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

|  |
| --- |
| **Public Document** |
| **N794** | 2022/07/19 |
| **Source** | Requirements (OSD) |
| **Title** | MPAI-OSD report and plans |
| **Target** | MPAI Members |

Contents

[1 Introduction 1](#_Toc108956272)

[2 Connected Autonomous Vehicles 1](#_Toc108956273)

[2.1 Autonomous Motion Subsystem 1](#_Toc108956274)

[2.2 Environment Sensing Subsystem 1](#_Toc108956275)

[2.3 Human-Cav Interaction 2](#_Toc108956276)

[3 Mixed-reality Collaborative Spaces 3](#_Toc108956277)

[3.1 Avatar-Based Videoconference 3](#_Toc108956278)

[3.1.1 MCS-ABV Transmitting Client 3](#_Toc108956279)

[3.1.2 MCS-ABV Virtual Secretary 4](#_Toc108956280)

[3.1.3 MCS-ABV Receiving Client 4](#_Toc108956281)

[3.2 Conversation About a Scene 5](#_Toc108956282)

[4 Multimodal Conversation 5](#_Toc108956283)

[4.1 Personal Status Extraction 5](#_Toc108956284)

# Introduction

This document analyses the Use Cases, and the Composite AIMs addressing areas covered by Requirements (OSD) activities. The relevant AIMs are depicted in ocra.

# Connected Autonomous Vehicles

*Connected Autonomous Vehicles* (MPAI-CAV) is an MPAI standard project seeking to define identified CAV standard components and their interfaces.

## Autonomous Motion Subsystem

Autonomous Motion Subsystem (AMS) issues commands to drive the CAV to the intended destination.

******

Figure 1 - CAV-AMS Reference Model

## Environment Sensing Subsystem

Environment Sensing Subsystem (ESS) acquires and processes information from the Environment via a variety of sensors.

The stack of AIMs on the left-hand side processes raw data from sensors to facilitate the subsequence process of Descriptor Extraction carried out by the AIM stach in the middle.



Figure 2 - CAV-EES Reference Model

## Human-Cav Interaction

Human-CAV Interaction (HCI) Handles human-CAV interactions.

The main processes are: Visual Scene Description, Object Description, Object Interpretation, Face Description and Face Recognition.



Figure 3 - CAV-HCI Reference Model

# Mixed-reality Collaborative Spaces

## Avatar-Based Videoconference

A system allowing human participants to attend a videoconference represented by their avatars who accurately represents their features and movements.

MCS-ABV is composed of 4 subsystems 5 of which use technologies falling within the Require­ments (OSD) activities.

### MCS-ABV Transmitting Client

The purpose of the MCS-ABV Transmitting Client is to create an accurate set of descriptors to be used by Receiving Clients.

The main processes are: Visual Scene Description, Gesture Description, Face Description1, and Face Description2.



Figure - MCS-ABV (Transmitting Client) Reference Model

### MCS-ABV Virtual Secretary

The purpose of the MCS-ABV Virtual Secretary is to create a summary of the utterances of the avatars using the Personal Status Descriptor to improve its understanding.



Figure - MCS-ABV Virtual Secretary Reference Model

### MCS-ABV Receiving Client

The purpose of the MCS-ABV Receiving Client is to allow participants to dispose the avatars participating in the videoconference preserving the source of their utterances.

The main processes are: Visual Scene Creation and AV Scene Viewer.



Figure - MCS-ABV (Receiving Client) Reference Model

## Conversation About a Scene

In MCS-CAS a machine converses with a human by understanding their gesture (finger pointing to an object) and Personal Status

The main processes are: Visual Scene Description, Object Description, Gesture Description, and Scene Presentation,



Figure - MMC-CAS Reference Model

# Multimodal Conversation

## Personal Status Extraction

The purpose of MMC-PSE is to extract the Personal Status of Text, Speech, Face, and Gesture.

The main processes are: PS-Face Description and PS-Gesture Description.



Figure 7 - MMC-PSE Reference Model