INTERNATIONAL STANDARD

**ISO/IEC 15938-9**

First edition
2005-04-01

**AMENDMENT 1**
2012-06-01



# Information technology — Multimedia content description interface —

## Part 9: Profiles and levels

## AMENDMENT 1: Extensions to profiles and levels

*Technologies de l'information — Interface de description du contenu multimédia —*

*Partie 9: Profils et niveaux*

*AMENDEMENT 1: Extensions aux profils et niveaux*

Reference number
ISO/IEC 15938-9:2005/Amd.1:2012(E)



ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

 

## Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

Amendment 1 to ISO/IEC 15938-9:2005 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

# Information technology — Multimedia content description interface —

## Part 9: Profiles and levels

## AMENDMENT 1: Extensions to profiles and levels

*After 4.4, add the following subclauses:*

### 4.5 AudioVisual Description Profile (AVDP)

#### 4.5.1 Application Areas

The description tools in this profile can be used to describe the results of various kinds of media analysis with visual and audio low-level features. Consequently the information resulting from media analysis, such as shot/scene detection, face recognition/tracking, speech recognition, copy detection and summarisation, can be used in a wide range of applications such as archive management, news, new services for new media, and many academic projects handling large scale video content, e.g. benchmarking initiatives. The goal is to support establishing an environment where users can exchange evaluation results of media analysis, hence helping the introduction of automatic media analysis tools into many application areas.

#### 4.5.2 Functionality

The main functionality of this profile is describing the results of automatic media analysis with low-, mid- and high-level features for audio-visual content.

The profile supports high-level functionalities such as (but not limited to) the following:

- Provides tools for describing feature extraction tool, version, contributors, and the date/time the tool applied
- Provides tools for describing several results with confidence levels in multiple timelines.
- Provides tools for describing the results of segmentation, text recognition, person identification, format detection, genre detection, keyword extraction, speech recognition, and subject classification.
- Provides tools for describing the results of copy detection and summarization with having several descriptions of audio-visual contents in one document.

#### 4.5.3 Tools in the Profile

This profile is defined with respect to the Version 2 schema as specified in ISO/IEC 15938-10. The namespace of the Version 2 schema providing a basis for this profile is “urn:mpeg:mpeg7:schema:2004.” The following table lists the ISO/IEC 15938 description tools (global elements, global attributes, attribute groups, complexTypes and simpleTypes) and the profile constraints that are part of this profile.

**Table AMD 1.1 — Description Profile Tool for AVDP**

| **Global Elements** | **Name** | **Constraint** |
|---|---|---|
| Mpeg7 | DescriptionUnit | element excluded |
| | Description | The document may contain an unlimited number of content descriptions (e.g. for describing a set of contents which contain copies and need to be cross-referenced, or for referencing source segments from a summary). At most one of these descriptions can be of type SummaryDescriptionType. |
| **Global Attributes** | **Name** | **Constraint** |
| dim | | |
| **Global Attribute Groups** | **Name** | **Constraint** |
| audioSpectrumAttributeGrp | | |
| mediaTimePropertyGrp | mediaTimeBase | The XPath shall not contain axis specifiers, i.e. it excludes elements [5] and [6] from the datatype definition in Annex A.1 of ISO/IEC 15938-5.<br><br>On the root a/v segment, mediaTimeBase may only point to a MediaLocator element in the same description, i.e. the value is limited to be one of<br><br>• MediaLocator<br><br>• MediaInformation/MediaProfile/MediaInstance[i]/MediaLocator where i is integer, and can be omitted if only one profile/instance exists. |
| referenceGrp | idref | attribute excluded |
| | href | attribute excluded |
| | xpath | ReferenceType is only used in Availability/InstanceRef. The value of the xpath attribute shall be "../../../MediaInformation/MediaProfile/MediaInstance[i]" where i is integer, and can be omitted if only one profile/instance exists. |

| Complex Types | Name | Constraint |
|---|---|---|
| AdvancedFaceRecognitionType | | |
| AgentType | Icon | element excluded |
| AudioBPMType | | |
| AudioDSType | | |
| AudioDType | | |
| AudioFundamentalFrequencyType | | |
| AudioHarmonicityType | | |
| AudioLLDScalarType | | |
| AudioLLDVectorType | | |
| AudioPowerType | | |
| AudioSegmentTemporalDecompositionType | AudioSegmentRef | element excluded |
| AudioSegmentType | MediaTime | minOccurs = "1" |
| | TemporalMask | element excluded |
| | MediaSourceDecomposition | element excluded |
| AudioSignalQualityType | | |
| AudioSignatureType | | |
| AudioSpectrumBasisType | | |
| AudioSpectrumCentroidType | | |
| AudioSpectrumEnvelopeType | | |
| AudioSpectrumFlatnessType | | |
| AudioSpectrumProjectionType | | |
| AudioSpectrumSpreadType | | |

| | | |
|---|---|---|
| AudioSummaryComponentType | Title | element excluded [a)] |
| | AudioSourceID | minOccurs = "1"; AudioSourceID is defined to reference MediaInformation/MediaProfile/ MediaInstance[i]/InstanceIdentifier, where i is integer, of the root content segment of a content description. |
| | AudioSourceLocator | minOccurs = "1" |
| | ComponentSourceTime | element excluded |
| | SoundLocator | element excluded |
| | SyncTime | element excluded |
| AudioTempoType | | |
| AudioVisualSegmentMediaSourceDecompositionType | AudioVisualSegment | element excluded |
| | AudioVisualSegmentRef | element excluded |
| | VideoSegmentRef | element excluded |
| | StillRegionRef | element excluded |
| | AudioSegmentRef | element excluded |
| AudioVisualSegmentTemporalDecompositionType | AudioVisualSegmentRef | element excluded |
| | AudioVisualRegion | element excluded |
| | AudioVisualRegionRef | element excluded |
| AudioVisualSegmentType | MediaTime | minOccurs = "1" |
| | TemporalMask | element excluded |
| | SpatialDecomposition | element excluded |
| | SpatioTemporalDecomposition | element excluded |
| AudioVisualType | | |
| AudioWaveformType | | |
| AvailabilityType | Financial | element excluded |
| | Rights | element excluded |
| BackgroundNoiseLevelType | | |
| BalanceType | | |
| BandwidthType | | |

| | | |
|---|---|---|
| BoxListType | | |
| CameraMotionSegmentType | | |
| CameraMotionType | | |
| ClassificationType | Release | element excluded |
| | Target | element excluded |
| | ParentalGuidance | element excluded |
| | MediaReview | element excluded |
| ColorLayoutType | | |
| ColorQuantizationType | | |
| ColorSamplingType | | |
| ColorSpaceType | | |
| ColorStructureType | | |
| ColorTemperatureType | | |
| CompleteDescriptionType | DescriptionMetadata | element excluded |
| | Relationships | element excluded |
| | OrderingKey | element excluded |
| ConfusionCountType | | |
| ContentAbstractionType | | |
| ContentDescriptionType | | |
| ContentEntityType | MultimediaContent | maxOccurs = "1" |
| ContourShapeType | | |
| ControlledTermUseType | | |
| CreationInformationType | | |
| CreationToolType | | |
| CreationType | TitleMedia | may only be used on the root a/v segment |
| CreatorType | | |
| CrossChannelCorrelationType | | |

| | | |
|---|---|---|
| DSType | Header | The following headers may be used:<br>*Description Metadata*<br>• A complete DescriptionMetadata DS is required on each top level temporal decomposition, and optional on the root a/v segment.<br>• A description metadata element that only specifies the Confidence of a description scheme, may be used on any segment whenever the description scheme does not provide any other means to specify the confidence of the description. It is assumed that the confidence levels are not normalised across tools and thus they are not directly comparable. The Confidence shall not occur in description metadata headers in summaries.<br>*Headers from part 3 and 4 types* may be used on any segment.<br>The use of Header is prohibited on all other elements. |
| | id | IDs shall be used on the following elements:<br>• All segments<br>• Top-level temporal decompositions<br>IDs may be used on any other description scheme. |
| | timePropertyGrp | attribute group excluded |
| | mediaTimePropertyGrp | mediaTimePropertyGrp shall be used on the root a/v segment, and allowed only on TemporalDecompositionType, SpatioTemporalDecomposition-Type SummarizationType and SummaryType.<br>If SummarizationType has mediaTimepropertyGrp, each of the SummaryType elements may have mediaTimePropertyGrp, otherwise each of the element shall have mediaTimePropertyGrp. |
| DType | | |

| | | |
|---|---|---|
| DcOffsetType | | |
| DescriptionMetadataType | Package | element excluded |
| | * | At least one element shall be present. |
| DescriptionProfileType | | |
| DisseminationType | | |
| DominantColorType | | |
| DoubleDiagonalMatrixType | | |
| DoubleMatrixType | | |
| EdgeHistogramType | | |
| ElectronicAddressType | | |
| ErrorEventType | | |
| ExtendedLanguageType | | |
| FaceRecognitionType | | |
| FigureTrajectoryType | | |
| FloatDiagonalMatrixType | | |
| FloatMatrixType | | |
| FocusOfExpansionType | | |
| FractionalPresenceType | | |
| GeographicPointType | | |
| GoFGoPColorType | | |
| GofGopFeatureType | | |
| GridLayoutType | | |
| HarmonicInstrumentTimbreType | | |
| HarmonicSpectralCentroidType | | |
| HarmonicSpectralDeviationType | | |
| HarmonicSpectralSpreadType | | |
| HarmonicSpectralVariationType | | |
| HeaderType | | |

| | | |
|---|---|---|
| HierarchicalSummaryType | SummaryThemeList | element excluded |
| | SummarySegmentGroupRef | element excluded |
| | components | attribute excluded |
| | hierarchy | fixed="dependent" |
| HomogeneousTextureType | | |
| IlluminationInvariantColorType | | |
| ImageLocatorType | MediaTimePoint | element excluded |
| | MediaRelTimePoint | element excluded |
| | BytePosition | element excluded |
| ImageTextType | | |
| InlineTermDefinitionType | Name | (See footnote) [b] |
| | Definition | element excluded |
| | Term | element excluded |
| InstrumentTimbreType | | |
| IntegerDiagonalMatrixType | | |
| IntegerMatrixType | | |
| IrregularVisualTimeSeriesType | | |
| KeyType | | |
| KeywordAnnotationType | | |
| LexiconType | | |
| LogAttackTimeType | | |
| MediaAgentType | AgentRef | element excluded |
| MediaFormatType | Medium | element excluded |
| | System | element excluded |
| | Bandwidth | element excluded |
| | TargetChannelBitRate | element excluded |
| | ScalableCoding | element excluded |
| | Emphasis | element excluded |
| | Presentation | element excluded |
| | SceneCodingFormat | element excluded |
| | GraphicsCodingFormat | element excluded |
| | OtherCodingFormat | element excluded |

| | | |
|---|---|---|
| MediaIdentificationType | AudioDomain | element excluded |
| | VideoDomain | element excluded |
| | ImageDomain | element excluded |
| MediaIncrDurationType | mediaTimeUnit | attribute excluded<br><br>The value of the mediaTimeUnit attribute on the closest parent TemporalDecomposition or SpatiotemporalDecomposition element applies. If the attribute is not present on any of the parent TemporalDecomposition or SpatiotemporalDecomposition elements, the value of the mediaTimeUnit attribute of the root a/v segment shall be used. |
| MediaInformationType | MediaProfile | maxOccurs = "1" |
| MediaInstanceType | LocationDescription | element excluded |
| MediaLocatorType | InlineMedia | element excluded |
| MediaProfileType | MediaTranscodingHints | element excluded |
| | master | attribute excluded |
| MediaQualityType | RatingInformationLocator | element excluded |
| MediaRelIncrTimePointType | mediaTimeUnit | attribute excluded<br><br>The value of the mediaTimeUnit attribute on the closest parent TemporalDecomposition or SpatiotemporalDecomposition element applies. If the attribute is not present on any of the parent TemporalDecomposition or SpatiotemporalDecomposition elements, the value of the mediaTimeUnit attribute of the root a/v segment shall be used. |
| | mediaTimeBase | attribute excluded<br><br>The value of the mediaTimeBase attribute on the closest parent TemporalDecomposition or SpatiotemporalDecomposition element applies. If the attribute is not present on any of the parent TemporalDecomposition or SpatiotemporalDecomposition elements, the value of the mediaTimeBase attribute of the root a/v segment shall be used. |

| | | |
|---|---|---|
| MediaSourceSegmentDecompositionType | | |
| MediaTimeType | MediaTimePoint | element excluded |
| | MediaRelTimePoint | element excluded |
| | MediaDuration | element excluded |
| MelodyContourType | | |
| MelodySequenceType | | |
| MelodyType | | |
| MeterType | | |
| MinusOneToOneMatrixType | | |
| MixtureAmountOfMotionType | | |
| MixtureCameraMotionSegmentType | | |
| MotionActivityType | | |
| MotionTrajectoryType | | |
| MovingRegionFeatureType | | |
| MovingRegionSpatioTemporalDecompositionType | MovingRegionRef | element excluded |
| | StillRegionRef | element excluded |
| MovingRegionType | SpatioTemporalMask | element excluded |
| | GofGopFeature | element excluded |
| | MultipleView | element excluded |
| | SpatialDecomposition | element excluded |
| | SpatioTemporalLocator | minOccurs = "1" |
| | TemporalDecomposition | element excluded |
| | MediaSourceDecomposition | element excluded |
| Mpeg7BaseType | | |
| Mpeg7Type | DescriptionProfile | minOccurs = "1" |
| | DescriptionMetadata | minOccurs = "1" |
| | timePropertyGrp | attribute group excluded |
| | mediaTimePropertyGrp | attribute group excluded |
| MultimediaContentType | | |

| | | |
|---|---|---|
| NameComponentType | | |
| NonMixtureAmountOfMotionType | | |
| NonMixtureCameraMotionSegmentType | | |
| OrganizationType | ContactRef | element excluded |
| | Jurisdiction | element excluded |
| | JurisdictionRef | element excluded |
| | AddressRef | element excluded |
| ParameterTrajectoryType | | |
| ParametricMotionType | | |
| PercussiveInstrumentTimbreType | | |
| PersonGroupType | MemberRef | element excluded |
| | Jurisdiction | element excluded |
| | JurisdictionRef | element excluded |
| | AddressRef | element excluded |
| PersonNameType | | |
| PersonType | OrganizationRef | element excluded |
| | PersonGroupRef | element excluded |
| | AddressRef | element excluded |
| PhoneLexiconType | | |
| PlaceType | | |
| ProbabilityMatrixType | | |
| RatingType | | |
| ReferenceType | | |
| RegionLocatorType | | |
| RegionShapeType | | |
| RegularVisualTimeSeriesType | | |

| | | |
|---|---|---|
| RelatedMaterialType | DisseminationFormat | element excluded |
| | MediaInformation | element excluded |
| | MediaInformationRef | element excluded |
| | CreationInformation | element excluded |
| | CreationInformationRef | element excluded |
| | UsageInformation | element excluded |
| | UsageInformationRef | element excluded |
| RelationType | type | use = "required" |
| | source | attribute removed [c] |
| | strength | Note: It is assumed that the strength levels are not normalised across tools and thus they are not directly comparable. |
| | target | restricted length of list to 1 [d] |
| RelativeDelayType | | |
| RightsType | | |
| ScalableColorType | | |
| ScalableSeriesType | | |
| SegmentDecompositionType | criteria | restricted to URI pattern [e]<br><br>criteria of the decomposition deeper than 2 level shall not use the same value of criteria previously used in their ancestor decomposition level, i.e. level 1 and 2. [f] |
| SegmentType | MediaInformation | MediaInformation shall be used on the root a/v segment and may not be used on any other segment |
| | MediaInformationRef | element excluded |
| | MediaLocator | MediaLocator can be optionally used on key frame segments to refer to extracted key frame images |
| | StructuralUnit | minOccurs = "1"<br><br>StructuralUnit of the segment deeper than 2 level shall not use the same value of StructuralUnit previously used in their ancestor segment level, i.e. level 1 and 2. [g] |
| | CreationInformationRef | element excluded |

| | | |
|---|---|---|
| | UsageInformation | UsageInformation can be optionally used on the root a/v segment and may not be used on any other segment |
| | UsageInformationRef | element excluded |
| | Semantic | element excluded |
| | SemanticRef | element excluded |
| | MatchingHint | element excluded |
| | PointOfView | element excluded |
| SequentialSummaryType | AudioVisualSummaryLocator | AudioVisualSummaryLocator is optional, but if used, it shall point to a segment representing the complete content, i.e. either the root a/v segment of a description, or a segment in a decomposition of the root a/v segment with the same start time and duration. |
| | VideoSummaryLocator | VideoSummaryLocator is optional, but if used, it shall point to a segment representing the complete content, i.e. either the root a/v segment of a description, or a segment in a decomposition of the root a/v segment with the same start time and duration. |
| | AudioSummaryLocator | AudioSummaryLocator is optional, but if used, it shall point to a segment representing the complete content, i.e. either the root a/v segment of a description, or a segment in a decomposition of the root a/v segment with the same start time and duration. |
| | TextualSummaryComponent | element excluded |
| | components | attribute excluded |
| SeriesOfScalarBinaryType | | |
| SeriesOfScalarType | | |
| SeriesOfVectorBinaryType | | |
| SeriesOfVectorType | | |
| Shape3DType | | |
| ShapeVariationType | | |

| | | |
|---|---|---|
| SilenceHeaderType | | |
| SilenceType | | |
| Spatial2DCoordinateSystemType | | |
| SpatialSegmentDecomposition Type | | |
| SpatioTemporalLocatorType | | |
| SpatioTemporalSegmentDecomp ositionType | | |
| SpeakerInfoType | | |
| SpectralCentroidType | | |
| SpokenContentHeaderType | | |
| SpokenContentIndexEntryType | | |
| SpokenContentLatticeType | | |
| SpokenContentLinkType | | |
| StillRegionFeatureType | | |
| StillRegionSpatialDecomposition Type | StillRegionRef | element excluded |
| StillRegionType | SpatialMask | element excluded |
| | MediaTimePoint | element excluded |
| | MediaRelTimePoint | element excluded |
| | IlluminationInvariantColor | element excluded [h)] |
| | MultipleView | element excluded |
| StructuredAnnotationType | | |
| SummarizationType | | |
| SummaryDescriptionType | Summarization | maxOccurs = "1" |
| SummarySegmentGroupType | Caption | element excluded |
| | SummarySegmentRef | element excluded |
| | SummarySegmentGroupRef | element excluded |
| | level | attribute excluded |
| | duration | attribute excluded |
| | numOfKeyFrames | attribute excluded |

| | | |
|---|---|---|
| | fidelity | fidelity will be used to define a relative score for using alternative segments instead of the parent segment. There can be groups with just one segment on the top level, the score is then 1.0. If it is necessary to have such a score for each segment, a binary tree with a segment and a group on each level can be created. If no values for fidelity are specified, the fidelity of the segments is assumed to be equal. It is assumed that the fidelity levels are not normalised across tools and thus they are not directly comparable. |
| | themeIDs | attribute excluded |
| SummarySegmentType | SourceID | SourceID is defined to reference MediaInformation/MediaProfile/MediaInstance/InstanceIdentifier of the root content segment of a content description. If it is not present, the value SummaryType/SourceID applies to the segment. |
| | KeyFrame | element excluded |
| | KeySound | element excluded |
| | themeIDs | attribute excluded |
| | order | (See footnote) [i)] |
| SummaryType | SourceID | SourceID is defined to reference MediaInformation/MediaProfile/MediaInstance[i]/InstanceIdentifier, where i is integer, of the root content segment of a content description. |
| | SourceLocator | element excluded |
| | SourceInformation | element excluded |
| TemporalCentroidType | | |
| TemporalInterpolationType | MediaDuration | element excluded |
| | MediaTimePoint | element excluded |
| | MediaIncrTimePoint | element excluded |
| TemporalSegmentDecompositionType | | |
| TemporalSegmentLocatorType | BytePosition | element excluded |

| | | |
|---|---|---|
| TermDefinitionBaseType | | |
| TermUseType | | |
| TextAnnotationType | DependencyStructure | element excluded |
| TextualBaseType | | |
| TextualType | | |
| TextureBrowsingType | | |
| TimeType | RelTimePoint | element excluded |
| | RelIncrTimePoint | element excluded |
| | IncrDuration | element excluded |
| TitleMediaType | | |
| TitleType | | |
| TransmissionTechnologyType | | |
| UniqueIDType | | |
| UsageInformationType | Rights | element excluded |
| | FinancialResults | element excluded |
| | UsageRecord | element excluded |
| VideoSegmentFeatureType | | |
| VideoSegmentSpatioTemporalDecompositionType | MovingRegionRef | element excluded |
| | StillRegionRef | element excluded |
| VideoSegmentTemporalDecompositionType | VideoSegmentRef | element excluded |
| | StillRegionRef | element excluded |
| VideoSegmentType | MediaTime | minOccurs = "1" |
| | TemporalMask | element excluded |
| | MultipleView | element excluded |
| | Mosaic | element excluded |
| | SpatialDecomposition | element excluded |
| | MediaSourceDecomposition | element excluded |
| VideoTextType | | |
| VisualDSType | | |
| VisualDType | | |
| VisualSummaryComponentType | VideoSourceID | minOccurs = "1"; VideoSourceID is defined to reference |

| | | |
|---|---|---|
| | | MediaInformation/MediaProfile/ MediaInstance[i]/InstanceIdentifier, where i is integer, of the root content segment of a content description. |
| | VideoSourceLocator | minOccurs = "1" |
| | ComponentSourceTime | element excluded |
| | ImageLocator | element excluded |
| | SyncTime | element excluded |
| | FrameActivity | element excluded |
| | Region | element excluded |
| VisualTimeSeriesType | | |
| WordLexiconType | | |
| **Simple Types** | **Name** | **Constraint** |
| PhoneLexiconIndexType | | |
| PhoneType | | |
| WordLexiconIndexType | | |
| WordType | | |
| basicDurationType | | |
| basicTimePointType | | |
| beatType | | |
| characterSetCode | | |
| contourType | | |
| countryCode | | |
| currencyCode | | |
| curvatureType | | |
| degreeAccidentalType | | |
| degreeNoteType | | |
| doubleVector | | |
| durationType | | |
| floatVector | | |
| integerVector | | |
| listOfPositiveIntegerForDim | | |

| | | |
|---|---|---|
| mimeType | | |
| minusOneToOneType | | |
| minusOneToOneVector | | |
| nonNegativeReal | | |
| phoneticAlphabetType | | |
| probabilityVector | | |
| regionCode | | |
| scaleType | | |
| termAliasReferenceType | | |
| termReferenceListType | | |
| termReferenceType | | |
| termRelationQualifierType | | |
| termURIReferenceType | | |
| textureListType | | |
| timePointType | | |
| unsigned1 | | |
| unsigned10 | | |
| unsigned11 | | |
| unsigned12 | | |
| unsigned13 | | |
| unsigned14 | | |
| unsigned15 | | |
| unsigned16 | | |
| unsigned17 | | |
| unsigned18 | | |
| unsigned19 | | |
| unsigned2 | | |
| unsigned20 | | |
| unsigned21 | | |
| unsigned22 | | |
| unsigned23 | | |

| | | |
|---|---|---|
| unsigned24 | | |
| unsigned25 | | |
| unsigned26 | | |
| unsigned27 | | |
| unsigned28 | | |
| unsigned29 | | |
| unsigned3 | | |
| unsigned30 | | |
| unsigned31 | | |
| unsigned32 | | |
| unsigned4 | | |
| unsigned5 | | |
| unsigned6 | | |
| unsigned7 | | |
| unsigned8 | | |
| unsigned9 | | |
| xPathAbsoluteSelectorType | | |
| xPathFieldType | | |
| xPathRefType | | |
| xPathSelectorType | | |
| xPathType | | |
| zeroToOneType | | |

[a] For conformance with VisualSummaryComponentType

[b] included to increase readability in addition to reference to term in CS, ControlledTermUseType has mandatory href attribute

[c] It is assumed that the segment on which the relation is annotated is the source of the relation.

[d] if there are relations between a segment and several others, several relation elements shall be used.

[e] Use reference to term in CS

[f] The level of segment is defined by the number of decompositions needed to reach the segment from the root a/v segment.

[g] The decomposition level is the number of decomposition appearance from the root a/v segment.

[h] Considered redundant to VisualDescriptor

[i] Alternative segments in the ground truth shall have the same order.

### 4.5.4 Levels for AVDP

No levels are defined for the AudioVisual Description Profile.

NOTE Levels for this profile may be defined in the future by means of an amendment to this part of ISO/IEC 15938.

### 4.5.5 Semantics for AVDP

This semantics defines the use of the selected tools for providing the interoperability. Figure AMD 1.1 indicates the Top level structure for content description supported by this profile. (a) is the most basic case to describe an audiovisual content. (b) is for describing some relations between more than two audiovisual contents. (c) is for describing a result of summarization itself. (d) is for describing a result of summarization with the related audiovisual contents.

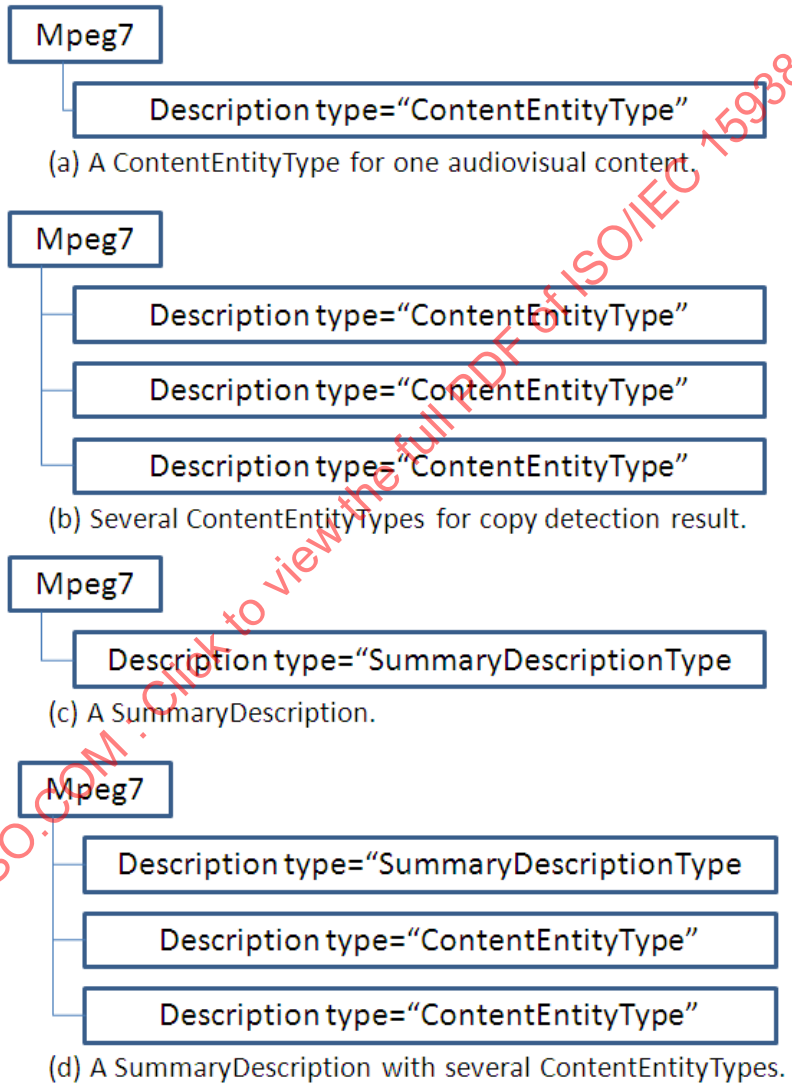


**Figure AMD 1.1 — Top level structure**

Figure AMD 1.2 indicates the high level structure for content description supported by this profile.

The root audiovisual segment (AudioVisual) expresses a whole content to which automatic media analysis is applied, and it is decomposed by Temporal Decompositions (TD) into AudioVisualSegment (AVS, henceforth referred as AVS-1st). Each temporal decomposition corresponds to an experimental result of media analysis, and it has a criteria (e.g. shot, ASR, …). Therefore each AVS-1st segment is a result of media analysis in term of specified criteria, and also take a role of container. If audio and/or video descriptors are needed, the AVS-1st segments are decomposed by means of Media Source Decomposition (MSD) into Video and Audio Segments (VS and AS), to which the descriptors are attached. These VS/AS segments shall have the same duration as the AVS-1st and appropriate structural unit values. There can be as many such video/audio segments (VS/AS) as there are video/audio channels. As for further decomposition into deeper AVS-*n*th in AVS-1st, temporal decomposition (TD) shall be used without the restriction on number of the levels. Levels deeper than 2 shall not use the same criteria/Structural Unit as previously used for higher levels. Spatial or spatiotemporal decompositions are not allowed for the AVS-1st segments. The level of segment is defined by the number of decompositions needed to reach the segment from the root a/v segment. The decomposition level is the number of decomposition appearance from the root a/v segment. Therefore in Figure AMD 1.2, TD under AudioVisual is the first-level decomposition and the level of AVS-1st is 1. As for AS, VS, VS-key, AS-key, SR cannot have “level 1”, they start from level 2. As for Moving Regions (MR), they can start on level 3.

There can be several MSD of an AVS-1st, e.g. one into modalities and one into key elements. If the MSD has the criteria *genericmsd* (from DecompositionCS, expressing non-specific MSD) it may hold all types of segments, however, in this case only this single MSD may be present. If more than one MSD is present, none of them may use the criteria *genericmsd*.

Key frames are represented in the same MSD of AVS-1st segment as the Video segment – (simplest solution, allows to access key frames always in the same way). Key frames could be extended to key clips (1 frame <= key frame duration < duration of AVS-1st segment). Therefore if the duration of the video segment is not equal to that of AVS-1st segment, the video segment is key frame or key clip (VS-key). In a similar way, key audio clips (AS-key) can be included in the same in this MSD. Key frame, key video and audio clips shall be annotated with appropriate structural units. Key frames/clips are not allowed to have any further decomposition and no feature descriptions.

Results of visual and audio feature extraction from the complete segment may be added to the video or audio segment under the MSD of the segment, i.e. AVS-1st. Results of visual feature extraction from single frames within the segment or regions of single frames shall be described using StillRegionType elements contained in the MSD of the segment and optionally further spatial decompositions of the still region. The still region element directly in the MSD shall represent a full frame, those in the decompositions below may refer to regions (e.g. face regions, image text). Results of feature extraction from the complete segment or a subsegment, which create region information with a temporal extent of more than one frame shall be described as moving regions in a spatiotemporal decomposition of the video segment. The moving regions may be further decomposed into moving regions by means of spatiotemporal decompositions.

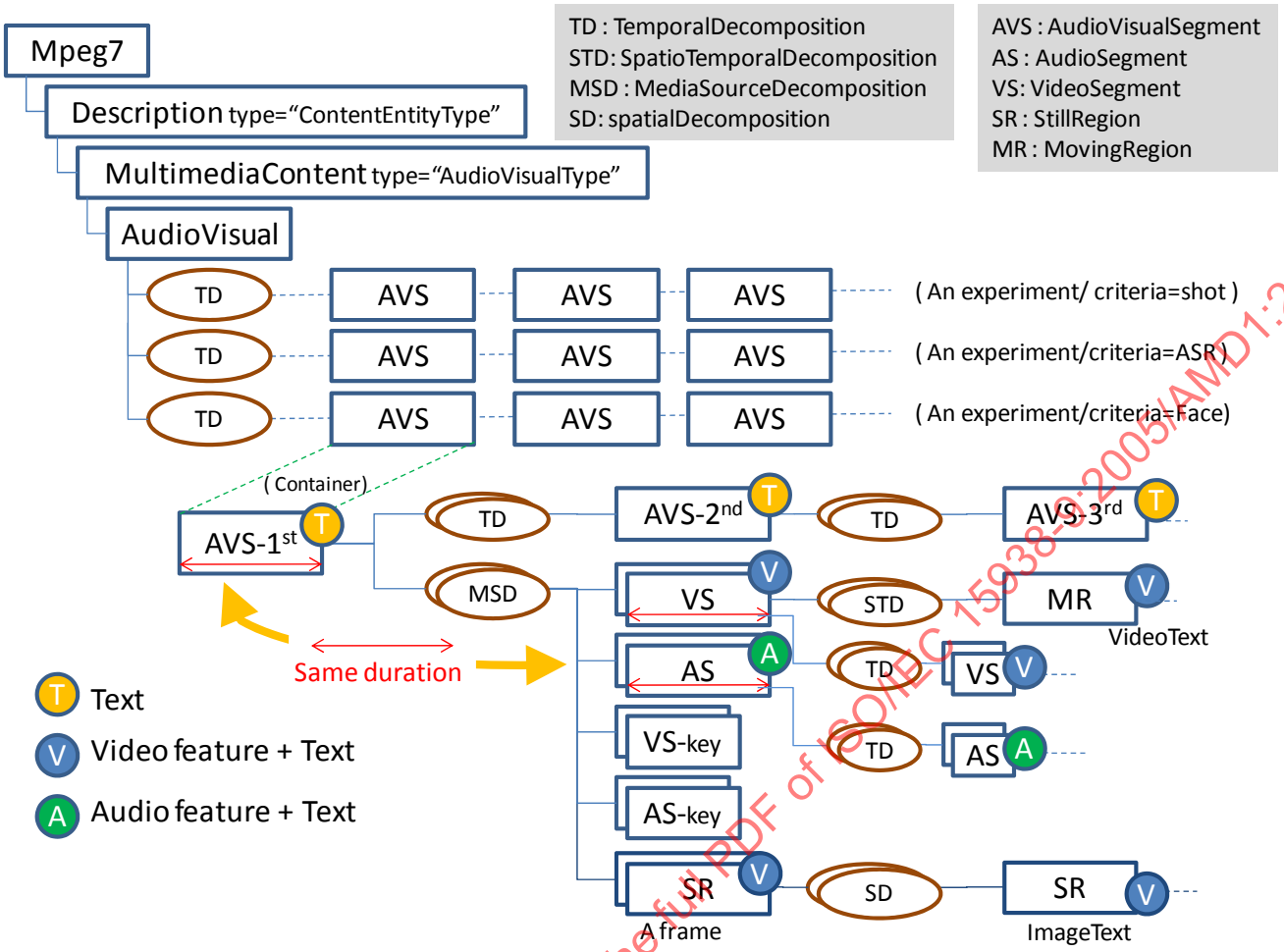


**Figure AMD 1.2 — Basic structure of AVDP descriptions**

### 4.5.6 Classification Schemes

AVDP makes use of reference data also known as controlled terms. It is strongly recommended to use controlled terms, represented e.g. as MPEG-7 classification schemes, whenever possible. While ClassificationSchemeType is not included in the profile, descriptions conforming to the profile can make reference to terms defined in MPEG-7 classification schemes. Different classification schemes can be used e.g. from MPEG or EBU (http://www.ebu.ch/metadata/cs) as their share a common format.

The following table lists all instances of (Controlled)TermUseType in the profile, for which it is considered best practice to reference terms from classification schemes. The table contains some element of other type (e.g. anyURI), for which classification schemes shall be used as well.

**Table AMD 1.2 — Examples of Classification schemes used in AVDP**

| Type | Element/Attribute | Recommended CS |
|---|---|---|
| MediaFormatType | FileFormat | ebu_FileFormatCS |
| | VisualCoding/Format | ebu_videoCompressionCodeCS |
| | AudioCoding/Format | ebu_AudioCompressionCodeCS |
| MediaAgentType | Role | ebu_RoleCodeCS |

| | | |
|---|---|---|
| ClassificationType | Genre | ebu_ContentGenreCS |
| | Format | ebu_EditorialFormatCodeCS |
| RelationType | Type | SegmentRelationCS |
| SegmentType | StructuralUnit | StructuralUnitCS |
| SegmentDecompositionType | Criteria | DecompositionCS |

*Replace Clause 5 with the following:*

## 5 MPEG-7 Profile URI Values and Schema Namespace

The following table contains ISO/IEC 15938 standard profile URI values:

**Table 4 — Profile URI Values**

| MPEG-7 Profile Name | ProfileAndLevelIndication URI Value | ISO/IEC 15938-10 Schema Namespace |
|---|---|---|
| Simple Metadata Profile and Level | urn:mpeg:mpeg7:profiles:2004:SMP | urn:mpeg:mpeg7:schema:2001 |
| User Description Profile and Level | urn:mpeg:mpeg7:profiles:2004:UDP | urn:mpeg:mpeg7:schema:2001 |
| Core Description Profile and Level | urn:mpeg:mpeg7:profiles:2004:CDP | urn:mpeg:mpeg7:schema:2001 |
| AudioVisual Description Profile and Level | urn:mpeg:mpeg7:profiles:2011:AVDP | urn:mpeg:mpeg7:schema:2004 |

*Add the following Annex:*

# Annex A
(informative)

# XM for AudioVisual Description Profile (AVDP)



## A.1 Introduction

This Annex contains AVDP's reference software in the form of metadata instances compliant with the AVDP profile. These instances have been validated during the AVDP evaluation phase using the AVDP validation tool developed by Joanneum Research and accessible at http://vamp.joanneum.at.

## A.2 Reference software: MediaAnalyse data, Joanneum Research, Austria

The MediaAnalyse tool automatically extracts metadata from video files for navigation and structuring (e.g. shot boundaries, key frames) and metadata which enables fast semi-automatic annotation, e.g. camera motion and face occurrences. Figure A.1 shows the MediaAnalyse user interface for batch analysis. An MPEG-7 document is producted for each of the input video files.

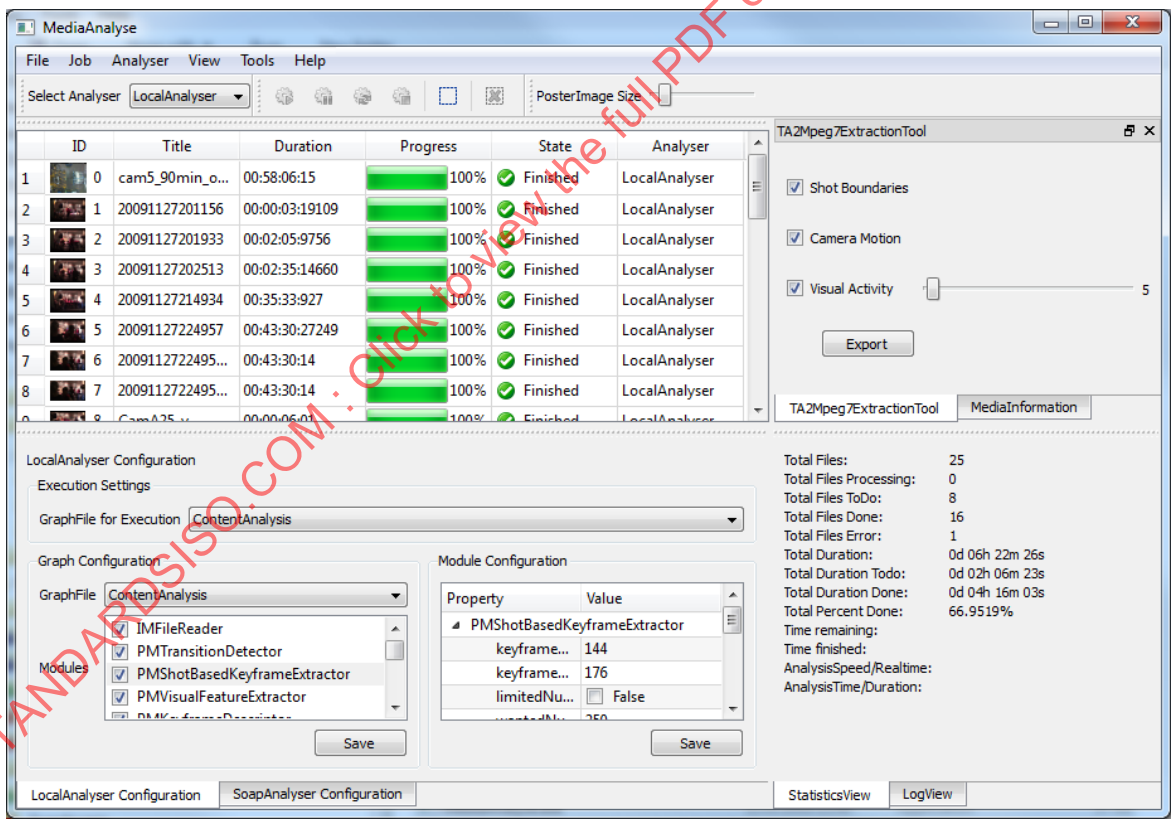


**Figure A.1 — MediaAnalyse User Interface**

The following content analysis modules are included in the tool:

- Shot boundary detection, detecting hard cuts and dissolves
- Key-frame extraction, extracting key frames for each shot
- Stripe images, providing an overview visualisation as an image composed from the center columns of each frame of a segment
- Camera motion, i.e. pan left, pan right, tilt up, tilt down, zoom in, zoom out, a value indicating the strength of the movement is calculated
- Camera motion, i.e. pan left, pan right, tilt up, tilt down, zoom in, zoom out, a value indicating the strength of the movement is calculated
- Face detection, annotating rectangles where a face has been detected in the video

### Description of analysis results

In order to represent the output of the tool, the following elements need to be included in the description:

- The shot structure of the video
- For each shot
    - One or more key frames representing the shot
    - The camera motion description
    - A spatiotemporal decomposition into regions representing face detections
- The stripe images as a separate list of segments

### MPEG-7 AVDP output

The tool generates output documents compliant to MPEG-7 AVDP. An example output of the tool can be found in below (reduced to one shot for readability).

```
<?xml version="1.0" encoding="utf-8"?>
<Mpeg7 xsi:schemaLocation="urn:mpeg:mpeg7:schema:2004 avdp-2010.xsd"
xmlns="urn:mpeg:mpeg7:schema:2004" xmlns:mpeg7="urn:mpeg:mpeg7:schema:2004"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <DescriptionProfile profileAndLevelIndication="urn:mpeg:mpeg7:profiles:2011:AVDP" />
  <DescriptionMetadata>
    <PrivateIdentifier>PID_1</PrivateIdentifier>
    <Instrument>
      <Tool>
        <Name>MediaAnalyse</Name>
      </Tool>
    </Instrument>
  </DescriptionMetadata>
  <Description xsi:type="mpeg7:ContentEntityType">
    <MultimediaContent xsi:type="mpeg7:AudioVisualType">
      <AudioVisual mediaTimeUnit="PT1N23F"
mediaTimeBase="mpeg7:MediaInformation/mpeg7:MediaProfile/mpeg7:MediaInstance/mpeg7:MediaLocator"
id="AVID 1">
        <MediaInformation>
          <MediaProfile>
```

```
            <MediaFormat>
              <Content href="uri:2">
                <Name>Audiovisual</Name>
              </Content>
              <FileFormat href="uri:3">
                <Name>mpeg</Name>
              </FileFormat>
              <FileSize>3385344</FileSize>
            </MediaFormat>
            <MediaInstance>
              <InstanceIdentifier>15122008060</InstanceIdentifier>
              <MediaLocator>
                <MediaUri>M:/TA2(KAR)/TA2-WOODBRIDGE-SCHOOL/(converted)/transcoded/NokiaMPEG4
transcoded/15122008060.mpeg</MediaUri>
              </MediaLocator>
            </MediaInstance>
          </MediaProfile>
        </MediaInformation>
        <StructuralUnit href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#1">
          <Name>Programme</Name>
        </StructuralUnit>
        <CreationInformation>
          <Creation>
            <Title>Undefined</Title>
          </Creation>
        </CreationInformation>
        <MediaTime>
          <MediaRelIncrTimePoint>0</MediaRelIncrTimePoint>
          <MediaIncrDuration>507</MediaIncrDuration>
        </MediaTime>
        <TemporalDecomposition id="VSID 1 TD 1"
criteria="http://www.ebu.ch/metadata/cs/mpeg/avdp/DecompositionCS#20" gap="true">
          <Header xsi:type="mpeg7:DescriptionMetadataType">
            <Instrument>
              <Tool>
                <Name>unknown</Name>
              </Tool>
            </Instrument>
          </Header>
          <AudioVisualSegment id="TRID 1 AV">
            <StructuralUnit href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10">
              <Name>Shot</Name>
            </StructuralUnit>
            <MediaTime>
              <MediaRelIncrTimePoint>0</MediaRelIncrTimePoint>
              <MediaIncrDuration>506</MediaIncrDuration>
            </MediaTime>
            <MediaSourceDecomposition
criteria="http://www.ebu.ch/metadata/cs/mpeg/avdp/DecompositionCS#10" id="TRID 1 MSD mod">
              <VideoSegment id="TRID_1">
                <Header xsi:type="mpeg7:DescriptionMetadataType">
                  <Confidence>1.000000</Confidence>
                </Header>
                <StructuralUnit href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10">
                  <Name>Shot</Name>
                </StructuralUnit>
                <MediaTime>
                  <MediaRelIncrTimePoint>0</MediaRelIncrTimePoint>
                  <MediaIncrDuration>506</MediaIncrDuration>
                </MediaTime>
```

```
                <VisualDescriptor xsi:type="mpeg7:CameraMotionType">
                  <Segment xsi:type="mpeg7:MixtureCameraMotionSegmentType">
                    <MediaTime>
                      <MediaRelIncrTimePoint>0</MediaRelIncrTimePoint>
                      <MediaIncrDuration>24</MediaIncrDuration>
                    </MediaTime>
                    <FractionalPresence>
                      <PanLeft>5</PanLeft>
                      <PanRight>31</PanRight>
                      <TiltDown>127</TiltDown>
                      <TiltUp>21</TiltUp>
                      <RollClockwise>5</RollClockwise>
                      <RollAnticlockwise>127</RollAnticlockwise>
                      <Fixed>47</Fixed>
                    </FractionalPresence>
                    <AmountOfMotion>
                      <PanLeft>0</PanLeft>
                      <PanRight>0</PanRight>
                      <TiltDown>4</TiltDown>
                      <TiltUp>0</TiltUp>
                      <RollClockwise>0</RollClockwise>
                      <RollAnticlockwise>2047</RollAnticlockwise>
                    </AmountOfMotion>
                  </Segment>
                </VisualDescriptor>
                <SpatioTemporalDecomposition criteria="faces">
                  <MovingRegion id="MRID_2">
                    <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#30">
                      <Name>MovingRegion</Name>
                    </StructuralUnit>
                    <SpatioTemporalLocator>
                      <ParameterTrajectory motionModel="still">
                        <MediaTime>
                          <MediaRelIncrTimePoint>55</MediaRelIncrTimePoint>
                          <MediaIncrDuration>1</MediaIncrDuration>
                        </MediaTime>
                        <InitialRegion>
                          <Polygon>
                            <Coords mpeg7:dim="2 4">197 86 0 -86 74 0 86 0</Coords>
                          </Polygon>
                        </InitialRegion>
                      </ParameterTrajectory>
                    </SpatioTemporalLocator>
                  </MovingRegion>
                </SpatioTemporalDecomposition>
              </VideoSegment>
            </MediaSourceDecomposition>
            <MediaSourceDecomposition
criteria="http://www.ebu.ch/metadata/cs/mpeg/avdp/DecompositionCS#11" gap="true" overlap="false"
id="TRID_1_MSD_key">
              <VideoSegment id="SHOT_KFID_1_AV">
                <StructuralUnit href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#11">
                  <Name>Keyframe</Name>
                </StructuralUnit>
                <MediaTime>
                  <MediaRelIncrTimePoint>3</MediaRelIncrTimePoint>
                  <MediaIncrDuration>1</MediaIncrDuration>
                </MediaTime>
              </VideoSegment>
```

```
            </MediaSourceDecomposition>
          </AudioVisualSegment>
        </TemporalDecomposition>
        <TemporalDecomposition id="VSID_1_TD_2" criteria="urn:x-davp_criteria:stripe_images"
gap="true">
          <Header xsi:type="mpeg7:DescriptionMetadataType">
            <Instrument>
              <Tool>
                <Name>unknown</Name>
              </Tool>
            </Instrument>
          </Header>
          <AudioVisualSegment id="SIID 1 AV">
            <StructuralUnit href="urn:x-mpeg-7-davp:cs:StructuralUnitCS:2005:vis.stripeimage">
              <Name>StripeImage</Name>
            </StructuralUnit>
            <MediaTime>
              <MediaRelIncrTimePoint>0</MediaRelIncrTimePoint>
              <MediaIncrDuration>504</MediaIncrDuration>
            </MediaTime>
            <MediaSourceDecomposition
criteria="http://www.ebu.ch/metadata/cs/mpeg/avdp/DecompositionCS#10" id="SIID 1 MSD mod">
              <VideoSegment id="SIID_1" xsi:type="mpeg7:VideoSegmentType">
                <StructuralUnit href="urn:x-mpeg-7-davp:cs:StructuralUnitCS:2005:vis.stripeimage">
                  <Name>StripeImage</Name>
                </StructuralUnit>
                <MediaTime>
                  <MediaRelIncrTimePoint>0</MediaRelIncrTimePoint>
                  <MediaIncrDuration>504</MediaIncrDuration>
                </MediaTime>
              </VideoSegment>
            </MediaSourceDecomposition>
          </AudioVisualSegment>
        </TemporalDecomposition>
      </AudioVisual>
    </MultimediaContent>
  </Description>
</Mpeg7>
```

## A.3 Reference software: ANTS data (Automatic Newscast Transcription System), RAI, Italy

### Functionality

RAI – Centre for research and Technological Innovation provides a platform for ingestion and annotation of news material named ANTS (Automated Newscast Transcription System). ANTS makes extensive usage of automated information extraction techniques. It can schedule material acquisition from digital terrestrial television, identify programme boundaries, detect shot boundaries, extract key frames, segment material into editorial units, transcribe speech into electronic text and classify transcribed text. It finally delivers extracted metadata regarding shot and editorial segmentation in compliance to MPEG-7 AVDP. One MPEG-7 document is produced for each elaborated programme.

Figure A.2 shows the ANTS browsing interface.

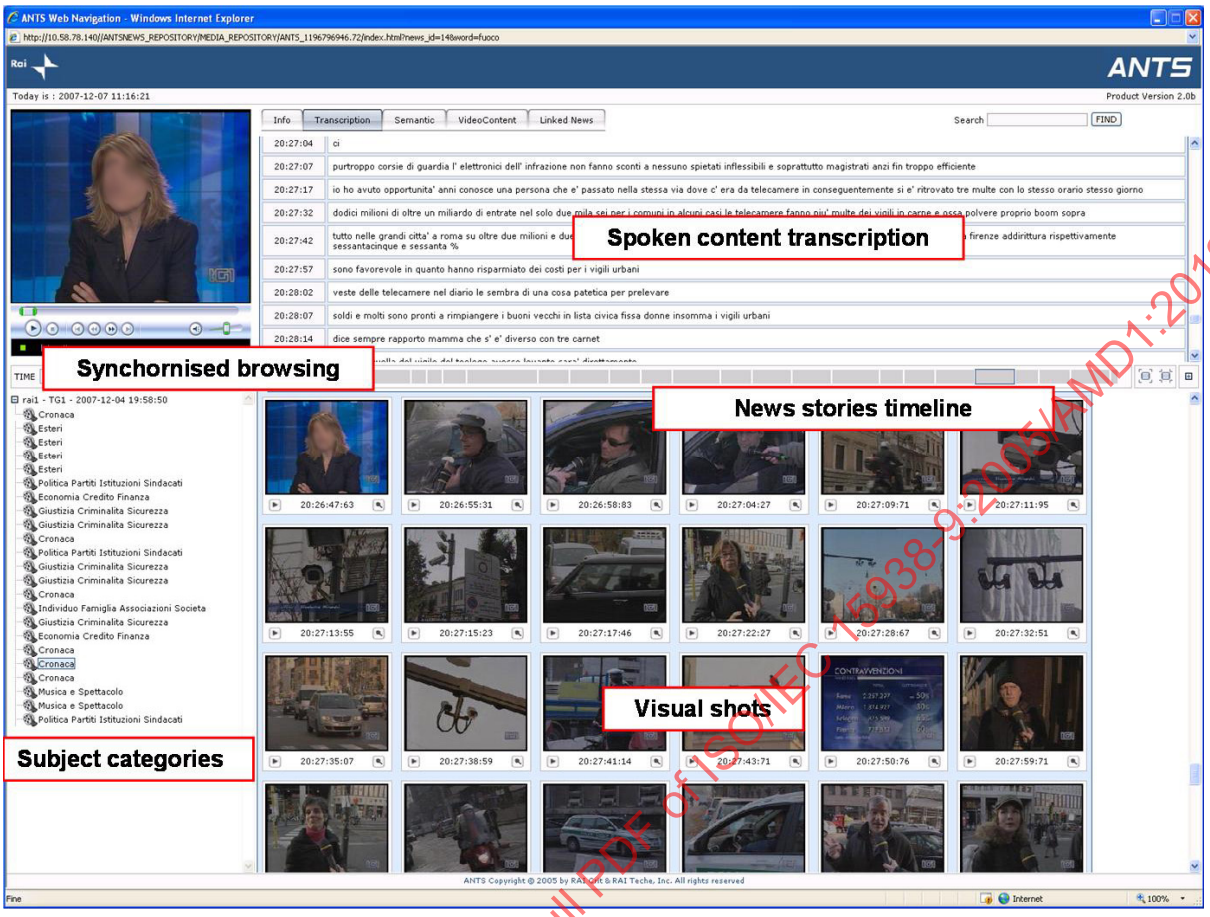


**Figure A.2 — ANTS Browsing Interface.**

The following analysis modules are included in ANTS:

- Shot boundary detection, detecting hard cuts and dissolves
- Key-frame extraction, extracting key frames for each shot
- Spoken content transcription
  - Editorial segmentation in elementary news stories
  - Classification of transcribed text according to subject categories (e.g., sports, politics, nature)

## Description of analysis results

To represent the analysis results concerning shot segmentation and editorial segmentation into elementary units of ANTS the following elements have to be used:

- Shot structure, including temporal references to the main audiovisual object's timeline;
- Editorial segments structure including temporal references to the main audiovisual object's timeline.

## Output in MPEG-7 AVDP format

- The serialisation component of ANTS produces documents compliant to MPEG-7 AVDP for what concerns shot detection and news story segmentation. An example output of the tool can be found below. The basic AVDP tool used is TemporalDecomposition of the root Audiovisual Segment using two distinct decomposition criteria.

```
<?xml version="1.0" encoding="UTF-8"?>
<Mpeg7 xmlns="urn:mpeg:mpeg7:schema:2004" xmlns:mpeg7="urn:mpeg:mpeg7:schema:2004"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:date="http://exslt.org/dates-and-times"
xsi:schemaLocation="urn:mpeg:mpeg7:schema:2004 ../profile.xsd">
    <DescriptionProfile profileAndLevelIndication="urn:mpeg:mpeg7:profiles:2011:AVDP"/>
    <DescriptionMetadata>
        <PrivateIdentifier>ANTS 1276512189.03</PrivateIdentifier>
        <Instrument>
            <Tool>
                <Name>Automatic Newscast Transcription System</Name>
            </Tool>
        </Instrument>
    </DescriptionMetadata>
    <Description xsi:type="mpeg7:ContentEntityType">
        <MultimediaContent xsi:type="mpeg7:AudioVisualType">
            <AudioVisual mediaTimeUnit="PT1N25F"
mediaTimeBase="mpeg7:MediaInformation/mpeg7:MediaProfile/mpeg7:MediaInstance/mpeg7:MediaLocator"
id="ANTS_1276512189.03">
                <MediaInformation>
                    <MediaProfile>
                        <MediaFormat>
                            <Content href="uri:2">
                                <Name>Audiovisual</Name>
                            </Content>
                            <FileFormat href="uri:3">
                                <Name>Windows Media 9</Name>
                            </FileFormat>
                        </MediaFormat>
                        <MediaInstance>
                            <InstanceIdentifier>ANTS_1276512189.03</InstanceIdentifier>
                            <MediaLocator>
                                <MediaUri>http://10.58.78.128/ANTS 1276512189.03.wmv</MediaUri>
                            </MediaLocator>
                        </MediaInstance>
                    </MediaProfile>
                </MediaInformation>
                <StructuralUnit href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#1">
                    <Name>News Programme</Name>
                </StructuralUnit>
                <CreationInformation>
                    <Creation>
                        <Title>TG12010-06-11</Title>
                    </Creation>
                </CreationInformation>
                <MediaTime>
                    <MediaRelIncrTimePoint>0</MediaRelIncrTimePoint>
                    <MediaIncrDuration>353500</MediaIncrDuration>
                </MediaTime>
                <TemporalDecomposition
criteria="http://www.ebu.ch/metadata/cs/mpeg/avdp/DecompositionCS#20" gap="true"
id="Shots ANTS 1276512189.03">
                    <Header xsi:type="mpeg7:DescriptionMetadataType">
```

```
                        <Instrument>
                            <Tool>
                                <Name>ANTS Shotfinder</Name>
                            </Tool>
                        </Instrument>
                    </Header>
                    <AudioVisualSegment id="ID-SHOT-4c124e7a54sk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>0</MediaRelIncrTimePoint>
                            <MediaIncrDuration>6040</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124e873csk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>6040</MediaRelIncrTimePoint>
                            <MediaIncrDuration>20000</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124e9b3csk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>26040</MediaRelIncrTimePoint>
                            <MediaIncrDuration>3520</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124e9f0csk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>29560</MediaRelIncrTimePoint>
                            <MediaIncrDuration>6720</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124ea554sk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>36280</MediaRelIncrTimePoint>
                            <MediaIncrDuration>3400</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124ea918sk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>39680</MediaRelIncrTimePoint>
                            <MediaIncrDuration>1600</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124eaa54sk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>41280</MediaRelIncrTimePoint>
                            <MediaIncrDuration>1960</MediaIncrDuration>
```

```
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124eac50sk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>43240</MediaRelIncrTimePoint>
                            <MediaIncrDuration>1920</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124eae48sk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>45160</MediaRelIncrTimePoint>
                            <MediaIncrDuration>960</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124eaf44sk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>46120</MediaRelIncrTimePoint>
                            <MediaIncrDuration>3600</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124eb31csk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>49720</MediaRelIncrTimePoint>
                            <MediaIncrDuration>7120</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124eba28sk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>56840</MediaRelIncrTimePoint>
                            <MediaIncrDuration>15000</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124ec928sk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>71840</MediaRelIncrTimePoint>
                            <MediaIncrDuration>2520</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124ecb5csk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>74360</MediaRelIncrTimePoint>
                            <MediaIncrDuration>2000</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124ecd5csk">
```

```
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>76360</MediaRelIncrTimePoint>
                            <MediaIncrDuration>2440</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124ed024sk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>78800</MediaRelIncrTimePoint>
                            <MediaIncrDuration>2240</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124ed23csk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>81040</MediaRelIncrTimePoint>
                            <MediaIncrDuration>2880</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124ed530sk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>83920</MediaRelIncrTimePoint>
                            <MediaIncrDuration>7400</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124edc58sk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>91320</MediaRelIncrTimePoint>
                            <MediaIncrDuration>1560</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124ede2csk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>92880</MediaRelIncrTimePoint>
                            <MediaIncrDuration>3640</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124ee208sk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>96520</MediaRelIncrTimePoint>
                            <MediaIncrDuration>1250</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124ee321sk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>97770</MediaRelIncrTimePoint>
```

```
                            <MediaIncrDuration>1070</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124ee428sk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>98840</MediaRelIncrTimePoint>
                            <MediaIncrDuration>4880</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124ee91csk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>103720</MediaRelIncrTimePoint>
                            <MediaIncrDuration>3240</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124eec34sk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>106960</MediaRelIncrTimePoint>
                            <MediaIncrDuration>16760</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124efd1csk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>123720</MediaRelIncrTimePoint>
                            <MediaIncrDuration>8920</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124f0614sk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>132640</MediaRelIncrTimePoint>
                            <MediaIncrDuration>6080</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124f0c1csk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>138720</MediaRelIncrTimePoint>
                            <MediaIncrDuration>5080</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124f1124sk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>143800</MediaRelIncrTimePoint>
                            <MediaIncrDuration>2930</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124f141dsk">
```

```
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>146730</MediaRelIncrTimePoint>
                            <MediaIncrDuration>3550</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124f1754sk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>150280</MediaRelIncrTimePoint>
                            <MediaIncrDuration>6720</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124f1e38sk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>157000</MediaRelIncrTimePoint>
                            <MediaIncrDuration>2840</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124f2128sk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>159840</MediaRelIncrTimePoint>
                            <MediaIncrDuration>2000</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124f2328sk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>161840</MediaRelIncrTimePoint>
                            <MediaIncrDuration>880</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124f241csk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>162720</MediaRelIncrTimePoint>
                            <MediaIncrDuration>19960</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124f3818sk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>182680</MediaRelIncrTimePoint>
                            <MediaIncrDuration>7120</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124f3f24sk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>189800</MediaRelIncrTimePoint>
```

```
                            <MediaIncrDuration>7720</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124f4708sk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>197520</MediaRelIncrTimePoint>
                            <MediaIncrDuration>1360</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124f482csk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>198880</MediaRelIncrTimePoint>
                            <MediaIncrDuration>1480</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124f495csk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>200360</MediaRelIncrTimePoint>
                            <MediaIncrDuration>800</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124f4a48sk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>201160</MediaRelIncrTimePoint>
                            <MediaIncrDuration>600</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124f4b20sk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>201760</MediaRelIncrTimePoint>
                            <MediaIncrDuration>1800</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124f4d0csk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>203560</MediaRelIncrTimePoint>
                            <MediaIncrDuration>3560</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124f5044sk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>207120</MediaRelIncrTimePoint>
                            <MediaIncrDuration>1920</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124f523csk">
```

```
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>209040</MediaRelIncrTimePoint>
                            <MediaIncrDuration>2840</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124f552csk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>211880</MediaRelIncrTimePoint>
                            <MediaIncrDuration>1360</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124f5650sk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>213240</MediaRelIncrTimePoint>
                            <MediaIncrDuration>1520</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124f5820sk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>214760</MediaRelIncrTimePoint>
                            <MediaIncrDuration>1720</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124f5a04sk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>216480</MediaRelIncrTimePoint>
                            <MediaIncrDuration>1200</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124f5b18sk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>217680</MediaRelIncrTimePoint>
                            <MediaIncrDuration>11440</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124f6644sk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>229120</MediaRelIncrTimePoint>
                            <MediaIncrDuration>1640</MediaIncrDuration>
                        </MediaTime>
                    </AudioVisualSegment>
                    <AudioVisualSegment id="ID-SHOT-4c124f6820sk">
                        <StructuralUnit
href="http://www.ebu.ch/metadata/cs/mpeg/avdp/StructuralUnitCS#10"/>
                        <MediaTime>
                            <MediaRelIncrTimePoint>230760</MediaRelIncrTimePoint>
```