2　標準化された技術

2-3 マルチメディア記述

柴田賀昭†

キーワード　メタデータ，MDS，MPEG-7スキーマ，コンテンツ管理，内容記述，要約記述

1．まえがき

MPEG-7 Part 5[1]は，Multimedia Description Schemes（以後，MDSパートと称す）と呼ばれ，ここでは，MPEG-7が規定したマルチメディアコンテンツ記述ツールのうち，Part 3 Visual（本小特集2-1節を参照），Part 4 Audio（本小特集2-2節を参照）に直接含まれない残りすべてのツールを規定している．このことからMDSパートは，MPEG-7コンテンツ記述ツールセットの根幹を規定したものとみなされ，Part 3およびPart 4で規定されたツールは，MDSパートで規定されたコンテンツの構造記述ツール（後述）において，コンテンツの詳細な特徴を記述するための補助ツールとして位置付けられている．

MDSパートの仕様書自体は700ページ以上にも及ぶため，限られた紙面でそれらすべてをカバーすることは到底できない．そこで本章では，MDSの概要紹介，MPEG-7メタデータの2つの形態および，MPEG-7を用いた具体的なコンテンツの内容記述例とその応用について簡単に紹介する．

2．MDSの概要

図1に，MDSパートで規定された各種ツール群の分類に基づくMDS概観図を示す．以下，図1のうち使用頻度の高いと考えられるツール群をいくつか紹介する．その他のツールについては参考文献2）などを参照されたい．

2.1　コンテンツ管理ツール（Content Metadata）

本カテゴリーでは，コンテンツの記録フォーマットやその品質などコンテンツの記録メディアに関する情報を記述するためのメディア関連記述ツール（Media），コンテンツ

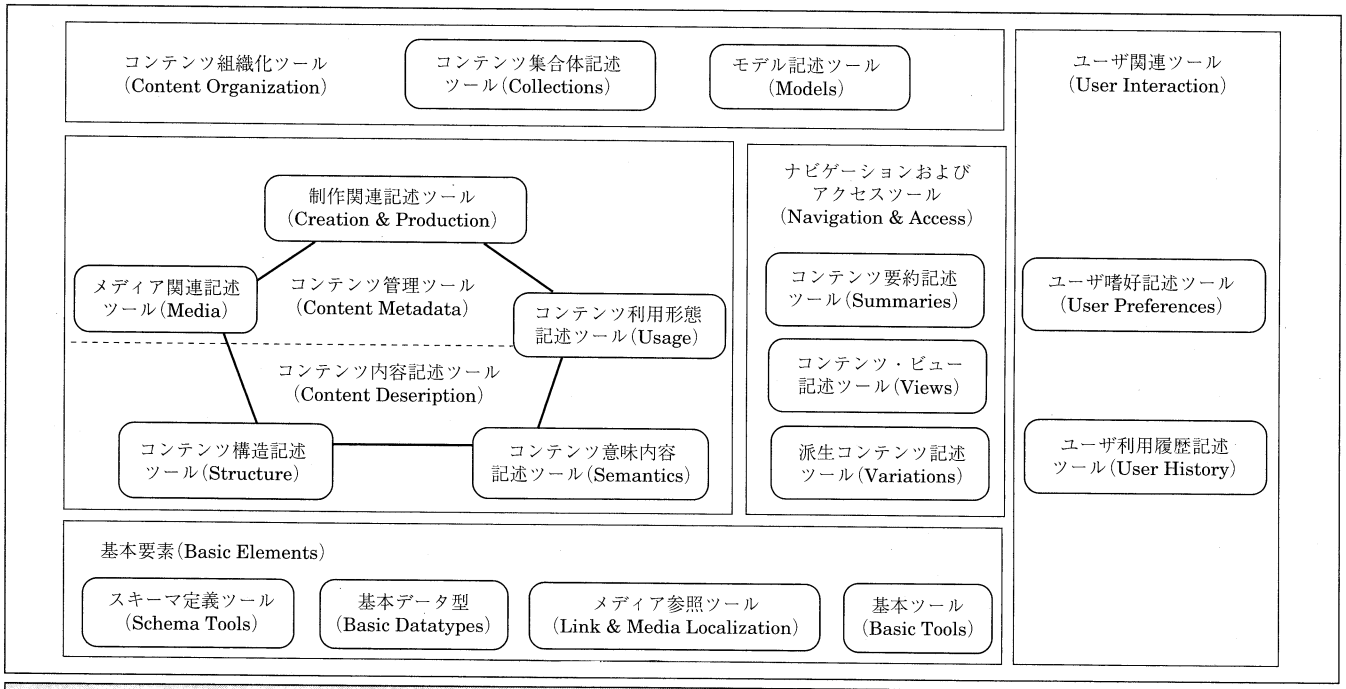

図1　MDS概観図

†ソニー株式会社 BSNC B&Pカンパニー
"Multimedia Contents Description" by Yoshiaki Shibata (Broadband Solutions Network Company, B&P Company, Sony Corporation, Kanagawa)

映像情報メディア学会誌 Vol. 56, No. 11, pp. 1720～1725 (2002)

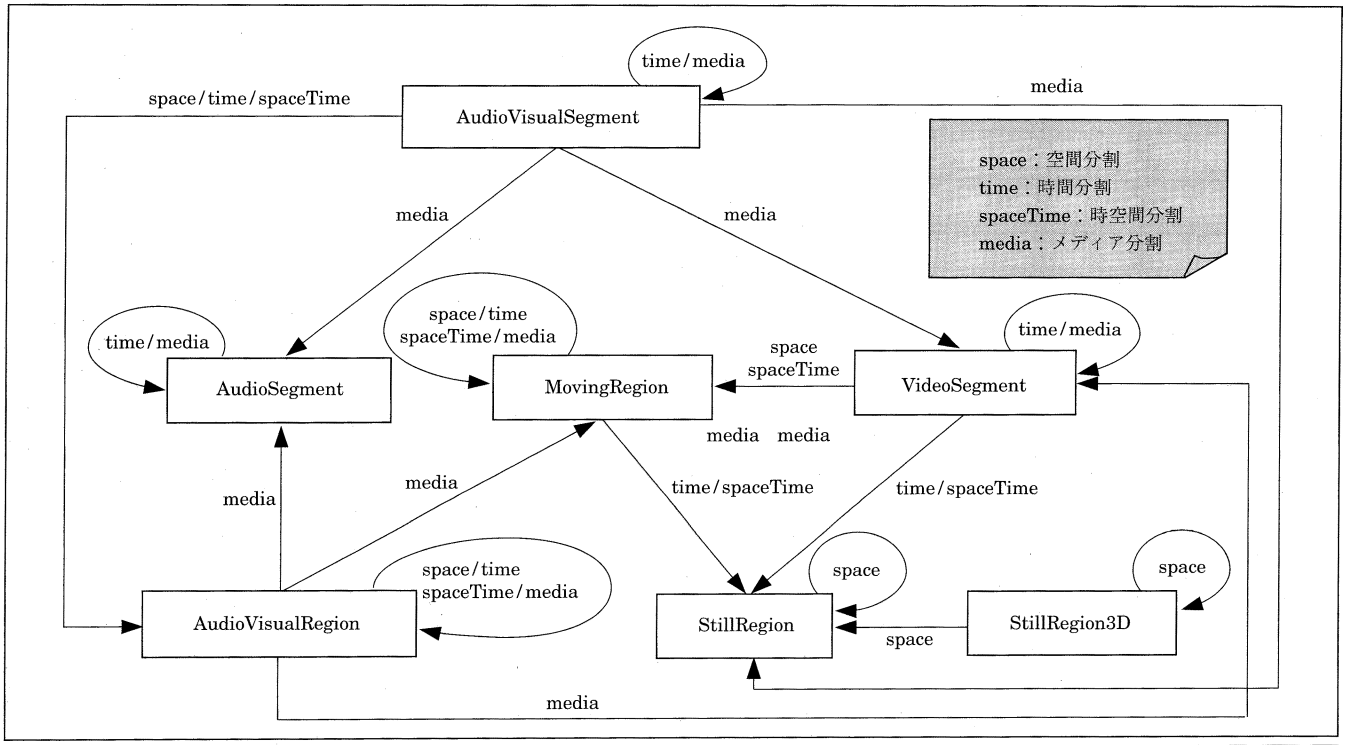

図2　セグメント分割関連図

のタイトル，制作者，制作場所などコンテンツの制作に関する情報を記述するための制作関連記述ツール（Creation & Production），そして，コンテンツの権利情報の記述*1やコンテンツへのアクセス条件など，コンテンツの利用に関する情報を記述するためのコンテンツ利用形態記述ツール（Usage）が規定されている．

2.2　コンテンツ内容記述ツール（Content Description）

本カテゴリーでは，ビデオ，オーディオ，イメージなど，各種マルチメディアコンテンツの物理的，あるいは論理的構造を記述するためのコンテンツ構造記述ツール（Structure）および，コンテンツが捉えた世界（Narrative world）の意味内容を，オブジェクト，イベントなどの意味的実体と，それらの相互関係として記述するためのコンテンツ意味内容記述ツール（Semantics）が規定されている．

ここでコンテンツの構造記述においては，その時空間分割要素を示すセグメント（Segment）という概念に基づき，表現メディアの種類に応じて，AudioVisualSegment DS（音声付動画セグメント），VideoSegment DS（動画セグメント），MovingRegion DS（動画部分領域セグメント），StillRegion DS（静止画領域セグメント），AudioSegment DS（音声セグメント）などが規定され，また，それらの間の分割（Decomposition）関係が厳密に定義されている．例えば，AudioVisualSegment DS（例えば音付きビデオ）は，それ自身に時間的に分割できるほか，そのVideoSegment DS（ビデオの映像のみ）とAudioSegment DS（ビデオの音声のみ）への分割（メディア分割）も可能であるが，AudioVisualSegment DSを直接StillRegion DS（静止画）へ時空間分割することは許可されていない．このような各種セグメント間の分割関係の定義をまとめたものを図2に示す．

なお，Part 3およびPart 4で規定されたツールは，構造化されたコンテンツの内部における詳細な視覚的，聴覚的特徴を記述する目的で，これらコンテンツ構造記述ツールの構成要素として利用されている．

2.3　ナビゲーションおよびアクセスツール（Navigation & Access）

本カテゴリーでは，要約コンテンツを記述するためのコンテンツ要約記述ツール（Summaries），信号処理結果としてのコンテンツの様々な見え方を記述するためのコンテンツビュー記述ツール（Views），コンテンツに関する様々な派生コンテンツを記述するための派生コンテンツ記述ツール（Variations）が規定されている．

ここでコンテンツの要約記述とは，例えばビデオの場合，その内容的に重要な部分を切出してまとめるといった手続きを記述したものである．具体例として，本カテゴリーのツールのひとつであるSummary DSを用いた場合のビデオ要約の記述例を模式的に図3に示す．

図3では，要約対象であるビデオの重要部分がSummary Segment要素によって参照されている様子が示されている．ここで灰色の個所は最重要部分，また黒い個所はそれに続く重要部分を表しており，それぞれがSummary

*1　MPEG-7が規定しているのは，外部にある権利情報記述への参照ツールである．

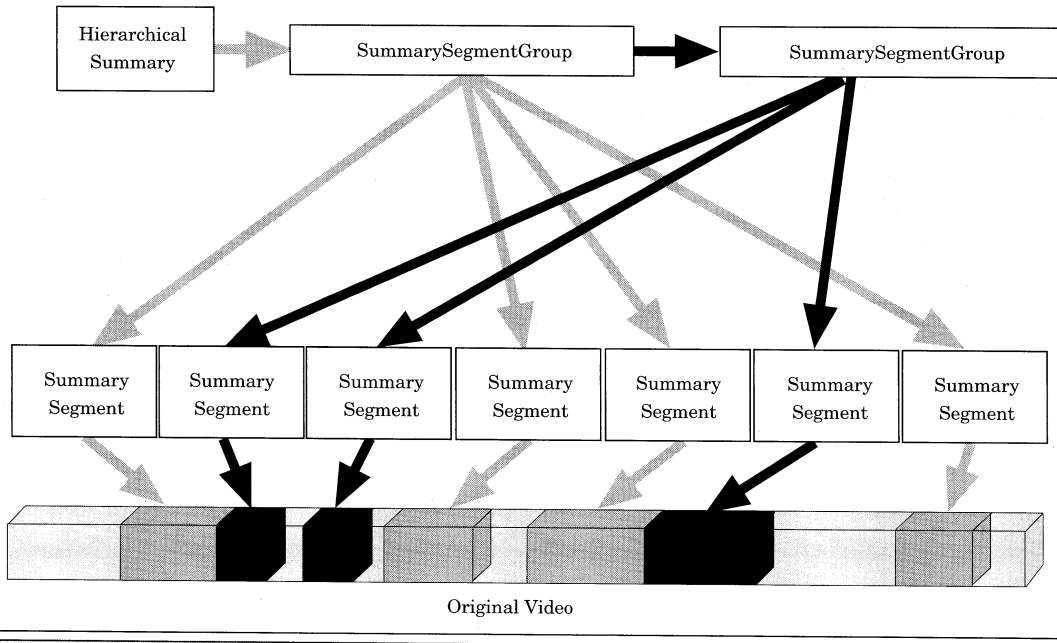

図3　Summary DSを用いたビデオの要約記述

SegmentGroup要素によってグループ化されている．このような記述データを得ることにより，時間的制約がある場合は灰色部分のみを連続再生し，また，さらに詳細な内容を得たい場合は，それに加えて黒い部分も連続再生するといった柔軟なビデオ要約を得ることができる（それ故，本ツールはHierarchicalSummary DSと呼ばれる）．

3．MPEG-7メタデータ

MPEG-7においては，すべてのメタデータは<Mpeg7>タグで始まるXML文書として表現される*2が，さらにそれは記述単位版メタデータ（Description Unit），あるいは完全記述版メタデータ（Complete Description）のいずれかに分類される．この区別を与える仕組みはMDSパートで規定されていることから，以下，これらについて説明する．

記述単位版メタデータとは，MPEG-7が標準記述スキーム（DS：記述の構造），あるいは記述子（D：DSの構成要素）として規定したツールのインスタンス（記述データ）が単体で生成されたものである．例として，Visualツールのひとつであ るScalableColor Dを単体で生成した場合の記述データを図4に示す．

図4において，<Mpeg7>開始タグ（ルート要素）内のxmlns属性値である“urn：mpeg：mpeg7：schema：2001”は，URN表記[3]に基づくMPEG-7の名前空間識別子[4]を表す．またそれに続く<DescriptionUnit>タグは，当該記述データが記述単位版メタデータであることを示す最上位要素である．ここで，DescriptionUnit要素は記述データの中間ラッパーとしての役割を果たしており，それが持つxsi：type属性値にMPEG-7が規定したツール（D/DS）の名称を指定することによって，そのツール単独に基づく記述データがDescriptionUnit要素内に展開される．なお図4では，DescriptionUnit要素内にあるべきScalableColor Dの内容モデル（タグ名と構造を規定したもの）に従ったXML文書が省略されていることに注意されたい．

```
<Mpeg7 xmlns="urn:mpeg:mpeg7:schema:2001" ...>
   <DescriptionUnit xsi:type="ScalableColorType" ...>
      :
   </DescriptionUnit>
</Mpeg7>
```

図4　記述単位版メタデータ（ScalableColor記述例）

このような記述単位版メタデータは，例えば，D/DSインスタンスと対象コンテンツとの関係がアプリケーションの内部テーブルとして管理されている場合や，後述する完全記述版メタデータにおいて，その構成要素の一部を変更あるいは更新したい場合などで利用される．

記述単位版メタデータが単一ツールのインスタンスであるのに対し，対象コンテンツに関する完全な記述に対応した記述データを完全記述版メタデータと呼ぶ．いま，静止画に対する完全記述版メタデータの記述例を図5に示す．

図5において，<Description>タグは完全記述版メタデータを示す最上位要素であり，それが持つxsi：type属性の値により，記述対象の大まかな区分け（図5の場合はコンテンツ実体に対する記述）を示している．さらに続く<Multi-

*2　MPEG-7ではXMLに基づくテキスト記述の他，本小特集2-4節で述べるバイナリー記述も規定されている．

```
<Mpeg7 xmlns="urn:mpeg:mpeg7:schema:2001" ...>
    <Description xsi:type="ContentEntityType">
        <MultimediaContent xsi:type="ImageType">
            <Image>
                :
            </Image>
        </MultimediaContent>
    </Description>
</Mpeg7>
```

図5　完全記述版メタデータ（静止画記述の場合）

mediaContent>タグが持つxsi：type属性値によって，その対象コンテンツが静止画（ImageType）であることを示し，続くImage要素内に静止画の具体的な記述データが展開される（図5では省略）．

図5で示したように，完全記述版メタデータにおいては，その上位要素において記述対象の分類を指定している．ここではコンテンツ実体に対する記述例を示したが，その他に要約記述やモデル記述などのコンテンツの抽象記述（Description要素のxsi：type属性値にそれぞれSummaryDescriptionType，ModelDescriptionTypeを指定）や，メディアや制作関連記述などのコンテンツの管理記述（同，MediaDescriptionType，CreationDescriptionType）が記述対象の分類として規定されている．

4．MPEG-7に基づくマルチメディア記述例

最後に，MPEG-7を用いたマルチメディアコンテンツ記述の具体例とその応用について，ニュース番組の構造記述を例に説明する．

4.1　MPEG-7に基づくニュース番組の構造記述

一般に，ひとつのニュース番組は複数のニュース項目から構成される．そして各ニュース項目においては，導入部としてアナウンサによりそのニュース項目の背景および概要が簡単に紹介され，その後，当該ニュース項目を具体的に説明するための，例えば現地特派員による状況説明や関係者のインタビューなどが続く．このモデルに従った場合，ショット検出技術を用いてあるニュース番組をショット分割し，各ニュース項目の導入部に相当する定期的に発生するアナウンサの類似ショット（以後，これをアンカーショットと呼ぶ）に着目することで，そのニュース番組に含まれる各ニュース項目の開始点，すなわちニュース番組の構造を同定することができる．図6に，このようにして得られたニュース番組の構造を，各ショットの代表フレームを用いて図示した様子を示す．

このようにニュース番組の構造が同定された場合，これはMPEG-7を用いて図7のように記述される．

図7において，2～4行目はこの記述データが映像音声コンテンツの実体に対する完全記述版メタデータあることを示している．それに続くAudioVisual要素は対象ニュース

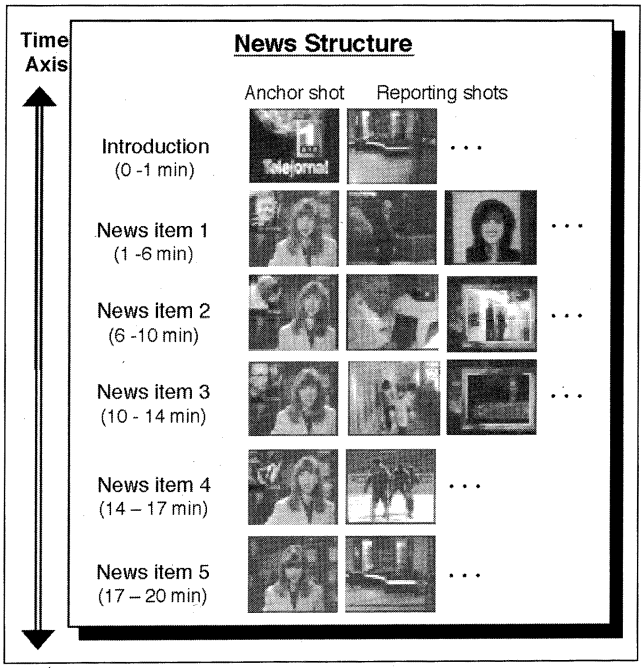

図6　典型的なニュース番組の構造

```
 1: <Mpeg7 xmlns="urn:mpeg:mpeg7:schema:2001" ...>
 2:   <Description xsi:type="ContentEntityType">
 3:     <MultimediaContent xsi:type="AudioVisualType">
 4:       <AudioVisual id="news1">
 5:         <MediaLocator> ... </MediaLocator>
 6:         <MediaTime>
 7:           <MediaTimePoint>T00:00:00:00</MediaTimePoint>
 8:           <MediaDuration>PT20M</MediaDuration>
 9:         </MediaTime>
10:         <TemporalDecomposition>
11:           <AudioVisualSegment id="introduction">
12:             <StructuralUnit ...>
13:               <Name>IntroductionItem</Name>
14:             </StructuralUnit>
15:                 :
16:           </AudioVisualSegment>
17:           <AudioVisualSegment id="newsItem1">
18:             <StructuralUnit ...>
19:               <Name>NewsItem</Name>
20:             </StructuralUnit>
21:             <MediaTime> ... </MediaTime>
22:             <TemporalDecomposition>
23:               <AudioVisualSegment id="anchor1">
24:                 <StructuralUnit ...>
25:                   <Name>AnchorShot</Name>
26:                 </StructuralUnit>
27:                 <MediaTime> ... </MediaTime>
28:               </AudioVisualSegment>
29:               <AudioVisualSegment id="reporting1-1">
30:                 <StructuralUnit ...>
31:                   <Name>Shot</Name>
32:                 </StructuralUnit>
33:                 <MediaTime> ... </MediaTime>
34:               </AudioVisualSegment>
35:                   :
36:             </TemporalDecomposition>
37:           </AudioVisualSegment>
38:           <AudioVisualSegment id="newsItem2">
39:               :
40:             <TemporalDecomposition>
41:               <AudioVisualSegment id="anchor2">
42:                 <StructuralUnit ...>
43:                   <Name>AnchorShot</Name>
44:                 </StructuralUnit>
45:                 <MediaTime> ... </MediaTime>
46:               </AudioVisualSegment>
47:                   :
48:             </TemporalDecomposition>
49:           </AudioVisualSegment>
50:               :
51:         </TemporalDecomposition>
52:       </AudioVisual>
53:     </MultimediaContent>
54:   </Description>
55: </Mpeg7>
```

図7　MPEG-7に基づくニュース番組の構造記述例

番組の全体を表し，5行目のMediaLocator要素において，例えばURLを用いた対象ニュース番組（ファイル）への参照を，続く6～9行目のMediaTime要素にて，その開始タイ

ムコードおよび全体の所要時間を記述している．

その後，10行目のTemporalDecomposition要素により，対象ニュース番組が番組の導入部（IntroductionItem）を先頭に，複数のニュース項目（NewsItem）へ時間分割されることを記述している．ここで分割された各ニュース項目もまた音声付動画セグメントであることから，それらはAudioVisualSegment DS（AudioVisualSegment要素）に基づいて記述される．なお，各分割セグメントの役割を示すStructuralUnit要素の値は，予め定められた制限用語に基づくものである*3．また各時間分割セグメントは，それぞれ対象ニュース番組内の時間軸上での開始タイムコードおよび所要時間を持つ．

図7中22行目のTemporalDecomposition要素は，各ニュース項目がさらに音声付動画セグメントとしてのショットへと時間分割される様子を示しており，その結果が23～28行目，29～34行目と続くAudioVisualSegment要素として記述されている．なお，各ショットに含まれるStructuralUnit要素もまた，それぞれのショットの役割を記述しており，先述のモデルに従い24～26行目では，最初のショットがアンカーショットであることを示している．なお各分割ショットにおいては，上位分割セグメントと同様，その開始タイムコードおよび所要時間が記述されている．

ところで，図7で示したようなニュース番組の構造記述データの生成にあたり，本節の冒頭では，対象とするニュース番組がすでに録画済みデータであるとの仮定の下，ショット検出技術や（定期的に出現する）類似ショットの抽出技術を用いること前提としてきたが，本小特集の第1章で説明されているように，MPEG-7ではその生成方法を何ら規定していない．すなわち，例えばニュース番組の制作段階において，新たなニュース項目へ進む際の収録現場でのキューや映像切替え時のスイッチャのタリーデータ（映像切替えを示すデータ），さらには編集時のカット情報などをまとめて管理することができれば，これを番組送出時のタイムコードと関連付けることで，図7で示したような構造記述データをリアルタイムで生成することも可能であると思われる．

4.2 構造記述データの応用

さて，図7で示したような構造記述データが一旦得られたならば，それは様々な応用に用いることができる．簡単な例では，図6で示したような，当該構造記述データに基づく対象ニュース番組構造のビジュアル表示に基づくニュース番組の“目次”の作成，また，これを用いた所望ニュース項目への高速アクセスなどである．さらに，映像アーカイブに保存されるすべてのニュース番組にこのような構造記述データが付加されていれば，例えば，日々放映されている当該ニュース番組の各放映日のトップニュース（多くの場合，これは最初に出現されるニュース項目であろう）のみを集めた新たなコンテンツを自動的に作成することも可能となる．

以下ではその応用の具体例のひとつとして，当該構造記述データを用いたニュースダイジェストの自動生成について述べる．

先述したように，一般にニュース番組では，各ニュース項目のアンカーショットにおいて，当該ニュース項目の背景および概要が簡単に紹介される．このことから，各ニュース項目から先頭のアンカーショットのみを切出して時間軸上に並べることで，そのニュース番組全体の内容を簡単に描写したニュースダイジェストを自動的に生成することができる[5]．

いま，このようにして得られるニュースダイジェストを，2章で紹介したSummary DSに基づいて記述した例を図8に示す．

図8では，Summary DSの特にHierarchicalSummary DSに従ったニュースダイジェストの記述データが，記述単位版メタデータとして生成された場合を示している（図8の2～3行目を参照のこと）．図8において6行目のSource Locatorは，対象とするニュース番組への参照を示しており，これは，図7の5行目のMediaLocator要素の値と同じものである．また8～9行目のSummaryTheme要素では，この要約のテーマ（ニュースダイジェスト）を与えている．

当該ニュース番組から切出されたアンカーショットは，まとめて11行目のSummarySegmentGroup要素内に記述される．すなわち，当該要素に含まれる12～21行目および22

```
 1: <Mpeg7 xmlns="urn:mpeg:mpeg7:schema:2001" ...>
 2:   <DescriptionUnit xsi:type="SummarizationType">
 3:     <Summary xsi:type="HierarchicalSummaryType"
 4:         hierarchy="independent"
 5:         components="keyAudioVisualClips">
 6:       <SourceLocator> ... </SourceLocator>
 7:       <SummaryThemeList>
 8:         <SummaryTheme id="newsSummary1">
 9:             News Digest</SummaryTheme>
10:       </SummaryThemeList>
11:       <SummarySegmentGroup>
12:         <SummarySegment>
13:           <Name>AnchorShot 1</Name>
14:           <KeyAudioVisualClip>
15:             <MediaTime>
16:               <MediaRelTimePoint>T00:01:00:00
17:                   </MediaRelTimePoint>
18:               <MediaDuration>PT25S</MediaDuration>
19:             </MediaTime>
20:           </KeyAudioVisualClip>
21:         </SummarySegment>
22:         <SummarySegment>
23:           <Name>AnchorShot 2</Name>
24:           <KeyAudioVisualClip>
25:             <MediaTime>
26:               <MediaRelTimePoint>T00:06:00:00
27:                   </MediaRelTimePoint>
28:               <MediaDuration>PT31S</MediaDuration>
29:             </MediaTime>
30:           </KeyAudioVisualClip>
31:         </SummarySegment>
32:             :
33:       </SummarySegmentGroup>
34:     </Summary>
35:   </DescriptionUnit>
36: </Mpeg7>
```

図8 MPEG-7に基づくニュースダイジェストの記述例

*3 ここでは“IntroductionItem”，“NewsItem”などの用語がコンテンツ構造上の同レイヤーに属することを，例えばClassificationScheme DSに基づく制限用語として階層的に定義されているものと仮定した．

～31行目のSummarySegment要素は，それぞれ最初および2番目のニュース項目におけるアンカーショットを表している．なお，各アンカーショットにおいては，その名前および対象ニュース番組内の時間位置がそれぞれ，Name要素，KeyAudioVisualClip要素内のMediaTime要素で記述されている．

なお，図8に示したようなニュースダイジェスト記述データは，図7が示す構造記述データを入力とした専用ツールなどで生成される他，XSLT（eXtensible Stylesheet Language Transformations）[6]を直接適用して生成することも可能である．

また本例では，予め定められたショット分類に基づいてニュースダイジェストを生成する例を示したが，MPEG-7ではテキスト注釈に関しても自由書式ツールに加え，5W1Hの制限用語に基づくツールやテキスト文の文法構造を考慮したツールなど，豊富なツール群を準備している．これより例えば，アナウンサが読み上げるスクリプトをテキスト注釈記述として導入することで，より深い内容に基づいたニュースダイジェストの生成も可能である．さらに，コンテンツ構造記述ツールの内部に定義されたPart 3 Visualおよび/あるいはPart 4 Audioのツールを駆使すれば，映像が捉えるオブジェクトの色，形，動きや，音声情報としてのテーマ音楽，発話音などに基づいたショットの分類や特定個所の切出しも可能となる．

5．むすび

本節では，MDSの概要を紹介したあと，MPEG-7メタデータの2つの形態について説明し，最後にMPEG-7メタデータの記述例およびその応用を，ニュース番組の論理構造記述を例に挙げて紹介した．

MPEG-7はまだ標準化が完了したばかりであり，他方で標準化の過程において広範囲のアプリケーションをサポートすべく膨大な数のツールが規定されたことから，現在では特定用途あるいは特定アプリケーションに向けて，これらのツールをどうやって使っていくべきかの議論が専門家の間で精力的に進められている．今しばらくはMPEG-7の有効な利用法についての探索が続くと予測されるが，近い将来には，これまでとは打って変わったマルチメディアコンテンツの様々な検索アプリケーションが，MPEG-7の利用によって実現されるものと期待される．

（2002年6月28日受付）

〔文　献〕

1）ISO/IEC 15938-5：2002 Information technology-Multimedia Content Description Interface-Part 5：Multimedia Description Schemes

2）B.S.Manjunath, et al.：“Introduction to MPEG-7：Multimedia Content Description Language”, John Wiley & Sons; ISBN：0471486787; Bk & Dvd edition（2002）

3）R. Moats：“URN Syntax”, RFC 2141（May 1997）

4）T. Bray, et al.：“Namespaces in XML”, W3C Recommendation, 14,（Jan. 1999）, http://www.w3.org/TR/REC-xml-names

5）例えば：“Q.Huang, et al.：“Automated Generation of News Content Hierarchy by Integrating Audio, Video, and Text Information”, in Proc. IEEE ICASSP '99, pp.3025-3028（1999）, など

6）J. Clark：“XSL Transformations（XSLT）Version 1.0”, W3C Recommendation, 16（Nov. 1999）, http://www.w3.org/TR/xslt

柴田　賀昭（しばた　よしあき）　1991年，大阪大学大学院工学研究科電子工学専攻修士課程修了．同年，ソニー（株）に入社．1997年から1年間，米イリノイ大学客員研究員として，画像処理の研究に従事．帰国後，MPEG-7標準化活動に関わり，各種作業グループ（Ad Hoc Group）議長の他，ISO/IEC 15938 Part 5（MDS）プロジェクト・エディタを歴任．現在，同社B&Pカンパニーにおいて，メタデータ戦略の立案および放送業務用機器へのメタデータ実装推進に従事．