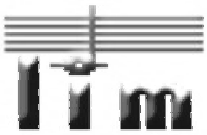
GIANCARLO VERCELLESI
(giancarlo.vercellesi@dico.unimi.it)

TUTORIAL SU MPEG-7 [ISO/IEC 15938]



L.I.M. - Laboratorio di Informatica Musicale
DI*Co* - Dipartimento di Informatica e Comunicazione
Università degli Studi di Milano
Via Comelico, 39/41
I-20135 Milano (Italy)

DICo

Indice

1. Introduzione

Lo standard MPEG-7 [ISO/IEC 15938], formalmente chiamato "*Multimedia Content Description Interface*", fornisce un set di Description Tools per la descrizione simbolica di contenuti multimediali audio-video (AV).
A differenza degli standard precedenti -i quali hanno avuto come obiettivo lo sviluppo di algoritmi di compressione (MPEG-1 ed MPEG-2) e l'organizzazione in *oggetti* di realtà multimediali eterogenee (MPEG-4)- MPEG-7 si pone come fine ultimo la descrizione di informazione multimediale attraverso una rappresentazione testuale (XML) che ne permetta una semplice ed immediata ricerca e navigazione rispetto ai contenuti, e non in funzione della propria struttura fisica (per esempio, un insieme di numeri rappresentanti la forma d'onda o lo spettro di un segnale audio).

Di seguito andremo a descrivere lo standard in tutte le sue parti, approfondendo particolarmente i tools per la descrizione audio.

2. Caratteristiche di MPEG-7

- Permette di descrivere le realtà multimediali nel loro complesso (audio, video, immagini, voce, modelli 3D, ecc.) fornendo un set di metodi e tools indipendenti tra loro
- MPEG standardizza solo la descrizione delle realtà multimediali, non i metodi per la loro estrazione o ricerca (pur fornendo del materiale informativo nella parte 8 dello standard)
- è indipendente dai formati
- offre diversi livelli di astrazione per la descrizione dell'informazione
- le descrizioni ed i relativi materiali possono essere collocati in posti fisicamente diversi. Esiste un meccanismo per il collegamento di tali informa zioni
- **è basato su XML**; la definizione delle strutture per la rappresentazione dei metadati (DDL - Data Description Language) è un'estensione dell'XML-Schema ed anche la rappresentazione testuale dei contenuti veri e propri è in XML; ove necessario (efficienza nella memorizzazione e/o trasporto), tali descrizioni possono essere eventualmente compresse in un formato binario sviluppato ad hoc da MPEG (BiM - BInary format for MPEG)

3. Obiettivi di MPEG-7

Lo scopo principale di MPEG-7 è quello di fornire una tipologia di rappresentazione che permetta di effettuare ricerche, navigazioni, ecc. per contenuti all'interno di sistemi Multimediali (per esempio, Database Multimediali). Vediamo degli esempi:

- Musica: l'utente fornisce una serie di note musicali attraverso una tastiera, ed in ritorno ottiene una lista di brani musicali, il più simili possibile al pezzo suonato.
- Voce: l'utente fornisce un estratto vocale di un particolare autore (cantante, giornalista, ecc.), ed in ritorno ottiene una lista di tutte le registrazioni audio che appartengono a questo personaggio.
- Grafica: l'utente disegna un insieme di linee ed in ritorno ottiene una lista di immagini contenenti loghi, ideogrammi o rappresentazioni grafiche il più simili possibile a quella fornita

4. Componenti Principali di MPEG-7

- *Description Tools*: è formato dai *Descriptors (D)* ed i *Descriptor Schemes (DS)*; i *D* definiscono la sintassi e la semantica di ogni feature AV (ossia l'elemento metadato), i *DS* specificano la struttura e la semantica delle relazioni tra i vari componenti (ossia le relazioni tra i vari metadati), che possono essere sia *D* che *DS*. Istanze di *D* e *DS* rappresentano una *Description*.
- *Data Description Language (DDL)*: definisce la sintassi dei *Description Tools (D* e *DS)* permettendo la creazione di nuovi *DS*, *D*, o permettendone una loro eventuale estensione.
- *System Tool*: fornisce i metodi per la rappresentazione dei contenuti in formato binario (per una maggiore efficienza nella trasmissione e memorizzazione), i meccanismi per la trasmissione (sia a livello testuale [TeX] che binario [BiM]), le tecniche per la sincronizzazione delle descrizioni rispetto ai contenuti, ecc.

In Fig. 1 sono mostrate le relazioni tra i vari elementi che compongono MPEG-7. I *Description Tools* permettono di creare la descrizione di una realtà multimediale (*Description)* con gli opportuni *Descriptors* e *Description Scheme*, definiti da un *Description Definition Language (DDL)* noto, opportunamente incorporato all'interno del documento MPEG-7. La descrizione simbolica può successivamente essere memorizzata, trasmessa, multiplexata o collegata ad eventuali materiali esterni tramite il *System Tool*.
E' altresì possibile estendere le funzioni dei *Description Tools* modificando (o creandone uno nuovo) opportunamente il *DDL* fornito da MPEG-7.

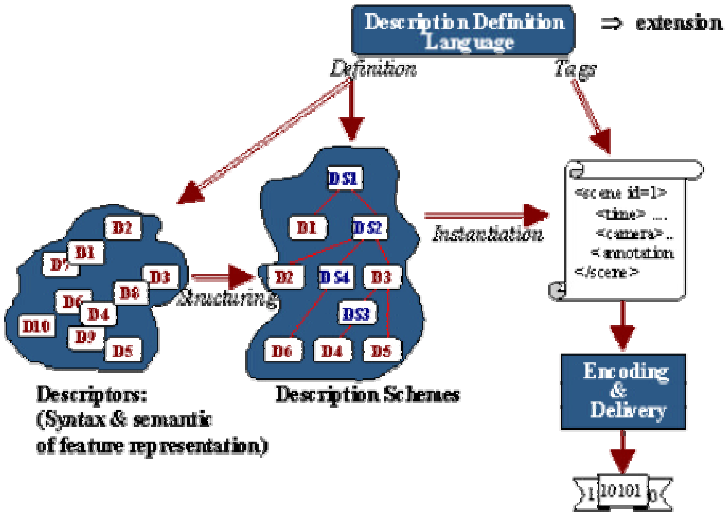


Figura 1: relazioni tra i componenti principali di MPEG-7 [1]

5. Parti di MPEG-7

MPEG-7 è composto dalle seguenti parti:

- MPEG-7 System [15938-1] - tool contenente il sistema per la codifica binaria BiM delle descrizioni MPEG-7 ed il 'Terminal Architecture', una sorta di applicazione standalone contenente i sistemi per l'encoding, il decoding, la memorizzazione, la trasmissione, ecc. dell'informazione MPEG-7.

- MPEG-7 Description Definition Language [15938-2]: contiene la descrizione del *Description Definition Language*, che in accordo con l'MPEG-7 Requirements Document è:
  "...un linguaggio che permette la creazione di nuovi Description Schemes (DS) e, possibilmente, i Descriptors (D). Esso inoltre permette l'estensione e la modifica di DS già esistenti"[1].
  DDL è basato su XML-Schema Language e ne rappresenta una sorta di estensione orientata al multimedia. DDL è composto da tre componenti logiche (normative) fondamentali:
  - XML-Schema structural language components
  - XML-Schema datatype language components
  - MPEG-7 specific extensions (per esempio il supporto delle matrici come tipo di dato)

- MPEG-7 Video [15938-3]: *Decription Tools* che forniscono strutture, *Descriptors (D)* e *Descriptor Schemes (DS)* per la descrizione di contenuti video (texture, shape, ecc.)

- MPEG-7 Audio [15938-4]: *Decription Tool* che forniscono strutture (che si appoggiano al *Multimedia Description Schemes*), *Descriptors (D)* e *Descriptor Schemes (DS)* per la descrizione di contenuti audio. I tools audio forniti possono essere suddivisi in due categorie: i *Low-Level Description Tools* (per la descrizione dello spettro, di features temporali quali l'inviluppo, ecc.) e gli *High Level Description Tools* (per la descrizione dei timbri e delle melodie musicali, del parlato, ecc.). MPEG-7 / Audio sarà trattato dettagliatamente più avanti nel documento.

- MPEG-7 Multimedia Description Schemes [15938-5]: MPEG-7 Multimedia Description Schemes, detto anche *MDS,* fornisce un set di strutture, *Descriptors (D)* e *Descriptor Schemes (DS)* per la descrizione di entità generiche e/o multimediali. I tools forniti possono essere raggruppati in cinque classi fondamentali (brevemente descritte al paragrafo seguente):
  - Content Description: rappresenta l'informazione percepibile (audio, video, ecc.);
  - Content Management: descrive informazioni sulla creazione, l'uso e le proprietà dei media;
  - Content Organization: permette di creare, modellare e classificare collezioni di media e relative descrizioni;
  - Navigation and Access: fornisce strumenti per facilitare la navigazione, l'accesso e la ricerca di materiali multimediali;
  - User Interaction Tools: permette di descrivere le preferenze degli utenti rispetto alla ricerca ed alla navigazione all'interno delle descrizioni multimediali.

- MPEG-7 Reference Software [15938-6]: implementazione SW delle parti normative più rilevanti di MPEG-7. Il SW, chiamato XM (eXperimental Model), permette di utilizzare i *Descriptors*, i *Descriptor Schema* ed il *Definition Description Language (DDL)* oltre che ad altre funzioni non normative ma necessarie per il corretto utilizzo di MPEG-7. L'architettura dell'applicazione è

Client/Server: la parte server permette l'estrazione delle informazioni, quella client la ricerca, il filtraggio e/o il transcoding dei dati multimediali.

- MPEG-7 Conformance Testing [15938-7]: procedure e linee guida su come validare documenti MPEG-7

- MPEG-7 Extraction and use of Descriptions [15938-8]: materiale informativo sull'uso di alcuni Description Tools e sulle tecniche di estrazione di informazione simbolica

6. MPEG-7 Description Tools

MEPG-7 fornisce una vasta quantità di tools per la descrizione dell'informazione multimediale, ivi compresi quelli inerenti l'audio ed il video. Qui di seguito verrà data loro una breve descrizione, raggruppandoli in accordo con le proprie funzionalità (Fig. 2).

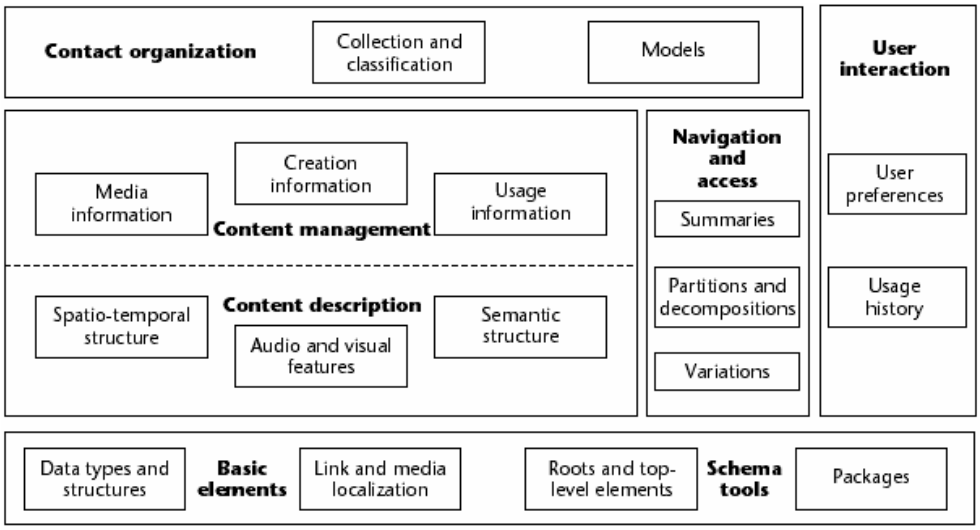


Figura 2: overview dei Description Tools di MPEG-7 [2]

- Basic Elements: fornisce tutti i *data types* utilizzati dai *Description Tools* come blocchi base per la rappresentazione dei dati (matrici, vettori, link, ecc.)
- Schema Tools: tool per il wrapping dei *Description* Tools tramite etichettatura, al fine di rendere fruibile alle applicazioni le *Descriptions* create.
- Content Management Tools: tools che permettono di descrivere tutte le informazioni riguardanti la creazione, l'uso ed il formato del media utilizzato.
  - *Creation Description Tools* permette di descrivere tutte le informazioni riguardanti il processo di creazione e classificazione di un media (per esempio data di creazione, genere, ecc.), ivi comprese le informazioni correlate.
  - *Media Description Tools* permette di descrivere i media, definendone il percorso al supporto fisico, il formato di codifica, il supporto di massa,ecc. Fornisce inoltre un descrittore per definire la qualità del media e le indicazioni di eventuali operazioni di transcoding per l'adattamento dei contenuti alle reti ed ai terminali.
  - *Usage Description Tools* permette la descrizione di tutte le condizioni d'uso del media (per esempio diritti d'autore, ecc.) e di mantenere l'history del suo utilizzo.
- Content Description Tools: fornisce i *Descriptor* ed i *Descriptor Schema* per la rappresentazione dell'informazione percepibile AV, sia da un punto di vista strutturale che da un punto di vista semantico.
  - *Structural Level*: le informazioni sono descritte in termini di segmenti (per esempio, *AudioSegment, VideoSegment,* ecc.) spazio-temporali organizzati gerarchicamente (o tramite

strutture a grafo), all'interno dei quali vengono inserite le descrizioni dei media (*Content Management Tools*) e dei rispettivi contenuti (AV tools).

- *Semantic Level*: descrive l'informazione multimediale dal punto vista del mondo reale, in termini di eventi, oggetti, concetti, posti, astrazioni e tempo (in senso narrativo). E' possibile creare una associazione diretta tra il livello semantico ed il livello strutturale tramite un opportuno set di link (*Relationship Elements*).

- Content Organization Tools: permette di creare e modellare collezioni di media e relative descrizioni
- Navigation and Access Tools: fornisce strumenti per facilitare la navigazione, l'accesso e la ricerca di materiali multimediali.
- User Interaction Tools: permette di descrivere le preferenze degli utenti rispetto alla ricerca ed alla navigazione all'interno delle descrizioni multimediali.

7. Approfondimento di MPEG-7 Multimedia Description Schemes [ISO/IEC 15938-5]

Multimedia Description Schemes rappresenta la parte più corposa di MPEG-7. Essa fornisce buona parte dei tools presenti nello standard, molti dei quali fungono da appoggio a quelli delegati alla descrizione dei contenuti AV ([15938-3], e [15938-4]).

Appartengono a questa parte dello standard le seguenti categorie di tools:

- Content Management Tools
- Content Organization Tools
- Navigation and Access Tools
- User Interaction Tools
- Structural Level (*Segments*) e Semantic Level in Content Description Tools

Nel prosieguo del paragrafo, tratteremo i tools più significativi rispetto alla descrizione dell'informazione musicale e dei media: lo *Structural Level*, il *Semantic Level* ed i *Content Management Tools.*

7.1. Structural Level

Alla base di questo livello c'è il *Segment DS* che permette di organizzare e descrivere i contenuti AV (*Content Entity)* sottoforma di segmenti temporali, spaziali, spazio-temporali, ecc. Lo standard fornisce una serie di *DS* derivati da *Segment DS*, ognuno dei quali è pensato per un particolare media (audio, video, audio-video, ecc.); tutti questi segmenti ereditano gli attributi previsti da *Segment DS* e possono essere messi in relazione tra loro. Per quanto riguarda l'audio, MPEG fornisce *AudioSegment DS*, che permette di suddividere l'informazione in zone temporalmente distinte.
E' possibile organizzare i segmenti gerarchicamente o attraverso una struttura a grafo (*SegmentRelation DS*), in modo tale da descrivere i contenuti multimediali in funzione della propria struttura. Generalmente, in tali situazioni, al solo root-element viene associato un media fisico con la relativa descrizione (*Management Description Tools*), supponendo che i figli ne ereditino tutte le informazioni.
Infine, ai segmenti sono associate le descrizioni simboliche dei media AV associati, effettuate tramite i vari tools previsti dallo standard come, per esempio, il *Melody Tool* (brevemente descritto al Cap. 7).

7.2. Semantic Level

I tools forniti a questo livello permettono di descrivere l'informazione AV dal punto di vista del mondo reale, attraverso strutture quali *Eventi, Oggetti, Concetti, Posti, Astrazioni e Tempo* (da un punto di vista *narrativo*), direttamente collegabili con i segmenti del livello strutturale tramite un opportuno set di link detto *Relationship Elements.*
Alla base c'è il *SemanticBase DS* da cui derivano una serie di *DS* dedicati alla descrizione delle varie entità semantiche che entrano in gioco in una realtà multimediale:

- *Object DS*: descrive un oggetto percepibile o astratto; per oggetto percepibile si intende un oggetto fisico ed esistente (esempio: la chitarra di Marco) mentre per oggetto astratto si fa riferimento ad un entità astratta che può esistere se applicata ad una realtà percepita (esempio: una qualunque chitarra);
- *AgentObjectDS*: DS derivato da Object DS che descrive una persona, un gruppo di persone od un qualunque oggetto personalizzato
- *Event DS*: descrive un evento percepibile o astratto mettendo opportunamente in relazione due oggetti (Object DS); per percepibile s'intende un evento con una propria dimensione spazio-

temporale (Adriano sta suonando la tromba) mentre per astratto s'intende un evento la cui applicazione nel mondo reale da un evento percepibile (esempio: qualcuno suona la tromba).

- *Concept DS*: rappresenta entità semantiche che non possono essere descritte come una generalizzazione o astrazione di un specifico posto, evento, ecc. Esso permette di esprimere una o più proprietà associate ad un'entità semantica come, per esempio, l'armonia o la pienezza di un suono.
- *SemanticPlace DS* e *SemanticTime DS*: descrivono rispettivamente, il luogo (esempio: Teatro alla Scala di Milano) e l'istante temporale (esempio: la data in GG/MM/AAAA) di un evento.
- *SemanticState DS*: permette di descrivere una serie di informazioni riguardanti un'entità percepibile (esempio: il pianoforte di Luca pesa 100Kg)
- *Abstraction*: esistono due differenti *Abstraction;* la *Media Abstraction* e la *Standard Abstraction.* La prima permette di raggruppare e descrivere semanticamente gruppi di media differenti i cui contenuti possiedono caratteristiche comuni; le seconde invece, permettono di raggruppare ulteriormente gruppi di Media Abstraction simili tra loro.

Infine, come per il livello strutturale, anche nel semantico è possibile creare una descrizione che metta in relazione tra loro le diversa entità semantiche organizzandole gerarchicamente o tramite strutture a grafo (*SemanticRelation DS*).

In Fig. 3 è mostrato un esempio. La descrizione narrativa è la seguente: Tom Daniels suona il Piano col proprio tutor. L'evento è caratterizzato da un SemanticTime ,7-8 PM on the 14th of October 1998, ed un SemanticPlace, Carnegie Hall. La descrizione è strutturata nel seguente modo: l'evento "eseguire", e quattro oggetti, Piano, Tom Daniels, il suo tutor e la generica nozione di musica. Tom Daniels ed il suo tutor appartengono più precisamente all'AgentObject DS.

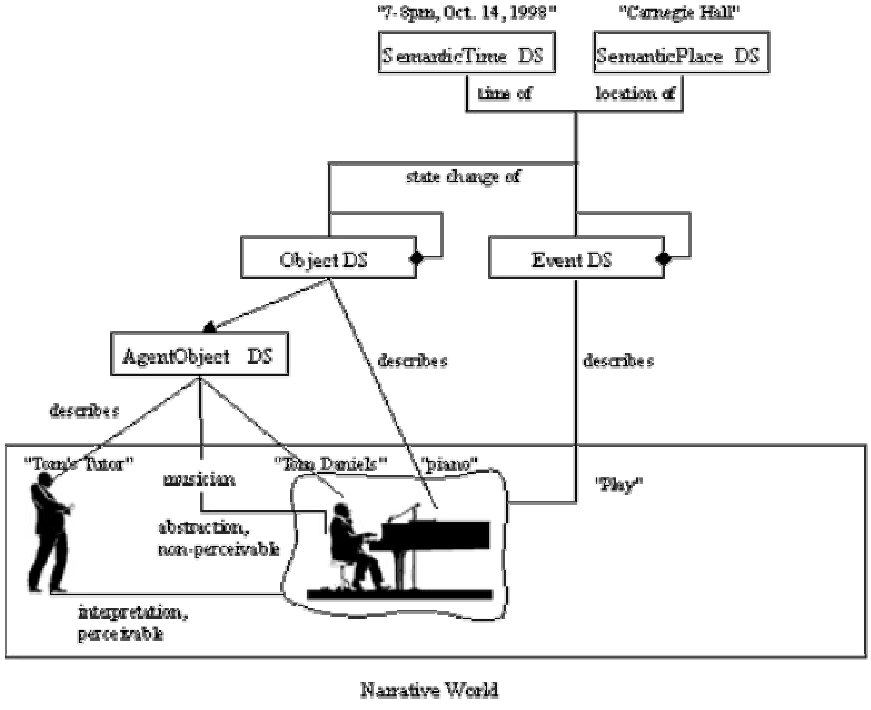


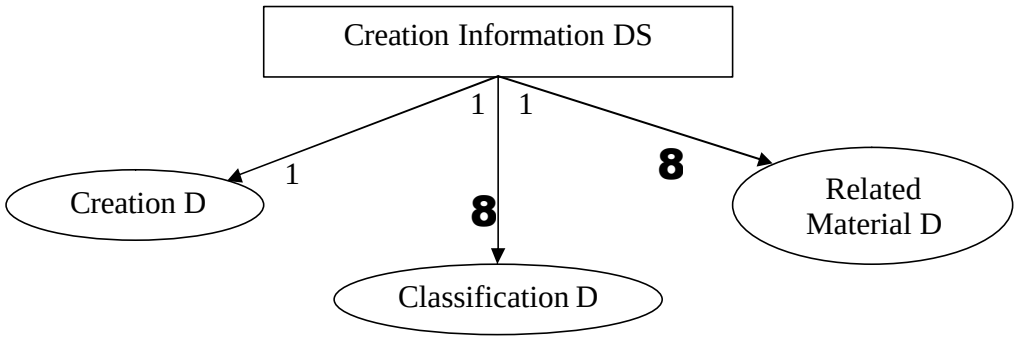
Figura 3: esempio di descrizione concettuale con Semantic DS

7.3. Content Management Tools

Fornisce una serie di *D* e *DS* per la completa descrizione dei media associati ai vari Segment.
L'idea di fondo è la possibilità di descrivere uno stesso contenuto con media differenti, i quali avranno conseguentemente diversi formati di codifica. In questo modo è per esempio possibile associare ad un segmento audio, sia un media in formato MP3 che uno in formato MIDI; oppure, se la descrizione MPEG-7 deve essere impiegata in ambienti streaming, si possono associare ad un segmento due media MP3, uno con valore di bitrate alto ed uno basso, in modo da trasmettere in rete quello più idoneo in funzione della disponibilità di banda. Un ultimo esempio riguarda le opere multimediali AV che possono essere codificate con codifiche prettamente audio, o con codifiche AV (per esempio codifiche AVI).
Content Management Tools fornisce tre differenti classi di DT attraverso i quali è possibile descrivere tutte le informazioni sulla creazione e produzione di un contenuto AV -*Creation and Production DT-*, sui detentori dei diritti di tali dati -*Content Usage DT-* e sui formati dei media fisici associati -*Media DT-*.

7.3.1. Creation and Production Description Tools

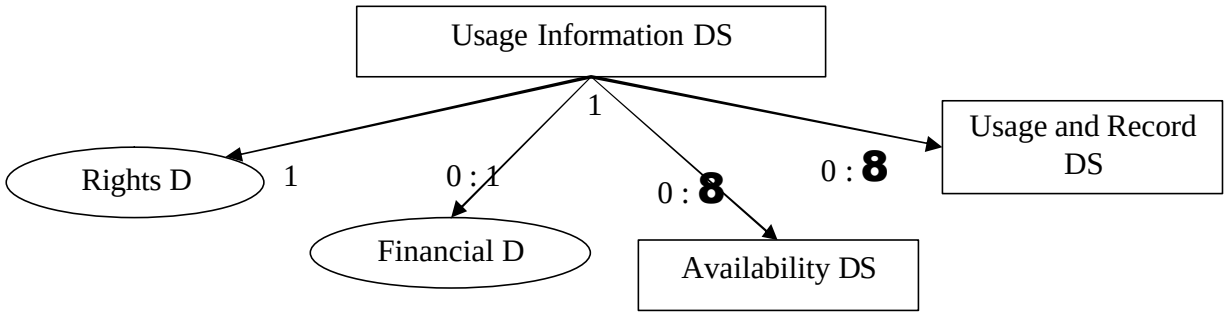
Permette la descrizione di tutte le informazioni riguardanti gli autori ed i luoghi dove il media AV è stato creato, generalmente non estraibili automaticamente dai contenuti AV.
Col *Creation D* è possibile definire il luogo e la data di creazione, i materiali utilizzati, lo staff le persone e l'organizzazione che ha lavorato alla costruzione del media, ecc.
Col *Classification D* si possono definire tutta una serie di informazioni utili per la catalogazione e classificazione del media come il genere, la lingua, lo stile, il mercato a cui è rivolto, ecc.
Col *Related Material D* infine, è possibile definire ulteriori informazioni riguardanti il media e la sua creazione.
Tale DT è in relazione 1<-->1 con i media associati ai Segments del Content Entity (ad ogni media corrisponde un solo Creation Information DT).



7.3.2. Content Usage Description Tools

Fornisce una serie di D e DS per la descrizione di tutte le informazioni riguardanti i processi d'uso del media AV: diritti d'autore (*Rights D*), informazioni sull'uso passato del media (*Usage and Records DS*), aspetti finanziari quali il costo di produzione, il prezzo di vendita, le percentuali spettanti ai vari

diritti, ecc. (*Financial D*) e le autorizzazioni sull'uso e la distribuzione del materiale mediatico (*Availability DS*).
Tale DT è in relazione 1<-->1 con i media associati ai Segments del Content Entity (ad ogni media corrisponde un solo Usage Information DT).

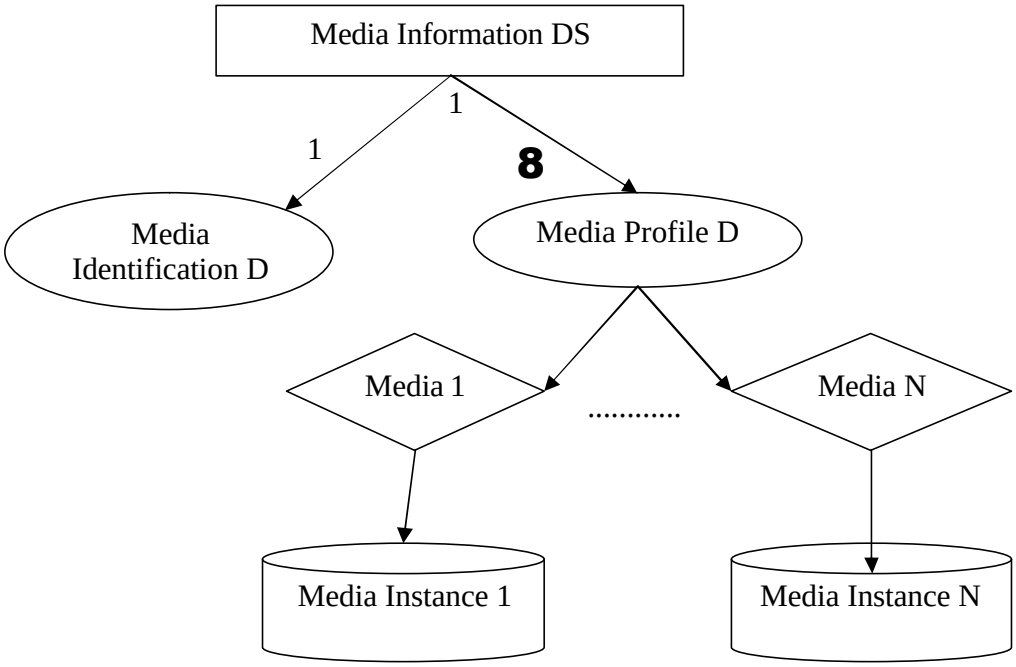


7.3.3. Media Description Tools

Descrive completamente il formato di codifica di un media ed il collegamento al dato mediale vero e proprio (*Media Instance*). E' possibile rappresentare differenti media fisici aventi il medesimo contenuto attraverso il Media Profile. In altre parole, ad uno stesso contenuto si possono far corrispondere differenti media fisici, aventi differenti formati di codifica. L'utilità di una tale struttura sta nell'utilizzo di MPEG-7 in ambiti quali lo streaming: associando ad un Segment dei media codificati con bitrate differenti, è possibile decidere dinamicamente quale trasmettere in funzione della quantità di banda disponibile.
Il *Media Information DS* fornisce un *Media Identificator D* ed un *Media Profile D*; il primo permette di identificare univocamente un contenuto AV, indipendentemente dai vari *Media Instance* disponibili. Il secondo invece permette di descrivere concretamente i formati dei vari media associati:

- Media Format
- Media Quality
- Media Transocoding Hints
- Media Instance link
- .....

Ad ogni Segment nel Content Entity, è possibile associare una sola descrizione effettuata con Media DT (relazione 1<-->1)


Media Information DS
1
1
8
Media Identification D
Media Profile D
Media 1
.....
Media N
Media Instance 1
Media Instance N

8. Approfondimento di MPEG-7 Audio Description Tool [ISO/IEC 15938-4]

Questo *Description Tool* fornisce una serie di *Descriptors* e *Description Schema* per la descrizione dell'audio ad alto livello, l'*High-Level Description Tools*, il quale si appoggia sul *Low-Level Description Tools*, un layer più basso che fornisce una serie di soli *Descriptors* per la rappresentazione di segnali audio. Il *Low Level Description Tools*, detto anche *Audio Framework*, si appoggia a sua volta sull'*MDS (Multimedia Description Schema)* il quale mette a disposizione le strutture dati necessarie per la rappresentazione dell'informazione audio; più precisamente, esso permette di rappresentare le features audio, o tramite i *Segments* forniti dallo *Structural Level* (per l'audio si usa l'*AudioSegment DS* che suddivide le tracce audio in sequenze di regioni distinte) o con valori campionati ad intervalli regolari. Inoltre, entrambe permettono di codificare i dati in due differenti tipologie (forniti dai *Basic Tools*): lo *scalar value* o il *vector value*.

8.1. Low-Level Description Tools - Audio Framework

Il *Low-Level Description Tools*, detto anche *Audio Framework*, fornisce una serie di soli *Descriptors* (circa 70) per la rappresentazione dei segnali audio, sia nel dominio temporale che frequenziale.

E' possibile suddividere i *Descriptors* nelle seguenti macro-categorie:

- Basic
- Basic Spectral
- Timbral Temporal
- Timbral Spectral
- Spectral Basis

In Fig. 4 sono evidenziati i principali *Descriptors*, raggruppati per classi di appartenenza (tranne il *Silence*), e brevemente descritti qui di seguito.

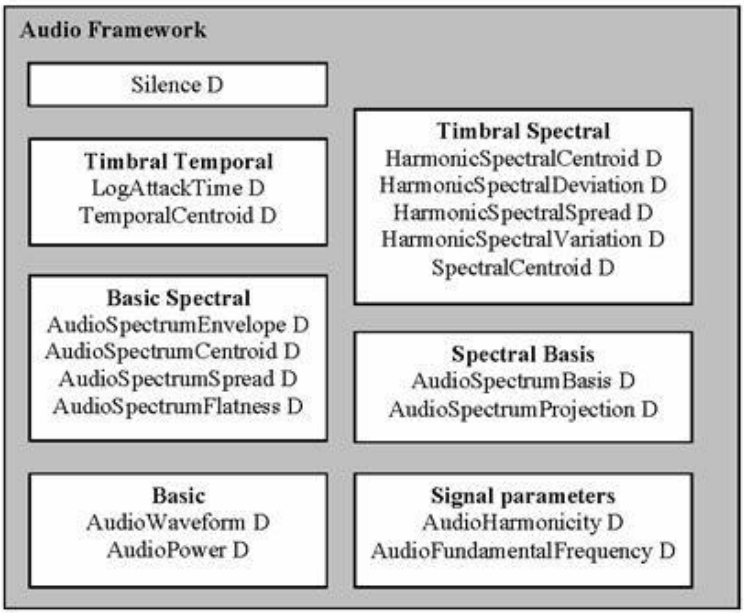


Figura 4: principali Descriptors forniti dall'Audio Framework di MPEG-7, raggruppati per classe di appartenenza [1]

Basic Descriptors: due *Descriptors* (*AudioWaveform, AudioPower*) che permettono di fornire una rappresentazione numerica dei campioni audio per descrivere l'inviluppo o l'andamento energetico (potenza) del segnale audio nel tempo.

Basic Spectral: quattro *Descriptors* che permettono di descrivere, in modi differenti, lo spettro di un segnale audio, utilizzando un dominio frequenziale logaritmico. *AudioSpectrumEnvelope* da una rappresentazione formale dello spettrogramma di una traccia audio, *AudioSpectrumCentroid* ne descrive il centro di gravità (il centroide appunto), *AudioSpectrumSpread* ne descrive il momento secondo rispetto al centroide ed infine, *l'AudioSpectrumFlatness* da una descrizione dei picchi dello spettro raggruppati per bande frequenziali.

Signal Parameters: due *Descriptors* utilizzati principalmente per segnali periodici e/o quasi-periodici, che ne descrivono la frequenza fondamentale (*AudioFundamentalFrequency*) e l'armonicità a livello spettrale (*AudioHarmonicity*), in modo da poter distinguere suoni armonici da suoni inarmonici.

Timbral Temporal: due *Descriptors* che permettono di descrivere l'andamento temporale di un segnale audio rispetto all'informazione timbrica in esso contenuta. *LogAttackTime* descrive l'inviluppo temporale del segnale nella fase di attacco mentre *TemporalCentroid* ne descrive il centro di massa, mostrando così ove è situato il massimo punto di energia

Timbral Spectral: cinque *Descriptors* (*HarmonicSpectralCentroid, HarmonicSpectralDeviation, HarmonicSpectralSpread, HarmonicSpectralVariation, SpectralCentroid*) che permettono di descrivere le caratteristiche timbriche in funzione di specifiche proprietà dello spettro di un segnale audio, utilizzando un dominio frequenziale lineare (in quanto pensati per segnali armonici). Il lavorare su di un dominio lineare, differenzia questi descrittori dagli omonimi presenti nei Basic Spectral.

Spectral Basis: due *Descriptors* (*AudioSpectrumBasis, AudioSpectrumProjection*) che permettono di descrivere tramite opportuna combinazione lineare di coppie di funzioni (ricavate in funzione dello spettro in potenza del segnale) lo spettro di un segnale audio, riducendo così drasticamente la quantità di spazio utilizzato (Fig. 5).

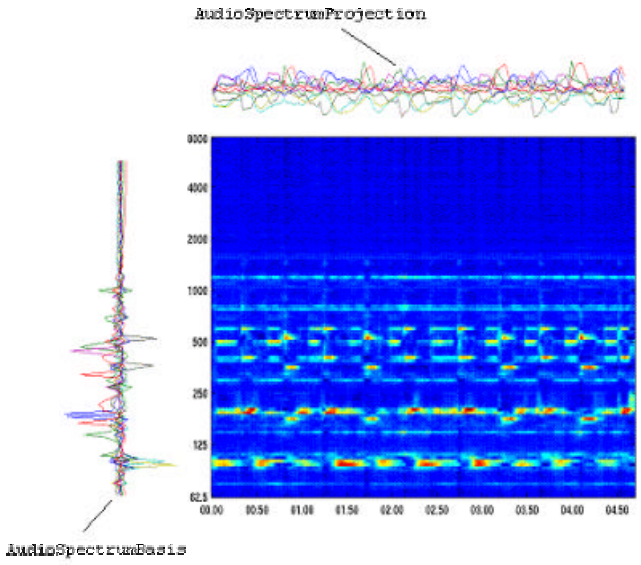


Figura 5: : Esempio d'uso dei descrittori AudioSpectrumBasis ed AudioSpectrumProjection [1]

Infine, è previsto un semplice *Descriptor* per la rappresentazione del silenzio (*Silence*) all'interno di un segmento audio.

8.2 High-Level Description Tools

L'*High-Level Description Tools*, fornisce cinque tools (comprensivi di *Descriptors* e *Description Schema*) per la rappresentazione dei suoni indipendentemente dal loro dominio di rappresentazione. I cinque tools, brevemente descritti di seguito, sono:

- Audio Signature Description Scheme
- General Sound Recognition and Indexing Description Tool (fortemente integrato con l'*MDS*)
- Spoken Content Description Tool (fortemente integrato con l'*MDS*)
- Musical Instrument Timbre Description Tool
- Melody Description Tool

Audio Signature: insieme di *DS* basati sui *Descriptors* dell'*Audio Framework*, che permettono una rappresentazione compatta ed univoca dei segnali audio al fine di crearne una sorta di identificatore unico, utile nelle applicazioni di riconoscimento automatico, sistemi di fingerprinting, ecc.

General Sound Recognition and Indexing: collezione di tools che permettono la rappresentazione delle fasi di riconoscimento ed indicizzazione di informazione audio, sia a basso (analisi segnale/spettro), intermedio (analisi statistiche - catene di Markov) o alto livello (analisi simbolica).

Spoken Content: collezione di tools che permettono la rappresentazione dell'intero parlato di una traccia audio. I due tools presenti forniscono *D e DS* per descrivere sia il parlatore (*SpokenContentHeader*) che l'output di un ASR (Automatic Speech Recognition) contenente la rappresentazione dell'intero parlato (*SpokenContentLattice*).

Musical Instrument Timbre: insieme di tools per la rappresentazione della timbrica di un pezzo musicale, sia ad alto che a basso livello. La prima permette di descrivere la timbrica di un segnale audio tramite un'organizzazione tassonomica degli strumenti musicali impiegati. La seconda invece ne da una rappresentazione matematica, basata sui *Descriptors* previsti dall'*Audio Framework* (*Timbral Spectral*, *Timbral Temporal*, *AudioFundamentalFrequency*, ecc.). E' altresì previsto un set di links per permettere, ove possibile, il collegamento tra la descrizione di basso e quella di alto livello.

Melody: insieme di tools per la rappresentazione dell'informazione melodica di segnali audio musicali monofonici. Sono previsti due differenti *Description Schema* per la descrizione della melodia: il *MelodyContour DS* ed il *MelodySequence DS*. Il primo *DS* rappresenta le differenze di intervalli tra note adiacenti tramite 5 valori quantizzati tra -2 e 2; il secondo *DS* invece definisce il preciso valore di pitch della nota con un livello di precisione pari al *cent*. Inoltre, entrambe le rappresentazioni prevedono una serie di *Descriptors* (molti dei quali, di fondamentale importanza nella descrizione del ritmo) per la metrica, la scala utilizzata, le battute, le alterazioni, la durata delle note ed i lyrics. E' infine possibile definire nell'*AudioSegment* associato alla descrizione melodica, il valore di BPM.

Vediamo ora come viene rappresentata l'informazione temporale in MPEG-7. Le strutture si trovano nell'*MDS* (*Multimedia Description Schema*), sono basate sullo standard ISO 8601 ed utilizzano tre strategie differenti (Fig. 7):

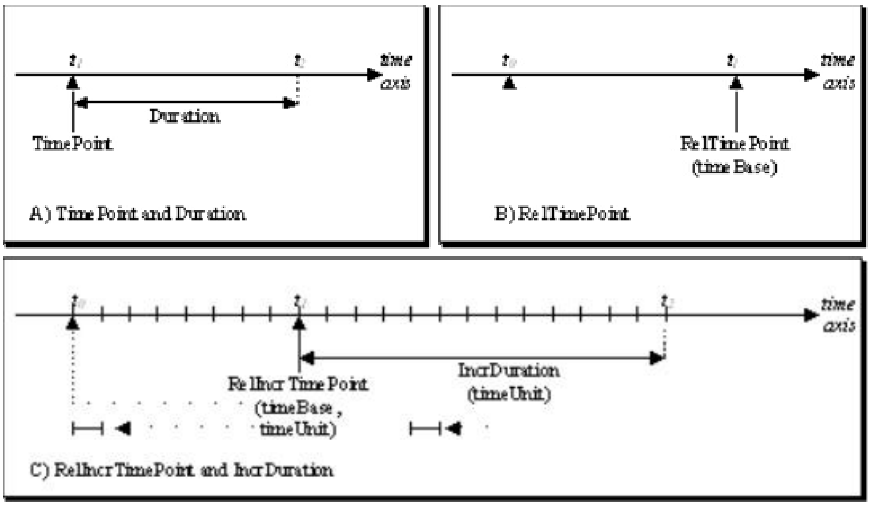


Figura 6: illustrazione delle tre descrizioni dell'informazione temporale in MPEG-7 [1]

Fig. 7.a) definizione del valore d'istante temporale iniziale T_1 ed un intervallo $T_{int} = T_2 - T_1$, con T_2 = istante temporale finale;

Fig. 7.b) definizione del valore d'istante temporale iniziale T_1 e finale T_2 come offset rispetto ad un istante T_i fissato;

Fig. 7.c) definizione di un intervallo temporale prefefinito T_U (Time Unit) da utilizzare come unità di misura per il valore d'istante temporale iniziale T_1 ed un intervallo $T_{int} = T_2 - T_1$, con T_2 = istante temporale finale.

9. Struttura di una MPEG-7 Descriptions

Esistono due tipi di descrizioni MPEG-7 valide: le *Description Units* e le *Complete Descriptions.* La prima permette di creare istanze di *Description Tools* (le *Descriptions*) come componenti indipendenti, permettendo così di inviare ad un'applicazione solo la parte di documento necessaria. La seconda invece, consiste nel descrivere l'intera realtà multimediale come un unico blocco monolitico, la cui struttura deve seguire uno o più rami della struttura gerarchica illustrata in Fig. 8.
Per ambedue le tipologie di documento è inoltre necessario costruire un wrapper utilizzando gli *Schema Tools*.

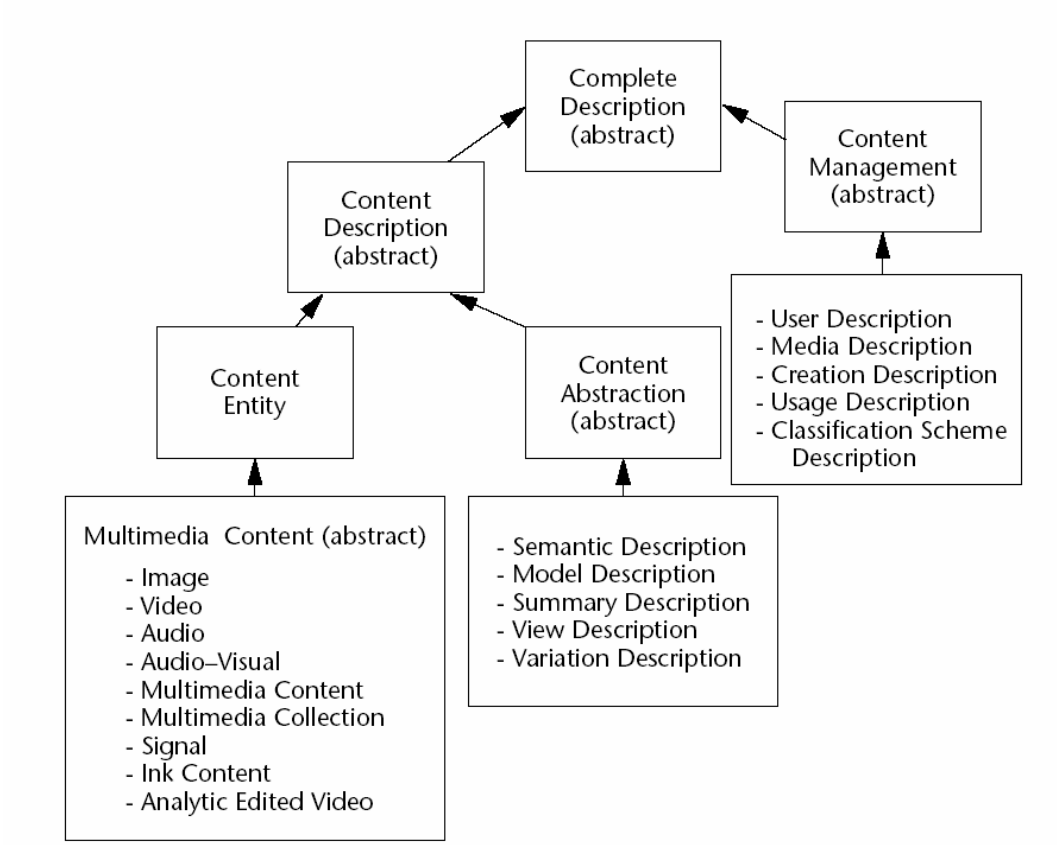


Figura 7: struttura gerarchica degli elementi principali di un documento MPEG-7

Analizziamo ora la struttura generale di un documento MPEG-7 in codifica testo XML.
L'elemento radice è il tag **<Mpeg7>** il quale prevede due sottoelementi: l'elemento **<DescriptionMetadata>** contenente metadati riguardanti la *Description*, e l'elemento **<Description>** (o **<DescriptionUnit>**, in funzione del tipo di descrizione MPEG-7 utilizzata) all'interno del quale viene posta l'intera descrizione della realtà multimediale che, come detto precedentemente, dovrà seguire uno o più rami della struttura gerarchica illustrata in Fig. 8.

Osservando la Fig. 8 si può dedurre come un documento MPEG-7 descriva una realtà multimediale attraverso due differenti rappresentazioni, tra loro associabili grazie ad un opportuno set di link forniti dallo standard (*Relationship Elements*): una di tipo strutturale (*Content Entity*) ed una di tipo semantico (*Content Abstraction*). Come già detto precedentemente, a livello strutturale i contenuti sono

organizzati come segmenti spazio-temporali organizzati gerarchicamente, all'interno dei quali si trovano le descrizioni simboliche dei contenuti AV (per esempio, tramite il Melody Tool); a livello semantico invece, l'informazione viene descritta dal punto vista del mondo reale, ossia in termini di eventi, oggetti, concetti, posti, astrazioni e tempo (in senso *narrativo*). E' possibile associare, ad ogni segmento del livello strutturale (in genere il root segment, in caso di strutture gerarchiche), una descrizione effettuata tramite i *Management Description Tools*, ossia l'assegnazione di un media (od un set di media, identici per contenuti ma con differenti formati), con le relative descrizioni riguardanti il formato (formato di codifica, formato di file, bitrate, ecc.), la creazione (chi l'ha creato, dove quando, ecc.) ed i diritti.
Focalizziamo ora la nostra attenzione sull'informazione musicale (Fig. 9); di fatto, lo standard MPEG-7 permette di descriverla in due dei suoi aspetti fondamentali, quelli strutturali e quelli semantici, associando loro un insieme di media audio -a loro volta ampiamente descritti col *Management Description Tools*- i quali contengono l'equivalente contenuto simbolico in un formato binario riproducibile tramite scheda audio (per esempio media in formato MP3) e/o visualizzabile su schermo (per esempio media in formato JPEG).

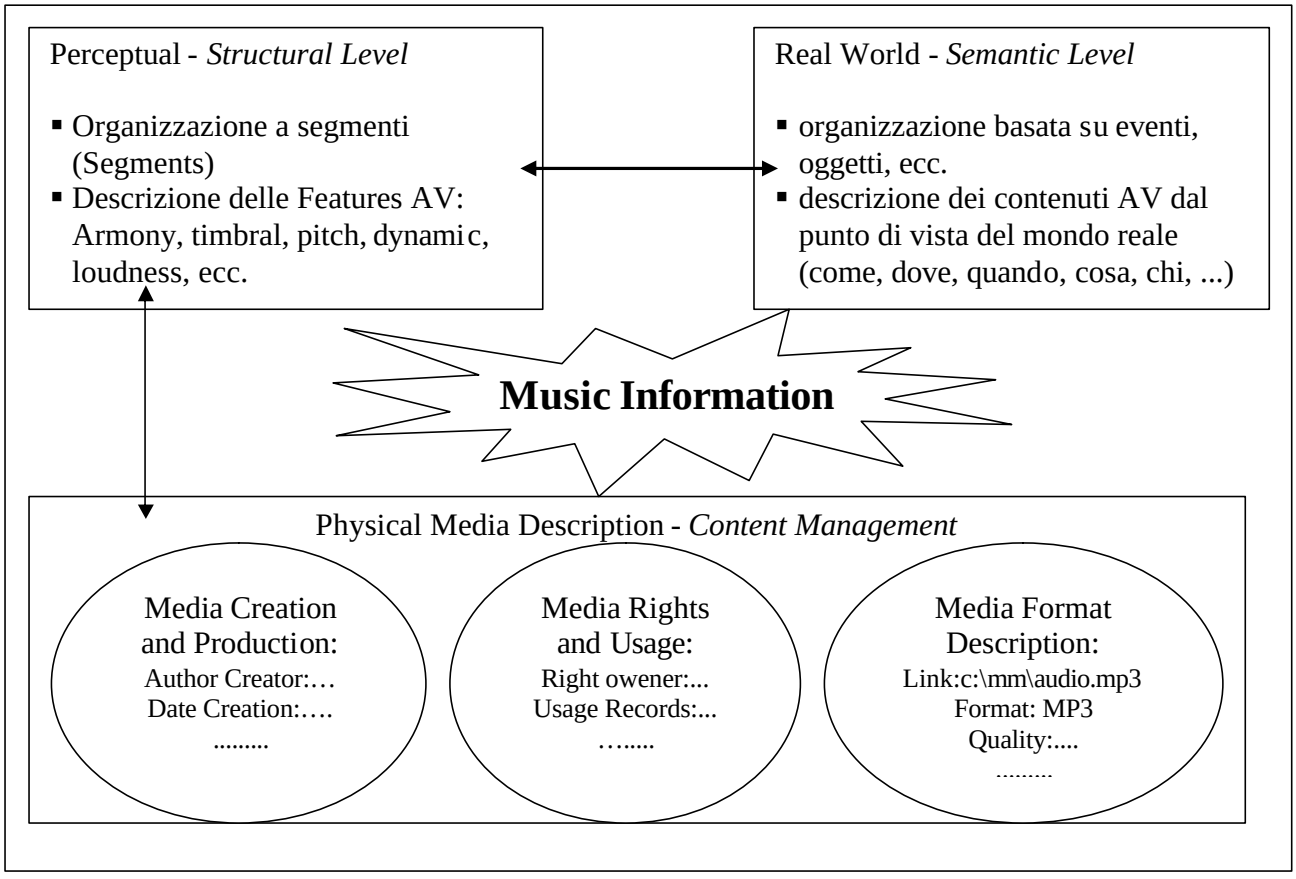


Figura 8: organizzazione dell'informazione musicale con MPEG-7

Per chiarire meglio i concetti trattati finora, analizziamo un semplice esempio pratico; in Fig. 10 viene descritta sommariamente la possibile rappresentazione di un brano musicale formato da due parti strumentali di chitarra e piano. Seguendo la logica MPEG-7, l'informazione viene suddivisa in segmenti temporali (*AudioSegment DS*), a loro volta organizzati gerarchicamente in funzione del loro

contenuto. Nell'esempio in figura, il pezzo è composto da due parti, una di piano ed una di chitarra, ognuna delle quali è composta da un insieme di note. A livello strutturale, AudioSegment1 ed AudioSegment2 conterranno la descrizione simbolica della melodia (*Melody Tool*) associata ai pezzi di chitarra e piano, ed i collegamenti -con le relative descrizioni di formato, ecc.- ai corrispondenti media audio fisici *(Management Content DT)*. A livello semantico invece, il pezzo viene descritto in termini *narrativi*, ossia in un linguaggio molto simile a quello impiegato nel mondo reale, ed ogni entità viene opportunamente associata all'equivalente segmento presente nel livello strutturale.

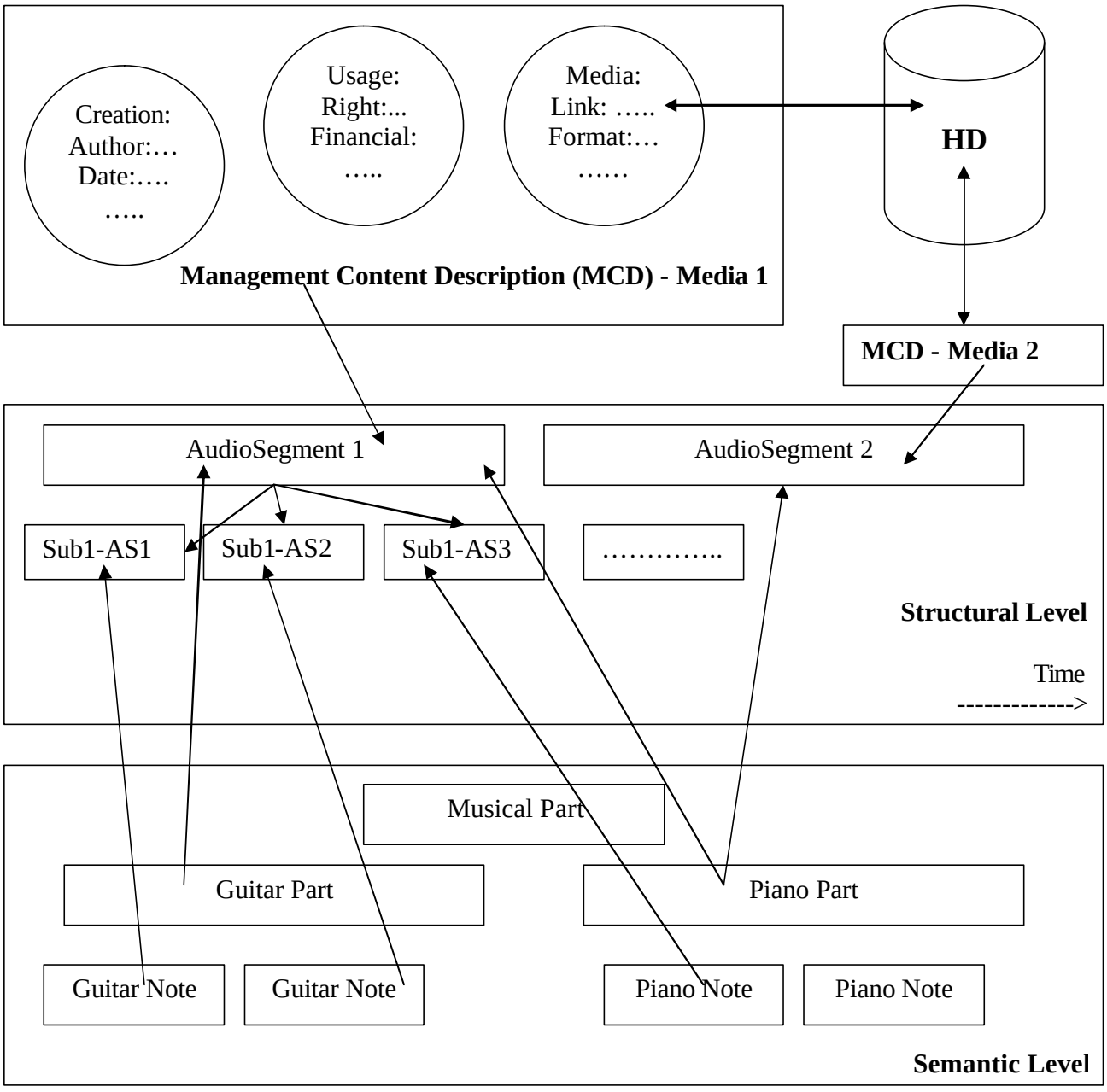


Figura 9: esempio di organizzazione dell'informazione musicale con MPEG-7

10. Esempi di file XML in formato MPEG-7

Qui di seguito presentiamo ed illustriamo alcuni esempi significativi di file XML basati su codifica MPEG-7 per la descrizione di informazione audio-musicale.

Esempio 1: Descrizione dei metadati ID3 (MP3) con MPEG-7 [3]

Nell'esempio qui di seguito è fornita una rappresentazione MPEG-7 degli ID3 di MP3 tramite l'uso di *Creation Description Tools* (tag <Description xsi:type="CreationDescriptionType">) e *Semantic Description Tools* (tag **<Description xsi:type="*SemanticDescriptionType*">**). Analizzando l'esempio, si possono effettuare le seguenti osservazioni:

- nel root tag <Mpeg7> è definito il DDL impiegato (in sostanza un XML-Schema ampliato verso il multimedia). Nell'esempio viene utilizzato il DDL di default fornito da MPEG.
- il tipo di tool utilizzato è definito nell'attributo 'type' del tag <Description>: *Creation Description Tools* per la parte sintattica e *Semantic Description Tools* per la parte semantica.
- è stata fornita una descrizione degli ID3 sia dal punto di vista sintattico che dal punto di vista semantico. Perciò sono stati seguiti due rami distinti di Fig. 8: rispettivamente, quello con top-element *Content Management* e quello con top-element *Content Description.*

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- ID3 V1.1 Example -->
<Mpeg7      xmlns="urn:mpeg:mpeg7:schema:2001"
            xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
            xsi:schemaLocation="urn:mpeg:mpeg7:schema:2001 C:\mpeg7\is\Mpeg7-2001.xsd">
    <Description xsi:type="CreationDescriptionType">
        <CreationInformation id="track-05">
            <Creation>
                <!-- ID3 Song Title -->
                <Title type="songTitle">If Ever You Were Mine</Title>
                <!-- ID3 Album Title -->
                <Title type="albumTitle">Celtic Legacy</Title>
                <!-- ID3 Comment -->
                <Abstract>
                    <FreeTextAnnotation>AG# 3B8308D8</FreeTextAnnotation>
                </Abstract>
                <!-- ID3 Artist -->
                <Creator>
                    <Role href="urn:mpeg:mpeg7:RoleCS:2001:PERFORMER"/>
                    <Agent xsi:type="PersonType">
                        <Name>
                                <FamilyName>MacMaster</FamilyName>
                                <GivenName>Natalie</GivenName>
                        </Name>
                    </Agent>
                </Creator>
                <!-- ID3 Year -->
                <CreationCoordinates>
                    <Date><TimePoint>1995</TimePoint></Date>
                </CreationCoordinates>
            </Creation>
            <!-- ID3 Genre (80 = Folk) -->
            <Classification>
                <Genre href=" urn:id3:cs:ID3genreCS:v1:80"><Name>Folk</Name></Genre>
            </Classification>
```

```
        </CreationInformation>
    </Description>
    <Description xsi:type="SemanticDescriptionType">
        <Semantics>
            <SemanticBase xsi:type="SemanticStateType">
                <!-- ID3 Track -->
                <AttributeValuePair>
                    <Attribute>
                        <TermUse href="urn:mpeg:maf:cs:CollectionElementsCS:assetNum"/>
                    </Attribute>
                    <IntegerValue>6</IntegerValue>
                </AttributeValuePair>
                <!-- ID3v2 TRCK /12-->
                <AttributeValuePair>
                    <Attribute>
                        <TermUse href="urn:mpeg:maf:cs:CollectionElementsCS:assetTot"/>
                    </Attribute>
                    <IntegerValue>12</IntegerValue>
                </AttributeValuePair>
                <!-- ID3v2 TPOS 1/2-->
                <AttributeValuePair>
                    <Attribute>
                        <TermUse href="urn:mpeg:maf:cs:CollectionElementsCS:volumeNum"/>
                    </Attribute>
                    <IntegerValue>1</IntegerValue>
                </AttributeValuePair>
                <AttributeValuePair>
                    <Attribute>
                        <TermUse href="urn:mpeg:maf:cs:CollectionElementsCS:volumeTot"/>
                    </Attribute>
                    <IntegerValue>2</IntegerValue>
                </AttributeValuePair>
            </SemanticBase>
        </Semantics>
    </Description>
</Mpeg7>
```

Esempio 2: Audio Spectrum Representation

Nell'esempio qui di seguito è fornita una rappresentazione MPEG-7 dello spettro di un segnale audio con coefficienti cepstrali reali utilizzando *Audio Description Tools*.
Seguendo la struttura descritta in Fig. 8, in questo esempio è stato creato un MediaEntity (<Description xsi:type="ContentEntityType">) di tipo audio (<MultimediaContent xsi:type="AudioType">). Successivamente, ne viene data la descrizione completa dello spettro sottoforma di coefficienti cepstrali reali, definendo anche ulteriori informazioni, tipiche dell'analisi cepstrale.

```
<?xml version="1.0" encoding="iso-8859-1"?>
<Mpeg7 xmlns="urn:mpeg:mpeg7:schema:2001"
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 xmlns:mpeg7="urn:mpeg:mpeg7:schema:2001"
 xmlns:xml="http://www.w3.org/XML/1998/namespace"
 xsi:schemaLocation="urn:mpeg:mpeg7:schema:2001 Mpeg7-2001.xsd">
    <Description xsi:type="ContentEntityType">
        <MultimediaContent xsi:type="AudioType">
```

```
            <Audio xsi:type="AudioSegmentType">
               <MediaTime>
                  <MediaTimePoint>T00:00:00</MediaTimePoint>
                  <MediaDuration>PT60S638N1000F</MediaDuration>
               </MediaTime>
               <AudioDescriptor xsi:type="AudioSpectrumProjectionType" loEdge="439.063"
   hiEdge="8000.000" octaveResolution="1/16" hopSize="PT10N1000F">
                  <SeriesOfVector totalNumOfSamples="242" scaleRatio="25">
                     <Raw mpeg7:dim="242 10">
                              615.201 -0.882  0.367 -0.185  0.140 -0.104  0.118 -0.066
                              .....................................................
                              .....................................................
                              .....................................................
                              505.073 -0.868  0.374 -0.207  0.159 -0.114  0.093 -0.107
                     </Raw>
                  </SeriesOfVector>
               </AudioDescriptor>
            </Audio>
         </MultimediaContent>
   </Description>
</Mpeg7>
```

Esempio 3: Musical Fragment Melody Description [4]

Nell'esempio qui di seguito è fornita una rappresentazione MPEG-7 della melodia di un frammento musicale utilizzando uno dei tool forniti da *Audio Description Tools*: il *Melody Description Tool*. Esso fornisce una serie di descrittori (*D*) e di schemi (*DS*) per la descrizione della melodia e delle sue principali caratteristiche (metrica, scala, chiave, ecc.). La descrizione della melodia vera e propria può essere effettuata in due modi diversi:

- la prima, chiamata *Melody Contour*, descrive la melodia tramite intervalli, utilizzando solo cinque livelli d'informazione [-2, 2];
- la seconda, chiamata *Melody Sequence*, descrive la melodia secondo intervalli precisi.

Nell'esempio vengono fornite entrambi le rappresentazioni.

```
<AudioDS xsi:type="MelodyType">
   <MelodyMeter>
      <Numerator>3</Numerator>
      <Denominator>4</Denominator>
   </MelodyMeter>
   <MelodyScale>1 2 3 4 5 6 7 8 9 10 11 12</MelodyScale>
   <MelodyKey mode="otherMode">
      <KeyNote display="sol">G</KeyNote>
   </MelodyKey>
   <Contour>
      <ContourData>2 1 1 1 1 1 1</ContourData>
   </Contour>
   <MelodySequence>
      <StartingNote>
         <StartingFrequency>391.995</StartingFrequency>
         <StartingPitch height="4">
            <PitchNote display="sol">G</PitchNote>
         </StartingPitch>
      </StartingNote>
      <NoteArray>
```

```
            <Note>
                <Interval>7</Interval>
                <NoteRelDuration>2.3219</NoteRelDuration>
                <Lyric>Moon</Lyric>
                <PhoneNGram>m u: n</PhoneNGram>
            </Note>
            <Note>
                <Interval>-2</Interval>
                <NoteRelDuration>-1.5850</NoteRelDuration>
                <Lyric>R-i</Lyric>
            </Note>
            <! Remaining notes are elided -->
        </NoteArray>
    </MelodySequence>
</AudioDS>
```

Esempio 4: Musical Fragment Melody Description [6]

Il *Classification Tool* è uno dei tools presenti in *Musical Instrument Timbre* e permette di dare una classificazione tassonomica degli strumenti musicali basata su di una descrizione testuale. Un esempio d'uso di tale tool è mostrato qui di seguito.

```
<ClassificationScheme term="0" scheme="Horbonstel-Sachs InstrumentTaxonomy">
    <Label>"HSIT"</Label>
    <ClassificationSchemeRefscheme="Cordophones" />
    <ClassificationSchemeRef scheme="Idiophones"/>
    <ClassificationSchemeRef scheme="Membranonphones" />
    <ClassificationSchemeRef scheme="Aerophones"/>
    <ClassificationSchemeRef scheme="Electrophones" />
</ClassificationScheme>
<ClassificationScheme term="1" scheme="Cordophones">
    <Label>"Cordophones"</Label>
    <ClassificationSchemeRef scheme= "Bowed" />
    <ClassificationSchemeRef scheme= "Plucked" />
    <ClassificationSchemeRef scheme= "Struck" />
</ClassificationScheme>
<ClassificationScheme term="2" scheme="Idiophones">
    <Label>"Idiophones"</Label>
    <ClassificationSchemeRef scheme="Struck" />
    <ClassificationSchemeRef scheme="Plucked" />
    <ClassificationSchemeRef scheme="Frictioned"/>
    <ClassificationSchemeRef scheme="Shakened" />
</ClassificationScheme>
<ClassificationScheme term="3" scheme="Membranophones">
</ClassificationScheme>
```

11. *Applicazioni e librerie SW MPEG-7*

Attualmente, ancora poche applicazioni, librerie SW e siti web utilizzano MPEG-7 come piattaforma per la rappresentazione dell'informazione multimediale. E' però possibile, per ogni classe applicativa, identificarne alcuni di una certa rilevanza:

- XM - eXperiment Software: prototipo SW sviluppato da MPEG (ISO/IEC 15938) che consente di lavorare su informazione simbolica codificata secondo questo standard.
- MPEG-7 Library - MPEG-7 C++ API Implementation: libreria SW che fornisce una serie di API per il trattamento diretto di file XML codificati in MPEG-7. Per esempio, è possibile leggere un file MPEG-7, caricarlo in memoria, validarlo, modificare (aggiungendo, modificando, eliminando, ecc.) un elemento ed infine riscriverlo su file.
- http://www.singingfish.com/: sito Web per la ricerca di informazione musicale con archivio MPEG-7 based.

12. Bibliografia

[1] "MPEG-7 Overview " - International Organization For Standardization Organisation Internationale Normalisation, ISO/IEC JTC 1/SC 29/WG 11, Coding of Moving Pictures and Audio - N5525, March 2003 Pattaya

[2] "MPEG-7, The Generic Multimedia Content Description Standard; Part 1" and "Overview of MPEG-7 Description Tools; Part 2" - Rob Koenen, Fernando Pereira - Siemens Corporate Research - Copyright @ 2002 IEEE

[3] "MPEG Music Player Application Format (MPEG-A)" - International Organization For Standardization Organisation Internationale Normalisation, ISO/IEC JTC 1/SC 29/WG 11, Coding of Moving Pictures and Audio - N6443, March 2004 Munchen (DE)

[4] "MusicNetwork" website - http://www.interactivemusicnetwork.org/

[5] "Automatic Stream Classification for the Singingfish Search Engine" - JeanRonan Vigouroux, Louis Chevallier and Robert Forthofer (Thomson Multimedia), Ted Diamond, Eric Rehm (Singingfish.com)

[6] "Using and enhancing the current MPEG-7 standard for a music content processing tool" - Emilia Gómez, Fabien Gouyon, Perfecto Herrera and Xavier Amatriain - Audio Engineering Society, Convention Paper Presented at the 114th Convention, 2003 March 22–25 Amsterdam, The Netherlands