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

MPAI Technical Specification

Object and Scene Description (MPAI-OSD)

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 the use of this Technical Specification.

Readers are invited to review **Error! Reference source not found.**Notices and Disclaimers.

Technical Specification
Object and Scene Description (MPAI-OSD) V1.1

Contents

1   Foreword

The international, unaffiliated, non-profit Moving Picture, Audio, and Data Coding by Artificial Intelligence (MPAI) organisation has been established in September 202i in the context of:

1. Increasing use of Artificial Intelligence (AI) technologies to a broad range of applications that have affected the life of millions of people and are expected to do so even more in the future.
2. The marginal use of standards in the development of those AI applications.
3. Most of what has happened in the AI domain has not been affected by standards.
4. The enormous and positive impact that standards have exerted on the digital media industry and billions of people,.

MPAI believes that AI-based data coding standards will have a similar positive impact on the Information and Communication Technology industry and has designed the organisation to develop AI-based Data Coding standards in pursuit of the following policies:

1. Publish from the start clear Intellectual Property Rights licensing frameworks.
2. Comply with a rigorous MPAI Standard Development Process.
3. *Be friendly* to the AI context but, to the extent possible, remain agnostic to the technology and allow that developers select the more appropriate – AI or Data Processing – for their implementations.
4. *Be attractive* to different industries, end users, and regulators.
5. *Address* five standardisation areas:
   1. Data Type, a particular type of MPAI-specified Data, e.g., Audio, Visual, Object, Scenes, and Descriptors with as clear semantics as possible.
   2. Qualifier, specialised Metadata conveying information on a Data Type's Sub-Type, Format, and Attributes of a Data Type.
   3. AI Module (AIM), MPAI-specified processing elements with identified functions and input/output Data.
   4. AI Workflow (AIW), MPAI-specified configurations of AIMs with identified functions and input/output Data.

5. <u>AI Framework</u> (AIF), an environment enabling dynamic configuration, initialisation, and control of AIWs.
6. *Provide* appropriate Governance of the ecosystem created by MPAI Technical Specifications enabling users to:
7.
    1. <u>Operate</u> Reference Software Implementations of MPAI Technical Specifications attached to Reference Software Specifications
    2. <u>Test</u> the conformance of an implementation with a Technical Specification based on the Conformance Testing Specification.
    3. <u>Assess</u> the performance of an implementation of a Technical Specification based on the Performance Assessment Specification.
    4. <u>Get</u> conforming implementations possibly with a performance assessment report from a trusted source through the MPAI Store.

The accomplishment of the MPAI mission is facilitated by two foundational Technical Specifications:

- *Technical Specification: Artificial Intelligence Framework (MPAI-AIF)* specifying an environment enabling initialisation, dynamic configuration, and control of AIWs in the standard AI Framework environment depicted in Figure 1. An AI Framework can execute AI applications called AI Workflows (AIW). An AIW can include interconnected AI Modules (AIM). MPAI-AIF supports small- and large-scale high-performance components and promotes solution with improved explainability.
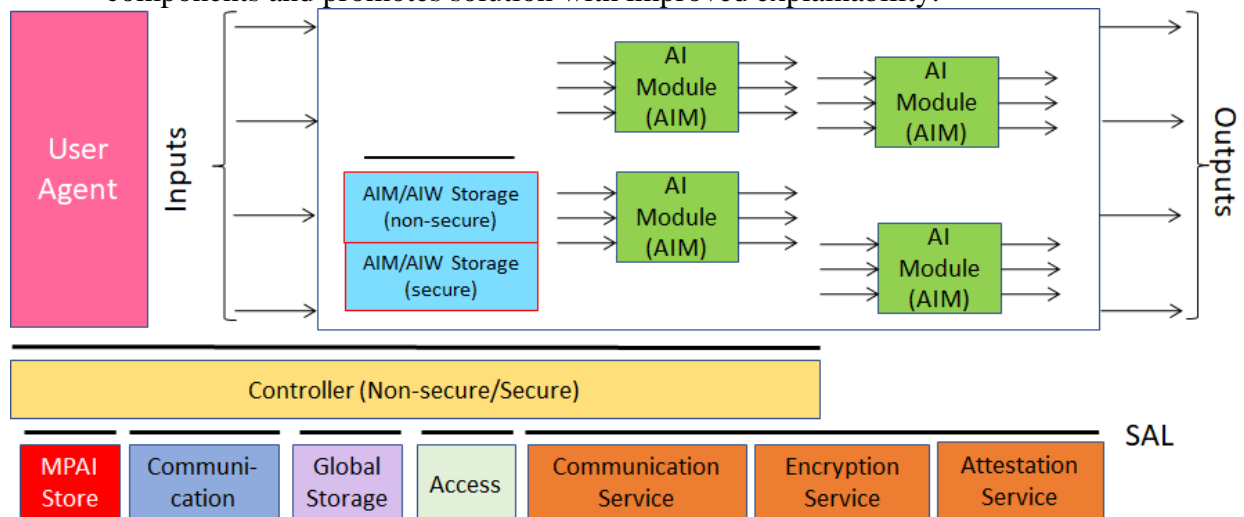

Figure 1 – The AI Framework (MPAI-AIF) V2 Reference Model

- *Technical Specification: Governance of the MPAI Ecosystem* defining the following elements:
    1. <u>Standards</u>, i.e., the ensemble of Technical Specifications, Reference Software, Conformance Testing, and Performance Assessment.
    2. <u>Implementers</u> of MPAI Technical Specifications.
    3. <u>MPAI Store</u> in charge of making AIMs and AIWs submitted by Implementers available to Integrators and End-Users.
    4. <u>Performance Assessors</u>, independent entities assessing the performance of implementations in terms of Reliability, Replicability, Robustness, and Fairness.
    5. <u>End Users</u>.
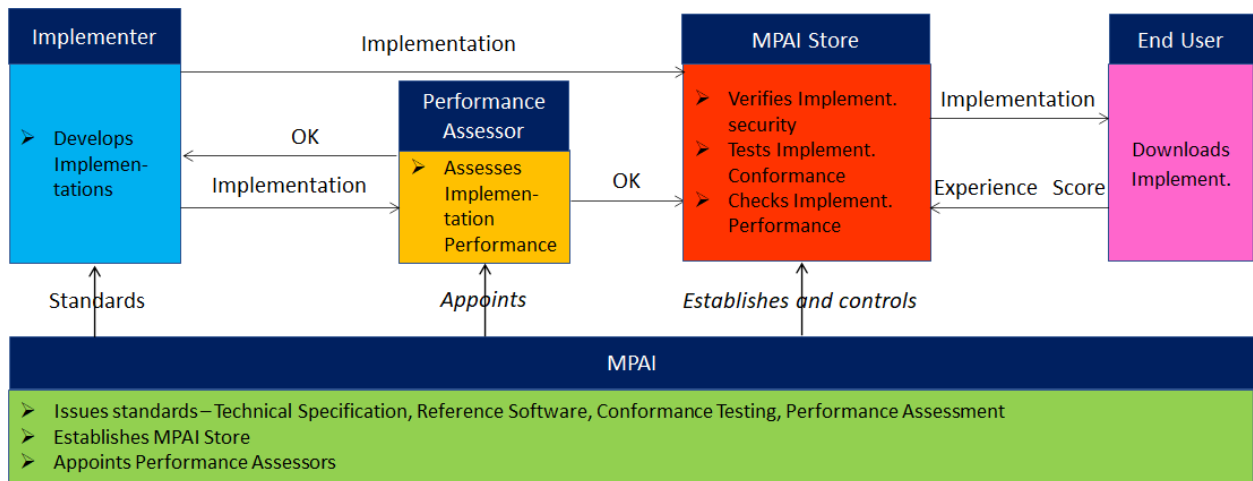
2   Introduction (informative)

Many MPAI Technical Specifications need data types that refer to Objects and Scenes that are uni- and multimodal and refer to locations that may be in virtual spaces.

Technical Specification: Object and Scene Description (MPAI-OSD) Version 1.1 (V1.1) – in the following also called MPAI-OSD V1.1 or MPAI-OSD – has been developed to support the Television Media Analysis (OSD-TMA) AI Workflow (AIW) and to provide a uniform and consistent normative use of Data Types and AI Modules (AIM) across MPAI Technical Specifications.

MPAI-OSD includes the OSD-TMA Reference Software Specification including the Open Source Software (BSD-3-vlause licence) Reference Software Implementation of the AIW and AIMs. The AIMs are sourced by MPAI-MMC, MPAI-OSD, and MPAI-PAF. The Reference Software Implementation offers users the opportunity to access Implementations that are Conforming with the Technical Specification.

In all Chapters and Sections, Terms beginning with a capital letter are defined in _Table 1_ if they are specific to this Technical Specification and in _Table 2_ if they are common to all MPAI Technical Specifications. All Chapters, and Sections are Normative unless they are labelled as Informative.

3   Scope

Technical Specification: Object and Scenes Description (MPAI-OSD) V1.1 specifies:

1. Data Types and associated Qualifiers that describe uni- and multi-modal Objects and Scenes in Virtual Environment with attributes of Space and Time.
2. The Television Media Analysis (OSD-TMA) AI Workflow.
3. The AI Modules required by the OSD-TMA AI Workflow and by Technical Specifications developed by other MPAI organisational groups.

MPAI-OSD V1.1 has been developed for uniform use across MPAI Technical Specifications with the cooperation of

1. The *AI Framework* (AIF-DC), *Context-based Audio Enhancement* (CAE-DC), *Multimodal Conversation* (MMC-DC), and *Portable Avatar Format* (PAF-DC).
2. The *Connected Autonomous Vehicle* (CAV) and *MPAI Metaverse Model* (MMM) groups of the Requirements Standing Committee.

In the future, MPAI may publish new Technical Specification in the MPAI-OSD scope that modify or extend the scope of this Technical Specification.

4   Definitions

Capitalised Terms have the meaning defined in *Table 1*. Terms applicable to all MPAI Technical Specifications are defined in *Table 2*. Non-capitalised terms letter have the meaning commonly defined for the context in which they are used or represent an entity in the real world. For instance,
1. Table 1 defines *Object, Scene,* and *User* but does not define *object, scene,* and *human*.
2. Object indicates an Item but object indicates an entity in the Universe commonly classified as object.

A dash "-" preceding a Term in Table 1 means the following:
1. If the font is normal, the Term in Table 1 without a dash and preceding the one with a dash should be placed <u>before</u> that Term. The notation is used to concentrate in one place all the Terms that are composed of, e.g., the word Data <u>followed</u> by one of the words Format and Type.
2. If the font is *italic,* the Term in the table without a dash and preceding the one with a dash should be placed <u>after</u> that Term. The notation is used to concentrate in one place all the Terms that are composed of, e.g., the word Descriptor <u>preceded</u> by one of the words Face and Body.

Table 1 – Terms and Definitions

Term	Definition
Attitude	
– *Spatial*	Position and Orientation and their velocities and accelerations of a Human and Visual Object in a Virtual Environment.
Audio	A Data Type an instance of which represents analogue signals – or is rendered to be perceived – in the human-audible range (16 Hz – 20 kHz).
Avatar	An Object rendered to represent a Human of a Machine in a virtual space.
– Model	An inanimate Avatar exposing animation interfaces.
– *Portable*	A Data Type including Avatar ID, Time, Audio-VisualScene Descriptors, Spatial Attitude, Avatar Model, Body Descriptors, Face Descriptors, Language Preference, Speech Coding, Speech Data, Text, and Personal Status [5].
Centre Point	The point of an Object selected to have coordinates (0,0,0).
Context	Additional information about a communication emitted by an Entity, such as language, culture etc..
Data	Information in digital form.
– Format	A specific digital representation of Data.

–　　Type	A recognised instance of Data.
Descriptor	The Digital Representation of a feature of an Object.
–　　*Body*	A Data Type including the digital representation of the features of the body of a real or digital human.
–　　*Face*	A Data Type including the digital representation of a feature of the face of a real or digital human.
Digital Representation	Data corresponding to and representing a physical entity.
Environment	A Virtual Space that may be null or may include an Audio-Visual Scene.
Human	A human being in a real space.
–　　*Digital*	A Digitised or a Virtual Human.
–　　*Digitised*	An Object that has the appearance of a specific human when rendered.
–　　*Virtual*	An Object 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.
Instance	An element of a set of entities – Objects, Digital Humans etc. – belonging to some levels in a hierarchical classification (taxonomy).
–　　*Audio*	The instance of an Audio Object.
–　　*Visual*	The instance of a Visual Object.
Object	A data structure that can be rendered to cause an Experience.
–　　*Audio*	An Object described by Audio Descriptors.
–　　*Audio-Visual*	An Object described by Audio-Visual Descriptors.
–　　*Body*	A digital representation of the body of a Human or a Machine.
–　　*Descriptor*	The digital representation of the feature of an Object.
–　　*Digital*	A Digitised or a Virtual Object.
–　　*Digitised*	The digital representation of a real object.
–　　*Face*	The digital representation of the face of a Human or a Machine.
–　　*Speech*	An Object described by Speech Descriptors.
–　　*Text*	A string of Text.
–　　*Virtual*	An Object not representing an object in the real environment.
–　　*Visual*	An Object described by Visual Descriptors.
Orientation	The 3 Euler angles of an Object in a Virtual Space.
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 composition of Objects located according to a Scene Geometry.
–　　*Audio*	A Scene composed of Audio Objects.

–	*Digital*	A digitised scene or a Virtual Scene
–	*Audio-Visual*	A Scene composed of Audio Objects, Visual Objects and co-located Audio-Visual Objects.
–	*Visual*	A Scene composed of Visual Objects.
Scene Descriptors		The digital representation of a feature of a scene.
–	*Audio*	A Data Type including the digital representation of the audio features of a digital scene.
–	*Audio-Visual*	A Data Type combining the Audio or Visual Scene Descriptors.
–	*Visual*	A Data Type including the digital representation of the visual features of a digital scene.
Scene Geometry		The digital representation of the Object arrangement of a Scene.
–	*Audio*	A Data Type describing the Spatial arrangement of the Visual Objects of a Scene.
–	*Audio-Visual*	A Data Type describing the Spatial arrangement of the Audio, Visual, and Audio-Visual Objects of a Scene.
–	*Visual*	A Data Type describing the Spatial arrangement of the Visual Objects of a Scene.
Speech		A Data Type an instance of which represents – or is rendered to be perceived – as an analogue signal with vocal characteristics.
Virtual Space		A space generated and maintained by a computing platform that can be rendered.

The Terms used in this standard whose first letter is capital and are not already included in *Table 1* are defined in *Table 2*. To concentrate in one place all the Terms that are composed of a common name followed by other words (e.g., the word Data <u>followed</u> 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 2 – 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 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	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: 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.

5 References

5.1 Normative Reference

1. MPAI; Technical Specification: Governance of the MPAI ecosystem (MPAI-GME) V1.1.
2. MPAI; Technical Specification: AI Framework (MPAI-AIF) V2.0.
3. MPAI; Technical Specification: Multimodal Conversation (MPAI-MMC) V2.2.
4. MPAI; Technical Specification: Portable Avatar Format (MPAI-PAF) V1.2.

5.2 Informative References

5. MPAI; The MPAI Statutes; N421
6. MPAI; Patent Policy
7. MPAI; Framework Licence: Object and Scene Description
8. MPAI; Technical Specification: Context-based Audio Enhancement (MPAI-CAE) V2.2.
9. MPAI; Technical Specification: Connected Autonomous Vehicles (MPAI-CAV) – Architecture (CAV-ARC) V1.1.
10. MPAI; Technical Specification: Connected Autonomous Vehicles (MPAI-CAV) – Technologies (CAV-TEC) V1.1.
11. MPAI; Technical Specification: MPAI Metaverse Model (MPAI-MMM) – Architecture V1.1.
12. MPAI; Technical Specification: MPAI Metaverse Model (MPAI-MMM) – Technologies V1.1.

6 AI Workflows

Technical Specification: Object and Scene Descriptors (MPAI-OSD) V1.1 assumes that implementations will be based on *Technical Specification: AI Framework (MPAI-AIF) V2.0* enabling dynamic configuration, initialisation, and control of AI Workflows (AIW) composed of interconnected AI Modules (AIM) in a standard AI Framework (AIF).

Table 1 displays the AIW specified by MPAI-OSD V1.1. Click the AIW to access its dedicated page, which includes a its functions, reference model, I/O Data, Functions of AIMs, I/O Data of AIMs, and a table providing links to the AIW-related AIW, AIMs, and JSON metadata.

	Name and Specification	**JSON**
OSD-TMA	**Television Media Analysis**	**X**

7 AI Modules

MPAI-OSD specifies the following AIMs. Some are copied or modified from MPA-OSD V1.0. Most are new AIMs specified by MPAI-OSD V1.1.

AIMs	Name and Specification	JSON
OSD-AVA	**Audio-Visual Alignment**	**X**
OSD-ABS	**Audio-Visual Basic Scene Description**	**X**
OSD-AVE	**Audio-Visual Event Description**	**X**
OSD-SDX	**Audio-Visual Scene Demultiplexing**	**X**
OSD-AVS	**Audio-Visual Scene Description**	**X**
OSD-SMX	**Audio-Visual Scene Multiplexing**	**X**
OSD-DVI	**Direct Visual Identification**	**X**
OSD-TVS	**Television Splitting**	**X**
OSD-VBD	**Visual Basic Scene Description**	**X**
OSD-VCD	**Visual Change Detection**	**X**
OSD-VDI	**Visual Direction Identification**	**X**
OSD-VII	**Visual Instance Identification**	**X**
OSD-VOE	**Visual Object Extraction**	**X**
OSD-VOI	**Visual Object Identification**	**X**
OSD-VSD	**Visual Scene Description**	**X**

8   Data Types

Table 1 provides the Data Types specified by MPAI-OSD V1.1. MPAI-OSD AIWs and AIMs also utilise Data Types specified by other MPAI Technical Specifications. The linked list of all MPAI Data Types is available.

Table 1 – Data Types specified by MPAI-OSD V1.1

Annotation	**Audio-Visual Basic Scene Descriptors**	**Audio-Visual Basic Scene Geometry**	**Audio-Visual Event Descriptors**
Audio-Visual Object	**Audio-Visual Scene Descriptors**	**Audio-Visual Scene Geometry**	**Basic Location**
BoundingBox	**Coordinates**	**Instance Identifier**	**Location**
Orientation	**Path**	**Perceptible Entity**	**Point of View**
Position	**Right Parallelepiped**	**Selector**	**SpaceTime**
Spatial Attitude	**Time**	**Visual Basic Scene Descriptors**	**Visual Basic Scene Geometry**
Visual Object	**Visual Scene Descriptors**	**Visual Scene Geometry**	