



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

MPAI Technical Specification

Portable Avatar Format MPAI-PAF

V1.1

WARNING

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

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

Readers are invited to review Annex 3 - Notices and Disclaimers.

Technical Specification

Portable Avatar Format (MPAI-PAF) V1

1	Introduction (Informative)	3
2	Scope	6
3	Terms and Definitions	7
4	References	9
4.1	Normative References	9
4.2	Informative References	10
5	Avatar-Based Videoconference	10
5.1	Introduction	10
5.2	Scope of Use Case	11
5.3	Videoconference Client Transmitter	13
5.3.1	Functions of Client Transmitter	13
5.3.2	Reference Model of Client Transmitter	13
5.3.3	Input and Output Data of Client Transmitter	14
5.3.4	Functions of Client Transmitter's AI Modules	14
5.3.5	I/O Data of Client Transmitter's AI Modules	15
5.3.6	Specification of Videoconference Client Transmitter AIMs and JSON Metadata ..	15
5.4	Avatar Videoconference Server	16
5.4.1	Functions of Avatar Videoconference Server	16
5.4.2	Reference Model of Avatar Videoconference Server	17
5.4.3	I/O Data of Avatar Videoconference Server	17
5.4.4	Functions of Avatar Videoconference Server AI Modules	18
5.4.5	I/O Data of Avatar Videoconference Server AI Modules	18
5.4.6	Specification of Avatar Videoconference Server AIMs and JSON Metadata	18
5.5	Virtual Meeting Secretary (Informative)	19
5.5.1	Functions of Virtual Meeting Secretary	19
5.5.2	Reference Model	19
5.5.3	I/O Data of Virtual Meeting Secretary	20
5.5.4	AIW and AIM Specifications and JSON Metadata	20
5.6	Videoconference Client Receiver	21
5.6.1	Functions of Client Receiver	21
5.6.2	Reference Model of Client Receiver	21
5.6.3	I/O Data of Client Receiver	22
5.6.4	Functions of Videoconference Client Receiver's AI Modules	22
5.6.5	I/O Data of Videoconference Client Receiver's AI Modules	22
5.6.6	Specifications of Videoconference Client Receiver's AIMs and JSON Metadata ..	23
6	Composite AI Modules	23
6.1	Personal Status Display (PSD)	23
6.1.1	Functions of Composite AIM	23
6.1.2	Reference Model	24
6.1.3	I/O Data of Personal Status Display	24
6.1.4	Functions of AI Modules of Personal Status Display	24
6.1.5	I/O Data of AI Modules of Personal Status Display	25
6.1.6	Specification of Personal Status Display AIMs and JSON Metadata	25
7	Data Formats	25

7.1	Space, Objects and Scenes	27
7.1.1	Coordinate Systems.....	27
7.1.2	Spatial Attitude and Point of View	27
7.1.3	Audio Objects and Scene.....	27
7.1.4	Visual Objects and Scene	27
7.1.5	Audio-Visual Objects and Scene	27
7.2	Portable Avatar	27
7.2.1	Syntax.....	27
7.2.2	Semantics.....	29
7.3	Body	29
7.3.1	Body Model	29
7.3.2	Body Descriptors.....	31
7.4	Face	31
7.4.1	Face Model	31
7.4.2	Face Descriptors.....	31
7.5	Avatar.....	32
7.5.1	Avatar Model	32
7.5.2	Avatar Descriptors	32
7.6	Other Data Types	32
7.6.1	Text.....	32
7.6.2	Language identifier	32
7.6.3	Meaning	32
7.6.4	Personal Status	32
Annex 1 - MPAI Basics (Informative)		33
1	General.....	33
2	Governance of the MPAI Ecosystem.....	33
3	AI Framework	34
4	Space, Objects, and Scenes	35
Annex 2 - General MPAI Terminology.....		36
Annex 3 - Notices and Disclaimers Concerning MPAI Standards (Informative).....		39
Annex 4 - Patent Declarations.....		41

1 Introduction (Informative)

There is a long history of computer-created objects called “digital humans”, i.e., digital objects having a human appearance when rendered. In most cases the underlying assumption of these objects has been that creation, animation, and rendering is done in a closed environment. Such digital humans had little or no need for standards.

In a communication and more so in a metaverse context, there are many cases where a digital human is not constrained within a closed environment thus requiring forms of standardisation. *Technical Specification: Portable Avatar Format (MPAI-PAF)* – in the following also called MPAI-PAF – is a first response to the requirements of new usage contexts. MPAI-PAF specifies a standard for Portable Avatar Format (PAF) enabling a receiving party to render a digital human as intended by the sending party.

MPAI-PAF has been developed by MPAI – Moving Picture, Audio, and Data Coding by Artificial Intelligence [12], the international, unaffiliated, non-profit organisation developing standards for Artificial Intelligence (AI)-based data coding with clear Intellectual Property Rights licensing frameworks in compliance with the rigorous MPAI Process [13] in pursuit of the following policies:

1. Be friendly to the AI context but, to the extent possible, agnostic to the technology – AI or Data Processing – used in an implementation.
2. Be attractive to different industries, end users, and regulators.
3. Address three levels of standardisation: data types, components (called AI Modules), configurations of components (called AI Workflows) all exposing standard interfaces with an aggregation level decided by the implementer.
4. Specify the data exchanged by components with a clear semantic to the extent possible.

As manager of the MPAI Ecosystem specified by Governance of MPAI Ecosystem (MPAI-GME) [1], MPAI ensures that a user can:

1. Operate the reference implementation of the Technical Specification, by providing a Reference Software Specification with annexed software.
2. Test the conformance of an implementation with the Technical Specification, by providing Conformance Testing Specification.
3. Assess the performance of an implementation of a Technical Specification, by providing the Performance Assessment Specification.
4. Get conforming implementations possibly with a performance assessment report from a trusted source through the MPAI Store.

Technical Specification: AI Framework (MPAI-AIF) V2 [2] enables an effective implementation of some of the policies. Figure 1 depicts the Reference Model.

MPAI-AIF specifies an environment called AI Framework (AIF) enabling the secure execution of AI Workflows (AIW) that can be constituted by AI Modules (AIM). Thus, users can have machines whose internal operation they understand to some degree, rather than machines that are just “black boxes” resulting from unknown training with unknown data and component developers can provide components with standard interfaces that can have improved performance compared to other implementations.

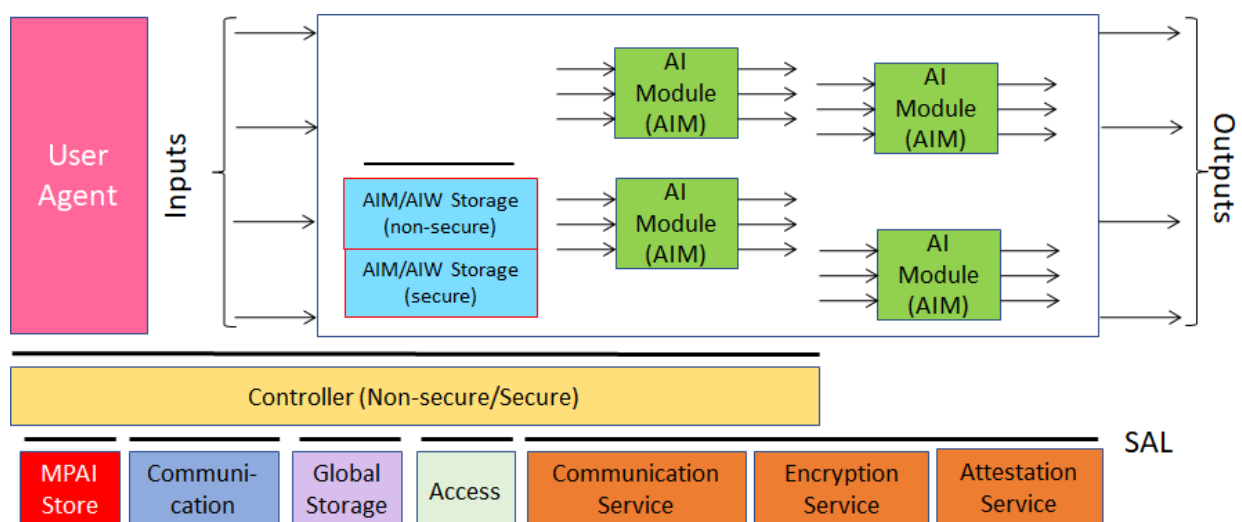


Figure 1 - The AI Framework (MPAI-AIF) V2 Reference Model

An AIW and its AIMs may have 3 interoperability levels:

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

Level 2 – Specified by an MPAI Application Standard.

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

MPAI offers Users access to the promised benefits of AI with a guarantee of increased transparency, trust and reliability as the Interoperability Level of an Implementation moves from 1 to 3.

AI Modules can execute data processing or Artificial Intelligence algorithms and can be implemented in hardware, software, or hybrid hardware/software.

The MPAI-PAF Technical Specification can be implemented in one of the following modalities:

1. As a specific AIW implementing a Use Case, as specified in this document.
2. As a specific AIM, as specified in this document.
3. As a specific data type, as specified in this document.

However, MPAI does not mandate the choice of modality, which remains the sole decision of the implementer.

The MPAI-PAF Technical Specification will be accompanied by the Reference Software, Conformance Testing, and Performance Assessment Specifications. Conformance Testing specifies methods enabling users to ascertain whether a data type generated by an AIM, an AIM, or an AIW conform with this Technical Specification.

The MPAI-PAF Technical Specification applies the technologies to the Avatar-Based Videoconference (PAF-ABV) Use Case where:

1. **Client Transmitters** send PAFs containing:
 - 1.1 At the beginning: Avatar Models and Language Preferences.
 - 1.2 Continuously: Avatar Descriptors, and Speech and Face Objects to a Server.
2. **Avatar Videoconference Server:**
 - 2.1 At the beginning:
 - 2.1.1 Selects an Environment, i.e., a meeting room and equips it with objects, i.e., meeting table and chairs.
 - 2.1.2 Places Avatar Models around the table.
 - 2.1.3 Distributes a PAF containing Environment, Avatar Models, and their positions to receiving clients.
 - 2.2 Continuously sends to receiving clients:
 - 2.2.1 Translated Speech from participants according to Language Preferences.
 - 2.2.2 Sends PAFs containing Avatar Descriptors and translated Speech.
3. **Client Receivers:**
 - 3.1 At the beginning: receive Environment and PAFs containing Avatar Models and Language Preferences
 - 3.2 Continuously:
 - 3.2.1 Receive PAFs containing Avatar Descriptors and translated Speech.
 - 3.2.2 Create Audio and Visual Scene Descriptors.
 - 3.2.3 Render the Audio-Visual Scene as seen from the human-selected Point of View.

MPAI-PAF utilises MPAI technologies specified by other MPAI standards to specify the Avatar-Based Videoconference Use Case. Similarly, other MPAI standards utilise standard MPAI-PAF technologies in other Use Cases such as Human-Connected Autonomous Vehicle (CAV) Interaction (CAV-HCI).

Chapters, Sections, and Annexes are Normative unless they are explicitly identified as Informative.

2 Scope

Technical Specification: Portable Avatar Format (MPAI-PAF) specifies:

1. The Portable Avatar Format and related Data Formats allowing a sender to enable a receiver to decode and render an Avatar as intended by the sender.
2. The Personal Status Display Composite AI Module allowing the conversion of a Text and a Personal Status to a Portable Avatar.
3. The AI Framework (MPAI-AIF)-conforming AI Workflows and AI Modules [2] composing the Avatar-Based Videoconference Use Case also using Data Types from other MPAI Technical Specifications.

The Use Case normatively defines:

1. The Functions of the AIWs and of the AIMs.
2. The Connections between and among the AIMs
3. The Semantics and the Formats of the input and output data of the AIW and the AIMs, in particular Avatar Descriptors.

The word *normatively* implies that an Implementation claiming Conformance to:

1. An *AIW*, shall:
 - a. Perform the function specified in the appropriate Section of Chapter 5.
 - b. Use AIMs connected with the topology and connections conform with the AIW Architecture specified in Chapter 5.
 - c. Have input and output data with the formats specified in Chapter 7.
2. An *AIM*, shall:
 - a. Perform the AIM function specified by the appropriate section of Chapter 5.
 - b. Be composed of AIMs as specified in Chapter 6.
 - c. Receive and produce the data specified in Chapter 7.
3. A data *Format*, the data shall have the format specified in Chapter 7.

Users of this Technical Specification should note that:

1. This Technical Specification defines Interoperability Levels but does not mandate any.
2. Implementers decide the Interoperability Level their Implementation satisfies.
3. Implementers can use the Reference Software of this Technical Specification to develop their Implementations.
4. The Conformance Testing specification can be used to test the conformity of an Implementation to this Standard.
5. Performance Assessors can assess the level of Performance of an Implementation based on the Performance Assessment associated with the MPAI-PAF Technical Specification.
6. Implementers and Users should consider the notices and disclaimers of Annex 2.

This Technical Specification includes the following elements:

1. Scope (This Chapter)
2. Terms and Definitions
3. References
4. Use Case (Avatar-Base Videoconference)
5. Composite AI Modules
6. Data Formats.

The current version of the Standard has been developed by the Portable Avatar Format Development Committee (PAF-DC). MPAI may issue new versions of MPAI-PAF that extend or replace the current Technical Specification.

3 Terms and Definitions

In this document, the following conventions apply:

1. Words beginning with a capital letter are defined in *Table 1*.
2. Words beginning with a small letter have the normal meaning consistent with the relevant context.

In *Table 1*, a dash “-” preceding a Term means the following:

1. If the font of the Term 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 of the Term 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.

Table 1 – Terms and Definitions

Term	Definition
Attitude	
- <i>Social</i>	A Factor of the Personal Status related to the way a human or Avatar intends to position vis-à-vis the Environment or subsets of it, e.g., “Respectful”, “Confrontational”, “Soothing”.
- <i>Spatial</i>	Position and Orientation and their velocities and accelerations of a Human and Physical Object in a Digital Environment.
Audio	Digital representation of an analogue audio signal sampled at a frequency between 8-192 kHz with a number of bits/sample between 8 and 32, and non-linear and linear quantisation.
Authentica- tion	The process of determining whether a device or a human is what it states it is.
Avatar	A rendered Digital Human.
- Model	An inanimate Avatar exposing animation interfaces.
- <i>Portable</i>	A Data Type including Avatar ID, Time, Audio-Visual Scene Descriptors, Spatial Attitude, Avatar Model, Body Descriptors, Face Descriptors, Language Preference, Speech Coding, Speech Data, Text, and Personal Status.
Body	A digital representation of a human body, head included, face excluded.
Centre Point	The point of an Object selected to have Local Coordinates (0,0,0).
Cognitive State	An element of the internal status reflecting the way a human or avatar understands the Environment, such as “Confused”, “Dubious”, “Convinced”.
Context	Additional information about a communication emitted by an Entity, such as language, culture etc.
Data	Information in digital form.
- Format	The syntax and semantics of a Data Type.
- Type	A particular type of Data.
Descriptor	The Digital Representation of a feature of an Object.
- <i>Body</i>	A Data Type including the digital representation of the features of the body of a real or digital human.
- <i>Face</i>	A Data Type including the digital representation of a feature of the face of a real or digital human.

Device	A piece of equipment used to interact and have Experience in a Digital Environment.
Digital Representation	Data corresponding to and representing a physical entity.
Emotion	The coded representation of the internal state resulting from the interaction of a human or avatar with the Environment or subsets of it, such as “Angry”, “Sad”, “Determined”.
Entity	A real or Digital Human
Environment	A Virtual Space that may be null or may include an Audio-Visual Scene.
Experience	The state of a human whose senses are continuously affected for a meaningful period.
Face	A digital representation of a human face.
Factor	One of Emotion, Cognitive State, and Spatial Attitude.
Gesture	A movement of a Digital Human or part of it, such as the head, arm, hand, and finger, often a complement to a vocal utterance.
Grade	The intensity of a Factor.
Human	
<i>Digital</i>	A Digitised or a Virtual Human in a Virtual Space.
<i>Digitised</i>	An Object in a Virtual Space that has the appearance of a specific human when rendered.
<i>Virtual</i>	An Object in a Virtual Space created by a computer that has a human appearance when rendered but is not a Digitised Human.
Identifier	The label uniquely associated with a human or an Object.
Modality	One of Text, Speech, Face, or Gesture.
Object	A data structure that can be rendered to cause an Experience.
- <i>Audio</i>	Coded representation of Audio information with its metadata. An Audio Object can include other Audio Objects.
- <i>Audio-Visual</i>	Coded representation of Audio-Visual information with its metadata.
- <i>Descriptor</i>	The digital representation of the feature of an Object.
- <i>Digital</i>	A Digitised or a Virtual Object.
- <i>Digitised</i>	The digital representation of a real object.
- <i>Visual</i>	Coded representation of Visual information with its metadata. A Visual Object can include other Visual Objects.
- <i>Virtual</i>	An Object not representing an object in a Real Environment.
Orientation	The 3 Euler angles of an Object in a Virtual Space.
Personal Status	A Data Type including three Factors – Cognitive State, Emotion and Social Attitude – conveyed by four Modalities – Text, Speech, Face, and Gesture and providing standard extensible labels for the three Factors.
Point of View	The Spatial Attitude of a Digital Human watching an Environment.
Portable Avatar	A Data Type representing an Avatar and its Context.
Position	The coordinates of a representative point for an object in a Virtual Space with respect to a set of coordinate axes.
Rendering	The process of instantiating a Virtual Space as a human-perceptible entity.
Scene	A Digital Environment populated by Objects.
- <i>Audio</i>	The Audio Objects of an Environment with Object metadata such as Spatial Attitude.

- <i>Audio-Visual</i>	(AV Scene) The Audio-Visual Objects of an Environment Object metadata such as Spatial Attitude.
- <i>Visual</i>	The Visual Objects of an Environment with Object metadata such as Spatial Attitude.
Scene Descriptors	The digital representation of a feature of a scene.
- <i>Audio</i>	A Data Type including the digital representation of the audio features of a real or digital scene.
- <i>Audio-Visual</i>	A Data Type combining the Audio or Visual Scene Descriptors.
- <i>Visual</i>	A Data Type including the digital representation of the visual features of a real or digital scene.
Representation	Data that digitally represents an entity of a real environment.
Scene Geometry	The digital representation of the object arrangement of a scene.
- <i>Audio</i>	A Data Type describing the spatial arrangement of the Visual Objects of a Scene.
- <i>Audio-Visual</i>	A Data Type describing the spatial arrangement of the Audio, Visual, and Audio-Visual Objects of a Scene.
- <i>Visual</i>	A Data Type describing the spatial arrangement of the Visual Objects of a Scene.
Speech	Digital representation of analogue speech sampled at a frequency between 8 kHz and 96 kHz with a number of bits/sample of 8, 16 or 24, and non-linear and linear quantisation or compressed. Data with characteristics of Speech may be synthetically produced.
Text	A sequence of characters represented according to [11].
Virtual Space	A space generated and maintained by a computing platform that can be rendered.

4 References

4.1 Normative References

Technical Specification: Portable Avatar Format (MPAI-PAF) normatively references the following documents, both from MPAI and other standards organisations. Referenced MPAI standards are publicly available at the URL indicated in the reference.

1. MPAI; Technical Specification: The governance of the MPAI ecosystem (MPAI-GME), V1.1; <https://mpai.community/standards/mpai-gme/>
2. MPAI; Technical Specification; AI Framework (MPAI-AIF) V2.0; <https://mpai.community/standards/mpai-aif/>
3. MPAI; Technical Specification: Context-based Audio Enhancement (MPAI-CAE) V2.1; <https://mpai.community/standards/mpai-cae/>
4. MPAI; Technical Specification; Multimodal Conversation (MPAI-MMC) V2.1; <https://mpai.community/standards/mpai-mmc/>
5. MPAI; Technical Specification; Object and Scene Description (MPAI-OSD) V1.0; <https://mpai.community/standards/mpai-osd/>
6. Khronos; Graphics Language Transmission Format (glTF); October 2021; <https://registry.khronos.org/glTF/specs/2.0/glTF-2.0.html>

7. 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>
9. ISO/IEC 19774-2:2019 Information technology – Computer graphics, image processing and environmental data representation – Part 2: Humanoid animation (HAnim) motion data animation; <https://www.web3d.org/documents/specifications/19774/V2.0/MotionDataAnimation/MotionDataAnimation.html>
10. ISO 639; Codes for the Representation of Names of Languages — Part 1: Alpha-2 Code.
11. ISO/IEC 10646; Information technology – Universal Coded Character Set

4.2 Informative References

These references are provided for information purposes.

12. MPAI; The MPAI Statutes; <https://mpai.community/statutes/>.
13. MPAI; The MPAI Patent Policy; <https://mpai.community/about/the-mpai-patent-policy/>.
14. MPAI; Framework Licence: Portable Avatar Format; <https://mpai.community/standards/mpai-paf/>
15. Ekman, Paul (1999), "Basic Emotions", in Dalgleish, T; Power, M (eds.), Handbook of Cognition and Emotion (PDF), Sussex, UK: John Wiley & Sons.

5 Avatar-Based Videoconference

5.1 Introduction

The Portable Avatar Format enables an implementation of the Avatar-Based Videoconference Use Case. Table 2 lists all Data Types required by the Portable Avatar Data Type and MPAI Technical Specifications supporting them. Chapter 7 provides the full specification including references of the MPAI-PAF Data Formats.

Table 2 - Data Types of Technical Specification: Portable Avatar Format

Section	Data Type	Technical Specification
7.1	Space, Objects and Scenes	
7.1.1	Coordinate Systems	MPAI-PAF
7.1.2	Spatial Attitude and Point of View	MPAI-OSD
7.1.3	Audio Scene	MPAI-CAE
7.1.4	Visual Scene	MPAI-OSD
7.1.5	Audio-Visual	MPAI-OSD
7.2	Portable Avatar Format	MPAI-PAF
7.3	Body	
7.3.1	Body Model	MPAI-PAF
7.3.2	Body Descriptors	MPAI-PAF
7.4	Face	
7.4.1	Face Model	MPAI-PAF
7.4.2	Face Descriptors	MPAI-PAF
7.5	Avatar	
7.5.1	Avatar Model	MPAI-PAF
7.5.2	Avatar Descriptors	MPAI-PAF

7.6	Other Data Types	
7.6.1	Text	MPAI-MMC
7.6.2	Language identifier	MPAI-MMC
7.6.3	Meaning	MPAI-MMC
7.6.4	Personal Status	MPAI-MMC

5.2 Scope of Use Case

The MPAI-PAF Avatar-Based Videoconference (PAF-ABV) Use Case enables a form of videoconference held in a Virtual Environment populated by Avatars representing humans showing their visual appearance and uttering their voices. Figure 2 depicts the system composed of four types of subsystems:

1. Videoconference Client Transmitters
2. Avatar Videoconference Server
3. Virtual Meeting Secretary
4. Videoconference Client Receivers.

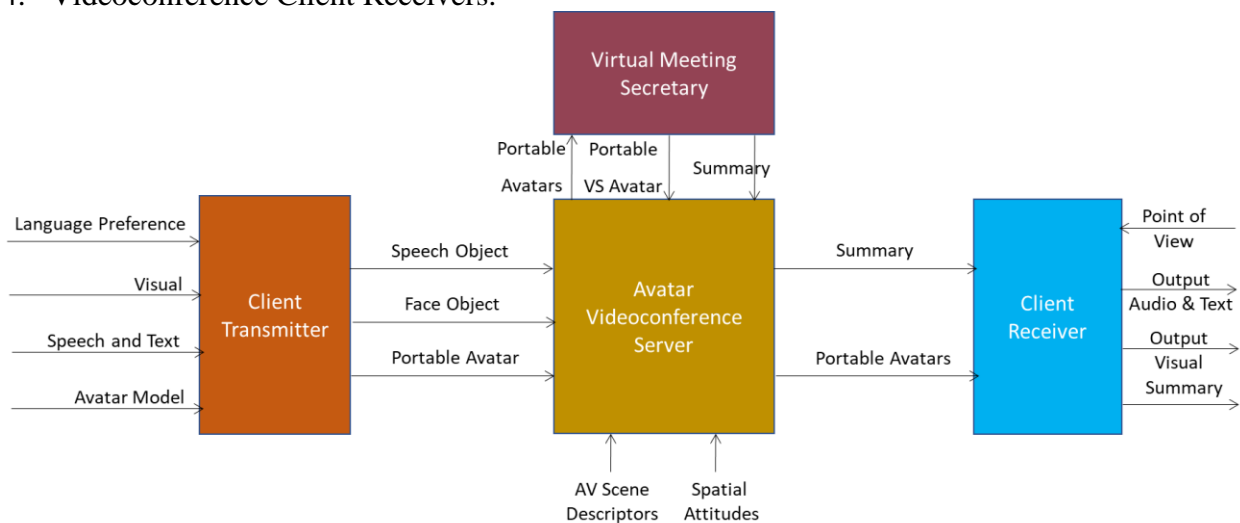


Figure 2 – Avatar-Based Videoconference end-to-end diagram

The components of the PAF-ABV system:

1. **Participant:** a human joining an ABV either individually or as a member of a group of humans in the same physical room.
2. **Audio-Visual Scene:** a virtual audio-visual space equipped with Visual Objects such as a table and an appropriate number of chairs and Audio Objects described by Audio-Visual Scene Descriptors.
3. **Portable Avatar:** represents a human participant represented in the Portable Avatar Format (PAF).
4. **Client Transmitter:**
 - 4.1. At the beginning of the conference,
 - 4.1.1. Receives from Participants and sends to the Server Portable Avatars containing the Avatar Models and Language Preferences.
 - 4.1.2. Sends to the Server Speech Object and Face Object for Authentication.
 - 4.2. Continuously sends to the Server Portable Avatars containing Avatar Descriptors and Speech.
5. **The Avatar Videoconference Server**
 - 5.1. At the beginning:
 - 5.1.1. Selects the Visual Descriptors, e.g., a meeting room.
 - 5.1.2. Equips the room with objects, i.e., meeting table and chairs.

- 5.1.3. Places Avatar Models around the table with a given Spatial Attitude.
- 5.1.4. Distributes Environment and Portable Avatars containing Avatars Models, and their Spatial Attitudes to all Receiving Clients.
- 5.1.5. Authenticates Speech and Face Objects and assigns IDs to Avatars.
- 5.1.6. Sets the common conference language.
- 5.2. Continuously:
 - 5.2.1. Translates Speech to Participants according to their Language Preferences.
 - 5.2.2. Sends Portable Avatars containing Avatar Descriptors, Speech, and Spatial Attitude of Participants and Virtual Meeting Secretary to all Receiving Clients and Virtual Meeting Secretary.
6. **Virtual Meeting Secretary** is an Avatar not corresponding to any Participant that continuously:
 - 6.1. Uses the common meeting language.
 - 6.2. Understands Avatars' utterances and extracts their Personal Statuses.
 - 6.3. Drafts a Summary of its understanding of Avatars' Text and Personal Status.
 - 6.4. Displays the Summary either to:
 - 6.4.1. Outside of the virtual meeting room for Participants to read and edit directly, or
 - 6.4.2. The Visual Space for Avatars to comment, e.g., via Text.
 - 6.5. Refines the Summary.
 - 6.6. Sends its Portable Avatar containing its Avatar Descriptors to the Server.
7. **Client Receiver:**
 - 7.1. At the beginning
 - 7.2. Receives Visual Scene Descriptors and Portable Avatars containing Avatar Models with Spatial Attitudes.
 - 7.3. Continuously:
 - 7.3.1. Receives Portable Avatars with Avatar Descriptors and Speech.
 - 7.3.2. Produces Visual and Audio Scene Descriptors.
 - 7.3.3. Renders the Audio-Visual Scene by spatially adding the participants' utterances to the Spatial Attitude of the respective Avatars' mouths. Rendering may be done from a Point of View, possibly different from the position assigned to their Avatars in the Visual Space, selected by participant who use a device of their choice (HMD or 2D display/earpad).

Each component of the Avatar-Based Videoconference Use Case is implemented as an AI Workflow (AIW) composed of AI Modules (AIMs). It includes the following elements:

Each Use Case is implemented as an AI Workflow (AIW) composed of AI Modules (AIMs) and includes the following elements:

1	Functions of the AIW	The functions performed by the AIW implementing the MPAI-PAF Subsystem.
2	Reference Model of the AIW	The Topology of AIMs in the AIW.
3	Input and Output Data of the AIW	Input and Output Data of the AIW.
4	Functions of the AIMs	Functions performed by all AIMs of the AIW.
5	Input and Output Data of the AIMs	Input and Output Data of all AIMs of the AIW.

6	Specification of AIMs and JSON Metadata	Links to summary specification on the web of each AIM used in the use case and the corresponding JSON Metadata as specified by MPAI-AIF [2].
---	---	--

5.3 Videoconference Client Transmitter

5.3.1 Functions of Client Transmitter

The function of a Client Transmitter is to:

1. Receive from a Participant:
 - 1.1. Input Audio from the microphone.
 - 1.2. Input Visual from the camera.
 - 1.3. Participant's Avatar Model.
 - 1.4. Participant's language preferences (e.g., EN-US, IT-CH).
2. Send to the Server:
 - 2.1. Speech Object (for Authentication).
 - 2.2. Face Object (for Authentication).
 - 2.3. Input Portable Avatars containing:
 - 2.3.1. Language preferences (at the start).
 - 2.3.2. Avatar Model (at the start).
 - 2.3.3. Speech.
 - 2.3.4. Avatar Descriptors.

5.3.2 Reference Model of Client Transmitter

Figure 3 gives the Reference Model of Client Transmitter AIW. Red text refers to data sent at meeting start.

At the start, each participant sends to the Avatar Videoconference Server:

1. Language preferences
2. Avatar Model.
3. Speech Object (for Authentication).
4. Face Object (for Authentication).

During the videoconference the following AIMs of the Client Transmitter produce:

AIM	Data
Audio-Visual Scene Description	Speech Objects, Face Descriptors, Body Descriptors, and Audio-Visual Scene Geometry
Automatic Speech Recognition	Recognised Text
Input Face Description	Face Descriptors
Input Body Description	Body Descriptors
Personal Status Extraction	Personal Status
Natural Language Understanding	Meaning
Portable Avatar Description	Avatar Descriptors

Client Transmitters send Portable Avatars to Avatar Videoconference Server that the Server processes and re-distributes to Receiving Clients.

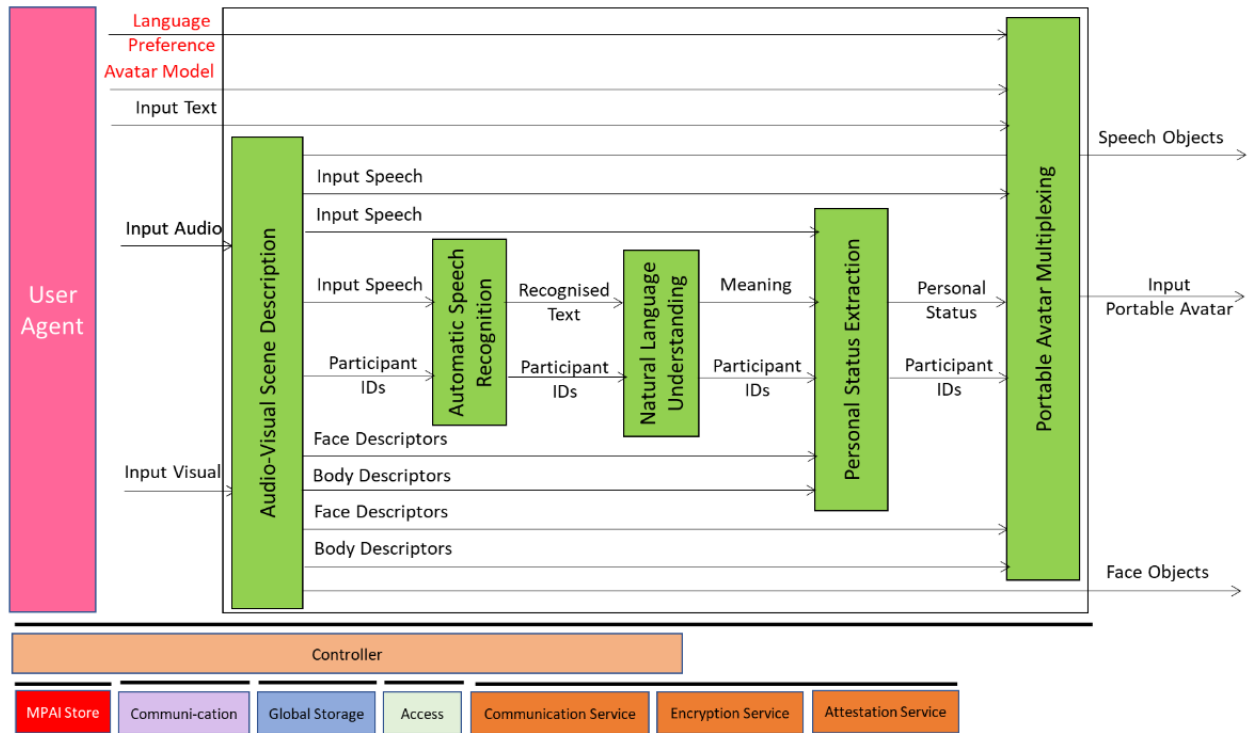


Figure 3 – Reference Model of Avatar Videoconference Client Transmitter

5.3.3 Input and Output Data of Client Transmitter

Table 3 gives the input and output data of the Client Transmitter AIW:

Table 3 – Input and output data of Client Transmitter AIW

Input	Description
Input Text	Chat text used by a human to communicate with Virtual Meeting Secretary or other participants
Language Preference	The language participant wishes to speak and hear.
Input Audio	Audio of Speech of participants in a meeting room.
Input Visual	Video of participants in a meeting room.
Avatar Model	The avatar model selected by the participant.
Output	Description
Speech Object	For authentication by Server.
Input Portable Avatar	Portable Avatar produced by Transmitting Client.
Face Object	For authentication by Server.

5.3.4 Functions of Client Transmitter's AI Modules

Table 5 gives the functions of AI Modules of the Client Transmitter AIW.

Table 4 – AI Modules of Client Transmitter AIW

AIM	Function
Audio-Visual Scene Description	<ol style="list-style-type: none"> 1. Receives Input Audio and Input Visual. 2. Provides Input Speech, Speech Object, Participant ID, Face Descriptors, Body Descriptors, Face Object.
Automatic Speech Recognition	<ol style="list-style-type: none"> 1. Receives Input Speech and Participant ID. 2. Provides Recognised Text and Participant ID.

Natural Language Understanding	<ol style="list-style-type: none"> 1. Receives Recognised Text, Participant ID. 2. Provides the Meaning of the Recognised Text.
Personal Status Extraction	<ol style="list-style-type: none"> 1. Receives Meaning, Speech, Face Descriptors, Body Descriptors, Participant ID. 2. Provides the Participant's Personal Status, Participant ID.
Portable Avatar Multiplexing	<ol style="list-style-type: none"> 1. Receives Language Preference, Avatar Model, Input Text, Input Speech, Personal Status, Participant ID, Face Descriptors, Body Descriptors. 2. Provides the Portable Avatars of Participant with Participant ID.

5.3.5 I/O Data of Client Transmitter's AI Modules

Table 5 gives the AI Modules of Client Transmitter AIW.

Table 5 – AI Modules of Client Transmitter AIW

AIM	Input	Output
Audio Scene Description	Input Audio	<ol style="list-style-type: none"> 1. Audio Scene Geometry 2. Input Speech 3. Speech Objects
Visual Scene Description	Input Visual	<ol style="list-style-type: none"> 1. Visual Scene Geometry 2. Face Descriptors 3. Body Descriptors 4. Face Objects
Audio-Visual Alignments	<ol style="list-style-type: none"> 1. Audio Scene Geometry 2. Visual Scene Geometry 	Participant IDs
Automatic Speech Recognition	Speech Objects	Recognised Text Participant ID
Natural Language Understanding	Recognised Text	Meaning Participant ID
Personal Status Extraction	Meaning Speech Face Object Human Object	Personal Status Participant ID
Portable Avatar Multiplexing	Language Preference Avatar Model Input Text Input Speech Personal Status Participant ID	Portable Avatars.

5.3.6 Specification of Videoconference Client Transmitter AIMs and JSON Metadata

Table 6 – AIMs and JSON Metadata

AIW	AIMs	Name	JSON
PAF-CTX		Videoconference Client Transmitter	X
-	OSD-AVS	Audio-Visual Scene Description	X

	-	CAE-ASD		Audio Scene Description	X
		-	CAE-AAT	Audio Analysis Transform	X
		-	CAE-ASL	Audio Source Localisation	X
		-	CAE-ASE	Audio Separation and Enhancement	X
		-	CAE-AST	Audio Synthesis Transform	X
		-	CAE-AMX	Audio Descriptor Multiplexing	X
	-	OSD-VSD		Visual Scene Description	X
	-	OSD-AVA		Audio-Visual Alignment	X
	-	MMC-ASR		Automatic Speech Recognition	X
	-	MMC-NLU		Natural Language Understanding	X
	-	MMC-PSE		Personal Status Extraction	X
		-	MMC-ITD	Input Text Description	X
		-	MMC-ISD	Input Speech Description	X
		-	PAF-IFD	Input Face Description	X
		-	PAF-IBD	Input Body Description	X
		-	MMC-PTI	PS-Text Interpretation	X
		-	MMC-PSI	PS-Speech Interpretation	X
		-	PAF-PFI	PS-Face Interpretation	X
		-	PAF-PGI	PS-Gesture Interpretation	X
		-	MMC-PMX	Personal Status Multiplexing	X
	-	MMC-PMX		Personal Status Multiplexing	X

5.4 Avatar Videoconference Server

5.4.1 Functions of Avatar Videoconference Server

The Server:

1. *At the start:*

- 1.1. Receives Speech Object and Speech Objects of each Participant.
- 1.2. Authenticates Participants.
- 1.3. Receives Portable Avatars each containing Language Preference and Avatar Model.
- 1.4. Selects an Audio-Visual Scene.
- 1.5. Selects the Spatial Attitudes of the Avatar Models in the Audio-Visual Scene.
- 1.6. Selects the common meeting language.
- 1.7. Distributes all Portable Avatars each containing: Audio-Visual Scene, Language Preference, Avatar Model, and Spatial Attitude.

2. *During the videoconference:*

- 2.1. Receives Participants' and Virtual Meeting Secretary's Avatar Descriptors.
- 2.2. Translates participants' Speech according to their Language Preferences.

- 2.3. Sends Portable Avatars containing Avatar ID, Text, Speech translated to the common meeting language, Face Descriptors and Gesture Descriptors to Virtual Meeting Secretary.
- 2.4. Receives Virtual Meeting Secretary's Portable Avatar containing Avatar ID, Text, Speech in the common meeting language, Face Descriptors and Gesture Descriptors.
- 2.5. Translates Virtual Meeting Secretary's Speech according to each participant's Language Preferences.
- 2.6. Sends Participants' and Virtual Meeting Secretary's Portable Avatars containing Avatar ID, Text, Translated Speech, Face Descriptors and Gesture Descriptors to Client Receivers.

5.4.2 Reference Model of Avatar Videoconference Server

Figure 5 gives the Reference Model of Avatar Videoconference Server AIW. Red text refers to data sent at meeting start.

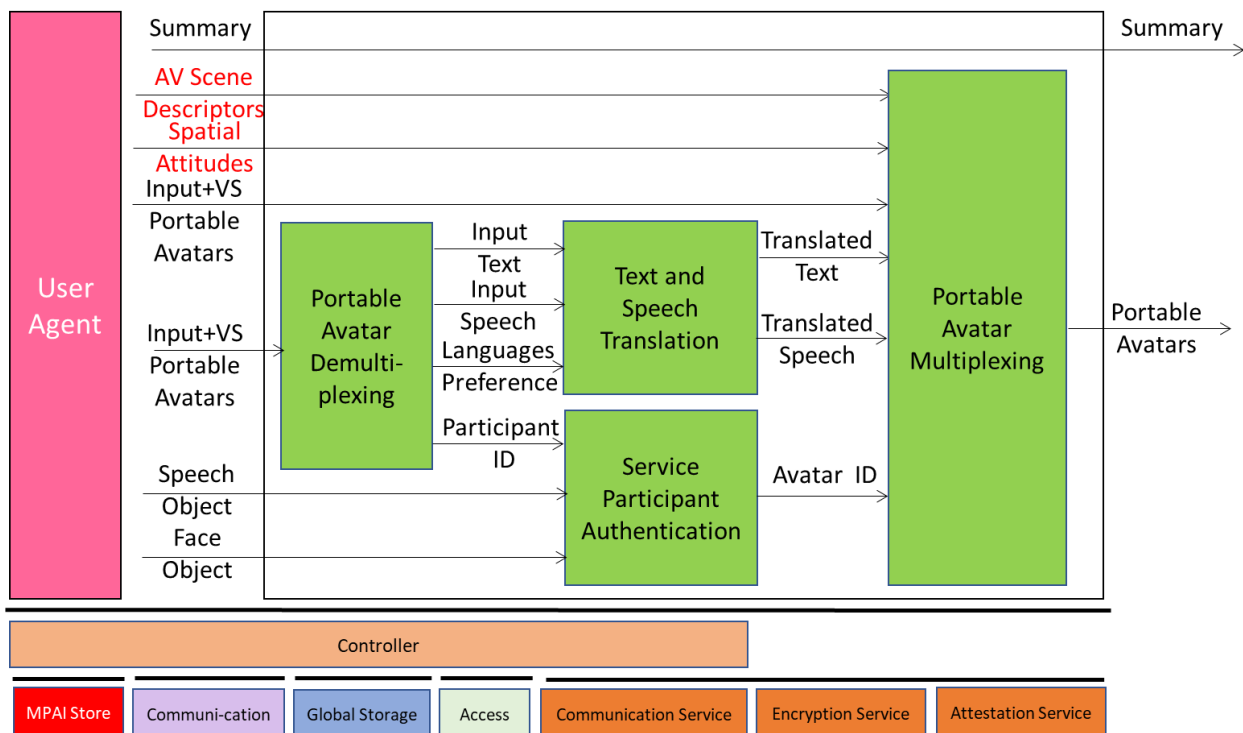


Figure 4 – Reference Model of Avatar Videoconference Server

5.4.3 I/O Data of Avatar Videoconference Server

Table 7 gives the input and output data of Avatar Videoconference Server AIW.

Table 7 – Input and output data of Avatar Videoconference Server AIW

Input	Description
Summary	From Virtual Meeting Secretary
AV Scene Descriptors	Set by Server
Spatial Attitudes	Set by Server
Input+VS Portable Avatars	From Transmitting Clients and Virtual Meeting Secretary
Speech Objects	Participants' Speech Object for Authentication
Face Objects	Participants' Face Object for Authentication
Outputs	Description

Summary	As above
Portable Avatars	As re-multiplexed by Server

5.4.4 Functions of Avatar Videoconference Server AI Modules

Table 8 gives the functions of the AI Modules of the Avatar Videoconference Server AIW.

Table 8 - AI Modules of Avatar Videoconference Server AIW

AIM	Functions
Portable Avatar De-multiplexing	Makes available Input Text, Input Speech, Language Preferences and Participant ID from Input and VS Portable Avatars.
Text and Speech Translation	Translates active Speech and Text of all Participants to the Selected Languages.
Service Participant Authentication	Authenticates Participants using Speech and Face Objects. Connects Participant ID to Avatar ID.
Portable Avatar Multiplexing	Multiplexes components of Portable Avatars for transmission to Receiving Clients.

5.4.5 I/O Data of Avatar Videoconference Server AI Modules

Table 9 gives the Input/Output Data of the AI Modules of the Avatar Videoconference Server AIW.

Table 9 - AI Modules of Avatar Videoconference Server AIW

AIM	Input	Output
Portable Avatar Demultiplexing	Input and VS Portable Avatars	1. Input Text 2. Input Speech 3. Language Preferences 4. Participant ID
Text and Speech Translation	1. Language Preferences 2. Text 3. Speech	1. Translates Text 2. Translated Speech
Service Participant Authentication	1. Speech Descriptors 2. Face Descriptors	Avatar ID
Portable Avatar Multiplexing	1. Audio-Visual Descriptors 2. Spatial Attitudes 3. Avatar Descriptors 4. Translated Text 5. Translated Speech 6. Avatar ID	Portable Avatars

5.4.6 Specification of Avatar Videoconference Server AIMs and JSON Metadata

Table 10 – AIW and AIMs Specification and JSON Metadata

AIW	AIMs	Name	JSON
PAF-AVS		Avatar Videoconference Server	X
	- PAF-PDX	Portable Avatar Demultiplexing	X
	- MMC-TST	Text and Speech Translation	X
	- PAF-SPA	Service Participant Authentication	X

5.5 Virtual Meeting Secretary (Informative)

Virtual Meeting Secretary is specified by *Technical Specification: Multimodal Conversation (MPAI-MMC)* [4]. Here only the Scope, Reference Model, Input/Output Data, and the JSON Metadata are reported for information.

5.5.1 Functions of Virtual Meeting Secretary

The functions of the Virtual Meeting Secretary are to:

1. Listen to each Avatar's Speech.
2. Demultiplex Avatar ID, and Face and Body Descriptors from Input Portable Avatars.
3. Extract Personal Status.
4. Draft a Summary using text in the meeting common language and graphics symbols representing the Personal Status.

The Summary can be handled in two different ways:

1. Transferred to an external application so that Participants can edit the Summary.
2. Displayed to avatars:
 - 2.1. Avatars make Speech or Text comments (e.g., via chat).
 - 2.2. The Virtual Meeting Secretary refines the Summary by interpreting Speech, Text, and the Avatars' Personal Statuses.

5.5.2 Reference Model

Figure 5 depicts the Reference Model of the Virtual Meeting Secretary. Summary and Edited Summary are data back and forth from Summarisation to Dialogue Processing to Summarisation. Summary is continuously sent in an updated form to Dialogue Processing which returns it updated by Avatars' comments in the form of Edited Summary.

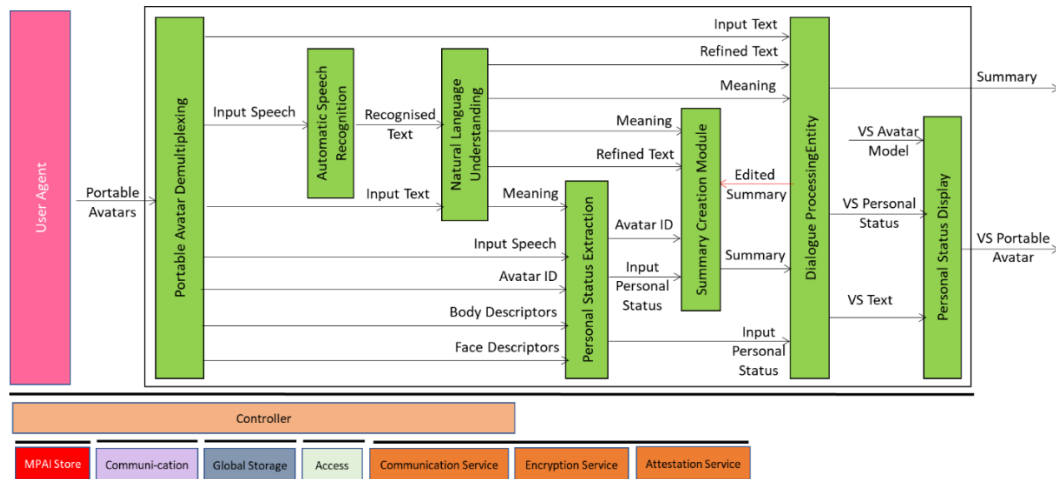


Figure 5 – Reference Model of Virtual Meeting Secretary

The Virtual Meeting Secretary workflow unfolds as follows:

1. Input Portable Avatar Demultiplexing produces: Input Text, Input Speech, Avatar ID, Body Descriptors, and Face Descriptors.
2. Automatic Speech Recognition produces Recognised Text from Input Speech.
3. Natural Language Understanding produces Refined Text and Meaning from Recognised Text.

4. Personal Status Extraction produces Input Personal Status from Meaning, Speech, Face Descriptors, and Body Descriptors.
5. Summary Creation Module produces Summary, i.e., Text in the meeting's common language and graphical symbols, from Refined Text, Personal Status, Meaning, and from Edited Summary received from Dialogue Processing.
6. Entity Dialogue Processing sends:
 - 6.1. Summary to external application.
 - 6.2. Edited Summary produced from Refined Text (from Speech), Avatar's Text (from chat), Meaning, and Summary back to Summary Creation Module.
 - 6.3. Produces VS Text and VS Personal Status.
7. Personal Status Display produces VS Portable Avatar containing VS Avatar Model, VS Text, VS Speech, and VS Avatar Descriptors.

5.5.3 I/O Data of Virtual Meeting Secretary

Table 8 gives the Input and Output Data of Virtual Meeting Secretary.

Table 11 - I/O data of Virtual Meeting Secretary

Input data	From	Description
Input Portable Avatars	Server	Contain: 1. Spatial Attitudes 2. Avatar Descriptors 3. Translated Text 4. Translated Speech 5. Avatar ID
Output data	To	Description
Summary	Server	Summary of Avatars' interventions
VS Avatar Model	Application	The
VS Speech	Avatars	Speech to avatars
VS Text	Avatars	Response to chat.
VS Portable Avatar	Server	Contains: 1. At start: VS Avatar Model 2. Subsequently: 1. VS Speech 2. VS Text 3. VS Avatar Descriptors

5.5.4 AIW and AIM Specifications and JSON Metadata

Table 12 – AIW and AIMs Specification and JSON Metadata

AIW	AIMs	Name	JSON
MMC-VMS		Virtual Meeting Secretary	X
	- PAF-PDX	Portable Avatar Demultiplexing	X
	- MMC-ASR	Automatic Speech Recognition	X
	- MMC-NLU	Natural Language Understanding	X
	- MMC-PSE	Personal Status Extraction	X
	- MMC-ITD	Input Text Description	X

- MMC-ISD	Input Speech Description	X
- PAF-IFD	Input Face Description	X
- PAF-IBD	Input Body Description	X
- MMC-PTI	PS-Text Interpretation	X
- MMC-PSI	PS-Speech Interpretation	X
- PAF-PFI	PS-Face Interpretation	X
- PAF-PGI	PS-Gesture Interpretation	X
- MMC-PMX	Personal Status Multiplexing	X
- MMC-SCM	Summary Creation Module	X
- MMC-EDP	Entity Dialogue Processing	X
- PAF-PSD	Personal Status Display	X
- MMC-TTS	Text-to-Speech	X
- PAF-IFD	Input Face Description	X
- PAF-IBD	Input Body Description	X
- PAF-PMX	Portable Avatar Multiplexing	X

5.6 Videoconference Client Receiver

5.6.1 Functions of Client Receiver

The Function of the Receiving Client is to:

1. Create the local Audio-Visual Scene by:
 - 1.1. Placing and animating the Avatar Models with their Spatial Attitudes.
 - 1.2. Adding Speech to Avatars' mouths.
2. Render the Audio-Visual Scene as seen from the Participant-selected Point of View.

5.6.2 Reference Model of Client Receiver

Figure 6 depicts the Reference Model of the Videoconference Client Receiver. Red text for data received at the start. This is the operation:

1. *At the start*

1.1. Receives Portable Avatars containing:

- 1.1.1. Audio-Visual Scene Descriptors
- 1.1.2. Avatar Models
- 1.1.3. Spatial Attitudes

1.2. Creates the initial Audio-Visual Scene.

2. *During the Videoconference:*

2.1. Receives the Avatar Models containing:

- 2.1.1. Speech
- 2.1.2. Body Descriptors
- 2.1.3. Face Descriptors

2.2. Creates the running Audio-Visual Scene using each Avatar's:

- 2.2.1. Body and Face Descriptors.
- 2.2.2. Speech.

3. Renders the Audio-Visual Scene based on the selected Point of View.

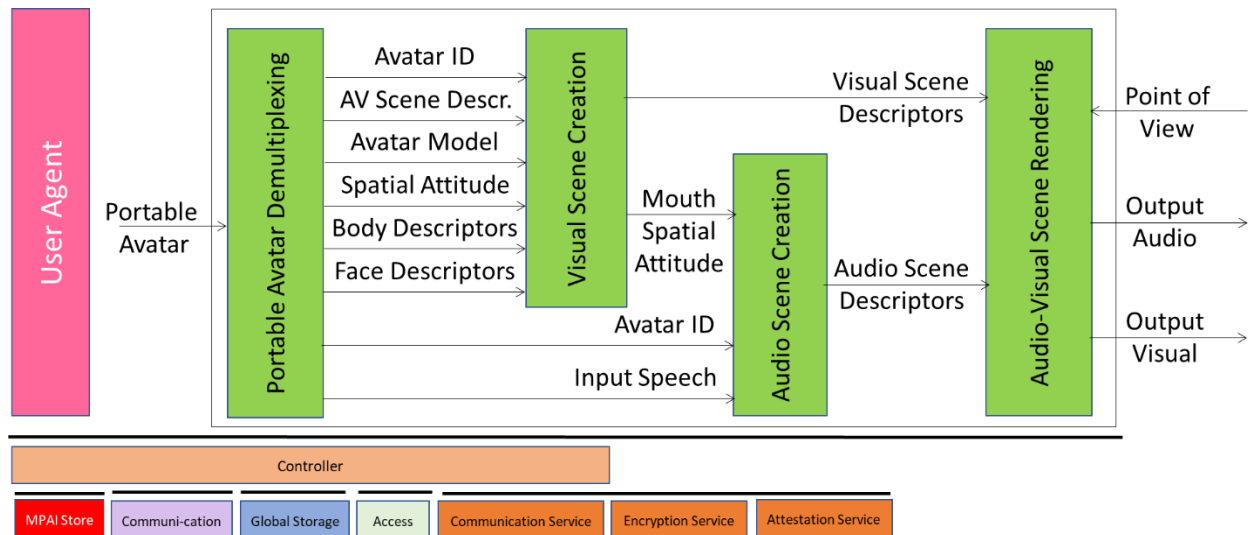


Figure 6 – Reference Model of Videoconference Client Receiver

Note: An implementation may decide to display text with a visual image for accessibility purposes.

5.6.3 I/O Data of Client Receiver

Table 13 gives the input and output data of Videoconference Client Receiver.

Table 13 – Input and output data of Videoconference Client Receiver AIW

Input	Description
Point of View	Participant-selected point of view to see the Audio-Visual Scene.
Portable Avatars	Portable Avatars from Videoconference Avatar Server.
Output	Description
Output Audio	Rendered by Audio-Visual Scene Rendering.
Output Visual	Rendered by Audio-Visual Scene Rendering.

5.6.4 Functions of Videoconference Client Receiver's AI Modules

Table 14 gives the AI Modules of Videoconference Client Receiver AIW.

Table 14 – Functions of Videoconference Client Receivers' AI Modules

AIM	Input
Portable Avatar Demultiplexing	Extracts Avatar ID, Audio-Visual Scene Descriptors, Avatar Model, Spatial Attitude, Body Descriptors, Face Descriptors, and Input Speech from Portable Avatars.
Visual Scene Creation	Creates the Visual Scene and provides the Spatial Attitudes of the mouths of all Avatars.
Audio Scene Creation	Creates the Audio Scene.
Audio-Visual Scene Rendering	Provides a ready-to-rendered AV Scene.

5.6.5 I/O Data of Videoconference Client Receiver's AI Modules

Table 15 gives the AI Modules of Videoconference Receiving Client AIW.

Table 15 – I/O Data of Videoconference Client Receivers' AI Modules

AIM	Input	Output
Portable Avatar Demultiplexing	Portable Avatars	1. Avatar ID 2. Audio-Visual Scene Descriptors 3. Avatar Model 4. Spatial Attitude 5. Body Descriptors 6. Face Descriptors 7. Input Speech
Visual Scene Creation	1. Avatar ID 2. Audio-Visual Scene Descriptors 3. Avatar Model 4. Spatial Attitude 5. Body Descriptors 6. Face Descriptors	1. Visual Scene Descriptors 2. Mouth Spatial Attitudes
Audio Scene Creation	1. Avatar ID 2. Input Speech 3. Mouth Spatial Attitudes	Audio Scene Descriptors
Audio-Visual Scene Rendering	1. Audio Scene Descriptors 2. Visual Scene Descriptors 3. Point of View	1. Output Audio 2. Output Visual

5.6.6 Specifications of Videoconference Client Receiver's AIMs and JSON Metadata

Table 16 – AIMs and JSON Metadata

AIW	AIMs	Name	JSON
PAF-CRX		Videoconference Client Receiver	X
-	PAF-PDX	Portable Avatar Demultiplexing	X
-	PAF-VSC	Visual Scene Creation	X
-	PAF-ASC	Audio Scene Creation	X
-	PAF-AVR	Audio-Visual Scene Rendering	X

6 Composite AI Modules

Some MPAI-PAF Use Cases need combinations of AI Modules called Composite AI Modules. This Chapter specifies the Personal Status Display Composite AIM using a format like the one adopted for Uses Cases.

6.1 Personal Status Display (PSD)

6.1.1 Functions of Composite AIM

Personal Status Display (PSD) is a Composite AIM receiving Text and Personal Status and generating the Face and Body Descriptors of an Avatar producing Text and uttering Speech with the intended Personal Status while the avatar's Face and Gesture show the intended Personal Status. The Personal Status driving the avatar can be extracted from a human or can be synthetically generated by a machine as a result of its conversation with a human or another avatar. Several Use

Cases from MPAI-MMC and other MPAI standards use this Composite AIM as a replacement for the combination of the AIMs depicted in Figure 7.

6.1.2 Reference Model

Figure 7 depicts the AIMs required to implement Personal Status Display.

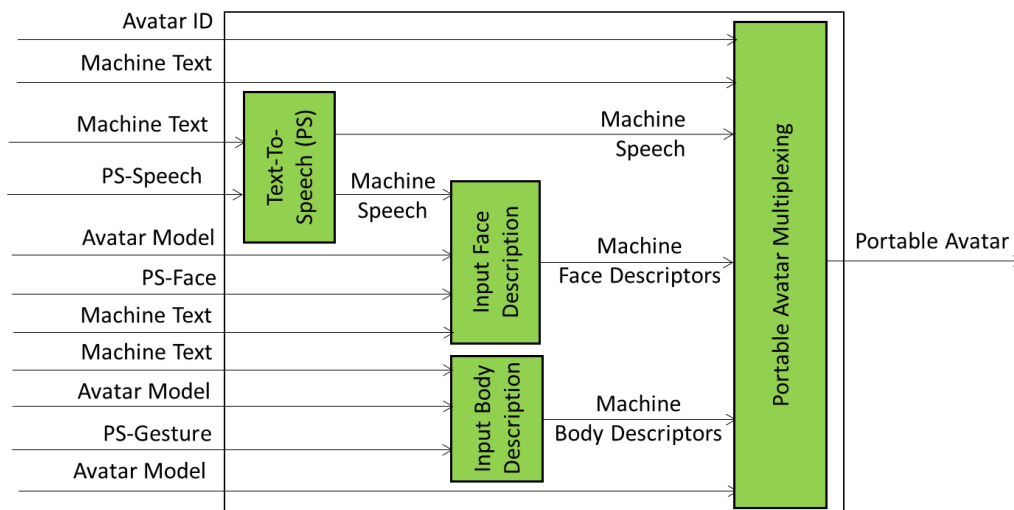


Figure 7 – Reference Model of Personal Status Display

The Personal Status Display operates as follows:

1. Avatar ID is the ID of the Portable Avatar.
2. Machine Text is synthesised as Speech using the Personal Status provided by PS-Speech.
3. Machine Speech and PS-Face are used to produce the Machine Face Descriptors.
4. PS-Gesture and Text are used for Machine Body Descriptors using the Avatar Model.
5. Portable Avatar Multiplexing produces the Portable Avatar.

6.1.3 I/O Data of Personal Status Display

Table 17 gives the Input/Output Data of Personal Status Display.

Table 17 – I/O Data of Personal Status Display

Input data	From	Description
Avatar ID	Upstream AIM	Portable Avatar's ID
Avatar Model	From upstream AIM or embedded	Part of Portable Avatar
Text Object	Keyboard or upstream AIM	Texts accompanying Portable Avatar
PS-Speech	Personal Status Extraction or Machine	To synthesise Speech
PS-Face	Personal Status Extractor or Machine	To describe Face
PS-Gesture	Personal Status Extractor or Machine	To describe Body
Output data	To	Description
Portable Avatar	Downstream AIM or renderer	As Portable Avatar

6.1.4 Functions of AI Modules of Personal Status Display

Table 18 gives the functions of the AIMs.

Table 18 – Functions of AI Modules of Personal Status Extraction

AIM	Functions
Text To Speech (PS)	Synthesises Text with Personal Status.
Input Face Description	Produces the Machine Face Descriptors with Personal Status.
Input Body Description	Produces the Machine Body Descriptors with Personal Status.
Portable Avatar Multiplexing	Multiplexes Data into Portable Avatar with Personal Status.

6.1.5 I/O Data of AI Modules of Personal Status Display

Table 19 gives the list of AIMs with their functions.

Table 19 – AI Modules of Personal Status Extraction

AIM	Receives	Produces
Text To Speech (PS)	1. Text 2. PS-Speech	Machine Speech
Input Face Description	1. Avatar Model 2. Machine Speech 3. PS-Face	Machine Face Descriptors
Gesture Description	1. Avatar Model 2. Text 3. Machine PS-Gesture	Machine Body Descriptors
Portable Avatar Multiplexing	1. Avatar ID 2. Machine Text 3. Machine Speech 4. Machine Body Descriptors 5. Machine Face Descriptors 6. Machine Avatar 7. Avatar Model	Portable Avatar

6.1.6 Specification of Personal Status Display AIMs and JSON Metadata

Table 20 –AIMs and JSON Metadata

AIMs	Name and Specification	JSON
PAF-PSD	Personal Status Display	X
- MMC-TTS	Text-to-Speech	X
- PAF-IFD	Input Face Description	X
- PAF-IBD	Input Body Description	X
- PAF-PMX	Portable Avatar Multiplexing	X

7 Data Formats

Technical Specification; Portable Avatar Format (MPAI-PAF) V1.1 specifies the Data Types listed in Table 21. The first column gives the name of the Data Type, the second the subsection where the Data Type is specified and the third the Use Cases making use of it.

Table 21 – Data Formats

Note: Entries in bold indicate Data Format categories.

Name of Data Type	Subsection	Use Case
Object and Scene		
Spatial Attitude	7.1.2	PAF-ABV
		CAV-ESS
		CAV-MAS
		MMC-CAS
Audio Scene Descriptors	7.1.3	PAF-ABV
		CAV-ESS
		CAV-MAS
		MMC-CAS
Visual Scene Descriptors	7.1.4	PAF-ABV
		CAV-ESS
		CAV-MAS
		MMC-CAS
Portable Avatar Format	7.1.5	PAF-CTX
		PAF-SRV
		PAF-CRX
		CAV-HCI
		MMC-VSV
Body		
Body Model	7.3.1	PAF-ABV
		CAV-HCI
		MMC-VSV
Body Descriptors	0	PAF-ABV
		CAV-HCI
		MMC-VSV
Face		
Face Model	7.4.1	PAF-ABV
		CAV-HCI
		MMC-VSV
Face Descriptors	7.4.2	PAF-ABV
		CAV-HCI
		MMC-VSV
Avatar		
Avatar Model	7.5.1	PAF-ABV
		CAV-HCI
		MMC-VSV
Conversation		
Text	7.6.1	PAF-ABV
		MMC
Language identifier	7.6.2	PAF-ABV
		MMC
Meaning	7.6.3	PAF-ABV
		MMC
Personal Status	7.6.4	PAF-ABV
		MMC

7.1 Space, Objects and Scenes

7.1.1 Coordinate Systems

Coordinate Systems enable the specification of the position of a point by three numbers. MPAI-PAF uses:

- Cartesian Coordinates specifying the position of a point by the signed distances from the point to three mutually perpendicular planes.
- Spherical Coordinates specifying the position of a point by:
 - The radial distance of that point from a fixed origin.
 - The polar angle measured from a fixed zenith direction.
 - The azimuthal angle of its orthogonal projection on a reference plane.

Coordinate Systems can be global or local. An Object in a Global Coordinate System may have a Local Coordinate Systems – Cartesian or Spherical. A rigid Object in a Virtual Space has a Spatial Attitude defined as the Position and Orientation and their velocities and accelerations. The Position of an Object composed of rigid Objects is that of a representative point in the Object. The notion of Spatial Attitude can also be applied to Audio Objects.

7.1.2 Spatial Attitude and Point of View

Spatial Attitude is the combination of Position and Orientation, and their Velocities and Accelerations. Spatial Attitude is specified in [5].

Point of View is represented by a Spatial Attitude that only includes Position and Orientation.

7.1.3 Audio Objects and Scene

Scene Geometry is specified in MPAI-CAE V2.1 [3].

Audio Object is specified in MPAI-CAE V2.1 [3].

Audio Scene Descriptors are specified in MPAI-CAE V2.1 [3].

7.1.4 Visual Objects and Scene

Visual Scene Geometry is specified in MPAI-OSD V1.0 [5].

Visual Object is specified in MPAI-OSD V1.0 [5].

Visual Scene Descriptors are specified in MPAI-OSD V1.0 [5].

7.1.5 Audio-Visual Objects and Scene

Audio-Visual Scene Geometry is specified in MPAI-OSD V1.0 [5].

Audio-Visual Object is specified in MPAI-OSD V1.0 [5].

Audio-Visual Scene Descriptors are specified in MPAI-OSD V1.0 [5].

7.2 Portable Avatar

Portable Avatar is a Data Type that enables a receiver to reproduce an Avatar and its Context as intended by the Transmitter.

7.2.1 Syntax

```
{
  "$schema": "http://json-schema.org/draft-07/schema#",
  "$id": "https://schemas.mpai.community/PAF/V1.0/PortableAvatarFormat.json",
  "title": "PortableAvatar",
  "type": "object",
  "properties": {
    "Header": {
      "type": "object",
      "properties": {
        "Standard": {
```

```

        "type": "string"
      },
      "Version": {
        "type": "integer"
      },
      "Subversion": {
        "type": "integer"
      }
    },
    "PAFID": {
      "type": "string"
    },
    "Time": {
      "type": "object",
      "properties": {
        "TimeType": {
          "type": "boolean"
        },
        "StartTime": {
          "type": "number"
        },
        "EndTime": {
          "type": "number"
        }
      }
    },
    "Visual": {
      "type": "object",
      "properties": {
        "AvatarID": {
          "type": "string"
        },
        "AvatarModel": {
          "$ref": "https://schemas.mpai.community/PAF/V1.0/data/AvatarModel.json"
        },
        "BodyDescriptors": {
          "$ref": "https://schemas.mpai.community/PAF/V1.0/data/BodyDescriptors.json"
        },
        "FaceDescriptors": {
          "1, 2, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20, 22, 23, 24, 25, 26, 27, 28, 41, 42, 43, 44, 45, 46, 61, 62, 63, 64"
        },
        "AudioVisualSceneDescriptors": {
          "$ref": "https://schemas.mpai.community/PAF/V1.1/data/ AudioVisu-
alSceneDescriptors.json"
        }
      }
    },
    "Audio": {
      "type": "object",
      "properties": {
        "LanguagePreference": {
          "type": "string",
          "minLength": 2,
          "maxLength": 2
        },
        "Speech": {
          "type": "object",
          "properties": {
            "Encoding": {
              "enum": [
                "MP3",
                "AAC"
              ]
            },
            "Utterance": {
              "type": "array",
              "contains": {
                "type": "integer"
              }
            }
          }
        }
      }
    }
  }

```

```

    }
  },
  "Text": {
    "type": "string"
  },
  "PersonalStatus": {
    "$ref": "https://schemas.mpai.community/MMC/V2.0/data/PersonalStatus.json"
  }
}

```

7.2.2 Semantics

Table 22 – Variables composing the Portable Avatar Format

Note: All elements in Table 22 are optional.

Label	Size	Description
HEADER	9 Bytes	
• Standard	7 Bytes	The string PAF-PAF
• Version	1 Byte	Major version
• Subversion	1 Byte	Minor version
PAFID	16 Bytes	UUID Identifier of Portable Avatar set.
Time	17 Bytes	Collects various data expressed with bits
• TimeType	0 bit	0=Relative: time starts at 0000/00/00T00:00 1=Absolute: time starts at 1970/01/01T00:00.
• Reserved	1-7 bits	reserved
• StartTime	8 Bytes	Start of current Portable Avatar (μs).
• EndTime	8 Bytes	End of current Portable Avatar (μs).
AudioVisualSceneDescriptors	N5 Bytes	Descriptors of the AV Scene.
Avatar		
AvatarID	N0 Bytes	ID of the Avatar
SpatialAttitude	N1 Bytes	Spatial Attitude of Avatar in AV Scene.
AvatarModel	N2 Bytes	Model of Avatar.
BodyDescriptors	N3 Bytes	Body Descriptors of Avatar.
FaceDescriptors	N4 Bytes	Face Descriptors of Avatar.
Language	2 Bytes	Language Preference of Avatar.
Speech		
SpeechType	3 Bytes	Speech Data representation method
Speech	N6 Bytes	Speech in the StartTime-EndTime period.
Text	N7 Bytes	Text in the StartTime-EndTime period.
PersonalStatus	N8 Bytes	Personal Status of Avatar.

7.3 Body

7.3.1 Body Model

MPAI adopts the Humanoid animation (HAnim) architecture [7] that gives access to the joint and end-effector hierarchy of a human figure. This allows a model-independent animation of a skeleton and related skin vertices associated with joints and geometry/accessories/sensors of individual body segments and sites.

The actual structure of the HAnim architecture depends on the selected element of the Level Of Articulations (LOA) hierarchy: LOA 1, LOA 2, LOA 3, or LOA 4. All joints of an HAnim figure are represented as a tree hierarchy starting with the humanoid_root joint. For an LOA 1 character, there are 18 joints and 18 segments in the hierarchy.

The bones of the body are described starting from position (x_0, y_0, z_0) of the root (neck or pelvis). The orientation of a bone attached to the root is defined by (α, β, γ) where α is the angle of the bone with the x axis, and so on. The joint of a bone attached to the preceding bone has a position (x_1, y_1, z_1) determined by the angles $(\alpha_1, \beta_1, \gamma_1)$ and the length of the bone.

The Body Model contains:

1. Pose composed by:
 - 1.1. The position of the root.
 - 1.2. The angles of the bones with the (x, y, z) coordinate axes.
 - 1.3. The orientation of the body defined by 3 angles.
2. The standard bone lengths.
3. Lengths of the bones of the specific model.
4. Surface-related
 - 4.1. Surface
 - 4.2. Texture
 - 4.3. Material
 - 4.4. Cloth (an integral part of the model).

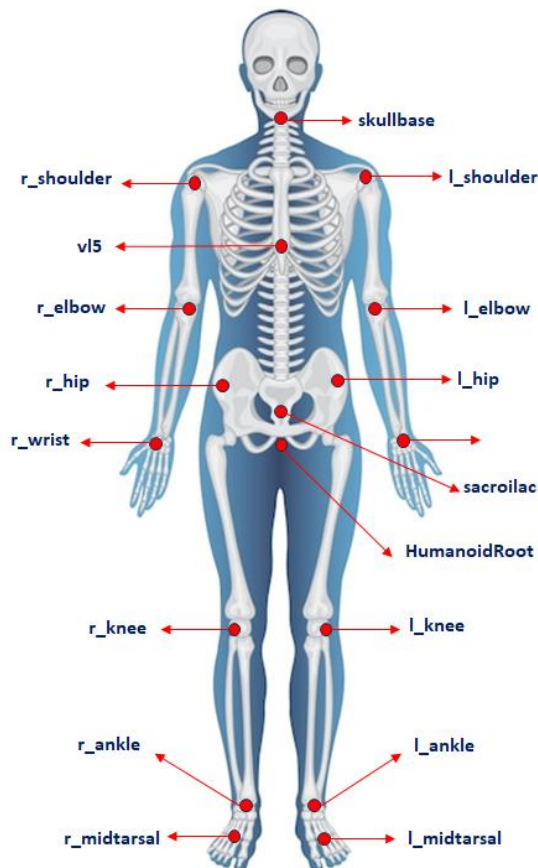


Figure 8 - Some joints of the Body Model

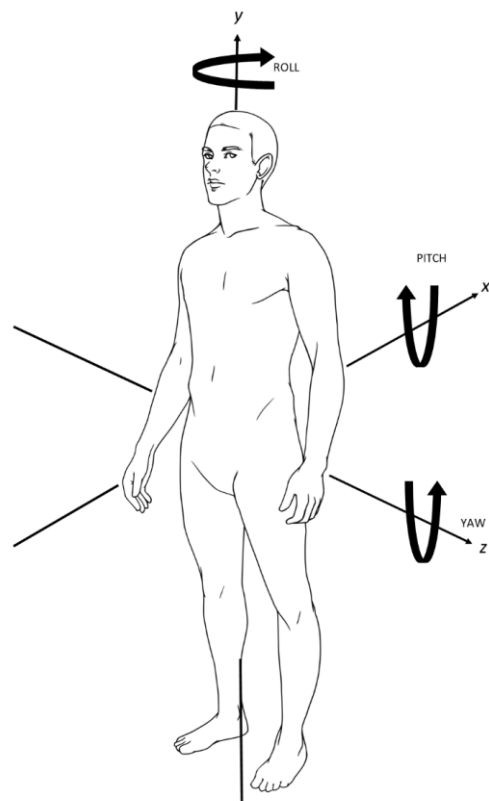


Figure 9 – Pitch, Roll, and Yaw of Body

The Body Model is represented in the glTF format and transmitted in the initial Portable Avatar at the beginning of the Avatar-Based Videoconference.

The Spatial Attitude of a Body is defined with respect to the global coordinate system defined by the Visual Scene.

7.3.2 Body Descriptors

Body Descriptors are included in the data set describing the root and joints as:

- 1 Position and Orientation of the root.
- 2 Rotation angles of the joints.

The rotation of the head is treated as any other joint.

Figure 9 depicts Roll (rotation around y), Pitch (rotation around y, and Yaw (rotation around z) of a Body.

7.4 Face

7.4.1 Face Model

A Face Model is represented according to the glTF syntax.

7.4.2 Face Descriptors

MPAI adopts as Face Descriptors the Actions Units of the Facial Action Coding System (FACS), originally developed by Carl-Herman Hjortsjö, adopted by Paul Ekman and Wallace V. Friesen (1978), and updated by Ekman, Friesen, and Joseph C. Hager (2002) [12].

AU	Description	Facial muscle generating the Action
1	Inner Brow Raiser	Frontalis, pars medialis
2	Outer Brow Raiser	Frontalis, pars lateralis
4	Brow Lowerer	Corrugator supercilii, Depressor supercilii
5	Upper Lid Raiser	Levator palpebrae superioris
6	Cheek Raiser	Orbicularis oculi, pars orbitalis
7	Lid Tightener	Orbicularis oculi, pars palpebralis
9	Nose Wrinkler	Levator labii superioris alaeque nasi
10	Upper Lip Raiser	Levator labii superioris
11	Nasolabial Deepener	Zygomaticus minor
12	Lip Corner Puller	Zygomaticus major
13	Cheek Puffer	Levator anguli oris (a.k.a. Caninus)
14	Dimpler	Buccinator
15	Lip Corner Depressor	Depressor anguli oris (a.k.a. Triangularis)
16	Lower Lip Depressor	Depressor labii inferioris
17	Chin Raiser	Mentalis
18	Lip Puckerer	Incisivii labii superioris and Incisivii labii inferioris
20	Lip stretcher	Risorius with platysma
22	Lip Funneler	Orbicularis oris
23	Lip Tightener	Orbicularis oris
24	Lip Pressor	Orbicularis oris
25	Lips part	Depressor labii inferioris or relaxation of Mentalis, or Orbicularis oris
26	Jaw Drop	Masseter, relaxed Temporalis and internal Pterygoid

27	Mouth Stretch	Pterygoids, Digastric
28	Lip Suck	Orbicularis oris
41	Lid droop	Relaxation of Levator palpebrae superioris
42	Slit	Orbicularis oculi
43	Eyes Closed	Relaxation of Levator palpebrae superioris; Orbicularis oculi, pars palpebralis
44	Squint	Orbicularis oculi, pars palpebralis
45	Blink	Relaxation of Levator palpebrae superioris; Orbicularis oculi, pars palpebralis
46	Wink	Relaxation of Levator palpebrae superioris; Orbicularis oculi, pars palpebralis
61	Eyes turn left	Lateral rectus, medial rectus
62	Eyes turn right	Lateral rectus, medial rectus
63	Eyes up	Superior rectus, Inferior oblique
64	Eyes down	Inferior rectus, Superior oblique

7.5 Avatar

7.5.1 Avatar Model

The Avatar Model combines the Body and Face Models, both represented in glTF [6].

7.5.2 Avatar Descriptors

The Avatar Descriptors combine the Body and Face Descriptors.

7.6 Other Data Types

7.6.1 Text

Specified in MPAI-MMC V2 [4].

7.6.2 Language identifier

Specified in MPAI-MMC V2 [4].

7.6.3 Meaning

Specified in MPAI-MMC V2 [4].

7.6.4 Personal Status

Specified in MPAI-MMC V2 [4].

Annex 1 - MPAI Basics (Informative)

1 General

In recent years, Artificial Intelligence (AI) and related technologies have been introduced in a broad range of applications affecting the life of millions of people and are expected to do so much 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. In addition, some AI technologies may carry inherent risks, e.g., in terms of bias toward some classes of users making the need for standardisation more important and urgent than ever.

The above considerations have prompted the establishment of MPAI, the international, unaffiliated, not-for-profit Moving Picture, Audio and Data Coding by Artificial Intelligence organisation with the mission to develop *AI-enabled data coding standards* to facilitate the development of interoperable AI-based products, applications, and services.

As a rule, MPAI standards include four documents: Technical Specification, Reference Software Specifications, Conformance Testing Specifications, and Performance Assessment Specifications. The last – and new in standardisation – type of Specification includes standard operating procedures that enable users of MPAI Implementations to make informed decision about their applicability based on the notion of Performance, defined as a set of attributes characterising a reliable and trustworthy implementation.

2 Governance of the MPAI Ecosystem

The technical foundations of the MPAI Ecosystem are currently provided by *Technical Specification: Governance of the MPAI Ecosystem (MPAI-GME)* [9] developed and maintained by MPAI:

1. Technical Specification.
2. Reference Software Specification.
3. Conformance Testing.
4. Performance Assessment.
5. Technical Report

An MPAI Standard is a collection of a variable number of the 5 document types.

Figure 10 depicts the MPAI ecosystem operation for conforming MPAI implementations.

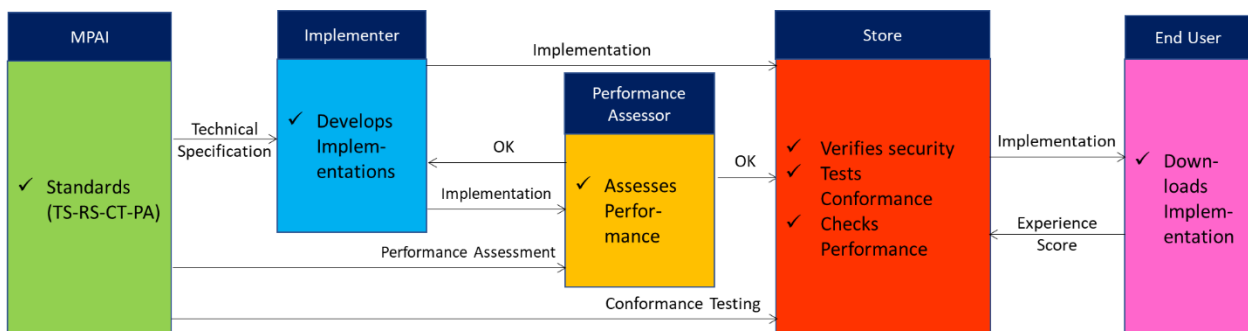


Figure 10 – The MPAI ecosystem operation

Table 23 identifies the following roles in the MPAI Ecosystem:

Table 23 - Roles in the MPAI Ecosystem

MPAI	Publishes Standards. Establishes the not-for-profit MPAI Store. Appoints Performance Assessors.
Implementers	Submit Implementations to Performance Assessors and the MPAI Store.
Performance Assessors	Inform Implementation submitters and the MPAI Store if Implementation Performance is acceptable.
MPAI Store	Assign unique ImplementerIDs (IID) to Implementers of MPAI Technical Specification in its capacity as ImplementerID Registration Authority (IIDRA) ¹ . Verifies security and Tests Implementation Conformance.
Users	Download Implementations and report their experience to MPAI.

3 AI Framework

In general, MPAI Application Standards are defined as aggregations – called AI Workflows (AIW) – of processing elements – called AI Modules (AIM) – executed in an AI Framework (AIF). MPAI defines Interoperability as the ability to replace an AIW or an AIM Implementation with a functionally equivalent Implementation.

Figure 11 depicts the MPAI-AIF Reference Model. Implementations of MPAI Application Standards and user-defined MPAI-AIF Conforming applications operate in an AI Framework [2].

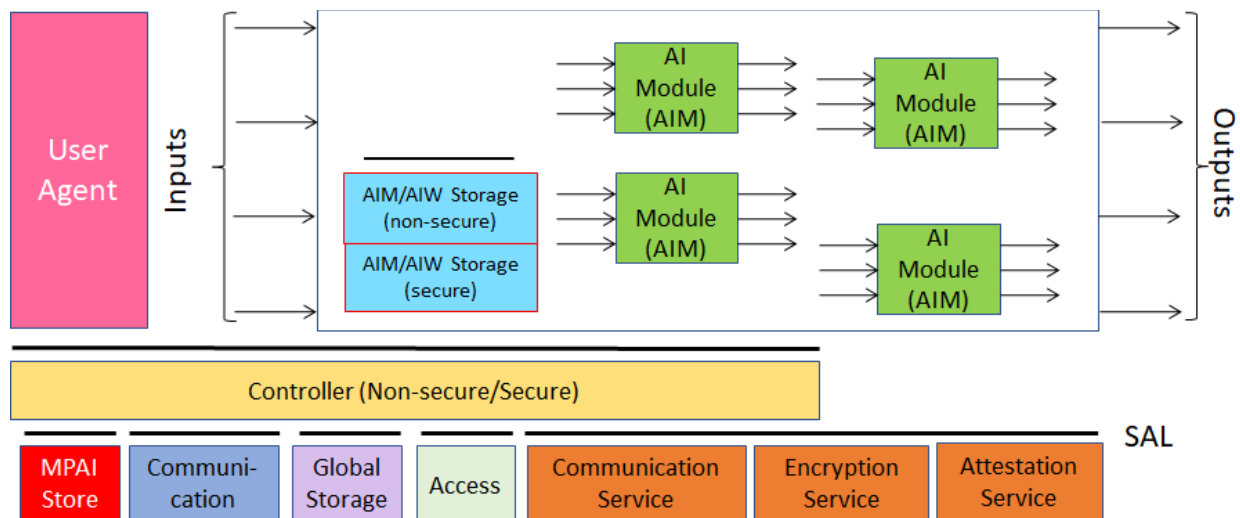


Figure 11 – The AI Framework (AIF) Reference Model

MPAI Application Standards normatively specify the Syntax and Semantics of the input and output data and the Function of the AIW and the AIMs, and the Connections between and among the AIMs of an AIW.

An AIW is defined by its Function and input/output Data and by the topology of its AIMs. Likewise, an AIM is defined by its Function and input/output Data. MPAI standard are silent on the technology used to implement the AIM which may be based on AI or data processing, and implemented in software, hardware or hybrid software and hardware technologies.

¹ At the time of publication of this Technical Report, the MPAI Store was assigned as the IIDRA.

AIW and its AIMs may have 3 interoperability levels:

Level 1 – Proprietary and satisfying the MPAI-AIF Standard.

Level 2 – Specified by an MPAI Application Standard.

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

4 Space, Objects, and Scenes

Technical Specification: Object and Scene Descriptors (MPAI-OSD) specifies Spatial Attitude defined as Position and Orientation, and their time derivatives up to the second order.

The ability to describe (i.e., digitally represent) an audio-visual scene is a key requirement of several MPAI Technical Specifications and Use Cases. The Audio, Visual and Audio-Visual Scene Descriptors are specified in [5]. MPAI has also developed Technical Specification: Context-based Audio Enhancement (MPAI-CAE) [3] that includes Audio Scene Descriptors and uses a subset of Graphics Language Transmission Format (gLTF) [6] to describe a visual scene.

Audio Scene Description is a Composite AI Module (AIM) specified by Technical Specification: Context-based Audio Enhancement (MPAI-CAE) [3]. The position of an Audio Object is defined by Azimuth, Elevation, Distance.

Annex 2 - General MPAI Terminology

The Terms used in this standard whose first letter is capital and are not already included in *Table 1* are defined in Table 24. To concentrate in one place all the Terms that are composed of a common name followed by other words (e.g., the word Data followed by one of the words Format, Type, or Semantics), the definition given to a Terms preceded by a dash “-” applies to a Term composed by that Term without the dash preceded by the Term that precedes it in the column without a dash.

Table 24 – 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 Model (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.
Component	One of the 7 AIF elements: Access, Communication, Controller, Internal Storage, Global Storage, Store, and User Agent
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.
- Testing	The normative document specifying the Means to Test the Conformance of an Implementation.
- Testing Means	Procedures, tools, data sets and/or data set characteristics to Test the Conformance of an Implementation.
Connection	A channel connecting an output port of an AIM and an input port of an AIM.
Controller	A Component that manages and controls the AIMs in the AIF, so that they execute in the correct order and at the time when they are needed
Data	Information in digital form.
- Format	The standard digital representation of Data.
- Type	An instance of Data with a specific Data Format.
- Semantics	The meaning of Data.
Descriptor	Coded representation of a text, audio, speech, or visual feature.
Digital Representation	Data corresponding to and representing a physical entity.

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.
AIM/AIW 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.
Instance ID	Instance of a class of Objects and the Group of Objects the Instance belongs to.
Interoperability	The ability to functionally replace an AIM with another AIW having the same Interoperability Level
- 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.
- Assessment	The normative document specifying the Means to Assess the Grade of Performance of an Implementation.
- Assessment Means	Procedures, tools, data sets and/or data set characteristics to Assess the Performance of an Implementation.
- 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 their 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	A set of Technical Specification, Reference Software, Conformance Testing, Performance Assessment, and Technical Report 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: <ol style="list-style-type: none"> 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.
Zero Trust	A cybersecurity model primarily focused on data and service protection that assumes no implicit trust.

Annex 3 - Notices and Disclaimers Concerning MPAI Standards (Informative)

The notices and legal disclaimers given below shall be borne in mind when [downloading](#) and using approved MPAI Standards.

In the following, “Standard” means the collection of four MPAI-approved and [published](#) 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](#). 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](#). 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](#) 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](#). Comments will be reviewed by the appropriate committee for their technical relevance. However, MPAI does not provide interpretation, consulting information, or advice on MPAI Standards. Interested parties are invited to join MPAI so that they can attend the relevant Development Committees.

Coverage and Applicability of MPAI Standards

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

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

IN NO EVENT SHALL MPAI BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO: THE NEED TO PROCURE SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR

TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE PUBLICATION, USE OF, OR RELIANCE UPON ANY STANDARD, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE AND REGARDLESS OF WHETHER SUCH DAMAGE WAS FORESEEABLE.

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

Users of MPAI Standards should consider all applicable laws and regulations when using an MPAI Standard. The validity of Conformance Testing is strictly technical and refers to the correct implementation of the MPAI Standard. Moreover, positive Performance Assessment of an implementation applies exclusively in the context of the [MPAI Governance](#) 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](#).

The Reference Software of an MPAI Standard is released with the [MPAI Modified Berkeley Software Distribution 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.

Annex 4 - Patent Declarations

The following table will include the entities declaring to agree to licence their standard essential patents reading on the Technical Specification: Portable Avatar Format (PAF) based on the MPAA-AIF Framework Licence [14]:

Entity	Name	Email address