

ブロードバンド向け複合コンテンツの動的生成・配信支援システム

増井誠生^{*1}、清水誠也^{*1}、藤後努^{*1}、森松映史^{*1}、
松田喜一^{*1}、松元雅美^{*2}、中島充^{*2}、永岡和子^{*3}

概要： ブロードバンド・ネットワークサービス向けの AV 複合コンテンツの作成・配信支援システムについて報告する。本システムでは、MPEG-7 メタデータを用いた Web ベースのリモート素材管理方式を採用し、コンテンツ登録・編集作業の簡易化・効率化を図った。また、コンテンツの動的組み上げ機能、および、複合コンテンツのシームレス再生機能によって、視聴者の意図に即したコンテンツを生成し、違和感なく提供することを可能とするフレームワークを構築した。

A System for Dynamic Creation and Delivery of Composite Multimedia Contents for Broadband Network Service

Motoo MASUI^{*1}, Seiya SHIMIZU^{*1}, Tsutomu TOGO^{*1}, Eishi MORIMATSU^{*1},
Kiichi MATSUDA^{*1}, Masami MATSUMOTO^{*2}, Mitsuru NAKAJIMA^{*2}, Kazuko NAGAOKA^{*3}

Abstract: We describe a system framework for dynamic creation and delivery of composite multimedia contents, operating on broadband network infrastructure. In our system, we introduce a web-based remote application for audiovisual data management, which uses MPEG-7 metadata to simplify the content-registration and content-authoring. We have implemented a system for providing web-based authoring software for dynamic content-building, and a web-based player that supports smooth and seamless playback of segmented streaming video.

1. 背景

ADSL や CATV を利用したネットワーク接続の普及、いわゆるブロードバンド化によって、一般家庭のユーザ向けに様々なジャンルの映像コンテンツ配信が普及しつつあるのは周知の通りである。また、企業 LAN を利用する業務ユーザにおいても、クライアントマシン及びネットワークインフラのブロードバンド対応によって映像を利用する環境が整いつつあり、実際に、e-Learning や、オンラインマニュアルなど、映像を用いた情報共有を目的に映像の業務利用が広く普及しつつある。

映像コンテンツを用いたサービスの提供者にとって、魅力のある映像コンテンツをいかに効率的に制作できる

かという点が課題となる。映像を用いたコンテンツのオーサリング作業を効率化し、作業コストを低減させるためには、既存の映像を素材として再利用することを含め、素材や、素材に付随するメタデータ、完成したコンテンツなど、可能な限りコンテンツの共有を図り、再利用を促進する仕組みが有用である。

映像コンテンツの利用者のための技術課題としては、映像の内容やメッセージ性を損なわずに、快適に配信を行う仕組みが必要である。また、映像の選択・映像構成などについて、利用者ごとに個別に映像をカスタマイズし、違和感なく提供する仕組みによって、映像コンテンツにより魅力を与えることができる。

*1: (株)富士通研究所 *2: 富士通(株) *3: (株)テクノプロジェクト
*1: Fujitsu Laboratories Ltd. *2: Fujitsu Limited. *3: Techno Project Co., Ltd
(E-mail: 増井: masui.motoo@jp.fujitsu.com, 森松: eishi@jp.fujitsu.com)

以上のような観点から、我々は、さまざまな視聴者の作業環境、時間制約、意図にしたがって、個人ごとにカスタマイズした映像を、より自然に提供することを目的として、複合コンテンツの動的生成・配信支援システムを開発した。

2. 個人カスタマイズの問題点

映像や音響を素材として利用する、マルチメディア複合コンテンツの動的生成・配信システムの構築において、映像の個人カスタマイズに関連する以下の2点について、予備的検討を行った。

- (1) コンテンツ編集の効率化、省力化をすすめる
オーサリング技術
- (2) 視聴時の快適さを追及するプレイヤー技術

2.1. オーサリングに関する検討

周知のように、映像コンテンツのオーサリングツールは既に数多く存在する。しかし、我々は、既存のオーサリングツールには、以下の問題点が存在すると考えた。

問題点1: ノンリニア編集方式の処理時間

映像素材をローカルに取り込み、ノンリニア編集によって完成した映像をファイル出力するオーサリングツールの場合、出力の際に時間長に応じた処理時間が必要となる。このようなツールで、利用者ごとに個別にカスタマイズした映像を用意する場合、作業時間や、データの保存領域などに無駄が生じる。

問題点2: 専用クライアントによる編集方式の制約

コンテンツ編集用のオーサリングツールを専用クライアントで動作させる場合に、専属のオペレータを確保することがある。この場合、オーサリング作業の変動に対して、オペレータの増員や作業マシンの追加などが必要となり、柔軟な運用が行えない。

問題点1に関して、我々は、ノンリニア編集のように映像素材を直接編集する方式は、高度な編集効果が要求される場合に適していると考えた。一方、利用者向けの個別カスタマイズの実現には、映像素材を仮想的に取り扱うオンライン編集方式が適していると考えた。具体的には、素材から得られるMPEG-7メタデータをオーサリングツールに読み込み、このメタデータを用いて複合コンテンツを論理的に編集し、XMLベースのコンテンツ記述として出力する。この結果、元の映像素材に手を加えることなく編集が可能となる。また、素材を集中的に管理する必要はなく、ネットワーク上に分散配置された素

材のメタデータを利用してコンテンツを生成し、利用することができる。

問題点2について、我々は、Webベースで提供するリモートオーサリング方式が妥当と判断した。複数のオペレータがオーサリングを同時に行うことも可能であり、オペレータを専属とする必要もない。作業量の増大にも、サーバの性能強化で対応できる。映像コンテンツ素材の提供者が、オーサリング機能を用いて自分自身でカスタマイズを行うことも可能である。

このように、コンテンツ編集におけるXMLメタデータの利用、および、Webベースのリモートオーサリングツールを組合せれば、様々な利用者を想定して個別カスタマイズしたコンテンツを、その場で視聴できる。

2.2. プレイヤに関する検討

オンライン編集方式によって生成した複合コンテンツの視聴時の課題について、以下の2つの問題点に対する解決手段の検討を行った。

問題点1: バッファリングの発生

ストリーミング再生技術の宿命として、メディアのスキップ時や、複数の素材の切り替え時などに、ストリームデータの読み直し、いわゆる「バッファリング」が発生し、その間、映像再生処理が停滞する。

問題点2: ブラウザとプレイヤーの時刻ずれ

複合コンテンツの同期管理においては、バッファリングの発生によって、Webブラウザ・メディアプレイヤーなどの同期を吸収できないケースがある。具体的には、メディアプレイヤーの再生起動処理を待つ結果、ブラウザとメディアプレイヤーの時刻管理にずれが生じる。

バッファリングの最小化とその課題

ストリーミングメディアの発生時に発生するバッファリングの発生時間を短くするための技術が、数社から提案されている[1][2][3]。これらの技術は、ストリーミングに必要なネットワーク帯域の数倍の余剰帯域を想定し、バッファリングの発生時にパースト転送を行うことで、短時間でプレイヤーを再生状態に復帰させる点を特徴とする。余剰帯域の利用により、映像のストリーミング再生中に利用者が再生をスキップした場合や、ネットワーク接続が切断された場合などに、バッファリングの発生時間を短縮させるといった効果がある。ただし、その代償として、相応のネットワークリソースを消費することには留意が必要である。

これらの方式は、プレイヤー上で一本の映像コンテンツをスムーズに再生させることを前提に、問題点1のバッファリングの発生時間を短縮するものである。問題点2に示した、ブラウザとプレイヤーの連携時に発生する時刻ずれを解決するものではない。

解決案：シームレス再生方式

我々は、問題点1と問題点2を同時に解決し、正確なシームレス再生を実現する方法を新たに考案した。具体的には、クライアント上で複数のメディアプレイヤーオブジェクトを待機させ、複合コンテンツ記述（3章を参照）に従って、映像や音などの時間メディアの切り替えと同時に、プレイヤー（デコーダ）を切り替える（図1）。

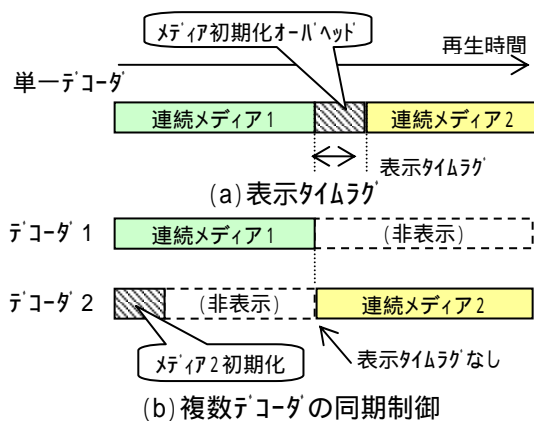


図1. 表示タイムラグの解消

複数のプレイヤーオブジェクトによってクライアントマシンへの負荷が発生し、バックグラウンドでプレイヤーを待機させるための帯域も必要であるが、ブラウザとプレイヤーの時刻管理を厳密に行うことができるという利点がある。よって、バッファリングの発生の影響を避け、確実に時刻ずれの発生を防ぐことができる。

2.3. 開発方針

以上の検討を経て、複合コンテンツの動的生成・配信システムに個人カスタマイズを導入するために、以下のアプローチをとることが妥当だと判断した。

(A) XML メタデータを採用

素材管理用メタデータ、複合コンテンツ記述

(B) Web ベースのオーサリング

どこでも（誰でも）コンテンツ編集が可能

(C) シームレス再生

バッファリング待ちを解消。厳密な時間管理

なお、XML メタデータ利用に関する先行研究として、

上村らによる「遠隔映像編集・検索・配信ネットワークサービスシステム」[4]がある。このシステムでは、ネットワーク上に分散された映像素材を利用するため、映像編集クライアント上に映像素材のネットワーク検索機能を実装している。また、コンテンツ記述として XML ベースの SMIL 1.0 [5] を採用し、GUI ベースで SMIL 記述の編集を可能としている。

我々も、上村らと同様に、

- ・ネットワーク上の映像を素材として利用する
- ・XML ベースのシナリオ記述言語を採用

を前提としたシステム設計を行った。

さらに、複合コンテンツの個人カスタマイズを実現するために必須となる、

- ・XML ベースの素材管理
- ・Web ベースのオーサリングツール
- ・複合コンテンツプレイヤーでのシームレス再生

などの機能を開発することで、個別カスタマイズしたコンテンツをスムーズに視聴できる、複合コンテンツの動的生成・配信システムを構築できると考えた。

3. 複合コンテンツ記述とプレイヤー

本節では、上記の技術課題の実現に必須となる複合コンテンツ記述、および、複合コンテンツ再生時のプレイヤーの動作について説明する。

3.1. 複合コンテンツ概要

複合コンテンツは、映像や音響・音声などの多種類の素材を、時間的、位置的、あるいは意味的に組み合わせたものである（図2）。一般的な用語で複合コンテンツの概念を説明すると無用な誤解を招くため、我々は、以下のように、複合コンテンツの構成要素を整理した。



図2. 複合コンテンツの表示

モノメディア

AudioVisual、Audio、Image、Text といった、複合メディアを構成する個々の単体メディア（素材）を指す。

セグメント

モノメディアを場面等の意味的な区切れ目で分割したもの。時間メディアである AudioVisual、Audio はセグメントに分割できる。しかし、静的メディアである Image、Text のセグメント分割はできない。

クリップ

各モノメディアをリンクさせて作成した複合メディアの最小単位を指す。1つのクリップ中に同種類のモノメディアが複数存在してもよい。

シーン

クリップを順番に組み合わせて完成した、1本のコンテンツを指す。

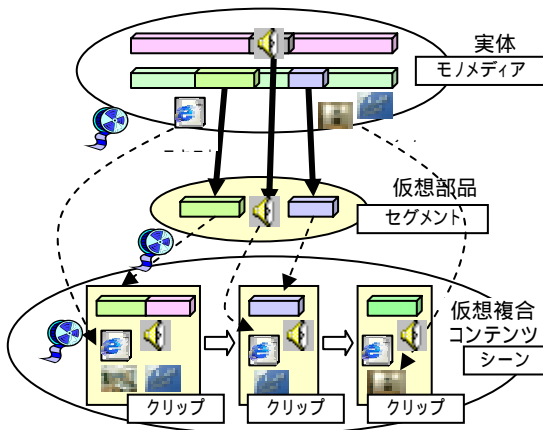


図3. 複合コンテンツの構成要素

モノメディアが、物理的に存在する単体メディアデータを指すのに対し、モノメディアから時間的に切り出したセグメントや、セグメントを複合メディアとして複合させたクリップや、複数のクリップを時間的に連続させたシーンは、いずれも論理的なデータである（図3）。

複合コンテンツの管理において、ネットワーク上に分散配置されたモノメディアに対し、その一部を参照するセグメントをベースとして複合コンテンツが構成される。実際のメディアデータには手を加えず、これを参照するメタデータの処理によって、動的なコンテンツの組み換えを即時的に実現できる。

3.2. WBPL

複合コンテンツを取り扱う言語として、W3C による SMIL 2.0 勧告[6]がよく知られている。我々は、この SMIL 2.0 仕様をベースとして、図3に示す複合コンテンツの構成要素を効率よく扱うための若干の改変を加えた、独自の複合コンテンツ記述言語 WBPL (Web-Based Presentation Language) を定義した。

具体的には、WBPL では SMIL 2.0 をベースに、

- ・アニメーション関連タグの削除、
- ・ページ単位でのレイアウトタグの追加

などを行っている。

我々は、コンテンツ制作経験のないユーザが、簡単に制作し、カスタマイズできるコンテンツ記述を目指し、WBPL の言語仕様をシンプルにまとめた。その上で、複合メディアの最小断片であるクリップごとに、HTML ページにおけるレイアウト構成をまとめて指定できることが、ブラウザベースの複合コンテンツの記述性を向上させると考えた。

WBPL の記述例として、映像、静止画、テキストの3種のモノメディアで構成されるクリップに対して、クリップに含まれるオブジェクトのレイアウト構成をまとめて記述できることを示す（図4）。

我々は、クリップ単位での動的カスタマイズが可能で、シームレス再生にも適した複合コンテンツ記述言語として WBPL を位置づけている。ただし、Web ブラウザでの GUI 操作によるオーサリングやプレイヤー制御を実現しているため、コンテンツの制作者や視聴者が WBPL によるコンテンツ記述を直接目にする機会はない。

```
<wbpl>
<head>
  <layout id="L_1">
    <region selector="#R_V_0" left="0" top="0" width="480" height="360"/>
    <region selector="#R_I_0" left="485" top="0" width="240" height="180"/>
    <region selector="#R_T_0" left="485" top="185" width="240" height="175"/>
  </layout>
  <layout> ... </layout>
</head>
<body>
  <seq>
    <par layout="L_1" dur="36s">
      <video region="R_V_0" src="video_file.wmv" type="video/x-wmv"/>
      
      <text region="R_T_0" src="text_file.html" type="text/html"/>
    </par>
    <par>
      ...
    </par>
  </seq>
</body>
</wbpl>
```

図4. WBPL によるクリップとレイアウトの記述

3.3. プレイヤ

プレイヤーはユーザクライアント PC での Web アプリケーションとして動作する HTML (JavaScript) ベースのプログラムである。WBPL による複合コンテンツ記述を読み込み、ストリーミングメディアプレイヤーのメディアデコードモジュールを同期制御しながら動作する。本プレイヤーは約 43KB の JavaScript プログラムであり、実行時に複合コンテンツと共にダウンロードされるため、プラグインや専用アプリのインストールは不要である。

4. 複合コンテンツサーバ

第 4 節では、複合コンテンツの生成・配信を支える、複合コンテンツサーバの構成の概要を述べる。複合コンテンツサーバは、素材 (モノメディア) からメタデータを抽出するインデキシングツール、抽出したメタデータを管理するコンテンツ管理 DB、複合メディアコンテンツを生成する Web ベースのオーサリングツールから構成される。複合コンテンツサーバとクライアントからなる、システム構成図を示す (図 5)

4.1. MPEG-7 ベースのコンテンツ記述メタデータ

本システムでは、モノメディア及び複合コンテンツの効率的な再利用を含めた統合管理を実現するために、モノメディア及び複合コンテンツに対し MPEG-7 ベースのメタデータを付加している。基本的には、MPEG-7 の標準に含まれる MDS (Multimedia Description Scheme) [7] のサブセットを使用している。具体的には、タイトル名や出演者、キーワードといったコンテンツ内容記述、符号化方式等のメディア属性記述、及び時間的な構造記述 (セグメント情報) といったタグセットを利

用している。複合コンテンツの記述については、MPEG-7 標準の中で適切な記述スキームが無いため、モノメディア間の時間同期情報を独自に拡張している。

4.2. コンテンツ管理 DB

コンテンツ管理 DB は、MPEG-7 ベースのメタデータを蓄積・管理するための DB ソフトウェアである。一般に、MPEG-7 データのような半構造データ (データ構造を事前に決定できない) を格納する目的には、オブジェクト指向 DB は RDB に比べて有利とされる。また MPEG-7 スキーマのクラス継承定義と同等な機能を、オブジェクト指向 DB は実装している。本システムでは、コンテンツ管理 DB の設計・構築の効率化を念頭に、オブジェクト指向 DB 製品である JASMINE[8] を使用している。

4.3. インデキシングツール

映像モノメディアデータの MPEG-7 メタデータを生成するためのツールである。本インデキシングツールは、ユーザクライアント PC での Web アプリケーションとして動作する、HTML (JavaScript) + ActiveX ベースのプログラムである。ストリーミングビデオや、ローカルストレージビデオについて時間的なセグメンテーションを行い、各セグメントにキーワード等を付加する。セグメンテーションは、手動もしくは自動 (指定した時間間隔での分割) である。作成した MPEG-7 データは、コンテンツ管理 DB で管理される。

4.4. オーサリングツール

映像や静止画、HTML ドキュメントといったモノメディアから、複合コンテンツを作成するためのオーサリングツールである。本ツールは、ユーザクライアント PC での Web アプリケーションとして動作する、HTML (JavaScript) ベースのプログラムである。本ツールにより、複合コンテンツを、必要なモノメディアを付箋のような感覚で、時間的及び空間的に自由に貼り合わせて作成することができる。本ツールは、コンテンツ管理 DB と連携しており、DB 検索により既存のものを簡単に再利用可能になっている。

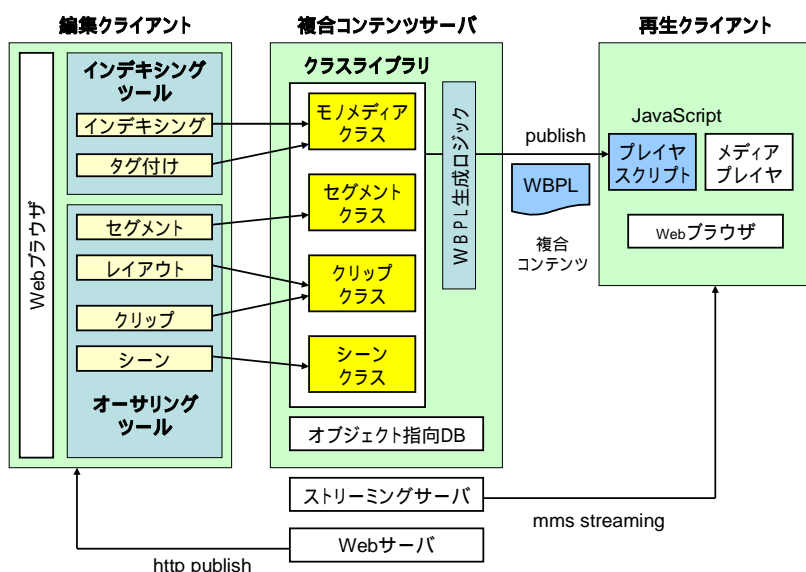


図 5. システム構成図

4.5. 映像配信サーバ

WBPL による複合コンテンツ記述では、映像配信機能を実現するために、汎用のストリーミングサーバの利用を前提としている。今回は、Windows 2000 Server 上で動作する Windows Media Service を利用した。

5. 評価実験

本システムの動作検証を目的に、技術セミナーの講演映像を、本システムに適用した事例を紹介する(図2)。まず、ビデオカメラで撮影した講演の映像を、ストリーミング形式にエンコードし、映像配信サーバに登録する。次にインデキシングツールによって、MPEG-7 メタデータを抽出し、映像の URL とともにコンテンツサーバに登録する。さらにオーサリングツール(図6)で、登録済の映像と発表スライドを利用して複合コンテンツを作成し、コンテンツサーバに登録する。このようにして制作した複合コンテンツが、WBPL 記述にしたがって、正確にシームレス再生されることを確認した。

このようなコンテンツ制作作業を実際に試みた結果、インデキシングツールでのセグメンテーションの際に、事前に映像素材の構成を把握しておく必要があるなど、作業時間を費やす要因があることがわかった。制作作業のコストを低減するため、メタデータ抽出の効率化が大きな研究課題だと考えている。

6. まとめ

蓄積映像やストリーミング映像から抽出した MPEG-7 メタデータの利用に適したサーバ構成に加えて、SMIL 2.0 ベースのコンテンツ記述言語 WBPL を定義し、WBPL の生成系と再生系を Web アプリケーションとして提供することによって、コンテンツの動的生成・配信システムにおける個人カスタマイズ機能を実現した。さらに、セミナー素材を利用したデモコンテンツを本システム上で作成し、シームレス再生機能が正常に動作することを確認した。

今後の課題として、前述のメタデータ抽出の効率化のほかに、コンテンツ記述におけるシナリオ分岐の導入、



図6. 複合コンテンツのオーサリング画面

モバイル機器向け配信、音声モダリティの利用などを考えている。

7. 参考文献・URL

- [1] Real One Player, Real Networks, (2002.3.5)
<http://www.realnetworks.com/company/press/releases/2002/realonegold.html>
- [2] Darwin Streaming Server, Apple Computer (2002)
<http://developer.apple.com/darwin/projects/streaming/>
- [3] "Fast Streaming" demonstration, Microsoft (2002)
http://www.microsoft.com/windows/windowsmedia/demos/producer/Producer_Fast_Streaming/Fast_Streaming_files/default.htm
- [4] 上村圭介, 遠隔映像編集・検索・配信ネットワークサーバシステム, ITX2001 - IPAプロジェクト成果発表展示会 (2001.2.14)
<http://www.ipa.go.jp/NBP/ITX2001-1/result/PDF/jisedai/036.pdf>
- [5] Synchronized Multimedia Integration Language (SMIL) 1.0 Specification, W3C Recommendation, 15 June 1998
<http://www.w3.org/TR/1998/REC-smil-19980615/>
- [6] Synchronized Multimedia Integration Language (SMIL 2.0), W3C Recommendation, 07 August 2001
<http://www.w3.org/TR/2001/REC-smil20-20010807/>
- [7] MPEG-7 Multimedia Description Schemes XM (Version 7.0), ISO/IEC JTC 1/SC 29/WG 11/N3964 March 2001, Singapore
http://mpeg.telecomitalia.com/working_documents/mpeg-07/mds/mds_xm.zip
- [8] JASMINE <http://software.fujitsu.com/jp/jasmine/>