

**To Respondents**

Respondents are requested to propose a single Odometer Data Format.

### Other Environment Data

The set of Environment data such as temperature, air pressure, humidity etc.

**To Respondents**

Respondents are requested to propose a set Environment Data Formats.

### Road Wheel Direction Command

The result of the interpretation of AMS Command to Road Wheel Direction.

**To Respondents**

Respondents are requested to propose a set of Road Wheel Direction Commands

### Road Wheel Direction Feedback

The feedback of Road Wheel Direction to Command Interpreter.

**To Respondents**

Respondents are requested to propose a set of Road Wheel Direction Feedbacks

### Road Wheel Motor Command

The result of the interpretation of AMS Command to Road Wheel Motor.

**To Respondents**

Respondents are requested to propose a set of Road Wheel Motor Commands

### Road Wheel Motor Feedback

The feedback of Road Wheel Motor to Command Interpreter.

**To Respondents**

Respondents are requested to propose a set of Road Wheel Motor Feedbacks

### Speedometer

An electronic sensor that measures the instantaneous speed of a CAV.

**To Respondents**

Respondents are requested to propose a single Speedometer data format.

# Data privacy

A CAV can generate or acquire data for which privacy is an important characteristic. Here are some of the functions potentially affected by data privacy or that are liable to become accessible to authorities, e.g., police and judiciary.

## Human-CAV Interaction (HCI)

By having interactions with humans, HCI becomes aware of potentially sensitive information, e.g.:

1. Result of monitoring the passenger cabin.
2. Minute requests from humans, e.g., go to a way point, display Full World Representation, turn off air conditioning, etc.
3. Dialogue with human

## Environment Sensing Subsystem (ESS)

ESS collects large among of environment data for the purpose of creating instantaneous Basic World Representations, e.g.:

1. GNSS gives the position of the CAV and of whatever is perceived by the CAV that is approximate, but sufficiently precise for the intended uses.
2. Radar, Lidar, Ultrasound give variously defined information about what is in the environment surrounding the CAV.
3. Cameras give a 360° panoramic view of the environment where all objects, save those occluded, are visible.
4. External microphones give a complete representation of the external sound field.

A user could create a permanent and certified recording of important data acquired by ESS.

The environment recorder could compress and record all data acquired for a limited amount of time. Some data could be recorded for a longer time.

## CAV to Everything (V2X)

V2X acquires the identity of the CAVs in range and communicates appropriate subsets of the Basic and Full Worlds Representations.

## Autonomous motion subsystem (AMS)

AMS knows the exact waypoints the CAV has passed through and all the commands given to the Motion Actuation Subsystem.

By integrating the Basic World Representations of all CAVs in range and its own, a CAV can create a pretty detailed and extended map of the environment.

Recording the decisions made by the Decision Recorder creates highly critical data.

## Motion Actuation Subsystem (MAS)

MAS acquires position information through its Inertial Measurements Unit.

1. – General MPAI Terminology (Normative)

The Terms used in this standard whose first letter is capital and are not already included in *Table 1* are defined in *Table 16.*

*Table 16 – MPAI-wide Terms*

|  |  |
| --- | --- |
| **Term** | **Definition** |
| Access | Static or slowly changing data that are required by an application such as domain knowledge data, data models, etc. |
| AI Framework (AIF) | The environment where AIWs are executed. |
| AI Module (AIM) | A processing element receiving AIM-specific Inputs and producing AIM-specific Outputs according to according to its Function. An AIM may be an aggregation of AIMs. |
| AI Workflow (AIW) | A structured aggregation of AIMs implementing a Use Case receiving AIM-specific inputs and producing AIM-specific inputs according to its Function. |
| AIF Metadata | The data set describing the capabilities of an AIF set by the AIF Implem­enter. |
| AIM Metadata | The data set describing the capabilities of an AIM set by the AIM Implem­enter. |
| Application Programming Interface (API) | A software interface that allows two applications to talk to each other |
| Application Standard | An MPAI Standard specifying AIWs, AIMs, Topologies and Formats suitable for a particular application domain. |
| Channel | A physical or logical connection between an output Port of an AIM and an input Port of an AIM. The term “connection” is also used as a synonym. |
| Communication | The infrastructure that implements message passing between AIMs. |
| Component | One of the 9 AIF elements: Access, AI Module, AI Workflow, Commun­ication, Controller, Internal Storage, Global Storage, MPAI Store, and User Agent. |
| Conformance | The attribute of an Implementation of being a correct technical Implem­entation of a Technical Specification. |
| Conformance Tester | An entity authorised by MPAI to Test the Conformance of an Implem­entation. |
| Conformance Testing | The normative document specifying the Means to Test the Conformance of an Implem­entation. |
| Conformance Testing Means | Procedures, tools, data sets and/or data set characteristics to Test the Conformance of an Implem­en­tation. |
| Connection | A channel connecting an output port of an AIM and an input port of an AIM. |
| Controller | A Component that manages and controls the AIMs in the AIF, so that they execute in the correct order and at the time when they are needed. |
| Data | Information in digital form. |
| Data Format | The standard digital representation of Data. |
| Data Semantics | The meaning of Data. |
| Device | A hardware and/or software entity running at least one instance of an AIF. |
| Ecosystem | The ensemble of the following actors: MPAI, MPAI Store, Implementers, Conformance Testers, Performance Testers and Users of MPAI-AIF Im­plem­en­tations as needed to enable an Interoperability Level. |
| Event | An occurrence acted on by an Implementation. |
| Explainability | The ability to trace the output of an Implementation back to the inputs that have produced it. |
| Fairness | The attribute of an Implementation whose extent of applicability can be assessed by making the training set and/or network open to testing for bias and unanticipated results. |
| Function | The operations effected by an AIW or an AIM on input data. |
| Global Storage | A Component to store data shared by AIMs. |
| Identifier | A name that uniquely identifies an Implementation. |
| Implementation | 1. An embodiment of the MPAI-AIF Technical Specification, or 2. An AIW or AIM of a particular Level (1-2-3). |
| Internal Storage | A Component to store data of the individual AIMs. |
| Interoperability | The ability to functionally replace an AIM/AIW with another AIM/AIW having the same Interoperability Level |
| Interoperability Level | The attribute of an AIW and its AIMs to be executable in an AIF Implem­en­tati­on and to be:   1. Implementer-specific and satisfying the MPAI-AIF Standard *(Level 1)*. 2. Specified by an MPAI Application Standard (*Level 2)*. 3. Specified by an MPAI Application Standard and certified by a Performance Assessor (*Level 3)*. |
| Knowledge Base | Structured and/or unstructured information made accessible to AIMs via MPAI-specified interfaces |
| Message | A sequence of Records. |
| Normativity | The set of attributes of a technology or a set of technologies specified by the applicable parts of an MPAI standard. |
| Performance | The attribute of an Implementation of being Reliable, Robust, Fair and Replicable. |
| Performance Assessment | The normative document specifying the procedures, the tools, the data sets and/or the data set characteristics to Assess the Grade of Performance of an Implementation. |
| Performance Assessment Means | Procedures, tools, data sets and/or data set characteristics to Assess the Performance of an Implem­en­tation. |
| Performance Assessor | An entity authorised by MPAI to Assess the Performance of an Implementation in a given Application domain |
| Port | A physical or logical communication interface of an AIM. |
| Profile | A particular subset of the technologies used in MPAI-AIF or an AIW of an Application Standard and, where applicable, the classes, other subsets, options and parameters relevant to that subset. |
| Record | Data with a specified structure. |
| Reference Model | The AIMs and theirs Connections in an AIW. |
| Reference Software | A technically correct software implementation of a Technical Specific­ation containing source code, or source and compiled code. |
| Reliability | The attribute of an Implementation that performs as specified by the Application Standard, profile and version the Implementation refers to, e.g., within the application scope, stated limitations, and for the period of time specified by the Implementer. |
| Replicability | The attribute of an Implementation whose Performance, as Assessed by a Performance Assessor, can be replicated, within an agreed level, by another Performance Assessor. |
| Robustness | The attribute of an Implementation that copes with data outside of the stated application scope with an estimated degree of confidence. |
| Scope | The domain of applicability of an MPAI Application Standard |
| Service Provider | An entrepreneur who offers an Implementation as a service (e.g., a recommendation service) to Users. |
| Specification | A collection of normative clauses. |
| Standard | The ensemble of Technical Specification, Reference Software, Confor­man­ce Testing and Performance Assessment of an MPAI application Standard. |
| Technical Specification | (Framework) the normative specification of the AIF.  (Application) the normative specification of the set of AIWs belon­ging to an application domain along with the AIMs required to Im­plem­ent the AIWs that includes:   1. The formats of the Input/Output data of the AIWs implementing the AIWs. 2. The Connections of the AIMs of the AIW. 3. The formats of the Input/Output data of the AIMs belonging to the AIW. |
| Testing Laboratory | A laboratory accredited by MPAI to Assess the Grade of Performance of Implementations. |
| Time Base | The protocol specifying how Components can access timing information |
| Topology | The set of AIM Connections of an AIW. |
| Use Case | A particular instance of the Application domain target of an Application Standard. |
| User | A user of an Implementation. |
| User Agent | The Component interfacing the user with an AIF through the Controller |
| Version | A revision or extension of a Standard or of one of its elements. |
| Zero Trust | A cybersecurity model primarily focused on data and service protection that assumes no implicit trust. |

1. - Notices and Disclaimers Concerning MPAI Standards (Informative)

The notices and legal disclaimers given below shall be borne in mind when [downloading](https://www.mpai.community/resources/) and using approved MPAI Standards.

In the following, “Standard” means the collection of four MPAI-approved and [published](https://www.mpai.community/resources/) documents: “Technical Specification”, “Reference Software” and “Conformance Testing” and, where applicable, “Performance Testing”.

Life cycle of MPAI Standards

MPAI Standards are developed in accordance with the [MPAI Statutes](https://mpai.community/statutes/). An MPAI Standard may only be developed when a Framework Licence has been adopted. MPAI Standards are developed by especially established MPAI Development Committees who operate on the basis of consensus, as specified in Annex 1 of the [MPAI Statutes](https://mpai.community/statutes/). While the MPAI General Assembly and the Board of Directors administer the process of the said Annex 1, MPAI does not independently evaluate, test, or verify the accuracy of any of the information or the suitability of any of the technology choices made in its Standards.

MPAI Standards may be modified at any time by corrigenda or new editions. A new edition, however, may not necessarily replace an existing MPAI standard. Visit the [web page](https://mpai.community/resources/) to determine the status of any given published MPAI Standard.

Comments on MPAI Standards are welcome from any interested parties, whether MPAI members or not. Comments shall mandatorily include the name and the version of the MPAI Standard and, if applicable, the specific page or line the comment applies to. Comments should be sent to the [MPAI Secretariat](mailto:secretariat@mpai.community). Comments will be reviewed by the appropriate committee for their technical relevance. However, MPAI does not provide interpretation, consulting information, or advice on MPAI Standards. Interested parties are invited to join MPAI so that they can attend the relevant Development Committees.

Coverage and Applicability of MPAI Standards

MPAI makes no warranties or representations concerning its Standards, and expressly disclaims all warranties, expressed or implied, concerning any of its Standards, including but not limited to the warranties of merchantability, fitness for a particular purpose, non-infringement etc. MPAI Standards are supplied “AS IS”.

The existence of an MPAI Standard does not imply that there are no other ways to produce and distribute products and services in the scope of the Standard. Technical progress may render the technologies included in the MPAI Standard obsolete by the time the Standard is used, especially in a field as dynamic as AI. Therefore, those looking for standards in the Data Compression by Artificial Intelligence area should carefully assess the suitability of MPAI Standards for their needs.

IN NO EVENT SHALL MPAI BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO: THE NEED TO PROCURE SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE PUBLICATION, USE OF, OR RELIANCE UPON ANY STANDARD, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE AND REGARDLESS OF WHETHER SUCH DAMAGE WAS FORESEEABLE.

MPAI alerts users that practicing its Standards may infringe patents and other rights of third parties. Submitters of technologies to this standard have agreed to licence their Intellectual Property according to their respective Framework Licences.

Users of MPAI Standards should consider all applicable laws and regulations when using an MPAI Standard. The validity of Conformance Testing is strictly technical and refers to the correct implementation of the MPAI Standard. Moreover, positive Performance Assessment of an implementation applies exclusively in the context of the [MPAI Governance](https://mpai.community/governance/) and does not imply compliance with any regulatory requirements in the context of any jurisdiction. Therefore, it is the responsibility of the MPAI Standard implementer to observe or refer to the applicable regulatory requirements. By publishing an MPAI Standard, MPAI does not intend to promote actions that are not in compliance with applicable laws, and the Standard shall not be construed as doing so. In particular, users should evaluate MPAI Standards from the viewpoint of data privacy and data ownership in the context of their jurisdictions.

Implementers and users of MPAI Standards documents are responsible for determining and complying with all appropriate safety, security, environmental and health and all applicable laws and regulations.

Copyright

MPAI draft and approved standards, whether they are in the form of documents or as web pages or otherwise, are copyrighted by MPAI under Swiss and international copyright laws. MPAI Standards are made available and may be used for a wide variety of public and private uses, e.g., implementation, use and reference, in laws and regulations and standardisation. By making these documents available for these and other uses, however, MPAI does not waive any rights in copyright to its Standards. For inquiries regarding the copyright of MPAI standards, please contact the [MPAI Secretariat](mailto:secretariat@mpai.community).

The Reference Software of an MPAI Standard is released with the [MPAI Modified Berkeley Software Distribution licence](https://mpai.community/about/licence/). However, implementers should be aware that the Reference Software of an MPAI Standard may reference some third party software that may have a different licence.

1. – The Governance of the MPAI Ecosystem (Informative)

**Level 1 Interoperability**

With reference to *Figure 1*, MPAI issues and maintains a standard – called MPAI-AIF – whose components are:

1. An environment called AI Framework (AIF) running AI Workflows (AIW) composed of inter­connected AI Modules (AIM) exposing standard interfaces.
2. A distribution system of AIW and AIM Implementation called MPAI Store from which an AIF Implementation can download AIWs and AIMs.

|  |  |
| --- | --- |
| Implementers’ benefits | Upload to the MPAI Store and have globally distributed Implementations of   * AIFs conforming to MPAI-AIF. * AIWs and AIMs performing prop­rietary functions executable in AIF. |
| Users’ benefits | Rely on Implementations that have been tested for security. |
| MPAI Store | * Tests the Conformance of Implementations to MPAI-AIF. * Verifies Implementations’ security, e.g., absence of malware. * Indicates unambiguously that Implementations are Level 1. |

A Level 1 Implementation shall be an Implementation of the MPAI-AIF Technical Specification executing AIWs composed of AIMs able to call the MPAI-AIF APIs.

**Level 2 Interoperability**

In a Level 2 Implementation, the AIW must be an Implementation of an MPAI Use Case and the AIMs must conform with an MPAI Application Standard.

|  |  |
| --- | --- |
| Implementers’ benefits | Upload to the MPAI Store and have globally distributed Implementations of   * AIFs conforming to MPAI-AIF. * AIWs and AIMs conforming to MPAI Application Standards. |
| Users’ benefits | * Rely on Implementations of AIWs and AIMs whose Functions have been reviewed during standardisation. * Have a degree of Explainability of the AIW operation because the AIM Func­tions and the data Formats are known. |
| Market’s benefits | * Open AIW and AIM markets foster competition leading to better products. * Competition of AIW and AIM Implementations fosters AI innovation. |
| MPAI Store’s role | * Tests Conformance of Implementations with the relevant MPAI Standard. * Verifies Implementations’ security. * Indicates unambiguously that Implementations are Level 2. |

**Level 3 Interoperability**

MPAI does not generally set standards on how and with what data an AIM should be trained. This is an important differentiator that promotes competition leading to better solutions. However, the performance of an AIM is typically higher if the data used for training are in greater quantity and more in tune with the scope. Training data that have large variety and cover the spec­trum of all cases of interest in breadth and depth typically lead to Implementations of higher “quality”.

For Level 3, MPAI normatively specifies the process, the tools and the data or the characteristics of the data to be used to Assess the Grade of Performance of an AIM or an AIW.

|  |  |
| --- | --- |
| Implementers’ benefits | May claim their Implementations have passed Performance Assessment. |
| Users’ benefits | Get assurance that the Implementation being used performs correctly, e.g., it has been properly trained. |
| Market’s benefits | Implementations’ Performance Grades stimulate the development of more Performing AIM and AIW Implementations. |
| MPAI Store’s role | * Verifies the Implementations’ security * Indicates unambiguously that Implementations are Level 3. |

**The MPAI ecosystem**

The following is a high-level description of the MPAI ecosystem operation applicable to fully conforming MPAI implementations:

1. MPAI establishes and controls the not-for-profit MPAI Store (step 1).
2. MPAI appoints Performance Assessors (step 2).
3. MPAI publishes Standards (step 3).
4. Implementers submit Implementations to Performance Assessors (step 4).
5. If the Implementation Performance is acceptable, Performance Assessors inform Implementers (step 5a) and MPAI Store (step 5b).
6. Implementers submit Implementations to the MPAI Store (step 6); The Store Tests Confor­mance and security of the Implementation.
7. Users download Implementations (step 7).

Text

Description automatically generated with low confidence

*Figure 10 – The MPAI ecosystem operation*

The Ecosystem operation allows for AIW and AIF Implementations to be:

1. Proprietary: security is verified and Conformance to MPAI-AIF Tested (Level 1).
2. Conforming to an MPAI Application Standard: security is verified and Conformance to the relevant MPAI Application Standard Tested (Level 2).
3. Assessed to be Reliable, Robust, Fair and Replicable (Level 3).

and have their Interoperability Level duly displayed in the MPAI Store.

1. – Datasets for CAV research

**nuScenes**

The nuScenes dataset (https://nuscenes.org/) is a large-scale autonomous driving dataset with 3d object annotations. It features:

* Full sensor suite (1x LIDAR, 5x RADAR, 6x camera, IMU, GPS)
* 1000 scenes of 20s each
* 1,400,000 camera images
* 390,000 lidar sweeps
* Two diverse cities: Boston and Singapore
* Left versus right hand traffic
* Detailed map information
* 1.4M 3D bounding boxes manually annotated for 23 object classes
* Attributes such as visibility, activity and pose
* New: 1.1B lidar points manually annotated for 32 classes
* New: Explore nuScenes on SiaSearch
* Free to use for non-commercial use

For a commercial license contact nuScenes@motional.com

nuImages is a large-scale autonomous driving dataset with image-level 2d annotations. It features:

* 93k video clips of 6s each (150h of driving)
* 93k annotated and 1.1M un-annotated images
* Two diverse cities: Boston and Singapore
* The same proven sensor suite as in nuScenes
* Images mined for diversity
* 800k annotated foreground objects with 2d bounding boxes and instance masks
* 100k 2d semantic segmentation masks for background classes
* Attributes such as rider, pose, activity, emergency lights and flying
* Free to use for non-commercial use

**Road Hazard data**

Otonomo real-time Road Hazard data from connected passenger vehicles powers diverse road safety use cases, including mapping, accident predictions, smart cities and many more. The Otonomo Vehicle Data Platform secures, cleanses and normalizes the hazard data to make it more valuable and accessible for diverse use cases.

https://info.otonomo.io/hazard-data-datasheet-lp

1. – ETSI Technical Report

ETSI specifies the Collective Perception Service (CPS) in its Technical Report [23]. The CPS includes the format and generation rules of the Collective Perception Message (CPM).

The CPM message format is (H=header, C=container, M=mandatory, O=optional).

*Table 17 – ETSI Collective Perception Message format*

|  |  |  |  |
| --- | --- | --- | --- |
| PDU header | H | M | protocol version, message ID and Station ID. |
| Management | C | M | transmitter type (e.g., vehicle or RSU) and position. |
| Station Data | C | O | transmitter heading, velocity, or acceleration etc. |
| Sensor Information | C | O | transmitter (e.g., speed, heading, or acceleration)  capabilities of the vehicle’s sensors. |
| Perceived Object | C | O | detected objects (e.g., distance, speed and dimensions)  time at which the measurements were done.  A CPM can report up to 128 detected objects |
| Free Space Addendum | C | O | free space areas/volume within the sensor detection areas |

Every 0.1s a CPM is generated if one of the 3 conditions is satisfied

no CPM has been generated in the last 1s

a new object has been detected

since last CPM sending info about a previously detected object (it must have an ID)

the following attributes have changed:

Absolute position ΔP > 4 m

Absolute speed ΔV > 0.5 m/s

more than 1s has passed (ΔT > 1 s).

ETSI makes use of a common coordinate system. A vehicle can communicate its absolute coordinates roll, pitch and yaw (Attitude).

Different CPM generation rules have been investigated [24].

1. – Some CAV Communication Technologies

The following categories of vehicular communication are part of the literature or industry effort:

|  |  |  |
| --- | --- | --- |
| V2V | Vehicle-to-Vehicle | communication between vehicles to exchange information about the speed and position of surrounding vehicles |
| V2I | Vehicle-to-Infrastructure | communication between vehicles and road infrastructure. |
| V2X | Vehicle-to-Everything | communication between a vehicle and any entity that may affect, or may be affected by, the vehicle |
| V2R | Vehicle-to-Roadside | communication between a vehicle and Road Side Units (RSUs). |
| V2P | Vehicle-to-Pedestrian | communications between a vehicle and (multiple) pedestrian device(s) and to other vulnerable road users, e.g., cyclists, in close proximity |
| V2S | Vehicle-to-Sensors | communication between a vehicle and its sensors on board |
| V2D | Vehicle-to-Device | communication between a vehicle and any electronic device that may be connected to the vehicle itself |
| V2G | Vehicle-to-Grid | communication with the power grid to sell demand response services by either returning electricity to the grid or by throttling their charging rate |
| V2N | Vehicle-to-Network | broadcast and unicast communications between vehicles and the V2X management system and also the V2X AS (Application Server) |
| V2C | Vehicle-to-Cloud | communication with data centres and other devices connected to the internet |

Technologies exist that support at least some aspects of the communication types of the table:

Radio access, e.g., visible light communication, mmWave, Cellular-V2X, and 5G

Radio resource management (RRM) for vehicular communication using cellular technology

3GPP Release 14: air interfaces and core network technologies to support V2X communic­ation.

Vehicular ad hoc network (VANET)

Dedicated Short-Range Communication (DSRC): 5.9 GHz band with a range of ~300 metres.

Software defined vehicular networks (SDVN)

Internet of vehicles (IoV)

Protocol stack of the intelligent transportation system (ITS)

Cooperative Awareness Messages (CAMs) messages related to the status of CAV’s sent via wireless broadcast in VANETs.

Cooperative or collective perception improve CAV’s perception beyond the sensors’ detection range.

Traffic situation can be extracted from Local dynamic map (LDM) that aggregates CAMs.

1. Traffic rules are those deriving from the semantic of the objects, e.g., speed limit 50 km/h, not on the reasoning made on the objects which this document assumes it is within the scope of the Autonomous Motion Subsystem. [↑](#footnote-ref-1)