Multimedia Representation for WWW

Andrzej Duda

IMAG-LSR,
BP 53, 38041 Grenoble Cedex 9, France

Keywords: *Building Web-based multimedia presentations; Time representation and media synchronization*

1 Introduction

Current WWW hypermedia technology allows composition of various media data such as text, image or audio/video coming from possibly distributed sources. It is possible to specify the logical structure and the layout of different types of media composing a hypermedia document. Currently, WWW integrates audio and video data within hypermedia documents as links to remotely stored objects. When the user follows a link, the browser prefetches the object before playing it. However, the WWW technology does not take into account temporal and synchronization aspects of multimedia presentations. For instance, it is not possible to define a temporal scenario for presenting different hypermedia documents.

Other existing audio/video technologies such as the Internet telephony or videophones consist of simple streaming of digitized data between a server and a client. The conceptual model of the technologies is relatively simple—a stream of data feeds the user browser or, which is the case of videophones, several streams are lip-sync synchronized. Composing complex temporal presentations is much more difficult. In order to extend WWW to what we would call an Information Superhighway, we need a new multimedia representation for temporal composition and synchronization.

2 Interval Expressions

We are interested in a new class of multimedia presentations for the WWW that include rich dynamic interactive scenarios. Such scenarios integrate dynamic user control over the flow of presentation, advanced processing of media content, and diverse sources of media streams such as live feeds or teleconferencing streams.

We have defined a model called *Interval Expressions* for structured temporal composition of interactive dynamic multimedia presentations [1, 2]. The model extends *Algebraic Video* [3] that used a set of basic operations on video segments to create a desired video stream. The elementary entities of our model are *media segments* that have inherent temporal behavior. They may be traditional multimedia objects containing video, audio, animations, images or text as well as new special objects such as live media streams, executable code or links to other presentations. For temporal composition, media segments are viewed as *time intervals* independently of their content or other characteristics. Media segments may have fixed durations, however, in some cases, their duration is unknown. For example, a segment may represent a teleconferencing stream or a user interaction may stop playback for some period of time.

Our model defines operators that express causal relations between intervals. Operators take time intervals as arguments and yield another interval as a result. They can be used to form nested interval expressions allowing specification of temporal compositions in a well-structured way. *Interval Expressions* provide a means of encapsulation and structuring: compound encapsulated intervals can be specified in terms of elementary media objects as building blocks. The functional approach to temporal composition allows integration of different media segments by specifying their temporal relations independently of their respective durations.

3 Existing models

Existing models of multimedia presentations allow static composition of different media having temporal behavior such as audio and video. Temporal composition defines synchronization among media segments according to some static temporal scenario. Usually, a presentation is played back linearly with simple VCR controls.

Temporal models for multimedia can be divided into two classes: *point-based* and *interval-based.* In point-based models, the elementary units are points in a time space. Each event in the model has its associated time point. The time points arranged according to some relations such as *precede, simultaneous* or *after* form complex multimedia presentations. An example of the point-based approach is *timeline*, in which media objects are placed on several time axes called *tracks*, one per each media type. All events such as the beginning or the end of a segment are totally ordered on the timeline. The timeline model is applied in HyTime. The model is well suited for temporal composition of media segments of known durations, however it falls short for unknown durations.

Interval-based models consider elementary media entities as time intervals ordered according to some relations. Existing models are mainly based on the relations defined by Allen for expressing the knowledge about time. Giving any two time intervals, they can be arranged according to seven relations : *before, meets, overlaps, finishes, during, starts, equals.* A relation expresses the inequalities between the end points of intervals. If the Allen relations are used for multimedia composition, then the duration of intervals may vary only within the limits defined by the inequalities of a given relation. This drawback makes the Allen relations not suitable for specifying composition of intervals with unknown duration.

Many existing representations such as OCPN are conceptually equivalent to the Allen relations. OCPN does not take into account possible unknown durations of intervals and to prepare an instantiated presentation (a presentation in which all interval end points are determined), the tree of interval relations must be traversed to determine deadlines used for scheduling of the presentation. If interval durations are modified, the schedule must be modified too.

MHEG defines an interchange format based on media objects. The objects are synchronized by means of two basic operators: *serial* and *parallel.* More complex temporal composition can be achieved by programming *link objects* that detect some events and trigger actions on other objects.

4 WWW multimedia documents

A WWW multimedia document specified using our model is composed of basic media segments that include executable code, live media streams or standard audio/video data. These media segment types are integrated in a seamless way within our temporal composition model. For example, some parts of a multimedia presentation can be composed of code that analyzes the content of media segments, processes on-line multimedia streams or evaluates some queries on a collection of media segments. In that way, the outcome of the execution can determine the flow of a presentation. This content-sensitivity becomes an important new characteristic of media segments—a segment can be made sensitive to its content and the flow of a presentation can change according to the detection of some content features. Executable code also allows taking into account user interactions. The user can control the flow of presentation, choose different paths, and even jump from one presentation to another. Live feeds become other interesting components of multimedia presentations. With the advent of television broadcasting coming into a computer, multimedia presentations may take advantage of rich sources of streams coming from the live feeds such as a satellite dish or a teleconferencing network connection.

WWW already provides support for down loadable code—Java applets. In our model, applets acquire temporal dimension. They are considered as having some inherent duration that is taken into account in global temporal composition of a document. A WWW multimedia document contains different media segments and their temporal composition is defined by an interval expression. The interval expression can be easily incorporated into HTML using a new tag or attributes of existing tags.

Our model enables multimedia authoring as a process of specifying pieces of a puzzle in which the pieces come from a large collection of media segments. The author looks for interesting media segments or parts of existing presentations and combines them together to form a new presentation. This authoring method takes advantage of the support for structuring and hierarchical composition—it is possible to define complex multimedia presentations in a well-structured manner by means of encapsulation and nesting of structured entities.

References

1. A. Duda and C. Keramane. Structured temporal composition of multimedia data. In *Proc. International Workshop on Multi-media Database Management Systems*, Blue Mountain Lake, NY, August 1995.
2. C. Keramane and A. Duda. Interval expressions - a functional model for interactive dynamic multimedia presentations. In *Proc. IEEE Int. Conference on Multimedia Computing and Systems*, Hiroshima, June 1996.
3. R. Weiss, A. Duda, and D.K. Gifford. Composition and search with a Video Algebra. *IEEE Multimedia*, 2(1), 1995.