Streaming SMIL Presentations via a Multimedia Semantic Model

Shu-Ching Chen[1] *Sheng-Tun Li*[2] *Hung-Chi Chen*[2] *Mei-Ling Shyu*[3] *Chengcui Zhang*[1]

[1]School of Computer Science, Florida International University, Miami, FL 33199, USA

[2]Department of Information Management,
National Kaohsiung First University of Science and Technology, Kaohsiung 811, Taiwan, R.O.C.

[3]Department of Electrical and Computer Engineering,
University of Miami, Coral Gables, FL 33124, USA

ABSTRACT

The Synchronized Multimedia Integration Language (SMIL) allows Web designers to design complicated and vivid multimedia presentations in a declarative manner. In this paper, an abstract semantic model called Multimedia Augmented Transition Network (MATN) is used to model the conceptual structure of SMIL, as well as the temporal relations and synchronization control. The advantages of using the MATN model are its simplicity and ease of modification (scalability). Users can easily generate their favorite multimedia presentations using MATNs and render them using the SMIL players. Since the SMIL specification is quite new to the Internet society, the functionality of the players is limited, i.e., they often acquire proprietary platforms or cannot deal with MPEG or RTP media objects. Another contribution of this paper is to propose a novel architecture based on Java JMF technology for tackling with such constraints.

1. INTRODUCTION

The cooperation of Web and multimedia technologies results in the Web-based multimedia presentations so that one can enjoy watching TV-like programs via general browsers. However, designing the multimedia presentations is a complicated task. It requires not only complex timer-based programming in a script language, but also a high-level conceptual model that can help to systematically model the temporal relations and user interactions in a multimedia presentation. In the perspectives of the languages, the Synchronized Multimedia Integration Language (SMIL) [6], built on XML and proposed by W3C, provides a simple way to design a multimedia presentation in a similar manner to HTML documents. The SMIL specification meets three requirements of the multimedia document models, namely temporal, spatial and interaction [2]. On the other hand, in order to choose a semantic model that is especially suitable for a multimedia presentation system, the abstract semantic model called the Multimedia Augmented Transition Network (MATN) [3, 4] is considered effective for this purpose. The contributions of this paper are described as follows:

(1) Propose a SMIL presentation system that can deal with synchronous video objects and RTP media streams based on the Java Media Framework (JMF) technology.

A SMIL document needs a SMIL player to render its contents on a Web browser following the temporal and spatial constraints. Although there are several SMIL-enabled players available, most of them rely on Microsoft Windows. This leads to the issue of platform neutrality that can be solved by the over-whelming methodology of Java computing. Similar presentation systems such as SOJA [10] and S2M2 [9] exist. However, either video (the most important media type) is excluded or streaming media objects is disallowed. In this study, we extend our previous work in JTViewer, a TV-like MPEG viewer over the Internet [8], and propose a SMIL presentation system that can deal with such constraints using JMF.

(2) Apply the MATN model to model the conceptual structure of SMIL, as well as the temporal relations and synchronization control.

As a promising graphical model, the MATN has been applied successfully in many application domains. It can describe and study multimedia systems that are characterized as being parallel, concurrent, and/or alternative. The advantages of using an MATN are its simplicity and scalability. It is simple in that it uses state-transition graph and regular-expression like grammars to represent the temporal and synchronization control information. It enables scalability in that an MATN can model user interactions in a single framework, unlike the timeline models [1] that need to have several timelines to model the same situation.

This remaining paper is organized as follows. Section 2 gives a brief review about XML, SMIL and JMF technologies. The system architecture of the proposed Web-based SMIL player is described in Section 3. Section 4 shows the effectiveness of using the MATN in the SMIL based multimedia presentation systems and gives the system demonstration. Section 5 concludes the paper.

2. XML, SMIL, and JMF

A SMIL player is essentially a Web-based application, which needs seamless integration of the emerging Web computing technology – Java and XML. Java is well recognized for its architecture neutrality. XML is designed for document neutrality that can facilitate universal message exchange. In this section, we briefly review the technologies used in the study.

2.1 XML

Similar to HTML, XML (Extensible Markup Language) is a metadata language for defining markup languages [5] such as SMIL, MathML (Mathematical Markup Language), and CML (Chemical Markup Language). XML differs from HTML in three key concepts, namely extensibility, structure, and validation [5]. In extensibility, XML allows authors to define new tags and attributes for specifying the syntax and semantics of data. In structure, XML supports nested document structures similar to regular expressions. In validation, XML allows a document to be validated via DTD (document type declaration) specification, which identifies what elements are allowed in a document, the sequence of these elements, and element attributes, etc. XML can be seamlessly integrated with Web in a variety of ways. The most often cited one is its ability in facilitating data exchange between heterogeneous databases due to its vendor-neutrality.

2.2 SMIL

The synchronized multimedia integration language (SMIL) provides a simple way for designing TV-like multimedia presentations on the Web [6]. The syntax of SMIL conforms to the XML standard so that one can write SMIL presentations using general text editors. There are several salient features making SMIL attractive in design Web-based multimedia presentations. Firstly, the presentation designer can indicate the spatial layout and temporal relationships of media, and where and when the objects are shown. Secondly, media objects can be distributed either from the same resource of the SMIL document or any place specified by the uniform resource locator (URL) such as HTTP, FTP, or RTSP. Thirdly, a set of navigation constructs is also defined to support the functionality of HTML-style hyperlinks. Finally, media objects can be presented on client's browsers according to the system and personal preferences.

Existing popular players for SMIL 1.0 includes Microsoft Internet Explorer 5.0, Apple QuickTime 4.15, RealNetworks Realplayer 8, Helio SOJA [10], and CWI GriNs. Most of them except SOJA can only run on proprietary platforms such as Microsoft Windows or Macintosh, which limits their usability. However, achieving smooth SMIL video presentation over the Internet is a non-trivial work that involves pre-fetching and buffering media objects.

2.3 Java Media Framework (JMF)

The JMF API provides three packages: Player, Capture, and Conference for manipulating time-based media including playback, capture and conferencing [7]. It supports most standard media content types such as AU, WAV, MIDI, MPEG-1, QuickTime, and AVI. JMF1.1FCS supports RTP for playing broadcast and multicast media for limited media types including DVI, GSM, G723, and G711 except MPEG-1. Although JMF2.0ea improves MPEG-1 in which Layer 1&2&3 (MP3) is also supported, MPEG-1 video is still not provided. Therefore, it is a great challenge for playing RTP multicast MPEG-1 using JMF.

3. JMF BASED SMIL MULTIMEDIA PRESENTAION SYSTEM

We utilize the Java computing and JMF technologies to develop the SMIL presentation system to meet two goals, namely RTP streaming and navigation of HTML-style hyperlinks. This section discusses our system development approach. The high-level system architecture is illustrated in Figure 1. It contains Web Server, SMIL Web Server, SMIL Document Server, RTP Server, Multimedia Database containing MPEG and MP3 items, and SMIL Player on the client side. The platform where each component runs can be different. The Web Server acts as an object repository for downloading the SMIL Player applet to eliminate the need for installing the component. Thus, the Web Server shows emerging Object Web of the evolving client/server model [8].

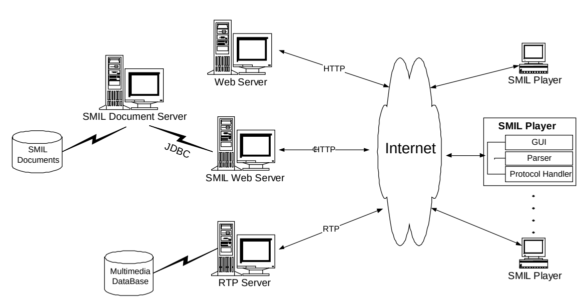

Figure 1. The system architecture

The player contains three major components. Via the GUI component, users are granted to retrieve SMIL presentations, managed by the SMIL Document Server connected to SMIL Web Server. The SAX (Simple Api for XML)-compliant Parser, an event-based parser, validates the specified SMIL document and retrieves the embedded elements and attributes. The Protocol Handler invokes the procedures for handling specific protocols such as RTP MPEG or HTTP/HTML. For the request of RTP MPEG streams, the client sends a subscription request to a mutlicast group for receiving contiguously transmitted demultiplexing RTP video/audio streams from RTP Server (see Figure 2). Raw video/audio data are then fed into buffers for efficiency consideration when playing. Depacketization and Video/Audio Buffer constitute Video/Audio Data Source that inherits from the class PullDataSource defined in JMF so JMF Player can treat video/audio raw data as coming from reliable data sources such as FILE or HTTP. Finally, the depacketized raw data in the buffer must be converted into bit streams for playing using JMF Player.

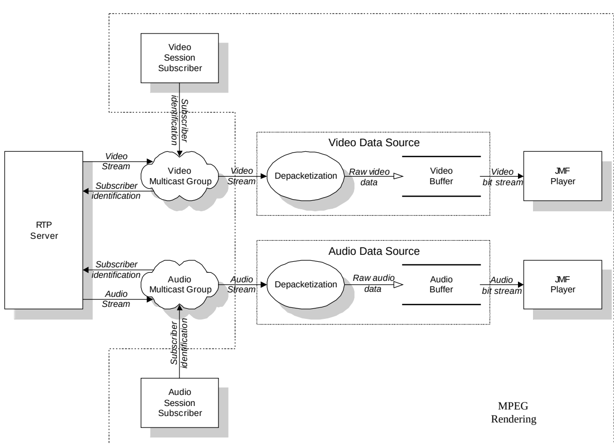

Figure 2. Rendering RTP-MPEG presentations

4. MODEL MULTMEDIA PRESENTATION USING MATN

The proposed system has been experimented on the National Kaohsiung First University of Science and Technology (NKFUST) campus network environment for supporting distance learning. We discuss the cases of navigation and RTP-MPEG streams in this section. Figure 3 shows the SMIL presentation that indicates the layout about the three synchronous media objects (one RTP-MPEG, one local HTML, and one remote HTML items). It shows the subtle syntax provided by SMIL when designing a presentation.

Figure 4 is the snapshot when the presentation is rendering. On the right-hand side of the panel, one can navigate the Web page as in a HTML browser when watching RTP-MPEG streams (as shown in left-top screen). The text window on the bottom provides the functionality of text caption. This greatly improves the interactivity of the presentation. Thus, the proposed system meets the requirements of spatiality, temporality, and interactivity for multimedia presentations.

```
<smil>
 <head>
  <layout>
   <root-layout
        width="640" height="420"   background-color="white"  />
    <region id="video_panel" left="10" top="20"
                   width="370" height="263"  />
    <region id="html_panel1" left="390" top="20"
                   width="240" height="390"  />
    <region id="html_panel2" left="10" top="300"
                   width="370" height="180"  />
  </layout>
 <head>
 <body>
  <par>
   <rtp   video="rtp://224.123.111.101:22224"
          audio="rtp://224.123.111.101:22225"
          region="video_panel"  />
   <html  src="http://www.cisco.com"  region="html_panel1"  />
   <html  src="file:d:/project/cisco.htm"   region="html_panel2"  />
  </par>
 </body>
</smil>
```

Figure 3. The experimental SMIL presentation

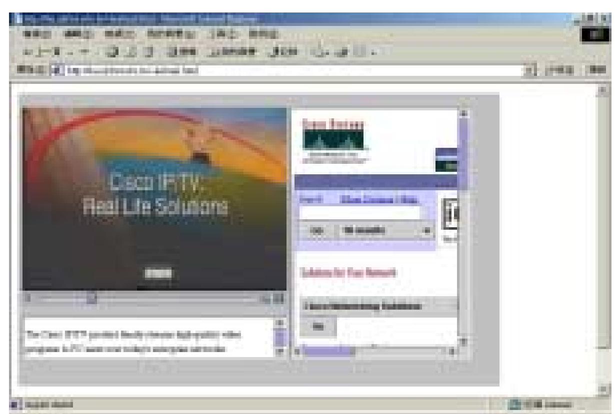

Figure 4. The rendering result of the SMIL presentation in Figure 3.

Regarding this example multimedia presentation written by SMIL, the MATN model can be applied to model its temporal relations and synchronous control. As shown in Figures 3 and 4, since there are 3 synchronous media objects, we denote them with symbols *V* (RTP-MPEG), *T* (local HTML) and *H* (remote HTML items). In this case, these media objects are displayed together. In fact, an MATN can be represented by a labeled directed graph called a transition graph. An arc represents an allowable transition from one state to another state.

Figure 5 gives an example to show how to use an MATN to model the temporal relations and synchronous control for this multimedia presentation case. As shown in Figure 5, the arc symbol ***V&T&H*** in arc number 1 represents that the three media objects will be displayed together on the layout screen. In fact, the ***&*** in the input string of MATN is corresponding to the SMIL construct <*par*> which schedules two media objects in parallel. In another word, the ***&*** symbol as defined in [3] denotes concurrent display. Considering another SMIL construct <*seq*>, though not used in this example presentation, can also be easily represented by the sequential transition in MATN model. In Figure 5, the start state is ***P/***, and after the transition arc ***V&T&H***, the control reaches a state node names ***P/V&T&H***, which is also the end state for this case. Furthermore, for each symbol in MATN that represents a media object, we can have a feature vector associated with it that includes all other information needed for presentation such as the spatial layout information, the source address, and the information of the duration time (that is corresponding to the time duration auxiliary attribute for synchronization control in SMIL).

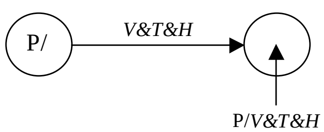

Figure 5. The MATN model for the SMIL multimedia presentation.

Through this example, we clearly show that an MATN can model the temporal and synchronous control of the multimedia presentations. Notice that, the control can go back to any state node it has reached before. For simplicity, this is not included in this example. It is clear to see that an MATN can model not only the sequential situations, but also the concurrent situations. At the current stage, each nonterminal symbol (such as ***V*** represents a RTP-MPEG video stream) in an MATN indicates a specific media object. In an MATN, these symbols can be further modeled by subnetworks that consist of either other multimedia presentation combinations or the semantic objects derived from the media streams. Furthermore, since any change in one of the subnetworks will automatically change any network that includes these subnetworks, we can deem the subnetwork as a well-defined package or a class that is easy to maintain and extend through the process of designing a complex multimedia presentation. One of the big advantages of MATN model is to enable the more complicated multimedia presentation system.

5. CONCLUSIONS AND FUTURE WORK

The work presented in this paper was motivated by the two needs in developing a modern multimedia presentation system. The first one is the need to develop the architecture-neutral SMIL players that can handle RTP streams. To meet such a demand, we propose a new architecture for playing SMIL RTP media by applying the Java-based distributed multimedia computing technology. To meet the reusability of software development, we integrated the component of RTP player developed in our previous work and SAX for parsing SMIL documents. Another increasing need is to find a systematic way to model the multimedia presentation using some conceptual semantic model. For this purpose, a discussion of how to use the Multimedia Augmented Transition Network (MATN) to model such a SMIL based multimedia presentation system is presented. Due to the effectiveness of the MATN model, the resulting system is at its initial stage but demonstrates an encouraging direction towards an efficient and powerful Java-based multimedia presentation system.

6. REFERENCES

[1] G. Blakowski, J. Huebel, and U. Langrehr, “Tools for Specifying and Executing Synchronized Multimedia Presentations,” *Proc. 2nd Int’l Workshop on Network and Operating System Support for Digital Audio and Video,* pp. 271-279, 1991.

[2] S. Boll, W. Klas, and U. Westermann, “Multimedia Document Models – Sealed Fate or Setting Out for New Shores?” *IEEE*, 1999.

[3] S-C. Chen and R. L. Kashyap, “Temporal and Spatial Semantic Models for Multimedia Presentations,” *1997 International Symposium on Multimedia Information Processing,* pp. 441-446, Dec. 11-13, 1997.

[4] S-C. Chen and R. L. Kashyap, “A Spatio-Temporal Semantic Model for Multimedia Database Systems and Multimedia Information Systems,” *IEEE Trans. on Knowledge and Data Engineering*, vol. 13, no. 4, pp. 607-622, July/August, 2001.

[5] “Extensible Markup Language (XML),” http://www.w3.org/XML.

[6] P. Hoschka (Ed.), “Synchronized Multimedia Integration Language,” World-Wide Web Consortium Recommendation, June 1998, http://www.w3.org/TR/REC-smil/.

[7] Java JMF homepage: http://java.sun.com/products/java-media/jmf/.

[8] S.-T. Li, T.-S. Li, and S.-W. Chou, “JTViewer: A Java JMF-based Viewer for Real-time TV-like Streams,” *International Workshop on Internet* in