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

|  |  |
| --- | --- |
| **N1712** | 2024/04/17 |
| **Source** | Requirements (CAV) |
| **Title** | Use Cases and Functional Requirements: MPAI-CAV – Technologies WD0.7 |
| **Target** | MPAI Members |

Contents

[1 Introduction 3](#_Toc163042474)

[2 Terms and definitions 3](#_Toc163042475)

[3 References 9](#_Toc163042476)

[3.1 Normative References 9](#_Toc163042477)

[3.2 Informative References 10](#_Toc163042478)

[4 Model of Connected Autonomous Vehicle 12](#_Toc163042479)

[4.1 Functions of Connected Autonomous Vehicle 12](#_Toc163042480)

[4.2 Reference Model of Connected Autonomous Vehicle 13](#_Toc163042481)

[4.3 I/O Data of Connected Autonomous Vehicle 14](#_Toc163042482)

[5 Human-CAV Interaction (HCI) 15](#_Toc163042483)

[5.1 I/O Data of Human-CAV Interaction 15](#_Toc163042484)

[5.2 Reference Model of Human-CAV Interaction 16](#_Toc163042485)

[5.3 Functions of Human-CAV Interaction’s AI Modules 16](#_Toc163042486)

[5.4 I/O Data of Human-CAV Interaction’s AI Modules 16](#_Toc163042487)

[5.5 I/O Data Types 17](#_Toc163042488)

[5.5.1 Personal Preferences 17](#_Toc163042489)

[5.5.2 Input Audio 18](#_Toc163042490)

[5.5.3 Input LiDAR 18](#_Toc163042491)

[5.5.4 Input Text 18](#_Toc163042492)

[5.5.5 Input Visual 18](#_Toc163042493)

[5.5.6 Point of View 18](#_Toc163042494)

[5.5.7 Audio-Visual Scene Descriptors 19](#_Toc163042495)

[5.5.8 Audio-Visual Scene Geometry 20](#_Toc163042496)

[5.5.9 Avatar Model 20](#_Toc163042497)

[5.5.10 Body Descriptors 20](#_Toc163042498)

[5.5.11 Face Descriptors 20](#_Toc163042499)

[5.5.12 Face ID 21](#_Toc163042500)

[5.5.13 Speaker ID 21](#_Toc163042501)

[5.5.14 Meaning 21](#_Toc163042502)

[5.5.15 Visual Object ID 21](#_Toc163042503)

[5.5.16 Personal Status 22](#_Toc163042504)

[5.5.17 Speech Model 22](#_Toc163042505)

[5.5.18 Output Audio 22](#_Toc163042506)

[5.5.19 Output Visual 22](#_Toc163042507)

[5.5.20 HCI-AMS Messages 22](#_Toc163042508)

[5.5.21 E-HCI to/from R-HCI messages 23](#_Toc163042509)

[5.5.22 E-AMS to/from R-AMS messages 23](#_Toc163042510)

[6 Environment Sensing Subsystem (ESS) 23](#_Toc163042511)

[6.1 Functions of Environment Sensing Subsystem 23](#_Toc163042512)

[6.2 Reference Architecture of Environment Sensing Subsystem 23](#_Toc163042513)

[6.3 I/O Data of Environment Sensing Subsystem 24](#_Toc163042514)

[6.4 Functions of Environment Sensing Subsystem’s AI Modules 25](#_Toc163042515)

[6.5 I/O Data of Environment Sensing Subsystem’s AI Modules 26](#_Toc163042516)

[6.6 Data functional requirements 27](#_Toc163042517)

[6.6.1 Sensor data 27](#_Toc163042518)

[6.6.2 Scene Descriptors 29](#_Toc163042519)

[6.6.3 Road Topology 32](#_Toc163042520)

[6.6.4 Basic Environment Representation 32](#_Toc163042521)

[7 Autonomous Motion Subsystem (AMS) 33](#_Toc163042522)

[7.1 Functions of Autonomous Motion Subsystem 33](#_Toc163042523)

[7.2 Reference Architecture of Autonomous Motion Subsystem 34](#_Toc163042524)

[7.3 I/O Data of Autonomous Motion Subsystem 35](#_Toc163042525)

[7.4 Functions of Autonomous Motion Subsystem’s AI Modules 35](#_Toc163042526)

[7.5 I/O Data of Autonomous Motion Subsystem’s AI Modules 36](#_Toc163042527)

[7.6 Data Types 36](#_Toc163042528)

[7.6.1 CAV Identifier 36](#_Toc163042529)

[7.6.2 Basic Environment Representation 37](#_Toc163042530)

[7.6.3 Full Environment Representation 37](#_Toc163042531)

[7.6.4 Offline map 37](#_Toc163042532)

[7.6.5 Route 37](#_Toc163042533)

[7.6.6 Traffic Signal Configuration 38](#_Toc163042534)

[7.6.7 Path 38](#_Toc163042535)

[7.6.8 Pose 38](#_Toc163042536)

[7.6.9 Trajectory 38](#_Toc163042537)

[7.6.10 Goal 38](#_Toc163042538)

[7.6.11 AMS-MAS Message 38](#_Toc163042539)

[7.6.12 MAS-AMS Response 39](#_Toc163042540)

[7.6.13 Road State 39](#_Toc163042541)

[8 Motion Actuation Subsystem (MAS) 39](#_Toc163042542)

[8.1 Functions of Motion Actuation Subsystem 39](#_Toc163042543)

[8.2 Reference Architecture of Motion Actuation Subsystem 40](#_Toc163042544)

[8.3 I/O Data of Motion Actuation Subsystem 40](#_Toc163042545)

[8.4 Functions of Motion Actuation Subsystem’s AI Modules 41](#_Toc163042546)

[8.5 I/O Data of Motion Actuation Subsystem’s AI Modules 41](#_Toc163042547)

[8.6 Data Types 41](#_Toc163042548)

[8.6.1 Odometer Data 41](#_Toc163042549)

[8.6.2 Speedometer Data 41](#_Toc163042550)

[8.6.3 Accelerometer Data 42](#_Toc163042551)

[8.6.4 Spatial Attitude 42](#_Toc163042552)

[8.6.5 Other Environment Data 42](#_Toc163042553)

[8.6.6 AMS-MAS Message 42](#_Toc163042554)

[8.6.7 MAS-AMS Message 42](#_Toc163042555)

[8.6.8 Messages between MAS and its subsystem 42](#_Toc163042556)

[9 CAV-to-Everything (V2X) (Informative) 42](#_Toc163042557)

[9.1 V2X Communication 43](#_Toc163042558)

[9.2 Input and output data 43](#_Toc163042559)

[9.2.1 CAVs within range 43](#_Toc163042560)

[9.2.2 CAV-aware equipment 44](#_Toc163042561)

[9.2.3 Other non-CAV vehicles 44](#_Toc163042562)

[9.2.4 Pedestrians 44](#_Toc163042563)

[Annex 1 - MPAI-wide terms and definitions 45](#_Toc163042564)

[Annex 2 - Notices and Disclaimers Concerning MPAI Standards (Informative) 48](#_Toc163042565)

# Foreword (Informative)

In recent years, Artificial Intelligence (AI) and related technologies have been applied to a broad range of applications, have started affecting the life of millions of people and are expected to do so even more in the future. As digital media standards have positively influenced industry and billions of people, so AI-based data coding standards are expected to have a similar positive impact. Moving Picture, Audio, and Data Coding by Artificial Intelligence (MPAI) has been established to develop standards that promote the efficient use of data especially using Artificial Intelligence technologies.

To establish a framework for the pursuit of its mission. MPAI has developed *Technical Specification: Governance of the MPAI Ecosystem* [1]. This defines the following elements: Standards, i.e., the ensemble of Technical Specifications, Reference Software, Conformance Testing, and Performance Assessment; Implementers of MPAI Technical Specifications; the MPAI Store in charge of making AIMs and AIWs submitted by Implementers available to Integrators and End-Users; Performance Assessors, independent entities assessing the performance of implementations in terms of Reliability, Replicability, Robustness, and Fairness; and End Users.

To facilitate the development of smaller-scale high-performance components and the availability of solution with improved explainability, MPAI has developed another foundational *Technical Specification: Artificial Intelligence Framework (MPAI-AIF)* [2]. This specifies an environment enabling initialisation, dynamic configuration, and control of AIWs in the standard AI Framework environment depicted in Figure 1.

A diagram of a computer server

Description automatically generated

Figure 1 - The AI Framework (MPAI-AIF) V2 Reference Model

An AI Framework can execute AI applications called AI Workflows (AIW). An AI Workflow can include interconnected AI Modules (AIM). AI Modules can be Composite if they include interconnected AIMs or Basic if they are not Composite.

# Introduction

MPAI has published a Technical Specification on Connected Autonomous Vehicle (MPAI-CAV) – Architecture that envisages the CAV composed of four Subsystem each of which is specified as an AI Workflow (AIW) composed of AI Modules (AIM) executed in an AI Framework [1] but does not specify the functional requirements that such Data should satisfy.

The purpose of this document is to collect and identify such functional requirements. They are intended to be published together with the MPAI-CAV Framework Licence and a Call for Technologies.

By developing some other MPAI Technical Specifications, namely MPAI-CAE [2], MPAI-MMC [3], MPAI-OSD [4], and MPAI-PAF [5], MPAI has been able to specify technologies that MPAI believes may be used – directly, with extensions, or with modifications – to develop the planned *Technical Specification: Connected Autonomous Vehicles (MPAI-CAV) – Technologies*. However, MPAI welcomes motivated proposals of alternative technologies for use in the MPAI-CAV Technical Specification.

# Terms and definitions

Table 1 and Table 16 define the CAV-specific Terms and the general MPAI Terms, respectively.

Table 1 - Terms and Definitions

Note A dash “-” preceding a Term in this Table means the following:

1. If the font is normal, the Term in the table without a dash and preceding the one with a dash should be placed before that Term. The notation is used to concentrate in one place all the Terms that are composed of, e.g., the word Audio followed by one of the words Object, Scene, and Scene Descriptors.
2. If the font is *italic,* the Term in the table without a dash and preceding the one with a dash should be placed after that Term. The notation is used to concentrate in one place all the Terms that are composed of, e.g., the word Attitude preceded by one of the words Social or Spatial.

|  |  |
| --- | --- |
| **Term** | **Definition** |
| Accelerometer Data | A Data Type representing the acceleration forces acting on a CAV. |
| Acceleration |  |
| * Coordinate | The acceleration measured in a coordinate system. |
| * Proper | The physical acceleration, i.e., measured by an accelerometer experienced by an object. |
| Accuracy | An estimate of how well the measurement of a physical entity corresponds to the actual value of that entity. |
| Alert | A Data Type representing environment-related elements that should be treated with priority by the Obstacle Avoider AIM. |
| AMS-MAS Message | A Data Type representing the command issued by the Autonomous Motion Subsystem instructing the Motion Actuation Subsystem to change the Ego CAV’s Spatial Attitude SAA at time tA to Spatial Attitude SAB at time tB. |
| AMS-HCI Message | A Data Type representing the response of the Motion Actuation Subsystem during and after the execution of an HCI-AMS Message. |
| Attitude |  |
| * *Social* | A Data Type representing an element of the internal status of a human or avatar related to the way they intend to position themselves vis-à-vis the context, e.g., “Confrontational”, “Respectful”. |
| * *Spatial* | A Data Type representing the CAV’s Position, Orientations and their velocities and accelerations at a given time. |
| Audio | A Data Type representing analogue signals generated by information perceptible by a human using the sense of sound. |
| * Data | A Data Type representing the output of a microphone array capturing the target environment to create the Audio Scene Description used to incorporate Environment Audio information in the Basic and Full Environment Representation. |
| * Object | A Data Type representing an audio object and associated metadata. |
| * Scene | A Data Type representing an audio scene. |
| * Scene Descriptors | A Data Type representing the Audio Data and produced by the Audio Scene Description AIM using previous Basic Environment Representations. |
| Audio-Visual | A Data Type representing analogue signals generated by information perceptible by a human using the sense of sound and sight. |
| * Object | A Data Type representing Audio-Visual information and associated metadata. |
| * Scene | (AV Scene) A Data Type representing an audio-visual scene. |
| * Scene Descriptors | A Data Type representing the outdoor and indoor Audio-Visual Scene in terms of Audio-Visual Objects that includes the following features:   1. Having a common time-base. 2. Associating co-located audio and visual objects if both are present. 3. Supporting the physical displacement and interrelation (e.g., occlusion) of Audio and Visual Objects over time. |
| Avatar | A Data Type representing a real or fictitious person in a virtual space rendered to a physical space. |
| * Model | A Data Type representing the model of a human that a User selects to impersonate the CAV’s HCI as rendered by the Personal Status Display AIM. |
| Basic Environment Representation (BER) | A Data Type representing Environment using information provided by a variety of sensors and including:   1. The Audio-Visual Scene Geometry. 2. The Ego CAV’s Spatial Attitude. 3. The Scene Description produced from the available Environment Sensing Technologies. 4. The Road Topology. 5. Other Environment Data. |
| Body Descriptors | A Data Type representing the motion and conveying information on the Personal Status of the body of a human or an avatar. |
| Brake |  |
| * Command | A Data Type representing the command issued to the Brakes by the Motion Actuation Subsystem after interpreting an AMS-MAS Message. |
| * Response | A Data Type representing the Brakes’ response to the AMS Command Interpreter in response to a Brake Command. |
| CAV | See Connected Autonomous Vehicle. |
| * Aware | An attribute of physical entities possessing some of the sensing and communication capabilities of a CAV without being a CAV, e.g., Roadside Units and Traffic Lights. |
| * Centre | The point in the Ego CAV selected to have coordinates (0,0,0). |
| * *Ego* | The Object in the Environment that the CAV recognises as itself. |
| * Identifier | A Data Type uniquely identifying a CAV and carrying information, such as:   1. Country code where the CAV has been registered. 2. Registration number in that country. 3. CAV manufacturer identifier. 4. CAV model identifier. |
| Cognitive State | A Data Type representing the internal status of a human or avatar reflecting their understanding of the context, such as “Confused” or “Dubious” or “Convinced”. |
| Connected Autonomous Vehicle | (CAV) The information technology-related components of a vehicle enabling it to autonomously reach a Pose by:   1. Understanding human utterances in the Subsystem (HCI). 2. Planning a Route (AMS). 3. Sensing and building a Representations of the external Environment (ESS). 4. Exchanging such Representations and other Data with other CAVs and CAV-aware entities (AMS). 5. Making decisions about how to execute the Route (AMS). 6. Acting on the MAS (AMS). |
| Data | Information in digital form. |
| * Format | The standard digital representation of Data. |
| * Semantics | The meaning of Data. |
| * Shape | A Data Type representing the volume occupied by a CAV. |
| * Type | An instance of Data with a specific Data Format. |
| Decision Horizon | The time interval within which a decision is planned to be implemented. |
| Descriptor | Coded representation of a text, audio, speech, or visual feature. |
| Digital Representation | Data corresponding to and representing a physical entity. |
| Emotion | A Data Type representing the internal status of a human or avatar resulting from their interaction with the context or subsets of it, such as “Angry”, and “Sad”. |
| Environment | The Digital Representation of the portion of the external world of current interest to a CAV. |
| * Representation | A Data Type representing the Environment such as produced by an Environment Sensing Technology or derived from the Basic or Full Environment Representation. |
| * Sensing Technology | (EST) One of the technologies used to sense the environment by the Environment Sensing Subsystem, e.g., RADAR, Lidar, Visual, Ultrasound, and Audio. The Offline Map is considered as an EST. |
| Face | The portion of a 2D or 3D Digital Representation corresponding to the face of a human. |
| * Descriptors | A Data Type representing the motion and conveying information on the Personal Status of the face of a human or an avatar. |
| * ID | The Identifier of a human belonging to a group of humans inferred from analysing the face of the human. |
| Factor | One of Cognitive State, Emotion, and Social Attitude |
| Full Environment Representation | (FER) A Data Type representing the Environment that extends the Basic Environment Representations of the Ego CAV with elements of those of other CAVs in range or Roadside Units. |
| * Audio | The ready-to-render Audio component of the Full Environment Representation Viewer. |
| * Commands | A Data Type representing the commands issued by a CAV passenger to the HCI to enable navigation in the Full Environment Representation, e.g., select a Point of View, zoom in/out, control sound level. |
| * Visual | The ready-to-render Visual component of the Full Environment Representation Viewer. |
| Gesture | A Data Type representing the movement of the body or part of it, such as the head, arm, hand, and finger, often a complement to a vocal utterance. |
| Global Navigation Satellite System (GNSS) | One of the systems providing global navigation information such as GPS, Galileo, Glonass, BeiDou, Quasi Zenith Satellite System (QZSS) and Indian Regional Navigation Satellite System (IRNSS). |
| Goal | The Spatial Attitude planned to be reached at the end of a Decision Horizon. |
| HCI-AMS Message | A Data Type representing high-level instructions issued by HCI to AMS to request it to reach a destination. |
| Identifier | A Label that is uniquely associated with an Object. |
| Inertial Measurement Unit | An inertial positioning device, e.g., odometer, accelerometer, speedometer, gyroscope etc. |
| Instance ID | An Instance of a class of Objects and the Group of Objects the Instance belongs to. |
| Latitude | A Data Type representing the angular distance of a place north of the earth's equator measured in degrees [4]. |
| LiDAR | A Data Type representing analogue signals generated by information captured by an active time-of-flight sensor operating in the µm range – ultraviolet, visible, or near infrared light (900 to 1550 nm). |
| * Data | A Data Type representing Data captured by a LiDAR sensor. |
| * Scene Descriptors | A Data Type representing the LiDAR Data and produced by the LiDAR Scene Description AIM using previous Basic Environment Representations. |
| Longitude | A Data Type representing the angular distance of a place west of the Greenwich meridian measured in degrees [4]. |
| Machine | Synonym with Human-CAV Interaction Subsystem. |
| * Avatar | The ready-to-render Avatar produced by the Personal Status Display. |
| * Speech | The ready-to-render synthetic speech produced by the Personal Status Display. |
| Map Scene Descriptors | A Data Type representing the Offline Map Data complemented by the previous Basic Environment Representations in the common Scene Description Format. |
| MAS Subsystems | The Brakes, the Steering Wheel, and the Wheel Motors of a CAV. |
| MAS-AMS Message | A Data Type representing the Messageof AMS Command Interpreter integrating the Responses from Brakes, Steering Wheel, and Wheel Motors. The MAS-AMS Message contains the value of the Spatial Attitudes of the Ego CAVat intermediate Poses with the corresponding Times. |
| Meaning | A Data Type representing an input text such as syntactic and semantic information. |
| Microphone Array Geometry | A Data Type representing the position of each microphone comprising a microphone array and characteristics such as microphone type, look directions, and array type. |
| Modality | One of Text, Speech, Face, or Gesture. |
| Model | A Data Type representing:   1. A ready-to-be animated object with its features, or 2. An Artificial Neural Network. |
| Object | A Data Type representing an object sensed by an EST and produced by an EST-specific Scene Description including the following elements:   1. Timestamp. 2. Identifier of the Scene Description AIM that has generated the Object. 3. Alerts. 4. Spatial Attitude of the Object and its estimated accuracy referred to the CAV Centre. 5. Bounding box. 6. Object type (2D, 2.5D, and 3D). |
| * ID | The Identifier uniquely associated with a particular class of Objects, e.g., hammer, screwdriver, etc. |
| Odometer Data | A Data Type representing the distance from the start up to the current Pose measured by the number of wheel rotations times the tire circumference (π x tire diameter). |
| Offline |  |
| * Map | A previously created digital map of an Environment and associated metadata. |
| * Map Data | A Data Type representing the Data provided by an Offline Map in response to a given set of coordinate values. |
| Orientation | A Data Type representing the set of the 3 roll, pitch, yaw angles indicating the rotation around the axes of a CAV as obtained by processing the data from the CAV sensors:   * Roll: around the principal axis X. * Pitch: around the Y axis. Y has an angle of 90˚ counterclockwise (right-to-left) with the x axis. * Yaw: around the Z axis perpendicular to and out of the ground. |
| Other Environment Data | Data Types representing information acquired by the Motion Actuation Subsystem and complementing the spatial data such as weather, temperature, air pressure, humidity, ice and water on the road, wind, fog etc. |
| Path | A Data Type representing a sequence of Poses. |
| Personal Status | A Data Type representing the ensemble of information internal to a person expressed by 3 Factors (Cognitive State, Emotion, Social Attitude) conveyed by one or more Modalities (Text, Speech, Face, and Gesture). |
| Point of View | The Spatial Attitude of a user looking at the Environment. |
| Portable Avatar | A Data Type providing context information and allowing the rendering of an Avatar as intended by the Portable Avatar producer. |
| Pose | A Data Type representing Position and Orientation of the CAV as obtained by processing the data from the CAV sensors. |
| Position | A Data Type representing the current coordinates of a CAV as obtained by processing the data from the CAV sensors. |
| RADAR | A Data Type representing analogue signals generated by information captured by an active time-of-flight sensor operating in the operating in the 24-81 GHz range. |
| * Data | A Data Type representing the Visual Data captured by a RADAR sensor. |
| * Scene Descriptors | A Data Type representing the Visual Data captured by RADAR and produced by the RADAR Scene Description AIM using previous Basic Environment Representations. |
| Remote AMS | AMS of a CAV or CAV-Aware entity in range. |
| Remote HCI | HCI of a CAV or CAV-aware entity in range. |
| Road |  |
| * Geometry | A Data Type representing the positioning of the physical elements of the roadway, e.g., traffic poles, road Signs, traffic tights, etc. |
| * State | A Data Type representing the state of the road the CAV is traversing such as detours and road conditions. Road State may be inferred by the AMS from internally available information or received from an external source via a communication channel. |
| * Topology | The geometric properties and spatial relationships of objects in traffic. |
| Roadside Unit | A wireless communicating device located on the roadside providing information to CAVs in range. |
| Route | A Data Type representing a sequence of Waypoints. |
| Scene Descriptors | A Data Type representing the combination of EST-specific 2D, 2.5D, or 3D Scene Descriptors used by an EST Scene Description in an EST-specific time window. |
| Speaker ID | The Identifier of a human belonging to a group of humans inferred from analysing the speech of the human. |
| Speech | A Data Type representing analogue speech sampled at a frequency between 8 kHz and 96 kHz with 8, 16 and 24 bits/sample, and linear quantisation. |
| Speech Model | A Data Type representing the Speech Descriptors characteristic of a speaker used to generate the Personal Status Display’s synthetic speech. |
| Speedometer Data | A Data Type representing the speed of a CAV as measured by the sensor. |
| Steering Wheel |  |
| * Command | A Data Type representing the command issued to the Steering Wheel by the Motion Actuation Subsystem after interpreting an AMS-MAS Message. |
| * Feedback | A Data Type representing the Steering Wheel’s response to the AMS Command Interpreter in response to a Steering Direction Command. |
| Subsystem | A high-level subdivision of a CAV. |
| Text | A series of characters drawn from a finite alphabet represented using a Character Set. |
| * *Recognised* | The Text produced by a speech recogniser. |
| * *Refined* | Text resulting from the Language Understanding AIM’s refinement of Recognised Text produced by a Speech Recognition AIM. |
| Traffic |  |
| * Rules | A Data Type representing the traffic rules applying to an Environment as extracted from the local Traffic Signals based on the local traffic rules. |
| * Signals | A Data Type representing the traffic signals on a road and around it, their Spatial Attributes, and the semantics of the traffic signals. |
| Trajectory | A Data Type representing a sequence of Spatial Attitudes and the expected time each Spatial Attitude will be reached. |
| Ultrasound | A Data Type representing analogue signals generated by information captured by an ultrasonic sensor, an active time-of-flight sensor typically operating in the 40 kHz to 250 kHz range. |
| * Data | A Data Type representing the Data provided by an ultrasonic sensor. |
| * Scene Descriptors | A Data Type representing the Visual Data captured by Ultrasound and produced by the Ultrasound Scene Description AIM using previous Basic Environment Representations. |
| Visual | A Data Type representing analogue signals generated by information perceptible by a human using the sense of sight. |
| * Data | A Data Type representing the Visual Data captured by a Visual sensor. |
| * Object | A Data Type representing a visual object and associated metadata. |
| * Scene | A Data Type representing a visual scene. |
| * Scene Descriptors | A Data Type representing the Visual Objects of an Environment, their Spatial Attitudes, and associated Visual Scene Geometry. |
| Waypoint | The coordinates of a point on an Offline Map. |
| Wheel Motor |  |
| * Command | A Data Type representing the command issued to the Wheel Motors by the Motion Actuation Subsystem after interpreting an AMS-MAS Message. |
| * Response | A Data Type representing the Wheel Motors’ response to AMS Command Interpreter in response to a Wheel Motor Command. |

# References

## Normative References

This document references the following normative documents:

1. Technical Specification: AI Framework (MPAI-AIF) V1.1; https://mpai.community/standards/mpai-aif/.
2. Technical Specification: Context-based Audio Enhancement (MPAI-CAE) VI.4, https://mpai.community/standards/mpai-cae/.
3. Technical Specification: Technical Specification: Multimodal Conversation (MPAI-MMC) V1.2; https://mpai.community/standards/mpai-mmc/.
4. Technical Specification: Object and Scene Description (MPAI-OSD) V1.0, https://mpai.community/standards/mpai-osd/.
5. Technical Specification: Portable Avatar Format (MPAI-PAF) V1.1, https://mpai.community/standards/mpai-cae/.
6. Universal Coded Character Set (UCS): ISO/IEC 10646; December 2020
7. ISO/IEC 14496-10; Information technology – Coding of audio-visual objects – Part 10: Advanced Video Coding.
8. ISO/IEC 23008-2; Information technology – High efficiency coding and media delivery in heterogeneous environments – Part 2: High Efficiency Video Coding.
9. ISO/IEC 23094-1; Information technology – General video coding – Part 1: Essential Video Coding.
10. ISO/IEC 19774-1:2019; Information technology - Computer graphics, image processing and environmental data representation, Part 1: Humanoid animation (HAnim) architecture; https://www.web3d.org/documents/specifications/19774-1/V2.0/index.html

## Informative References

This document references the following informative documents:

1. Technical Specification: The Governance of the MPAI Ecosystem (MPAI-GME) V1.1; https://mpai.community/standards/mpai-gme/.
2. https://www.etsi.org/deliver/etsi\_en/302800\_302899/30289002/02.01.01\_20/en\_30289002v020101a.pdfNavigation Data Standards, https://nds-association.org/
3. SharedStreets; <https://sharedstreets.io/>
4. 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
5. 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.
6. P. Avgeriou, "Describing, Instantiating and Evaluating a Reference Architecture: A Case Study," Enterprise Architecture Journal, June 2003.
7. S. Behere and M. Törngren, "A functional reference architecture for autonomous driving," Information and Software Technology 73, 2016, 136-150.
8. 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.
9. K. Jo, M. Lee, D. Kim, J. Kim, C. Jang, E. Kim, and M. Sunwoo, "Overall Reviews of Autonomous Vehicle A1-System
10. Architecture and Algorithms," The International Federation of Automatic Control, 46(10), 2013, 114-119.
11. 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
12. 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
13. 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
14. ISO 8855: "Road vehicles -- Vehicle dynamics and road-holding ability -- Vocabulary"
15. Rodolfo W. L. Coutinho and Azzedine Boukerche, Guidelines for the Design of Vehicular Cloud Infrastructures for Connected Autonomous Vehicles, IEEE Wireless Communications - August 2019
16. 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
17. 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
18. ETSI TR 103 562 V2.1.1 (2019-12), Analysis of the Collective Perception Service (CPS); Release 2.
19. 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
20. CAR 2 CAR Communication Consortium, https://www.car-2-car.org/
21. 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
22. 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
23. https://phantom.ai/assets/uploads/PAI%20Renesas%20Partnership%20Announcement%20(1).pdf
24. 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>
25. 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
26. LAS (LASer) File Format, Version 1.4 ,  
     https://www.loc.gov/preservation/digital/formats/fdd/fdd000418.shtml
27. https://www.lesliesikos.com/pcap/
28. https://pointclouds.org/documentation/tutorials/hdl\_grabber.html
29. https://www.mathworks.com/help/vision/ref/velodynefilereader.html
30. 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
31. https://pro.arcgis.com/en/pro-app/latest/help/data/data-interoperability/supported-formats-with-the-data-interoperability-extension.htm
32. 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>
33. Matthias Schreier; Environment Representations for Automated On-Road Vehicles; https://www.researchgate.net/publication/323105152\_Environment\_Representations\_for\_Automated\_On-Road\_Vehicles
34. Toh Chai Keong, Juan Carlos Cano, Carlos-Javier Fernandez-Laguia, Pietro Manzoni; Wireless Digital Traffic Signs of the Future; IET Networks, September 2018
35. 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
36. Bojarski, Mariusz, et al. “End to end learning for self-driving cars.” arXiv preprint arXiv:1604.07316 (2016).
37. 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>
38. Integrated Public Alert & Warning System; <https://www.fema.gov/gu/emergency-managers/practitioners/integrated-public-alert-warning-system/public>
39. OASIS Standard; Common Alerting Protocol Version 1.2; <https://docs.oasis-open.org/emergency/cap/v1.2/CAP-v1.2-os.pdf>
40. 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>
41. <https://youtu.be/nCaomvjPIok>
42. Stefano Mariani and Franco Zambonelli; Degrees of Autonomy in Coordinating Collectives of Self-Driving Vehicles; ISoLA 2021
43. 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/>
44. 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.
45. 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.
46. 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>
47. Maximilian Zipfl, J. Marius Zöllner; Towards Traffic Scene Description: The Semantic Scene Graph; https://arxiv.org/pdf/2111.10196.pdf

# Model of Connected Autonomous Vehicle

## Functions of Connected Autonomous Vehicle

A Connected Autonomous Vehicle is defined as a physical system that:

1. Converses with humans by understanding their utterances, e.g., a request to be taken to a destination.
2. Senses the environment where it is located or traverses like the one depicted in *Figure 1*.

A picture containing icon

Description automatically generated

*Figure 1 - An environment of CAV operation*

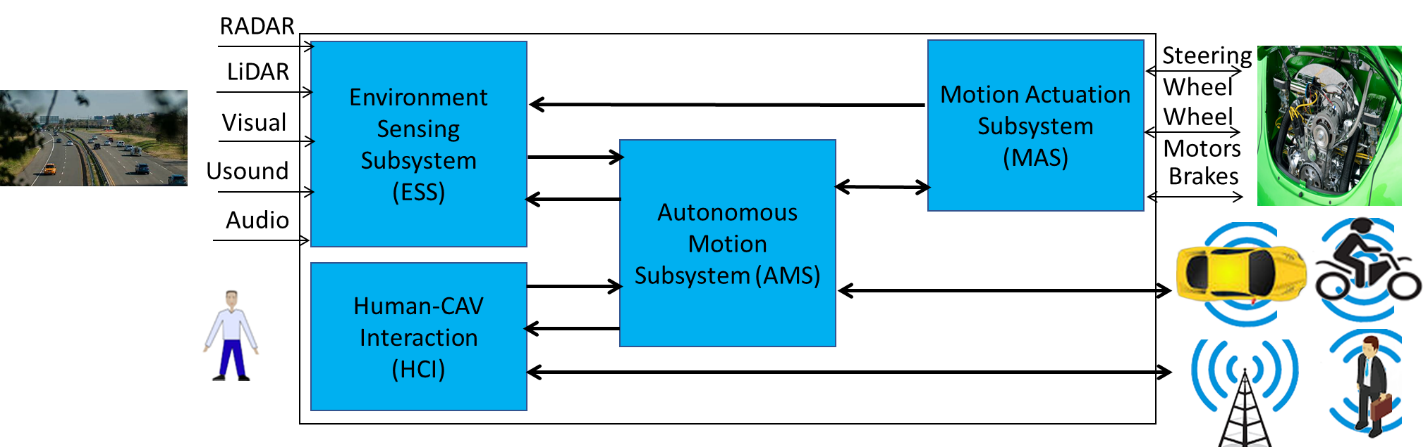
1. Plans a Route enabling the CAV to reach the requested destination.
2. During the travel it informs passengers and converses with them.
3. Autonomously reaches the destination by:
   1. Moving in the physical environment.
   2. Building Digital Representations of the Environment.
   3. Exchanging elements of such Representations with other CAVs and CAV-aware entities.
   4. Making decisions about how to execute the Route.
   5. Actuating the CAV motion to implement the decisions.
   6. Reconsidering motion decisions based on actuator feedbackpronomi personali in .

## Reference Model of Connected Autonomous Vehicle

The MPAI-CAV Reference Model is composed of four Subsystems:

1. Human-CAV Interaction (HCI).
2. Environment Sensing Subsystem (ESS),
3. Autonomous Motion Subsystem (AMS).
4. Motion Actuation Subsystem (MAS).

The Subsystems are represented in *Figure 2* where the arrows refer to the exchange of information between Subsystems and between a Subsystem and other CAVs or CAV-aware systems. The sensing of the Environment and the Motion Actuation are represented by icons.



*Figure 2 – The MPAI-CAV Subsystems*

The ESS uses the CAV Sensors to develop a representation of the external environment called Basic Environment Representation (BER). The AMS receives the BER and may share part of it with CAVs in range. By receiving similar information from CAVs in range, the AMS can refine its BER and create the Full Environment Representation (FER)

.

The operation of a CAV unfolds according to the following workflow:

|  |  |
| --- | --- |
| *Human* | Requests the CAV, via *HCI*, to take the human to a destination. |
| *HCI* | Authenticates humans.  Interprets the request of humans.  Issues commands to the *AMS*. |
| *AMS* | Requests *ESS* to provide the current Pose. |
| *ESS* | Computes and sends the BER to *AMS.* |
| *AMS* | Computes and sends Route(s) to *HCI*. |
| *HCI* | Sends travel options to Human. |
| *Human* | May integrate/correct their instructions.  Selects and communicates a Route to *HCI*. |
| *HCI* | Communicates Route selection to *AMS*. |
| *AMS* | 1. Sends the BER to the *AMS*s of other CAVs. 2. Computes the FER. 3. Decides best motion to reach the destination. 4. Issues appropriate commands to *MAS*. |
| *MAS* | 1. Executes the command via Steering Wheel, Wheel Motors and Brakes. 2. Sends response to *AMS*. |
| *Human* | 1. Interacts and holds conversation with other humans on board and the *HCI*. 2. Issues commands to *HCI*. 3. Requests *HCI* to render the FER. 4. Navigates the FER. 5. Interacts with humans in other CAVs. |
| *HCI* | Communicates with *HCI*s of other CAVs. |

## I/O Data of Connected Autonomous Vehicle

Table 2 gives the input/output data of the Connected Autonomous Vehicle. Note that E and R prefixed to CAV, HCI, and AMS stand for Ego and Remote.

Table 2 - I/O data of Human-CAV Interaction

|  |  |  |
| --- | --- | --- |
| **Input data** | **From** | **Comment** |
| Input Audio | Environment | For User authentication, command/ interaction with HCI, etc. |
| Input RADAR | Environment | Environment perception |
| Input LiDAR | Environment, Passenger Cabin | Environment perception and command/interaction with HCI |
| Input Visual | Environment, Passenger Cabin | Environment perception, User authentication, command/interaction with HCI, etc. |
| Input Ultrasound | Environment | Perception of Environment |
| Input Text | User | Text complementing/replacing User input |
| R-HCI to E-HCI Message | Remote HCI | Remote HCI to Ego HCI |
| R-AMS to E-AMS Message | Remote AMS | Remote AMS to Ego AMS |
| Global Navigation Satellite System (GNSS) Data | ~1 & 1.5 GHz Radio | Various GNSS Data sources |
| Other Environment Data | Environment | Temperature, Air pressure, Humidity, etc. |
| Steering Wheel Response | Ego CAV | Response of Steering Wheel |
| Wheel Motors Response | Ego CAV | Response of Wheel Motors |
| Brake Response | Ego CAV | Response of Brake |
| **Output data** | **To** | **Comment** |
| Output Audio | Cabin Passengers | HCI’s avatar Audio |
| Output Visual | Cabin Passengers | AMS’s avatar Visual |
| Steering Wheel Command | Ego CAV | Action on Steering Wheel |
| Wheel Motors Command | Ego CAV | Action on Wheel Motors |
| Brake Command | Ego CAV | Action on Brake |
| R-HCI to E-HCI Message | Remote HCI | Ego HCI to Remote HCI |
| R-AMS to E-AMS Message | Remote AMS | Ego AMS to Remote AMS |

# Human-CAV Interaction (HCI)

## I/O Data of Human-CAV Interaction

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

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

|  |  |  |
| --- | --- | --- |
| **Input data** | **From** | **Comment** |
| Input Audio | Environment | For User authentication, command/ interaction with HCI, etc. |
| Input RADAR | Environment | Environment perception |
| Input LiDAR | Environment, Passenger Cabin | Environment perception and command/interaction with HCI |
| Input Visual | Environment, Passenger Cabin | Environment perception, User authentication, command/interaction with HCI, etc. |
| Input Ultrasound | Environment | Perception of Environment |
| Input Text | User | Text complementing/replacing User input |
| R-HCI to E-HCI Message | Remote HCI | Remote HCI to Ego HCI |
| R-AMS to E-AMS Message | Remote AMS | Remote AMS to Ego AMS |
| Global Navigation Satellite System (GNSS) Data | ~1 & 1.5 GHz Radio | Various GNSS Data sources |
| Other Environment Data | Environment | Temperature, Air pressure, Humidity, etc. |
| **Output data** | **To** | **Comment** |
| Output Audio | Cabin Passengers | HCI’s avatar Audio |
| Output Visual | Cabin Passengers | AMS’s avatar Visual |
| R-HCI to E-HCI Message | Remote HCI | Ego HCI to Remote HCI |
| R-AMS to E-AMS Message | Remote AMS | Ego AMS to Remote AMS |

## Reference Model of Human-CAV Interaction

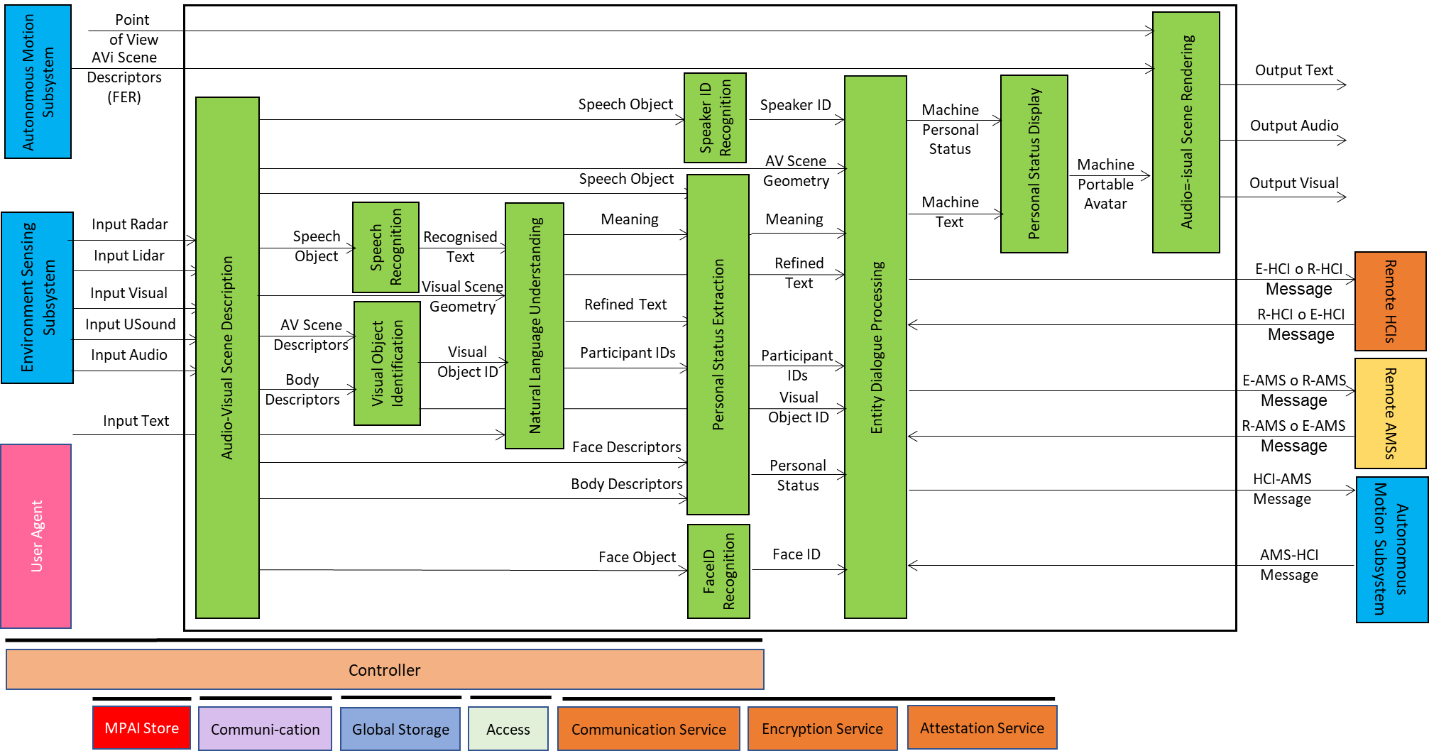


Figure 3 - Reference Model of the CAV-HCI Subsystem

## Functions of Human-CAV Interaction’s AI Modules

Table 4 gives the functions of all Environment Sensing Subsystem AIMs.

Table 4 - Functions of Human-CAV Interaction’s AI Modules

|  |  |
| --- | --- |
| **AIM** | **Function** |
| **Audio-Visual Scene Description** | Produces the Audio-Visual Scene Descriptors using (indoor or outdoor) Input Audio, Input Visual and LiDAR. |
| **Automatic Speech Recognition** | Converts speech into Recognised Text. |
| **Visual Object Identification** | Provides the ID of the class of objects of which the Visual Object is an Instance |
| **Natural Language Understanding** | Refines the Recognised Text by using context information (e.g., Instance ID of object). |
| **Speaker Identity Recognition** | Provides Speaker ID from Speech Object. |
| **Personal Status Extraction** | Provides the Personal Status of a passenger with Participant ID. |
| **Face Identity Recognition** | Provides Face ID from Face Object. |
| **Entity Dialogue Processing** | Provides:   1. Machine Text containing the HCI response. 2. Machine Personal Status. |
| **Personal Status Display** | Produces Machine Personal Avatar. |
| **Audio-Visual Scene Rendering** | Converts AV Scene Descriptors or Portable Avatar into perceptible Output Text, Output Audio and Output Visual. |

## I/O Data of Human-CAV Interaction’s AI Modules

Table 5 gives the input/output data of the Human-CAV Interaction AIMs.

Table 5 - I/O Data of Human-CAV Interaction’s AI Modules

|  |  |  |
| --- | --- | --- |
| **AIM** | **Input** | **Output** |
| **Audio -Visual Scene Description** | Input Audio  Input Visual  Input LiDAR | AV Scene Descriptors |
| **Automatic Speech Recognition** | Speech Object | Recognised Text |
| **Visual Object Identification** | AV Scene Descriptors  Visual Objects | Visual Object Instance ID |
| **Natural Language Understanding** | Recognised Text  Visual Scene Geometry  Visual Object ID  Input Text | Refined Text  Meaning |
| **Speaker Identity Recognition** | Speech Object | Speaker ID |
| **Personal Status Extraction** | Meaning  Input Speech  Face Descriptors  Body Descriptors | Personal Status |
| **Face Identity Recognition** | Face Object | Face ID |
| **Entity Dialogue Processing** | R-HCI to E-HCI Message  R-AMS to E-AMS Message  AMS-HCI Message  Speaker ID  Meaning  NLU Text  Personal Status  Face ID | E-HCI to R-HCI Message  R-AMS to R-AMS Message  HCI-AMS Message  Machine Text  Machine Personal Status |
| **Personal Status Display** | Machine Personal Status  Machine Text | Machine Personal Avatar |
| **Audio-Visual Scene Rendering** | AV Scene Descriptors  Machine Personal Avatar  Point of View | Output Text  Output Audio  Output Visual |

## I/O Data Types

### Personal Preferences

Personal Preferences includes passenger-specific preferences that enable an HCI to have access to information that facilitates human-HCI interaction. This is particularly useful when the passenger uses a rented CAV.

The data in the Personal Preferences should include:

1. Seat position.
2. Mirror position.
3. Display characteristics.
4. Preferred driving style.
5. Preferential routes.
6. …

**To Respondents**

Respondents are invited to:

1. Comment and elaborate on or extend the functional requirements of the Personal Preferences identified above.
2. Propose coded representation formats of the Personal Preferences.

### Input Audio

Multichannel Audio provided by a Microphone Array used to:

1. Create Audio Scene Descriptors to:
   1. Enable extraction of speech addressed by humans outside or inside the HCI.
   2. Incorporate outdoor Audio information into the Basic Environment Representation.
2. Suppress noise and individual sound sources outside the passenger cabin.

MPAI has developed specifications for [Multichannel Audio Stream](https://mpai.community/standards/data-types/multichannel-audio-stream/) and [Microphone Array Geometry](https://mpai.community/standards/data-types/microphone-array-geometry/).

**To Respondents**

Respondents are invited to:

1. Comment or elaborate on the applicability of the above-mentioned two standards.
2. Propose motivated extensions, or new solutions.

### Input LiDAR

A scene captured by LiDAR sensors is used to create LiDAR Scene Descriptors.

**To Respondents**

Respondents are invited to:

1. Identify functional requirements of the output data produced by LiDAR sensors for indoor (cabin) use.
2. Propose a standard format for the output data of a LiDAR sensor for indoor use.

### Input Text

Textual information represented using a Character Set able to represent the characters for most of the currently used languages.

MPAI has selected the format for [Text](https://mpai.community/standards/text/).

### Input Visual

A visual scene can be captured by an array of visual sensors characterised by:

1. Number and position of sensing devices.
2. Number of horizontal and vertical sensors in a sensing device.
3. Frame frequency.
4. Colour space.
5. Bit-depth information.
6. Depth (distance from scene pixel) information

**To Respondents**

Respondents are invited to:

1. Reference standard formats for the output data of a 2D, 2D+ depth, or 3D visual sensors.
2. Propose new formats.

### Point of View

Information provided by a passenger to the HCI to navigate the Full Environment Representation, e.g.:

1. Select a Point of View.
2. Zoom in/out.
3. Control sound level.

MPAI has specified a format for [Point of View](https://mpai.community/standards/data-types/spatial-attitude/#PointofView) and [Spatial Attitude](https://mpai.community/standards/data-types/spatial-attitude/). The latter is required when a passenger requests to view an area while the CAV is travelling.

**To Respondents**

Respondents are invited to:

1. Comment or elaborate on the applicability of the above-mentioned two formats.
2. Propose motivated extensions, or new solutions.

### Audio-Visual Scene Descriptors

The Environment Sensing Subsystem (ESS) uses its sensors to develop a representation of Environment called Basic Environment Representation (BER) that can be external or internal (cabin) to the CAV. The representation of the External Environment is passed to the Autonomous Motion Subsystem (AMS) that may improve it by exchanging BER elements with Remote CAVs (R-AMSs) and produce the E-AMS’s Full Environment Representation (FER). All Environment Representations are Audio-Visual Scene Descriptors.

The Audio Scene Descriptors enable the HCI to interact with:

1. *Outdoor humans* to enable the HCI to understand the outdoor sound field, detect and separate speech sources, recognise the identity of the speakers, and possibly converse with them.
2. Indoor humans to enable the HCI to have general and one-to-one conversation with passengers in the cabin.

The Audio Scene is assumed to be composed of Audio Objects that have a Spatial Attitude with an identifiable source origin.

The Visual Scene Descriptors enable the HCI to interact with:

1. *Outdoor humans* to enable the HCI to understand the outdoor visual field, detect approaching humans, and recognise their identity.
2. *Indoor humans* to enable the HCI to locate and identify humans conversing with the HCI.

The Visual Scene is assumed to be composed of Visual Objects that have a Spatial Attitude. A Visual Scene Description AIM may produce Descriptors for a Visual Scene that includes occluded objects, and the Descriptors should signal that information, if detected.

Visual Scene Descriptors describe the structured composition of Visual Objects in a Visual Scene.

The Audio-Visual Scene Descriptors enable the description of an audio-visual scene – that can be outdoor and indoor – in terms of Audio-Visual Objects:

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

MPAI has developed specifications for:

1. [Audio Object](https://mpai.community/standards/data-types/audio-object/) and [Audio Scene Descriptors](https://mpai.community/standards/data-types/audio-scene-descriptors/)
2. [Visual Object](https://mpai.community/standards/data-types/visual-object/) and [Visual Scene Descriptors](https://mpai.community/standards/data-types/visual-scene-descriptors/)
3. [Audio-Visual Object](https://mpai.community/standards/data-types/audio-visual-object/) and [Audio-Visual Scene Descriptors](https://mpai.community/standards/data-types/audio-visual-scene-descriptors/).

**To Respondents**

Respondents are invited to:

1. Comment or elaborate on:
   1. The assumptions in this subsection.
   2. The use of the MPAI Audio, Visual, and Audio-Visual Object and Scene Descriptions.
2. Propose extensions to or alternative formats for the six data type referenced by this subsection.

### Audio-Visual Scene Geometry

Represents the spatial arrangement of the Audio, Visual, and Audio-Visual Objects of a Scene. AV Scene Geometry does not contain the Objects in the scene, only their arrangement.

MPAI has developed specifications for:

1. [Audio Scene Geometry](https://mpai.community/standards/data-types/audio-scene-geometry/)
2. [Visual Scene Geometry](https://mpai.community/standards/data-types/visual-scene-geometry/)
3. [Audio-Visual Scene Geometry](https://mpai.community/standards/data-types/audio-visual-scene-geometry/).

**To Respondents**

Respondents are invited to:

1. Comment or elaborate on:
   1. The assumptions in this subsection.
   2. The use of the MPAI Audio, Visual, and Audio-Visual Object and Scene Geometry.
2. Propose extensions to or alternative formats for the three data type referenced by this subsection.

### Avatar Model

The Model of an Avatar selected by a user as the human representation of the CAV’s HCI.

Avatar Model is an element of MPAI’s [Portable Avatar](https://mpai.community/standards/data-types/avatar/#PortableAvatar) generated by the [Personal Status Display](https://mpai.community/standards/mpai-paf/mpai-paf-composite-aims/personal-status-display-paf-psd/) AIM that is rendered by the [Audio-Visual Scene Rendering](https://mpai.community/standards/aiws-and-aims/audio-visual-scene-rendering-hmc-avr/) AIM.

**To Respondents**

Respondents are invited to:

1. Comment or elaborate on the MPAI-specified technologies identified in this subsection.
2. Propose extensions to the identified technologies or new ones.

### Body Descriptors

Body Descriptors digitally represent the body of:

1. A human having a multimodal conversation with the HCI when it estimates the human’s Personal Status.
2. The Avatar embodying the HCI in the Portable Avatar generated the [Portable Avatar](https://mpai.community/standards/data-types/avatar/#PortableAvatar) generated by the [Personal Status Display](https://mpai.community/standards/mpai-paf/mpai-paf-composite-aims/personal-status-display-paf-psd/) AIM that is rendered by the [Audio-Visual Scene Rendering](https://mpai.community/standards/aiws-and-aims/audio-visual-scene-rendering-hmc-avr/) AIM.

MPAI has specified the format of [Body Descriptors](https://mpai.community/standards/data-types/body-descriptors/).

**To Respondents**

Respondents are invited to:

1. Comment or elaborate on the MPAI Body Descriptors.
2. Propose extensions to the identified technologies or new ones.

### Face Descriptors

The Face Descriptors are the Descriptors of the face of:

1. A human having a multimodal conversation with the HCI when it estimates the human’s Personal Status.
2. The Avatar embodying the HCI in the Portable Avatar generated the [Portable Avatar](https://mpai.community/standards/data-types/avatar/#PortableAvatar) generated by the [Personal Status Display](https://mpai.community/standards/mpai-paf/mpai-paf-composite-aims/personal-status-display-paf-psd/) AIM that is rendered by the [Audio-Visual Scene Rendering](https://mpai.community/standards/aiws-and-aims/audio-visual-scene-rendering-hmc-avr/) AIM.

MPAI has specified the format of [Face Descriptors](https://mpai.community/standards/data-types/face-descriptors/).

**To Respondents**

Respondents are invited to:

1. Comment or elaborate on the MPAI Face Descriptors.
2. Propose extensions to the identified technologies or new ones.

### Face ID

The Identifier of the human inferred from a Face Object captured from a target human. The scope of the ID could cover the members of an authorised group, such as the members of a family, specific employees of a company or the customers of a car renting company.

MPAI has specified the format of [Instance ID](https://mpai.community/standards/data-types/Instance-Identifier/) that is agnostic of the specific case and considers Face in the same way as any other object that is identified as a member of a class of objects.

**To Respondents**

Respondents are invited to:

1. Comment or elaborate on the MPAI Instance ID format.
2. Propose extensions to the identified technologies or new ones.

### Speaker ID

The Identifier of the human inferred from their utterances. The Speaker ID may be derived by analysing speech segments of the speaker under consideration. The scope of the ID may cover the members of an authorised group, such as the members of a family, specific employees of a company, or the customers of a car renting company.

MPAI has specified the format of [Instance ID](https://mpai.community/standards/data-types/Instance-Identifier/) that is agnostic of the specific case and considers Speech in the same way as any other object that is identified as a member of a class of objects.

**To Respondents**

Respondents are invited to:

1. Comment or elaborate on the MPAI Instance ID format.
2. Propose extensions to the identified technologies or new ones.

### Meaning

Representing an input text as syntactic and semantic information.

MPAI has specified [Meaning](https://mpai.community/standards/data-types/text-descriptors/#Meaning).

**To Respondents**

Respondents are invited to:

1. Comment or elaborate on the MPAI specification of Meaning.
2. Propose extensions to it or a new specification.

### Visual Object ID

The Identifier uniquely associated with a particular class of objects, e.g., hammer, screwdriver, etc.

MPAI has specified the format of [Instance ID](https://mpai.community/standards/data-types/Instance-Identifier/) that is agnostic of the specific case and considers Face in the same way as any other object that is identified as a member of a class of objects.

**To Respondents**

Respondents are invited to:

1. Comment or elaborate on the MPAI Instance ID format.
2. Propose extensions to the identified technologies or new ones.

### Personal Status

Personal Status (PS) indicates a set of three Factors (Cognitive State, Emotion, Social Attitude) conveyed by one or more Modalities (Text, Speech, Face, and Gesture Modalities). PS is used to assign a label, according to a given classification, to the internal state of an Entity – human of Machine.

MPAI has developed the full specification of [Personal Status](https://mpai.community/standards/data-types/personal-status/).

**To Respondents**

Respondents are invited to:

1. Comment or elaborate on the MPAI Personal Status format.
2. Propose extensions of the labels or new sets of labels for the Factors, or new technologies.

### Speech Model

A Neural Network trained to generate utterances with specific Speech Descriptors. MPAI has added a Speech Model to the Personal Status Display Composite AIM.

**To Respondents**

Respondents are invited to comment on the use of Speech Model to the Personal Status Display Composite AIM and requested to propose Speech Model technology.

### Output Audio

[Output Audio](https://mpai.community/standards/data-types/audio/#OutputAudio) is Audio information produced by the Audio-Visual Rendering AIM.

### Output Visual

[Output Visual](https://mpai.community/standards/data-types/audio/#OutputVisual) is Visual information produced by the Audio-Visual Rendering AIM.

### HCI-AMS Messages

An HCI can send a Message to the AMS based on a human message or an R-HCI message to:

1. Provide options of Routes connecting the current place and the destination. The request may include:
   1. Stops in between place and destination.
   2. Desired arrival time, e.g., during the travel.
2. Select and execute one Route out of a few.
3. Suspend a Route.
4. Resume a Route.
5. Change a Route.

An AMS can send the HCI:

1. A message in response to a message received.
2. Information, such as:
   1. List of Road options.
   2. Unexpected delay in implementing the original request.
   3. Road problems.
   4. Equipment failure.
   5. Target reached.

The HCI may convert the received information to a format – such as Text, Audio, and Visual – that is understandable by humans.

**To Respondents**

Respondents are invited to:

1. Comment or elaborate on the functional requirements AMS-HCI and HCI-AMS Messages identified above.
2. Propose coded representation formats of AMS-HCI Messages and HCI-AMS.

### E-HCI to/from R-HCI messages

The E-HCI may send or receive messages to/from an R-HCI:

1. An E-HCI may send a Message to be forwarded to a passenger of an R-HCI.
2. An E-HCI may send or ask an R-HCI, a CAV-Aware entity (e.g., Roadside Unit), or a Store and Forward entity to send the Road State.

**To Respondents**

Respondents are invited to:

1. Comment or elaborate on the functional requirements identified E-HCI to/from R-HCI message above.
2. Propose coded representation formats of E-HCI to/from R-HCI messages.

# Environment Sensing Subsystem (ESS)

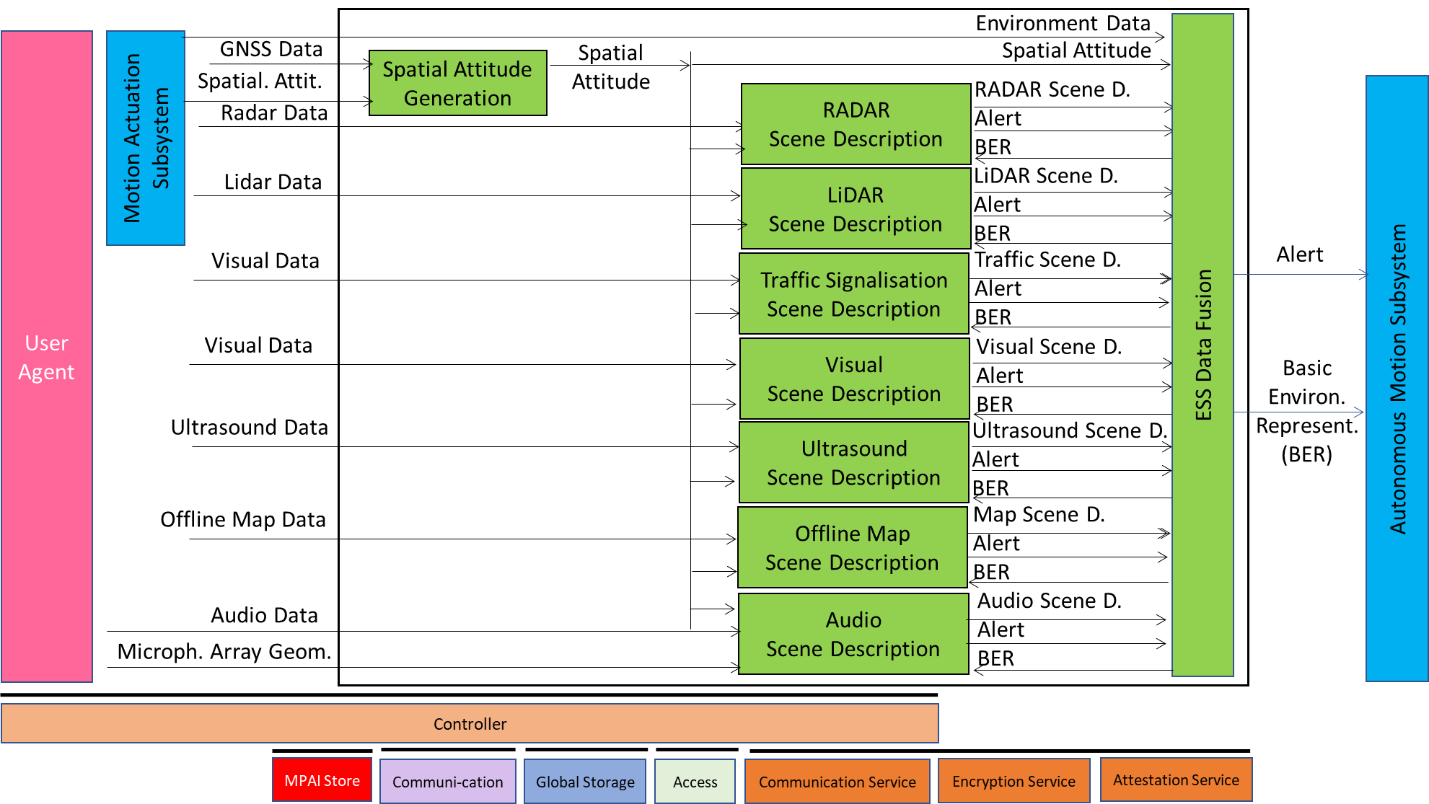
## Functions of Environment Sensing Subsystem

The Environment Sensing Subsystem (ESS) of a Connected Autonomous Vehicle (CAV):

1. Acquires electromagnetic and acoustic data from the Environment using its sensors.
2. Receives Environment Data (e.g., temperature, pressure, humidity, etc.) from the Motion Actuation Subsystem.
3. Receives an initial estimate of the Ego CAV’s Spatial Attitude generated by the Motion Actuation Subsystem
4. Produces a sequence of Basic Environment Representations (BER) for the journey.
5. Passes the BER to the Human-CAV Interaction Subsystem (HCI) and Autonomous Motion Subsystem (AMS).
6. Requests elements of the Full Environment Representations (FER) produced by AMS.

## Reference Architecture of Environment Sensing Subsystem

*Figure 4* gives the Reference Model of Environment Sensing Subsystem.



*Figure 4 – Environment Sensing Subsystem Reference Model*

The typical sequence of operations of the Environment Sensing Subsystem is:

1. Compute the CAV’s Spatial Attitude using the initial Spatial Attitude provided by the Motion Actuation Subsystem and GNSS Data.
2. Receive Environment Sensing Technology (EST)-specific Data, e.g., RADAR Data provided by the RADAR EST.
3. Produce and send EST-specific Alert, if necessary, to Autonomous Motion Subsystem.
4. Access the Basic Environment Representation at previous times, if necessary.
5. Produce EST-specific Scene Descriptors, e.g., the RADAR Scene Descriptors.
6. Integrate the Scene Descriptors from different ESTs into the Basic Environment Representation.

Note that *Figure 4* assumes that:

1. Traffic Signalisation Recognition produces the Road Topology by analysing Visual Data. The model of *Figure 4* can easily be extended to the case where Data from other ESTs is processed to compute or help compute the Road Topology.
2. Environment Sensing Technologies are individually processed. However, an implementation may combine two or more Scene Descriptors AIMs handling two or more ESTs, provided the relevant interfaces are preserved.

## I/O Data of Environment Sensing Subsystem

The currently considered Environment Sensing Technologies (EST) are:

1. GNSS – Global Navigation Satellite System (~1 & 1.5 GHz Radio).
2. Spatial Attitude of the Ego CAV – Geographical Position and Orientation and their time derivatives up to 2nd order.
3. Visual Data in the visible range, possibly supplemented by depth information (400 to 700 THz).
4. LiDAR Data (~200 THz infrared).
5. RADAR Data (~25 & 75 GHz).
6. Ultrasound Data (> 20 kHz).
7. Audio Data in the audible range (16 Hz to 20 kHz).
8. Spatial Attitude (from the Motion Actuation Subsystem).
9. Other environmental data (temperature, humidity, ...).

The Offline Map data can be accessed either from stored information or online.

Table 6 gives the input/output data of the Environment Sensing Subsystem.

Table 6 - I/O data of Environment Sensing Subsystem

|  |  |  |
| --- | --- | --- |
| **Input data** | **From** | **Comment** |
| Radar Data | ~25 & 75 GHz Radio | Environment Capture with Radar |
| Lidar Data | ~200 THz infrared | Environment Capture with Lidar |
| Visual Data | Video (400-800 THz) | Environment Capture with visual cameras |
| Ultrasound Data | Audio (>20 kHz) | Environment Capture with Ultrasound |
| Offline Map Data | Local storage or online | cm-level data at time of capture |
| Audio Data | Audio (16 Hz-20 kHz) | Environment or cabin Capture with Microphone Array |
| Microphone Array Geometry | Microphone Array | Disposition of microphones in the array |
| Global Navigation Satellite System (GNSS) Data | ~1 & 1.5 GHz Radio | Get Pose from GNSS |
| Spatial Attitude | Motion Actuation Subsystem | To be fused with Pose from GNSS Data |
| Other Environment Data | Motion Actuation Subsystem | Temperature, Humidity, etc. |
| **Output data** | **To** | **Comment** |
| Alert | Autonomous Motion Subsystem | Critical information from an EST. |
| Basic Environment Representation | Autonomous Motion Subsystem | ESS-derived representation of Environment |

## Functions of Environment Sensing Subsystem’s AI Modules

Table 7 gives the functions of all AIMs of the Environment Sensing Subsystem.

Table 7 - Functions of Environment Sensing Subsystem’s AI Modules

|  |  |
| --- | --- |
| **AIM** | **Function** |
| **RADAR Scene Description** | Produces RADAR Scene Descriptors and Alert. |
| **LiDAR Scene Description** | Produces LiDAR Scene Descriptors and Alert. |
| **Traffic Signalisation Recognition** | Produces Road Topology. |
| **Visual Scene Description** | Produces Visual Scene Descriptors and Alert. |
| **Ultrasound Scene Description** | Produces Ultrasound Scene Descriptors and Alert. |
| **Online Map Scene Description** | Produces Online Map Data Scene Descriptors. |
| **Audio Scene Description** | Produces Audio Scene Descriptors and Alert. |
| **Spatial Attitude Generation** | Computes the CAV Spatial Attitude from CAV Center using GNSS and Motion Actuation Subsystem information. |
| **Environment Sensing Subsystem Data Fusion** | Receives Scene Descriptors and critical Environment Representation as Alert from the different ESTs.  Produces Alert and Basic Environment Representation (BER) including all available information from ESS:   1. The different Scene Descriptors generated by the different EST-specific Scene Description AIMs. 2. Environment Data. 3. The Spatial Attitude of the Ego CAV (Figure 5). |

A car with directions and lines

Description automatically generated with medium confidence

Figure 5 – Spatial Attitude in a CAV

## I/O Data of Environment Sensing Subsystem’s AI Modules

For each AIM (1st column), Table 8 gives the input (2nd column) and the output data (3rd column) of the Environment Sensing Subsystem. Note that the Basic Environment Representation in column 2 refers to the previously produced BER.

Table 8 - I/O Data of Environment Sensing Subsystem’s AI Modules

|  |  |  |
| --- | --- | --- |
| **AIM** | **Input** | **Output** |
| **Radar Scene Description** | Radar Data  Basic Environment Representation | Alert  Radar Scene Descriptors |
| **Lidar Scene Description** | Lidar Data  Basic Environment Representation | Alert  Lidar Scene Descriptors |
| **Traffic Signalisation Recognition** | Visual Data  Basic Environment Representation | Alert  Road Topology |
| **Visual Scene Description** | Visual Data  Basic Environment Representation | Alert  Lidar Scene Descriptors |
| **Ultrasound Scene Description** | Ultrasound Data  Basic Environment Representation | Alert  Ultrasound Scene Descriptors |
| **Map Scene Description** | Offline Map Data  Basic Environment Representation | Alert  Map Scene Descriptors |
| **Audio Scene Description** | Audio Data  Basic Environment Representation | Alert  Audio Scene Descriptors |
| **Spatial Attitude Generation** | GNSS Data  Spatial Attitude form MAS | Spatial Attitude |
| **Environment Sensing Subsystem Data Fusion** | RADAR Scene Descriptors  LiDAR Scene Descriptors  Road Topology  Lidar Scene Descriptors  Ultrasound Scene Descriptors  Map Scene Descriptors  Audio Scene Descriptors  Spatial Attitude  Other Environment Data | Alert  Basic Environment Representation |

## Data functional requirements

### Sensor data

#### Audio Data

Microphone (arrays) are used to capture the sound both outdoor and indoor for the purpose of creating Audio Scene Description to:

1. Provide the location of sound sources.
2. Enable extraction of speech addressed to CAV by humans.
3. Remove unwanted noise from the passenger cabin.
4. Incorporate Audio information into the Basic Environment Representation.

MPAI specifies: [Microphone Array Geometry](https://mpai.community/standards/data-types/microphone-array-geometry/) [Multichannel Audio](https://mpai.community/standards/data-types/audio/#TMultichannelAudio) and [Multichannel Audio Stream](https://mpai.community/standards/data-types/multichannel-audio-stream/).

**To Respondents**

Respondents are requested to comment on the usability of the specified technologies [2] for MPAI-CAV and/or propose and justify functional requirements for an Audio Array Format suitable to create a 3D sound field representation of the Environment for the stated purposes.

#### Visual Data

A single camera or an array of cameras providing pixel value, time and potentially depth are used to create the Visual Scene Description that:

1. Provides the position and orientation of individual visual objects.
2. Adds Visual information into the Basic Environment Representation.
3. Enables identification, tracking and representation of relevant visual objects, especially humans.

**To Respondents**

Respondents are invited to comment or elaborate on the functional requirements of Visual Data.

#### 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 human next to a vehicle), while a can may produce an image out of proportion to its size.

The main features of Radar Data are:

1. Suitability to measure distance (short range radar in the 25 GHz band).
2. Low resolution images.
3. Ability to detect objects and measure speed @ ≤ 250 m (long range radar in the 76-77 GHz).
4. Possibility to use a small antenna because wavelength is ~3.5-4 mm.
5. Limited interference with other systems because of atmospheric absorption.
6. Ability to provide a resolution of ~25-cm radial and ~1.5 degrees angular.

**To Respondents**

Respondents are invited to comment or elaborate on the functional requirements of RADAR images formats with the goal of achieving tracking and representation of objects for the RADAR 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 objects 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 in the order of 1 Mpoint/s with a frame frequency between 10 and 30 frame/s.
8. Angular resolution is in the order of 0.1º vertical and 1º horizontal with a maximum capture of field is 40º vertical.
9. A LiDAR produces a data rate in the order of 100 Mbit/s.

**To Respondents**

Respondents are invited to comment or elaborate on the functional requirements of LiDAR images formats with the goal of achieving tracking and representation of objects for the LiDAR Scene Description.

#### Ultrasound Data

The main features of Ultrasound are:

1. Ability to monitor the immediate surroundings of the vehicle ((≤ 10 m)
2. Operation frequency above 30 kHz.
3. Low-resolution images.

**To Respondents**

Respondents are invited to comment or elaborate on the functional requirements of Ultrasound images formats with the goal of achieving tracking and representation of objects for the Ultrasound Scene Description.

#### GNSS Data

Global Navigation Satellite Systems (GNSS) provide spatial information with different accuracies.

Some data formats are:

1. GPS Exchange Format (GPX): an XML schema providing a common GPS data format that can be used to describe waypoints, tracks, and routes.
2. Environment Geodetic System (WGS): definition of the coordinate system's fundamental and derived constants, the ellipsoidal (normal) Earth Gravitational Model (EGM), a description of the associated Environment Magnetic Model (WMM), and a current list of local datum transfor­mations.
3. International GNSS Service (IGS) SSR: 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 set of functional requirements for a single GNSS data format that is ideally able to represent the features of all GNSS types.

#### Offline Map Data

An Offline Map or HD map or 3D map 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.

The features of an Offline Data Format used by a CAV should consider the features of data formats such as:

1. Navigation Data Standards [12] calls itself “The Environment wide standard for map data in automotive eco-systems”. Their NDS specification covers data model, storage format, interfaces, and protocols.
2. SharedStreets [13] Referencing System calls itself a global non-proprietary system for describing streets.

**To respondents**

Respondents are requested to propose functional requirements of the Offline Map Data Format to support the most common offline map formats.

#### Spatial Attitude from MAS

The CAV Spatial Attitude is a snapshot of the estimated CAV spatial attitude received from the Motion Actuation Subsystem containing:

1. Three spatial coordinates of the Centre Point of the CAV.
2. The Orientation of the CAV.
3. Derivatives of Position and Orientation up to the second order.
4. Accuracies of estimated Position and Orientations.

MPAI has specified the [Spatial Attitude](https://mpai.community/standards/data-types/spatial-attitude/) Data Type.

The Ego CAV sets z=0 to its Center Point.

**To Respondents**

Respondents are requested to comment on the use of the MPAI-specified Spatial Attitude for CAV purposes or to propose the coding of alternative functional requirements for the said set of data, and to comment on the [Spatial Attitude](https://mpai.community/standards/data-types/spatial-attitude/) specification or to propose a different definition.

### Audio-Visual Scene Descriptors

#### Introduction

This Section develops the functional requirements of the different Data Formats that contribute to the description of the Environment. The term Scene will be used to indicate the outcome of the process undergone by:

1. Using a specific Environment Sensing Technology (EST).
2. Obtaining *sensed data* (EST Data).
3. Process the EST Data to represent the environment with a Scene.

To the extent possible, a Scene created from Data of a specific EST should have a compatible format to facilitate the fusion of the individual EST-based Scenes into the Basic Environment Representation to be passed to the Autonomous Motion Subsystem.

The operation of the Environment Sensing Subsystem unfolds as follows:

1. A given EST produces EST Data at discrete Δt time increments that depend on the EST operating frequency. Different ESTs may use different Δt values.
2. EST-specific Data are passed to the EST-specific Scene Description AIM.
3. An EST-specific Scene Description AIM produces EST-specific Scene Descriptors. These may have a complex data structure that includes several elementary Data Types each having their own Data Formats.
4. EST-specific Scene Descriptors enable an object-based, time-dependent, and constantly updated Scene Descriptors that may contain Objects potentially with different resolutions, e.g., an object at 100m and another object at 10m may be represented with different spatial and temporal resolutions.
5. Scene Descriptors#1 produced from EST#1 Data may include Data Types not included in Scene Descriptors#2 produced from EST#2 Data. However, the Environment Sensing Subsystem (ESS) Data Fusion AIM is cognisant of both Data Formats.
6. Scene Descriptors#1 from EST#1 Data may not represent the environment with the same Accuracy as or may provide values that conflict with the environment representation provided by Scene Descriptors#2 from EST#2 Data.
7. The format of the Offline Maps should allow for lossless transformation of its EST Data into Scene Descriptors without loss of information so as to enable the fusion of its Scene Descriptors into the Basic Environment Representation produced by the ESS Data Fusion AIM.
8. EST Scene Descriptors SD(t) at time t are obtained from:
   1. Using sensed EST Data at time t and previously computed Scene Descriptors SD(t-Δt), SD(t-2Δt) etc.
   2. Updating the Objects inherited from preceding SDs.
   3. Removing objects present in previous SDs and no longer present in SDs(t).
   4. Adding and assigning attributes to new Objects, i.e., entirely new Objects, the merge of two or more Objects, or the splitting of a previously merged Object.
9. SD(t) is a list objects detected and confirmed at time t with their attributes.
10. EST Scene Description AIMs keep memory of past Scene Descriptors. Recent Objects may retain all attributes while Objects in a farther past may have coarser attributes or not be available at all.
11. EST-specific Scene Descriptors allow for the description of Object using one of a limited number of MPAI-standardised formats:
    1. The coordinates of a centre of gravity of an Object.
    2. The Bounding Box of the Object.
    3. 2D Scene Objects
       1. Static environment:
       2. Parametric free space representation represented as a single object.
       3. Alternative representations as individual static objects.
       4. Dynamic environment: object-based representation.
    4. 2.5D Scene Objects
       1. Static components of the scene
       2. Grid-based (elevation maps or Stixel Environment), represented as a single object.
       3. Object-based for traffic poles and signals (e.g., Stixel Environment, Multi-level surface map).
       4. Object-based for the dynamic parts (e.g., Stixel Environment, Multi-level surface map).
    5. 3D (Volumetric) Scene Objects
       1. Static components of the scene
       2. Voxel grids, meshes, possibly as a single.
       3. Object-based for traffic poles and signals (voxel grids, meshes).
       4. Dynamic components of the scene (point clouds, voxel grids, meshes, …)
12. An EST-specific Scene can contain Objects with different formats.
13. At a given time that depends on the operating frequency of a specific EST, the Scene described by the Audio-Visual Scene Descriptors represents an EST-specific snapshot of the environment.

MPAI has developed specification of [Audio Object](https://mpai.community/audio-object/), [Visual Object](https://mpai.community/visual-object/), [Audio-Visual Object](https://mpai.community/standards/data-types/audio-visual-object/), and of [Audio Scene Descriptors](https://mpai.community/standards/data-types/audio-scene-descriptors/), [Visual Scene Descriptors](https://mpai.community/standards/data-types/visual-scene-descriptors/), and [Audio-Visual Scene Descriptors](https://mpai.community/standards/data-types/audio-visual-scene-descriptors/) supporting the functional requirements identified above.

**To Respondents**

Respondents are

1. Invited to comment on the functional requirements identified above and on the MPAI specifications that provide the information identified in 6.6.2.1.
2. Requested to propose motivated extensions or new technologies.

#### Visual Scene Descriptors

The Visual Scene Description AIM

1. Receives the Spatial Attitude from MAS
2. Retrieves the current Spatial Attitude.
3. Receives or retrieves a specified subset of a prior Basic Environment Representation
4. Provides Visual Scene Descriptors, a machine-readable description of the Visual Scene’s:
5. Spatial Attitudes of the Visual Objects.
6. Visual Objects.

**To Respondents**

Respondents are requested to propose functional requirements of Visual Scene Descriptors that provide the information identified in 6.6.2.1.

#### Lidar Scene Descriptors

The LiDAR Scene Description AIM receives LiDAR Data, Spatial Attitude from MAS, and a portion of a prior Basic Environment Representation and provides LiDAR Scene Descriptors.

**To Respondents**

Respondents are requested to propose functional requirements of LiDAR Scene Descriptors that provide the information identified in 6.6.2.1.

#### RADAR Descriptors

The RADAR Scene Description AIM receives RADAR Data, Spatial Attitude from MAS, and a portion of a prior Basic Environment Representation and provides RADAR Scene Descriptors.

**To Respondents**

Respondents are requested to propose functional requirements of RADAR Scene Descriptors that provide the information identified in 6.6.2.1.

#### Ultrasound Scene Descriptors

The Ultrasound Scene Description AIM receives Ultrasound Data, Spatial Attitude from MAS, and a portion of a prior Basic Environment Representation and provides Ultrasound Scene Descriptors.

**To Respondents**

Respondents are requested to propose functional requirements of Ultrasound Scene Descriptors that provide the information identified in 6.6.2.1.

#### Offline Maps Scene Descriptors

The Offline Map Scene Description AIM receives Offline Map Data, Spatial Attitude from MAS, and a portion of a prior Basic Environment Representation and provides Offline Map Scene Descriptors.

**To Respondents**

Respondents are requested to propose functional requirements of Offline Map Scene Descriptors that provide the information identified in 6.6.2.1.

#### Audio Scene Descriptors

The Audio Scene Description AIM receives Audio Data, Spatial Attitude from MAS, and a portion of a prior Basic Environment Representation and provides Audio Scene Descriptors.

**To Respondents**

Respondents are requested to propose functional requirements of Audio Scene Descriptors that provide the information identified in 6.6.2.1.

### Road Topology

For the sake of simplicity, it is assumed that Road Topology uses Visual Data. The content of this Subsection can be easily extended to apply to other EST Data.

The Traffic Signalisation Recognition AIM receives:

1. Visual Data containing:
   1. Road signs (informative)
   2. Traffic signs (to be mandatorily obeyed)
   3. Traffic lights
   4. Walkways
   5. Lanes
2. Portions of a prior Basic Environment Representation.

The MPAI-specified Scene Descriptors can represent the components of the Road Topology and provide support to a rich semantic description of Scene and Object Attributes.

**To Respondents**

Respondents are requested to:

1. Comment on the MPAI Object and Space Description solution for Road Topology needs.
2. Propose additional functional requirements for a set of Traffic Signalisation Descriptors.
3. Propose new Road Topology solutions.

### Basic Environment Representation

The Scene Descriptors of the environment traversed by a CAV is called Basic Environment Representation (BER). The BER results from the integration of all data sensed by a CAV:

1. Spatial information (e.g., GNSS, odometry).
2. Scene Descriptors obtained from EST-specific sensed data.
3. Road Topology.
4. Environmental data (e.g., weather, temperature, air pressure, ice and water on the road, wind, fog etc.)

The functional requirements of the BER format are:

1. Includes all available information that enables the Autonomous Motion Subsystem (AMS) to define a Path to be executed in a Decision Horizon Time.
2. Describes the Environment in terms of Scene Descriptors (including static objects, e.g., from Offline Maps) and Topology (e.g., roads and lanes).
3. Enables object tracking, inference of motion vectors, etc. by referencing the BERs of sufficient prior snapshots.
4. Describes each Object with the following attributes:
   1. Time. This is specified as Start and End Time of the validity of the Object Description.
   2. Object Identifier. An Identifier is assigned to an Object that is retained until the Object disappears.
   3. AIM Identifier. This identifies the AIM that provided the initial Data used to represent the Object.
   4. Object Format ID. MPAI is identifying a set of Object Format specifications that enable unambiguous reference to an Object Format.
   5. Identifiers of parent Objects corresponding to the current Object.
   6. Identifier of a parent Object that has spawned more than one current Object.
   7. ID of spatially corresponding Object of different Type.
   8. Spatial Attitude of Object.
   9. Object dimensionality (2D, 2.5D and 3D). Applicable only to Visual Objects.
   10. Visual Object shape.
   11. Semantic relationship with other Objects, e.g., identification of groups of Objects (platoon). The components of a platoon may broadcast Platooning Information, or a CAV may be able to deduce it by observing the behaviour of a group of CAVs over a period.
   12. Accuracy of all Object values.
5. Allows for easy verification 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 BER).
6. Has a scalable representation, i.e., it allows for:
   1. Gradual refinement of a BER when new EST-specific Scene Descriptors are added.
   2. Extraction of part of the BER 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. Shape (e.g., bounding box for a Visual Object).
   5. Selected (read) access to data required by different AIMs, e.g., the RADAR Scene Description AIM accesses the current BER to improve its description.
   6. Easy update of Objects and Scene from one snapshot to another.
   7. Possibility that a CAV communicates a subset of its BER to another CAV. E.g., Objects have different degrees of details, starting from bounding boxes and their Position Attributes, depending on the available bandwidth.

**To Respondents**

Respondents are requested to propose functional requirements for the following Data Types:

1. Timestamp.
2. Type (Audio or Visual).
3. Object IDs.
4. Motion State (Static, Dynamic, Unknown).
5. Spatial Attitude.
6. Object dimensionality (2D, 2.5D and 3D).
7. Visual Object shape
8. Semantics of groups of objects (IDs, relative distances, etc.)

# Autonomous Motion Subsystem (AMS)

## Functions of Autonomous Motion Subsystem

The functions of the Autonomous Motion Subsystem (AMS) are:

1. Receive a request to reach a destination as instructed by Human-CAV Interaction (HCI).
2. Request current Pose to Environment Sensing Subsystem (ESS).
3. Converse with HCI and settle on final Route.
4. Receive Basic Environment Representation (BER) from ESS.
5. Broadcast appropriate BER subsets to Remote AMSs.
6. Respond to specific Remote AMS requests.
7. Produce Full Environment Representation.
8. Generate Paths (Plath Planner).
9. Generate Goal and Trajectory (Motion Planner).
10. Check whether Trajectory can be implemented (Obstacle Avoider).
11. Issue Command to Motion Actuation Subsystem.

## Reference Architecture of Autonomous Motion Subsystem

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

This is the operation of the Reference Model:

1. A human requests the Human-CAV Interaction to take them to a destination.
2. HCI interprets request and passes interpretation to the AMS.
3. The AMS activates the Route Planner to generate a set of Waypoints starting from the current Pose, obtained from the Full Environment Representation, up to the destination.
4. The Waypoints enter the Path Planner which generates a set of Poses to reach the next Waypoint.
5. For each Path, the Motion Planner generates a Trajectory to reach the next Pose.
6. Obstacle Avoider receives the Trajectory and checks if an Alert was received.
7. If an Alert was received, Obstacle Avoider checks whether the implementation of the Trajectory creates a collision.
   1. If a collision is indeed detected, Obstacle Avoider requests a new Trajectory from the Motion Planner.
   2. If no collision is detected, Obstacle Avoider issues a Command to Motion Actuation Subsystem.

A screenshot of a computer

Description automatically generated

*Figure 6 – Autonomous Motion Subsystem Reference Model*

1. The Motion Actuation Subsystem sends MAS-AMS Response about the execution of the Command.
2. The AMS, based on the MAS-AMS Responses received potentially conveying changes in the Environment, can decide to discontinue the execution of the earlier Command and issue another AMS-MAS Message instead.
3. The decision of each element of the said chain may be recorded in the Decision Recorder (“black box”).

## I/O Data of Autonomous Motion Subsystem

Table 9 gives the input/output data of Autonomous Motion Subsystem.

Table 9 - I/O data of Autonomous Motion Subsystem

|  |  |  |
| --- | --- | --- |
| **Input data** | **From** | **Comment** |
| Basic Environment Representation | Environment Sensing Subsystem | CAV’s Environment representation. |
| Alert | Environment Sensing Subsystem | Critical information from an EST in ESS. |
| HCI-AMS Message | Human-CAV Interaction | Human commands, e.g., “take me home” |
| Environment Representation | Remote AMSs | Other CAVs and vehicles, and roadside units. |
| MAS-AMS Message | Motion Actuation Subsystem | Message sent by the AMS to the MAS. |
| **Output data** | **To** | **Comment** |
| AMS-HCI Message | Human-CAV Interaction | AMS’s message to HCI-AMS. |
| AMS-MAS Message | Motion Actuation Subsystem | Message to MAS, e.g., “in 5s assume a given Spatial Attitude”. |
| Environment Representation | Remote AMSs | For information to other CAVs |

## Functions of Autonomous Motion Subsystem’s AI Modules

Table 10 gives the AI Modules of the Autonomous Motion Subsystem.

Table 10 - Functions of Autonomous Motion Subsystem’s AI Modules

|  |  |
| --- | --- |
| **AIM** | **Function** |
| **Full Environment Representation Fusion** | Creates an internal representation of the Environment by fusing infor­mation from itself, Remote AMSs, and other CAV-aware entities. |
| **Route Planner** | Computes a Route, through a road network, from the current to the target destination. |
| **Path Planner** | Generates a set of Paths, considering:   1. Route. 2. Spatial Attitude. 3. Full Environment Representation. 4. Traffic Rules. |
| **Motion Planner** | Defines a Goal and a Trajectory to reach the Goal using the Spatial Attitude satisfying the CAV’s kinematic and dynamic constraints and considering passengers’ comfort. |
| **Obstacle Avoider** | Checks whether the Trajectory is compatible with any Alert information. If it is, it passes the Trajectory to the Command Issuer. If it is not, it requests a new Trajectory. If Command Issuer informs Obstacle Avoider that AMS-MAS Message cannot be implemented, Obstacle Avoider may issue a “discontinue previous Command” and forward to the appropriate upstream AIM, possibly up to the Route Planner. This may decide to communicate the Road State to the Human-CAV Interaction Subsystem. |
| **Command Issuer** | Instructs the MAS to execute the Trajectory considering the Environment conditions and receives MAS-AMS Responses about the execution. Based on the Response the Road State may be communicated to the Obstacle Avoider. |
| **Decision Recorder** | Records decisions by Route Planner, Path Planner, Motion Planner, Obstacle Avoider, Command Issuer. |

## I/O Data of Autonomous Motion Subsystem’s AI Modules

Table 11 gives, for each AIM (1st column), the input data (2nd column) and the output data (3rd column) of Autonomous Motion Subsystem.

Table 11 - CAV Autonomous Motion Subsystem data

|  |  |  |
| --- | --- | --- |
| **CAV/AIM** | **Input** | **Output** |
| **Environment Representation Fusion** | Basic Environment Representations  from ESS or Remote AMSs  Other data from Remote AMSs | Full Environment Representation |
| **Remote AMSs** | See Table 15 | See Table 15 |
| **Route Planner** | Full Environment Representation  Offline maps  Road State | Route  Estimated time of arrival |
| **Path Planner** | Route  Full Environment Representation  Offline maps  Road State | Set of Paths |
| **Motion Planner** | Set of Paths  Full Environment Representation  Road State | Trajectory |
| **Obstacle Avoider** | Trajectory  Full Environment Representation  Alert  Road State | Trajectory |
| **Command Issuer** | Trajectory  MAS-AMS Message | AMS-MAS Message |
| **Decision Recorder** | Route  Path  Trajectory  AMS-MAS Message  MAS-AMS Message | Recorded Data |

## Data Types

### CAV Identifier

A code uniquely identifying a CAV.

The CAV identification system may 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**

Respondents are requested to comment on or extend or reformulate the functional requirements.

### Basic Environment Representation

Defined in the Environment Sensing Subsystem.

**To Respondents**

Respondents are requested to comment on the functional partitioning of Basic and Full Environment Representation.

### Full Environment Representation

The Autonomous Motion Subsystem produces the Full Environment Representation (FER) by *integrating*:

1. The BER from the Ego CAV.
2. BER-related information received from other CAVs in range or Roadside Units.

The AMS may harvest available bandwidth and utilise it to send a version of the Full Environment Representation that is compatible with the available mobile bandwidth.

The AMS reconciles the different values by considering, e.g., the distance of the ego CAV from the object vs the distance of the other CAV from the object.

The FER format shall satisfy the following requirements:

1. The FER shall be an extension of the BER.
2. Each element of the FER shall provide:
   1. The Identifier, Spatial Attitude, and Shape of the other CAVs.
   2. The value of major discrepancies between the elements of the Ego CAV’s BER and those of another CAV.
3. The Objects in the FER may include additional semantics, such as:
   1. Flags (e.g., warning coming from other CAVs and RSUs).
   2. Platooning information indicating which objects belong to a platoon.
   3. Priority (e.g., the vehicle is a police car, an ambulance, a vehicle carrying hazardous material).
4. The local Traffic Signal Configuration.
5. The FER data structure shall allow for actions that are:
   1. deliberative, e.g.: the Obstacle Avoider should have fast access to data in the FER to get information useful to assess whether a Trajectory to be implemented collides with the estimated trajectories of an object.
   2. reactive, e.g., the Obstacle Avoider needs to find a Trajectory that avoids a new object whose trajectory is estimated to collide with the CAV.

**To Respondents**

Respondents are requested to comment on or extend or reformulate the functional requirements.

### E-AMS to/from R-AMS messages

The E-AMS sends or receives messages to/from an R-AMS to .

The E-AMS may send a Message to be forwarded to a passenger of a Remote CAV.

**To Respondents**

Respondents are invited to:

1. Comment or elaborate on the HCI-Remote HCI message functional requirements identified above.
2. Propose coded representation formats of E-AMS to/from R-AMS messages.

### Offline map

Defined in the Environment Sensing Subsystem.

### Route

Route is a sequence of Way Points on the Offline Map.

### Traffic Signal Configuration

A Traffic Signal Configuration is the digital representation of the traffic signalisations used in the Environment, including:

1. Position and Orientation of the traffic signals in the Environment.
2. Semantics of the traffic signals.

**To Respondents**

Respondents are requested to comment on or extend or reformulate the functional requirements.

### Path

Path is a sequence of Poses 𝑝𝑖 = (xi, yi, zi, αi, βi, γi) in the Environment.

### Pose

Pose is the Position and Orientation of the CAV in the Environment, i.e., the Spatial Attitude without Velocity and Acceleration.

### Trajectory

Trajectory is the sequence of Spatial Attitudes *SAi* (*SA1, SA2,…, SAi*) and corresponding Times *ti (t1, t2, tj)* planned to reach a Goal.

The Spatial Attitudes composing the Trajectory shall be designed so that the time between two SAs of a Trajectory shall:

1. Require a level of resources affordable to the CAV.
2. Comply with the Traffic Rules.
3. Ensure a level of passenger comfort.

**To Respondents**

Respondents are requested to comment on or extend or reformulate the functional requirements of the Trajectory.

### Goal

A Goal is the Spatial Attitude planned to be reached at the end of a Decision Horizon.

### AMS-MAS Message

The AMS-MAS Message performs the function of instructing the Motion Actuation Subsystem (MAS) to:

1. *Change* its Spatial Attitude SAA at time tA to Spatial Attitude SAB at time tB by passing to the AMS the following data:
   1. Times: TA, TB
   2. Spatial Attitudes: SAA, SAB
2. *Discontinue* the Command as a result of a receiving an MAS-AMS Message containing information on the difficulty to execute an AMS-MAS Message expressed by a Road State.

**To Respondents**

Respondents are requested to comment on or extend or reformulate the functional requirements of the AMS-MAS Message.

### MAS-AMS Response

After receiving an AMS-MAS Message, the MAS may issue a series of MAS-AMS Messages at intermediate Poses and corresponding Times to inform the AMS about Command execution. An MAS-AMS Response contains the Spatial Attitude and Time until:

1. The Goal is reached at the prescribed Time TB with the CAV having the prescribed SAB, or
2. The MAS signals back to the AMS that:
   1. The AMS-MAS Message cannot be implemented, and
   2. The MAS discontinues the Implementation of the Message and sends back the Road State to the AMS.

Based on MAS-AMS Message, Command Issuer may send information to the Obstacle Avoider AIM that there is an anomalous situation. Depending on the severity of the situation, the Command Issuer:

1. Issues a new AMS-MAS Message, or
2. Sends the Road State to Obstacle Avoider and requests a new Trajectory.

Obstacle Avoider provides a new Trajectory or, depending on the severity of the situation, send the “Road State” to Motion Planner. This may send it to Path Planner which may send it to Route Planner which communicates the “Road State” to the Human-CAV Interaction Subsystem by issuing an AMS-HCI Command.

### Road State

Road State is information about the state of the road the CAV is going through inferred by the AMS from internally available information or received from an external source via a communication channel.

Road State includes the following information:

1. Time
2. The SA reached at Time.
3. Steering Wheels or Brakes do not respond to a command.
4. Information about Road state
   1. Ice.
   2. Wind.
   3. Water.
   4. Pothole.
5. Additional information that may be added to Road State is:
   1. Road blocked at waypoint wi
   2. Traffic jam at waypoint wi
   3. The road is slippery at waypoint wi
   4. Weather conditions at waypoint wi

**To Respondents**

Respondents are requested to comment on or extend or reformulate the functional requirements of Road State.

# Motion Actuation Subsystem (MAS)

## Functions of Motion Actuation Subsystem

The Motion Actuation Subsystem:

1. Transmits spatial and environmental information gathered from its sensors and mechanical subsystems to the Environment Sensing Subsystem.
2. Receives AMS-MAS Message from the Autonomous Motion Subsystem.
3. Translates AMS-MAS Message into specific Commands to its own mechanical subsystems, e.g., Brakes, Steering Wheel, and Wheel Motors.
4. Receives Responses from its mechanical subsystems.
5. Packages Responses into an MAS-AMS Message.
6. Sends MAS-AMS Messages to Autonomous Motion Subsystem.
7. During the execution, the MAS may send MAS-AMS Message containing a Road State.

## Reference Architecture of Motion Actuation Subsystem

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

The operation of the Motion Actuation Subsystem unfolds as follows:

1. Other Environment Data are transferred to the Environment Sensing Subsystem.
2. MAS Spatial Attitude Generation sends to Environment Sensing Subsystem the Ego CAV Spatial Attitude computed using Odometer, Speedometer, and Accelerometer.
3. AMS Command Interpreter interprets AMS-MAS Messages received from Autonomous Motion Subsystem and issues commands to Wheel Motors, Steering Wheel, and Brakes.
4. MAS Response Analyser interprets the responses received from Wheel Motors, Steering Wheel, and Brakes and sends MAS-AMS Responses to the Autonomous Motion Subsystem.

A diagram of a vehicle

Description automatically generated

*Figure 7 – Motion Actuation Subsystem Reference Model*

## I/O Data of Motion Actuation Subsystem

Table 12 gives the input/output data of Motion Actuation Subsystem.

Table 12 - I/O data of Motion Actuation Subsystem

|  |  |
| --- | --- |
| **Input** | **Comments** |
| Odometer | Provides distance data. |
| Speedometer | Provides instantaneous velocity. |
| Accelerometer | Provides instantaneous acceleration. |
| Other Environment data | Other environment data, e.g., humidity, pressure, temperat­ure. |
| AMS-MAS Message | High-level motion command. |
| Wheel Motor Response | Forces wheels rotation, gives feedback. |
| Steering Wheel Response | Moves wheels by an angle, gives feedback. |
| Brake Response | Acts on brakes, gives feedback. |
| **Output** | **Comments** |
| Spatial Attitude | Position, Orientation and their velocity and acceleration vectors. |
| Other Environment data | Environment data such as humidity, pressure, temperat­ure. |
| Wheel Motor Command | Forces wheels rotation, gives feedback. |
| Steering Wheel Command | Moves wheels by an angle, gives feedback. |
| Brakes Command | Acts on brakes, gives feedback. |
| MAS-AMS Message | Feedback from MAS Response Analyser during and after AMS-MAS Message ex­ecution. |

## Functions of Motion Actuation Subsystem’s AI Modules

Table 13 gives the AI Modules of Autonomous Motion Subsystem.

Table 13 - Functions of Motion Actuation Subsystem’s AI Modules

|  |  |
| --- | --- |
| **AIM** | **Function** |
| **MAS Spatial Attitude Generation** | Computes Ego CAV’s Spatial Attitude using odometer, speedometer, and accelerometer data. |
| **AMS Command Interpreter** | Receives, analyses, and actuates AMS-MAS Message into specific commands to Brakes, Steering Wheels, and Wheel Motors. |
| **MAS Response Analyser** | Receives and analyses responses from Brakes, Steering Wheel, and Wheel Motors.  Assembles and forwards the MAS-AMS Response to AMS. |

## I/O Data of Motion Actuation Subsystem’s AI Modules

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

Table 14 - I/O Data of Motion Actuation Subsystem’s AI Modules

|  |  |  |
| --- | --- | --- |
| **CAV/AIM** | **Input** | **Output** |
| **MAS Spatial Attitude Generation** | Odometer  Speedometer  Accelerometer | Spatial Attitude |
| **AMS Command Interpreter** | AMS-MAS Message | Brake Command  Wheel Motor Command  Steering Wheel Command |
| **MAS Response Analyser** | Brake Response  Steering Wheel Response  Wheel Motor Response | MAS-AMS Message |

## Data Types

### Odometer Data

The Odometer converts the number of wheel rotations times the tire circumference (π x tire diameter) from the start up to the point being considered and provides the distance travelled.

### Speedometer Data

Data produced by a sensor that measures the instantaneous velocity of the CAV.

### Accelerometer Data

Dara provided by the electronic accelerometer sensor measuring the acceleration forces acting on the CAV. An accelerometer measures proper acceleration (the acceleration of a body in its own instantaneous rest frame), not the coordinate acceleration (the acceleration in a fixed coordinate system). A CAV accelerometer at rest measures an acceleration g straight upwards of ~9.81 m/s2. In free fall (falling toward the centre of the Earth at ≈ 9.81 m/s2) the acceleration measures zero.

### Spatial Attitude

The MAS may be able to produce an estimate of the Spatial Attitude of the ego CAV using its sensors.

### Other Environment Data

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

### AMS-MAS Message

The Message issued by the Autonomous Motion Subsystem (see AMS-MAS Message).

### MAS-AMS Message

The Message of Command Interpreter summarising the Feedback from the MAS Subsystems.

### Messages between MAS and its subsystem

#### Brakes Command

The result of the interpretation of AMS-MAS Message to Brakes.

#### Brakes Response

The feedback of Brakes to MAS Response Analyser.

#### Steering Wheel Command

The result of the interpretation of AMS-MAS Message to Steering Wheel.

#### Steering Wheel Response

The feedback of Steering Wheel to MAS Response Analyser.

#### Wheel Motor Command

The result of the interpretation of AMS-MAS Message to Wheel Motors.

#### Wheel Motor Response

The feedback of Wheel Motor to MAS Response Analyser.

# CAV-to-Everything (V2X) (Informative)

To enhance its own capabilities to perceive the Environment, a CAV exchanges information via radio with other entities, e.g., Remote AMSs may be achieved via secure channels.

V2X is the CAV component that allows the CAV Subsystems – Human-CAV Interaction and Autonomous Motion Subsystems – to communicate to peer entities external to the Ego CAV. For instance, the HCI of a CAV may send/request information to/from the HCI of another CAV or an AMS may send/request the BER to/from the AMS of another CAV.

## V2X Communication

V2X may be achieved via different communication protocols. A CAV may use multicast or unicast to send CAVs in range information it has become aware of by using the V2X communication interface. For instance, while executing a Command, the MAS of CAV may become aware of ice on the road. The AMS may decide to broadcast that information to CAVs in range. Unicast may be used for communication between the Ego CAV and a specific CAV.

Multicast Communication may be used when a CAV broadcasts large amounts of data such as the Basic Environment Representation (BER). The BER data format should be scalable to enable transmission of the more important elements of the BER in case the number of CAVs is large.

Communication with Remote CAVs is managed by a Communication Device. Here is an example of the operation flow of the Communication Device communicating with Remote CAVs:

1. Receive identities broadcasted by Remote AMSs.
2. Establish unicast sessions with Remote AMSs.
3. Create a list of Remote CAVs with which it has established a session.
4. Sends the Ego CAV’s information to Remote AMSs.

## Input and output data

### CAVs within range

Table 15 gives the Data Types a CAV broadcasts to CAVs in range via its Communication Device.

Table 15 - I/O data of CAV’s Communication Device

|  |  |  |
| --- | --- | --- |
| **Input Data** | **From** | **Comments** |
| CAV Identity | Remote AMSs | In principle, this should be the digital equivalent of today’s plate number including Manufacturer and Model information. |
| Basic Environment Representation | Remote AMSs | A scalable subset of the digital representation of the Environment created by the ESS. |
| Information Messages | Remote AMSs and/or HCIs | Possible messages received by a CAV:   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** |
| CAV Identity | Remote AMSs | In principle, this should be the digital equivalent of today’s plate number including Manufacturer and Model information. |
| Basic Environment Representation | Remote AMSs | Same as input for all other input data. BER accuracy depends on available bandwidth. |
| Information Messages | Remote AMSs and/or HCIs | Typical messages as above. |

### CAV-aware equipment

Examples are traffic lights, roadside units, and vehicles with CAV communication capabilities. The following data may be included in an exchange between CAVs:

1. Identity and Spatial Attitude.

Static Full Environment Representation that is regularly updated.

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.)

CAV-aware equipment:

1. Can act as any other CAV in range.
2. May have the authority to organise motion of CAVs in range.

### Other non-CAV vehicles

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

### Pedestrians

Their smartphones can transmit their Spatial Attitudes. No response capability is expected.

1. MPAI-wide terms and definitions

The Terms used in this standard whose first letter is capital and are not already included in **Error! Reference source not found.** 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 Modules (AIM) | A data 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 AIW-specific inputs and producing AIW-specific outputs according to the AIW Function. |
| Application Standard | An MPAI Standard designed to enable a particular application domain. |
| Channel | A connection between an output port of an AIM and an input port of an AIM. The term “connection” is also used as synonymous. |
| Communication | The infrastructure that implements message passing between AIMs |
| Composite AIM | An AIM aggregating more than one AIM. |
| Component | One of the 7 AIF elements: Access, Communication, Controller, Internal Storage, Global Storage, Store, and User Agent |
| Conformance | The attribute of an Implementation of being a correct technical Implem­entation of a Technical Specification. |
| Conformance Tester | An entity Testing 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 Format | The standard digital representation of data. |
| Data Semantics | The meaning of data. |
| Ecosystem | The ensemble of actors making it possible for a User to execute an application composed of an AIF, one or more AIWs, each with one or more AIMs potentially sourced from independent implementers. |
| 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. |
| Internal Storage | A Component to store data of the individual 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) conforming with a Use Case of an MPAI Applic­ation Standard. |
| Implementer | A legal entity implementing MPAI Technical Specifications. |
| ImplementerID (IID) | A unique name assigned by the ImplementerID Registration Authority to an Implementer. |
| ImplementerID Registration Authority (IIDRA) | The entity appointed by MPAI to assign ImplementerID’s to Implementers. |
| Interoperability | The ability to functionally replace an AIM with another 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:   1. Be proprietary (Level 1) 2. Pass the Conformance Tes­ting (Level 2) of an Applic­ation Standard 3. Pass the Perform­ance Testing (Level 3) of an Applic­ation Standard. |
| Knowledge Base | Structured and/or unstructured information made accessible to AIMs via MPAI-specified interfaces |
| Message | A sequence of Records transported by Communication through Channels. |
| 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 Means 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 Assessing the Performance of an Implementation. |
| 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 | A data structure 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. |
| 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 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. |

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 of any kind 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/standards/mpai-gme/) 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.