

A SMIL-based Lesson Recording System for Distance Learning

Chun-Chuan Yang, Yi-Zheng Yang, Kwin-Yee Lin

Multimedia and Communications Laboratory
Department of Computer Science and Information Engineering
National Chi Nan University, Taiwan, R.O.C.
ccyang@csie.ncnu.edu.tw

Abstract—For Web-based distance learning, two types of lectures should be supported, live and on-demand lectures. In this paper, a SMIL-based recording system is proposed to support the preparation of the on-demand lessons. The lesson recording system records the lecturing materials, such as audio, video, HTML pages, and the annotation actions of the teacher, and converts them in the SMIL format. SMIL is selected as the internal format of the lessons because of its popularity. Implementation of the system is also presented in the paper.

Index terms—SMIL, Lesson recording, Distance learning

I. INTRODUCTION

With the popularity of Internet and WWW, people are getting used to acquire information from Internet. The trend is introducing new styles in every part of life that includes shopping, entertainment, education, etc. For the application of WWW in education [1-4], many online courses based on HTML pages have been offered in Internet. Integration of audio and video materials with HTML pages provides a new manner for distance lectures over WWW [4] such that the efficacy of the Web-based lectures is significantly improved.

In order to support the on-demand lectures for distance learning, a lesson recording system is required to record the information regarding the lecturing process, which includes the slides (HTML pages) and the explanation for each slide. The media data about the explanation by the teacher for each slide may include video, audio and the annotation (via drawing) on the slide. The lesson recording system records the materials for the lecture as well as the temporal relationship among the recorded materials so that the live lecturing process could be completely recorded. With the lesson recording system, the teacher does not have to spend extra time on composing the presentation of the lecture via handwriting tools, and the overhead of preparing the distance lessons is significantly reduced.

The proposed lesson recording system adopts *SMIL* (*Synchronized Multimedia Integration Language*) [5, 6] for recording the temporal relationship. The reason for the adoption of SMIL as the recording language is explained in the following. In order to make on-demand retrieval of the lectures conveniently, the recording language is better interpretable by popular browsers like *IE* or *Netscape*

Navigator so that the students could easily browse the lessons. From the popularity point of view, HTML seems to be the best candidate. However, the lack of the ability in integrating synchronized multimedia for HTML makes it improper to be the lecture recording language. SMIL was developed by the *WWW Consortium (W3C)* to address the lack of HTML for multimedia over WWW. It provides an easy way to compose multimedia presentations. With the efforts of W3C, SMIL is becoming the most popular language in authoring multimedia presentations and it is going to be supported by the newest versions of the browsers. We briefly introduce SMIL in the following.

SMIL could be used to describe both the spatial relationship and temporal relationship of a multimedia presentation. The spatial relationship is concerning with the visual layout of media objects in the presentation, while the temporal relationship is concerning with the timing control of media objects.

The synchronization elements in SMIL for temporal relationship include the `<seq>` element, the `<par>` element, and the class of media object elements such as ``, `<video>`, `<audio>` and `<text>`, etc. The `<seq>` element defines a sequence of elements in which elements play one after the other. The `<par>` element defines a simple parallel time grouping in which multiple elements can play back at the same time. The media object elements allow the inclusion of media objects into an SMIL presentation. Media objects are included by reference (using a *URL*).

The remainder of the paper is organized as follows. First of all, the overview of the system is presented in section II. The recording process for the system is presented in section III, and the converting process that converts the recorded information to SMIL is presented in section IV. In section V, we describe the implementation for the system. Finally, section VI concludes the paper.

II. SYSTEM OVERVIEW

The paradigm for distance learning adopted in the system is that the teacher uses slides for the lecture and explains the slides one after the other. The slides are composed of HTML pages. The media data of the explanation for each slide includes video and audio as well as the annotation data on the slide. More specifically, the teacher could add extra comments on each slide via the annotation tool, which is

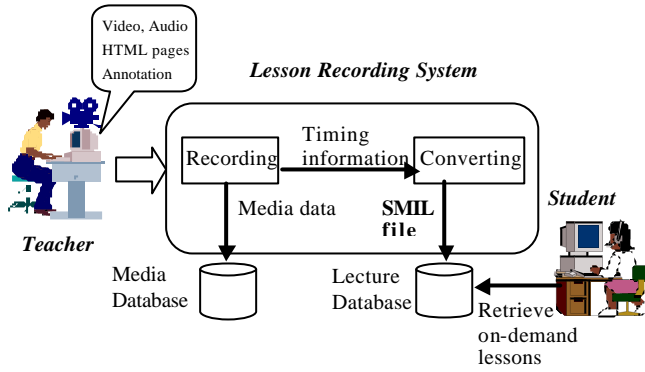


Figure 1. Overview of the Lesson Recording System

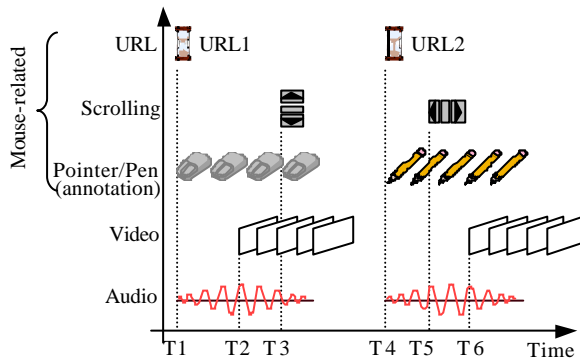


Figure 2. Media types used in the recording system

actually a drawing tool on the slide area. Thus, the visual layout for the proposed system includes two primary windows: the HTML window (for slide show and annotation) and the video window.

The main function of the lesson recording system is to record all the media data of the lecture in the database and to represent the temporal relationship among the media data in the SMIL format. There are two steps for lesson recording. First, the lecturing materials, which includes the HTML pages, video, audio, and annotations, are stored in the database, and the timing information for these materials are also recorded in a temporal format. The process is called the *recording process*. Second, the *converting process* is exercised to convert the timing information to the SMIL format. The overview of the system is illustrated in Figure 1. The recording process and the converting process are presented respectively in the following sections.

III. RECORDING PROCESS

The recording process records all the information of the lecture given by the teacher. There are two kinds of information that need to be recorded: *media data* and the *temporal information* [7, 8] about the media data. Media data is concerning about the content of each medium used in the lecturing process, which includes audio, video and some mouse-related events as illustrated in Figure 2. The mouse-related events include the loading of the HTML page, the scrolling action with the HTML window, and the

(a) Audio/Video

Format: (Start time, End time, URL)

E.g. (10s, 15s, rtsp://mcl1.csie.ncnu.edu.tw/audio1.wav)
(20s, 30s, rtsp://mcl2.csie.ncnu.edu.tw/video1.avi)

(b) Mouse-related events

URL event: (Loading time, URL)

Scrolling event: (Action time, Direction, Offset),
Direction = vertical/horizontal

Annotation event: (Action time, Type, Coordination, Color),
Type = pointer/pen, Coordination = <x, y>

E.g. URL event: (10s, http://mcl3.csie.ncnu.edu.tw/slide2.html)

Scrolling event: (12s, horizontal, 50)

Annotation event: (15s, pen, <100, 250>, blue)

Figure 3. Temporal information for each medium

annotation action on the HTML window. Moreover, there are two kinds of annotation actions defined in the system, mouse move (*pointer*) and the drawing action (*pen*).

When the recording process is started, the lesson recording system saves the media data into the media database, and at the same time it records the occurring time of the media data as its temporal information. Note that the temporal information is actually a relative time value to the start time (time base) of the lecture. The temporal information is temporarily saved in a file and converted to SMIL by the converting process. Since the characteristics of audio/video and mouse-related events are quite different, we discuss them respectively in the following sub-sections.

A. Audio/Video

Since the teacher may stop the video or audio and resume them later in the lecturing process, many video clips and audio segments may occur during the recording process. We have to record the presentation time for each video clip and audio segment. A video clip or an audio segment represents a media file in the media database. The temporal information for the video clip or audio segment is the playback time of the corresponding media file. Hence, the recording process records the playback duration for audio/video that includes the start time and end time of the media data. The format of the temporal information for audio/video is shown in Figure 3-(a).

B. Mouse-related events

There are three kinds of mouse-related events in the system as shown in Figure 2, which includes *URL events*, *scrolling events*, and *annotation events*. Since the amount of data for these events is little, we embed the media data of the events in the temporal information. An URL event occurs at the moment that the teacher loads a new HTML page by inputting a new URL, activating a hyperlink, or using the backward/forward function of the browser. The loading time and the URL address are both recorded for the URL event.

The scrolling event occurs at the moment that the teacher clicks the scroll bars of the HTML window. Since there are a vertical scroll bar and a horizontal scroll bar for the

```

<smil>
  <head>
    <layout>
      <root-layout height="600" width="800" />
      <region id="v" top="0" left="0"
        height="120" width="160" />
      <region id="h" top="0" left="170"
        height="500" width="600" />
    </layout>
  </head>
  <body>
    .....
  </body>
</smil>

```

Figure 4. The SMIL script for the layout template

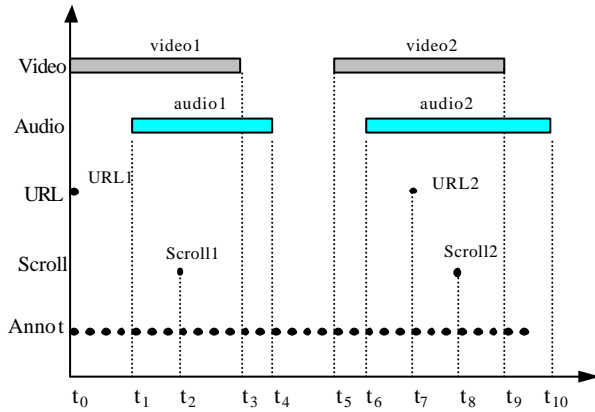


Figure 5. Example of temporal information

HTML window, the direction for the scrolling event must be also recorded. Hence, the recording process records the action time, the scrolling direction, and the scrolling offset for the scrolling event.

When the teacher moves his/her mouse pointer into the HTML window, the recording process continuously and periodically records (tracks) the state and the position of the mouse. The action of the mouse within the HTML window is defined as the annotation event. Two types of the annotation events are defined in the system: *<pointer>* and *<pen>*. *<Pointer>* represents the move of the mouse, while *<pen>* represents the drawing action on the HTML window. For the annotation event, the recording process records the action time of the event, type of the event, as well as the coordination of the mouse position. The format of the temporal information for the mouse-related events and some examples are shown in Figure 3-(b).

IV. CONVERTING ALGORITHM

The converting process analyzes all the temporal information obtained from the recording process and represents the information in the SMIL format. Since the visual layout for the lesson recording system is static as mentioned in section II, a layout template in the format of SMIL is adopted in the converting algorithm. The SMIL

```

<par>
  <video src="rtsp://video1.avi" begin="t0" />
  <video src="rtsp://video2.avi" begin="t5" />
  <audio src="rtsp://audio1.wav" begin="t1" />
  <audio src="rtsp://audio2.wav" begin="t6" />
  <text src="http://url1.html" begin="t0" />
  <text src="http://url2.html" begin="t7" />
  <scroll offset="x=10" begin="t2" />
  <scroll offset="y=15" begin="t8" />
  <annot src="rtsp://annot.data" begin="t0" />
</par>

```

Figure 6. The straightforward conversion for the example

script about the visual layout for the system is shown in Figure 4.

As mentioned in section III, media types in the temporal information include audio, video, the URL event, the scrolling event, and the annotation event. Since the current SMIL specification (version 1.0) [12] does not support the scrolling event and the annotation event, we extend two new tags, *<scroll>* and *<annot>*, to SMIL for the converting process.

The temporal information consists of a set of media objects with timestamp. An example for the temporal information, which will be used as the sample input for the converting process, is shown in Figure 5. The most straightforward way to convert the temporal information is to treat all media objects as the children of a *<par>* element, and the value of the "begin" attribute for each object is assigned to the playback time of the object. The result of the straightforward conversion for the example is shown in Figure 6. Note that the annotation event occurs periodically in the temporal information, so it is better treat the annotation event as a continuous medium and collect them in a media file such that the streaming technique could be applied.

The straightforward conversion is simple but introduces more overhead to the browser while presenting the SMIL file. The reason is that the straightforward conversion makes all media objects the children of a *<par>* element, and the *<par>* element, by the definition, requires the browser to deal with all its children concurrently. Hence, more processing overhead and more buffers are required for browsing the resulted SMIL file of the straightforward conversion.

Therefore, from the processing point of view of the browser, more sequential parts in the resulted SMIL file make the browsing more efficient. A revised converting algorithm is proposed to increase the percentage of the sequential part in the SMIL file. By observing the sequential property of the objects of the same medium type, the revised algorithm groups all the objects of a medium as the children of a *<seq>* element. In other words, the sequence of the objects for each medium type forms a *<seq>* element. Furthermore, the *<seq>* elements of all medium types form a *<par>* element. The result of the revised algorithm for the example in Figure 5 is shown in

```

<par>
  <seq>
    <video src=".../video1.avi" begin="t0" />
    <video src=".../video2.avi" begin="t5-t3" />
  </seq>
  <seq>
    <audio src=".../audio1.wav" begin="t1" />
    <audio src=".../audio2.wav" begin="t6-t4" />
  </seq>
  <seq>
    <text src=".../url1.html" begin="t0" />
    <text src=".../url2.html" begin="t7-t0" />
  </seq>
  <seq>
    <scroll offset="x=10" begin="t2" />
    <scroll offset="y=15" begin="t8-t2" />
  </seq>
  <annot src=".../annot.data" begin="t0" />
</par>

```

Figure 7. Result of the revised algorithm for the example

Figure 7. Because the time base of a child element in `<seq>` is the end of the previous element, the value of the “begin” attribute of a child element in `<seq>` is the child’s playback time minus the end time of its previous element. For instance, the “begin” attribute of video2 is its playback time (t_5) minus the end time (t_3) of the previous element (video1), i.e., t_5-t_3 as shown in Figure 7.

V. IMPLEMENTATION

A prototype of the lesson recording system has been implemented. A snapshot of the implementation is shown in Figure 8. The sampling rate for the annotation events within the HTML window is set to 20 times per second. In order to differentiate the `<pointer>` type and the `<pen>` type of the annotation events, we define the mouse action with the `<ctrl>` key pressed within the HTML window as the `<pen>` type of the annotation event. We also define some function keys to allow the teacher to change the color of the `<pen>` for annotation.

The previewing function for the SMIL file is also included in the prototype system since two new tags `<scroll>` and `<annot>` are extended to the SMIL specification 1.0. The interpretation of the new tags for commercial browsers requires a plug-in installed in the browser.

VI. CONCLUSION

In this paper, a SMIL-based lesson recording system is proposed to support the on-demand lectures for web-based multimedia distance learning. The lesson recording system records the live lecture given by the teacher and converts the materials of the lecture to SMIL.

The lesson recording system makes the preparation of the on-demand lectures more efficiently since the teacher does not have to spend extra time on composing the presentation of the lecture via some handwriting tools. Moreover, since

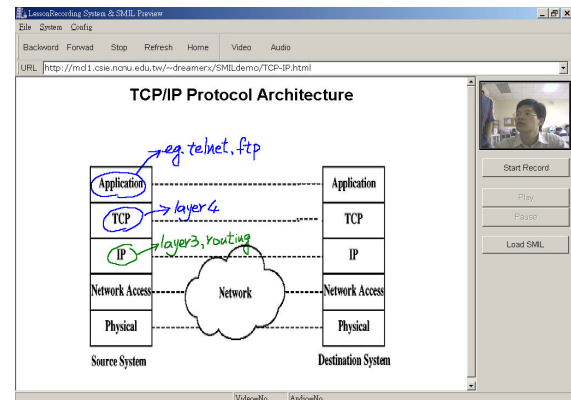


Figure 8. A snapshot of the lesson recording system

SMIL is becoming popular in authoring multimedia presentation and will be supported in the newest version of commercial browsers, the SMIL-based on-demand lectures enable the students to retrieve the lessons conveniently.

References

- [1] Wei-Hsiu Ma, Yen-Jen Lee, David H.C. Du, and Mark P. McCahill, “Video-Based Hypermedia for Education-on-Demand,” IEEE Multimedia, vol. 5, pp. 72-83, Jan.-Mar., 1998.
- [2] M. P. Hui and S. L. Hung, “Using distributed hypermedia educational systems for distance education,” Proc. IEEE International Conference on Multimedia Engineering Education, pp. 135-140, 1996.
- [3] W. Doube, “A browser-based system to support and deliver DE [distance education],” Proc. 28th Annual Frontiers in Education Conference (FIE '98), Vol. 1, pp. 479-484, 1998.
- [4] H.Y. Chen and G.Y. Chen, “Design of a Web-based Synchronized Multimedia Lecture System,” Proc. IEEE International Conference on Multimedia Computing and Systems, Vol. 2, pp. 887-891, 1999.
- [5] W3C, “Synchronized Multimedia Integration Language (SMIL) 1.0 Specification,” W3C Recommendation June-1998; <http://www.w3.org/TR/REC-smil>.
- [6] W3C, “Synchronized Multimedia Integration Language (SMIL) Boston Specification,” W3C working Draft 20-August-1999; <http://www.w3.org/TR/smil-boston>
- [7] F. Cruz and P.S. Mahalley, “Temporal Synchronization in Multimedia Presentations,” Proc. IEEE International Conference on Multimedia Computing and Systems, pp. 851-856, 1999.
- [8] Z. Yang, et al., “A New Look at Multimedia Synchronization in Distributed Environments”, Proceedings., 4th International Symposium on Parallel Architectures, Algorithms, and Networks, (I-SPAN '99), pp. 322–327, 1999