A Multimedia Synchronization Model Described by Boolean Expressions

Tatsuo SATO, Joung-hoon LIM, Kenichi OKADA, Yutaka MATSUSHITA
Faculty of Science and Technology
Department of Instrumentation Engineering
Keio University
3-14-1 Hiyoshi, Kohoku-ku, Yokohama 223, Japan

Abstract

In this paper, a multimedia synchronization model is proposed. The purpose of this model is description for temporal relationships(i.e. time ordering and synchronization for multimedia presentation) among various media. In the model, user interactions, such as conditional branching and menu selection, are also considered. For the description of the temporal relationships, Boolean Expressions are used. The concept of *media objects* would be a great help to manipulate multimedia data. The functions of the media objects are defined in this paper. A media object is an entity that contains associated data and a media variable representing its presentation state. Various types of media are classified into two types: OUTPUT and INPUT media objects. OUTPUT media objects have data perceived by the user such as graphics, image, drawings, animation, audio and video. INPUT media objects such as buttons and menus can handle input from the user.

Temporal relationships among media objects can be represented by a set of *presentation expressions* consisting of a boolean expression; three logical operators(*and*, *or*, and *not*) and *media variables*. The value of the media variable indicates the state of media(e.g., in the case of video object, it has two states; *playing* and *not playing*.) Furthermore, by describing temporal relationships between input media objects and output media objects, user interaction can be successfully described in our model.

A multimedia animal guide system, prototype system base on the model, was implemented. The system can synchronously present the media such as video, audio, image, and text, and can explain the feature and characteristics of animals, and the sequence of the presentation can be controlled by the user.

1 Introduction

The increasing power of computer processors makes possible the use of new representation media such as graphics, still pictures, sound sequences and even audio-visual sequences. While conventional paper-based applications consist of text, pictures, and drawings, multimedia applications consist of various *media*. A medium denotes a type of information such as text, graphics, image, animation, audio, and video.

For the management and processing of such multimedia information, data of different media types must be combined for use in presentation to the user. The process of combining data in this manner is commonly known as media composition. Composition generally takes spatial and temporal forms[1].

Spatial composition is concerned with the combination of media components in space, such as images overlay or text with image. The techniques in this type of composition must consider size, rotation and placement of participating components. There has been an effort to standardize the aspects of data integration with respect to spatial types. The Office Document Architecture(ODA)[2] is one such standardization whose efforts have resulted in both logical and layout representations for multimedia documents. These representations develop a hierarchy of objects with object-oriented paradigm. A similar methodology is defined for exchange of temporal relationships occurring within documents, including parallel and sequential relationships [4].

For temporal composition, a time ordering assigned to the presentation of the elements of the multimedia object is required. Consider a multimedia slide presentation in which a series of verbal annotations, or audio, must accompany a set of images presented. The presentation of the annotations is sequential, as it is for the images. Points of synchronization correspond to the change of image segments and the end of a verbal annotation. Figure 1 shows an example of rather coarse synchronization among objects. A multimedia system must preserve the timing relationships among the elements of the object presentation at these points of synchronization by the process of temporal composition.

On the other hand, The component consists of a sequence of frames. During the display of one video frame, the corresponding sequence of audio samples does not comprise a logical information unit for which we can apply synchronization points. Some other form of coordination, therefore, is necessary to satisfy a tight synchronization requirement (e.g., lip synchronization[6]) between two streams during playback. Specific synchronization points among these types of multimedia objects do not exist except at the beginning and the end of the data sequences; rather, an ongoing form of synchronization is necessary.

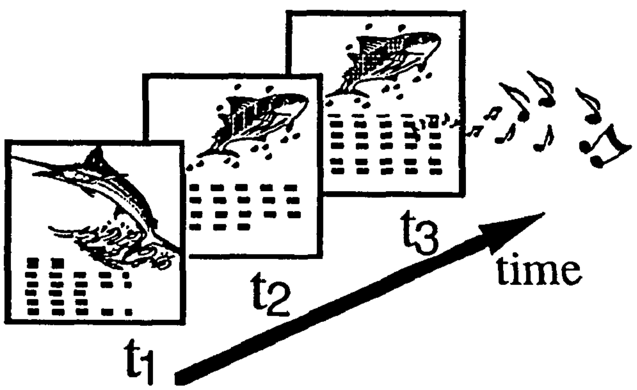

Figure 1: Time Dependent Document

Temporal relationships among the media may be implied, as in the simultaneous acquisition of voice and video, or may be explicitly formulated, as in the case of a multimedia document including voice annotated text. In either situation, in order to schedule properly the synchronization of media with vastly different presentation requirements, the characteristics of media and relationships among them must be formally described and manipulated.

Although the problem of multimedia synchronization has been addressed by various applications, there is no general expression or notation for describing synchronization relationships among units of multimedia information. To present, there have been some efforts to standardize these aspects, but their synchronization model can not specify user interactions. For the purpose of integration of multimedia synchronization and user interaction, we propose a multimedia synchronization model in which user interaction can be successfully specified and various types of inter-media synchronization can be generalized.

Furthermore, we implemented a prototype of multimedia animal guide system based on our model. It can control the presentation of various media(video, audio, image, and text) to explain the feature and characteristics of animals, and the user can also control the sequence of the presentation.

This paper is organized as follows; in section 2 the concept of media objects is described, sections 3 and 4 give the definition and semantics of presentation expressions and its presentation algorithm, and section 5 describes the description of inter-media synchronization. Finally in section 6, the prototype of an animal guide system implemented is explained.

2 Concept of Media Objects

2.1 Classification of Media Objects

A media object is an entity that contains associated data and a media variable representing its presentation state. Multimedia information is composed of combination of media objects, therefore, an object is also used as an unit of multimedia information; the word "media" simply means type of media object.

Figure 2 shows the classification of media objects. All media objects are classified into the following three types:

1. OUTPUT media object

   An OUTPUT media object is one which has data perceived by the user such as audio, video, still images and graphics. OUTPUT media objects are classified according to their associated data. OUTPUT media objects that have *time dependent media* (TDM for short) are called TDM objects, while those that have *time independent media* (TIM for short) are called TIM objects. TDM, such as video and audio, can be expressed as function of time and takes into account the characteristics of continuous data streams, while text, still picture and graphics are called TIM in this sense.

2. INPUT media object

   An INPUT media object is one which can handle input from a pointing device such as a mouse, pen, or touch sensor. By introducing INPUT media objects and describing temporal relationships among INPUT and OUTPUT media objects, user interaction can be successfully described in our model.

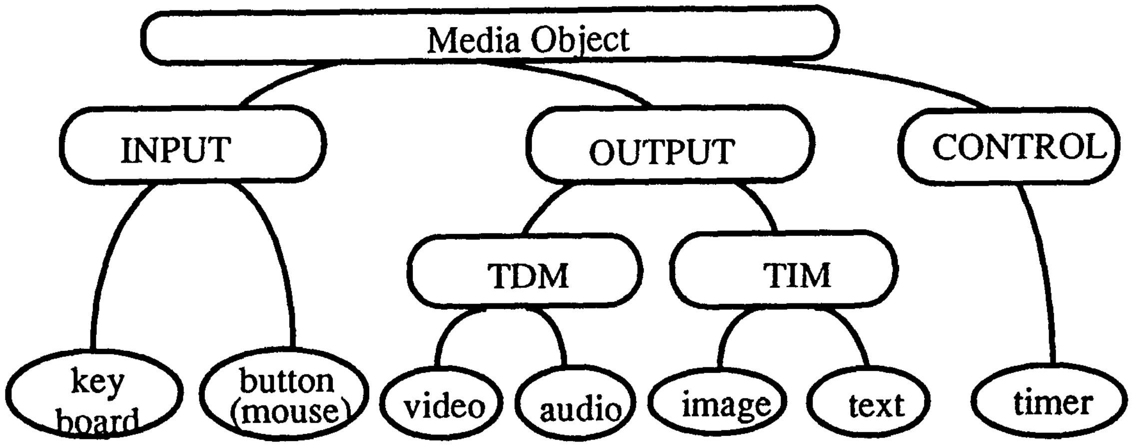

Figure 2: Classification of media objects

3. CONTROL media object

   A CONTROL media object is introduced in order to specify temporal attributes. The CONTROL media objects have no data to present to the user while the OUTPUT and INPUT media objects have data perceived by the user. They all have a timer which can indicate the presentation period to the TIM objects. An attribute 'duration', the initial value of the timer, is used in a certain multimedia application which the TIM objects sequentially presented such as an automatic slide show application (see 5.5).

2.2 Definition of a Media Variable

Each media objects has a *media variable* which indicates its internal presentation state and takes a boolean value (i.e. 0 or 1). The meaning of the value is different according to the type of media objects:

1. In the case of OUTPUT media objects,

   Its media variable indicates whether there remains the allowable presentation period or not. If an audio object has 3-second-long audio sequence, for instance, the allowable presentation period in this object indicates 3 seconds. At the beginning of presentation, the value of all media variables are initialized to 1. When the audio object is completely finished, the media variable changes to 0.

2. In the case of INPUT media objects,

   Its media variable indicates whether it is selected or not by the user. The variable is initialized to 1 and if the INPUT media object is selected by the user, the value changes from 1 to 0. Once the value becomes 0, two cases exit according to the type of INPUT media objects. First, the value never changes and second the value changes from 0 to 1 when the object is selected again.

3. In the case of CONTROL media objects,

   Its media variable indicates whether its own timer expires or not. As mentioned before, a CONTROL media object has a timer and an attribute 'duration' which indicates initial value of the timer. As for CONTROL media objects, 'presentation' means counting down its own timer from the initial value indicated by 'duration'. As long as the timer does not expire, the value of its media variable remains 1. When it expires, the value changes from 1 to 0.

3 Semantics of Presentation Expressions

Temporal relationships among media objects are described by a set of logical expressions. Our multimedia presentation model is based on these expressions. We call them *presentation expressions*. The presentation expression consists of logic operators(*and*, *or*, *not*) and some logical variables called media variables as mentioned before. Each media variables takes boolean value; the presentation expression also yields boolean value which takes 1 or 0.

In the case of INPUT media, the presentation variable indicates whether or not INPUT media object can be handled through the input devices.

In the case of OUTPUT and CONTROL media, when the presentation value changes from 0 to 1, the media object can be presented unless the media value is 0.

If the presentation value turns into 0, the presentation is suspended. For example, in the case of a video sequence, if the media value changes from 1 to 0 and the presentation value remains 1, the presentation of the video object cannot be continued. At this moment, the video object may show the last frame image of the video sequence until the presentation value has ones.

4 Presentation Algorithm

The algorithm for the presentation is simple. Assume that n is the number of media objects included in the multimedia presentation, and that the i-th media object has a media value m_i and presentation value of i-th presentation expression is denoted by p_i. The logical expression L_i generally takes n media variables $m_1, \ldots, m_n$ as arguments;

$$p_i = L_i(m_1, \ldots, m_n)$$

Presentation algorithm is shown as follows:

1) Initialize m_i into 1 and initialize p_i into 0 for all i.

2) Evaluate the logical expression L_i for all i, and then set the result of the evaluation to the presentation variable p_i. If p_i changes from 0 to 1, start or resume presenting the i-th media object. If p_i changes from 1 to 0, suspend the presentation of i-th media object.

3) During the presentation, if any changes of the value m_i for all i occur because of finishing the presentation of OUTPUT media objects or by selecting INPUT media objects, then go to 2).

4) go to 3).

5 Description of Basic Intermedia Synchronization

Three types of intermedia synchronization exist, in general, that are *sequential type, simultaneous type* and *independent type*[7]. In this paper, we discuss only about sequential and simultaneous types; both of them are significant to describe explicitly, because independence means that no relationships exist and no synchronization occur between two or more media. Any temporal relationships can be expressed by the combination of sequential and simultaneous type relationships. Sequential relationship is also called *"series"* in [5], and synchronous one is called *"parallel"* in [1]. If the relationship between media objects is sequential type, the presentation of each media objects occurs after another. On the other hand, if the relationship is simultaneous type, the presentation starts together at the same time.

The following examples show how to describe the basic intermedia synchronization using presentation expressions and media variables.

5.1 Simultaneous Relationships between TDM

Inter-TDM synchronization is explained. Let us consider an example of simultaneous relationships between the audio and video streams. In this case, each presentation of two media starts together at the same time. Note that there are two case about how to complete the presentation; (1)first, they complete the presentation simultaneously, that is, the presentation of both two media finish simultaneously when the presentation of one media completes even if the other is in progress. (2)Second, on the contrary, even if the presentation of one media complete, the other cannot be interrupted.

Assume that the media value of the video and the audio are m_{video} and m_{audio} respectively. In the case of (1), the relationship can be described as follows;

$$p_{video} = m_{video} \text{ and } m_{audio}$$
$$p_{audio} = m_{video} \text{ and } m_{audio}$$

At the first time, both m_{video} and m_{audio} are initialized as $m_{video} = 1$ and $m_{audio} = 1$. Each logical expression (that is, presentation expression) is evaluated, and as a result, $p_{video} = 1$ and $p_{audio} = 1$. Therefore, the presentation of each media starts. When the presentation of audio complete, the value of m_{audio} changes into 0. This change is detected, and each presentation expression is evaluated again, then $p_{video} = 0$ and $p_{audio} = 0$. Thus, the presentation of these media simultaneously finish.

In the case of (2), the relationship can be described as follows;

$$p_{video} = m_{video} \text{ or } m_{audio}$$
$$p_{audio} = m_{video} \text{ or } m_{audio}$$

Both p_{video} and p_{audio} change into 0, only when $m_{video} = 0$ and $m_{audio} = 0$. This causes each presentation to finish independently.

5.2 Simultaneous Relationships between TDM and TIM

Consider, for example, the case that the image media object as TIM is displayed on the screen during same presentation period of TDM such a audio media object. The media value of image media object denoted by m_{image} is always 1. This type of synchronization is described as follows;

$$p_{image} = m_{video}$$
$$p_{video} = m_{video}$$

Both p_{image} and p_{video} change into 0 when m_v changes into 0. So each media object has the same presentation period.

5.3 Sequential Relationships between TDM

Let us consider that there are sequential relationships between two TDM objects. The presentation of one media object starts after the end of the other one. In the case of the sequential relationships between two video sequences, for example, this type of synchronization is described as follows;

$$p_{video1} = m_{video1}$$
$$p_{video2} = (\text{ not } m_{video1}) \text{ and } m_{video2}$$

At the first time, both m_{video1} and m_{video2} are initialized as $m_{video1} = 1$ and $m_{video2} = 1$. Each presentation expression is evaluated, and as a result, $p_{video1} = 1$ and $p_{video2} = 0$. Therefore, the presentation of video1 begins to start, but the presentation of video 2 is not invoked. When the presentation of video1 complete, the value of m_{video1} changes into 0, but m_{video2} remains 1. This change is detected, and each presentation expression is evaluated. Then the value of p_{video2} changes from 0 to 1. These events occur sequentially.

5.4 Temporal Relationships with User Interaction

In order to integrate the user interaction to the model, we introduce INPUT media. There are various types of synchronization between INPUT media and any other media. One example of such synchronization is shown as follows.

When the user select the button displayed on the screen, in other words, clicks it with mouse, the media value of the button media object denoted by m_{button} changes from 1 to 0. The following expressions show the case that the presentation of video begins when the button is clicked.

$$p_{video} = (\text{ not } m_{button}) \text{ and } m_{video}$$
$$p_{button} = m_{button}$$

At the first time, both m_{video} and m_{button} are initialized as $m_{video} = 1$ and $m_{button} = 1$. Each presentation expression is evaluated, and as a result, $p_{video} = 0$ and $p_{button} = 1$. Therefore, the presentation of video media does not start. When the user clicks the button, the value of m_{button} changes into 0. This change is detected, and each presentation expression is evaluated again. The presentation value of p_{video} changes from 0 to 1, and the value of p_{button} changes from 1 to 0. Thus, the presentation of video begins, and it becomes impossible for the user to select the button.

Furthermore, in order to implement a *menu selection* — a user can select several items displayed as buttons which the user clicks with the mouse, e.g., if 3 items in the menu, 3 buttons required — we use the following expressions;

$$p_{video,1} = (\text{ not } m_{button,1}) \text{ and } m_{video,1}$$
$$p_{video,2} = (\text{ not } m_{button,2}) \text{ and } m_{video,2}$$
$$\vdots$$
$$p_{video,n} = (\text{ not } m_{button,n}) \text{ and } m_{video,n}$$
$$p_{button,1} = m_{button,1} \text{ and } \ldots \text{ and } m_{button,n}$$
$$p_{button,2} = m_{button,1} \text{ and } \ldots \text{ and } m_{button,n}$$
$$\vdots$$
$$p_{button,n} = m_{button,1} \text{ and } \ldots \text{ and } m_{button,n}$$

When the user clicks the i-th button$(1 \leq i \leq n)$, the value of $m_{button,i}$ changes into 0. Therefore, the presentation value $p_{video,i}$ changes from 0 to 1, and all of $p_{button,k}(1 \leq k \leq n)$ changes from 1 to 0. For the above reasons, the presentation of i-th video begins, and all of buttons are disappeared, thus user can not select any buttons.

5.5 An Example of the Description of Temporal Relationships

Let us consider a slide show as shown in Figure 3. First, during D1, an image of a lady(image 1) and a text of its explanation (text 1) are displayed. Next, the figure including a picture of a man wearing a hat (image 2), its voice annotation (audio1), and explanation(text 2) is presented during the period of the voice annotation (D2). Finally, the figure of a dancing man (image 3) with voice annotation(audio 2) and its explanation(text 3) is presented during the presentation period of audio2(D3).

The temporal relationships of this slide show are represented on time/media space as shown in Figure 4. In the presentation of the lady, no media object indicating the presentation period exists. Thus, a control object which has arbitrary duration is required.

The temporal relationships can be described by the set of presentation expressions.

First, three presentation expressions:

$$p_{control} = m_{control}$$
$$p_{image1} = m_{control}$$
$$p_{text1} = m_{control}$$

indicate that, the values of $p_{control}$, p_{image1} and p_{text1} are set to the value of $m_{control}$ (the media variable of the control object). This means that image1 and text1 are

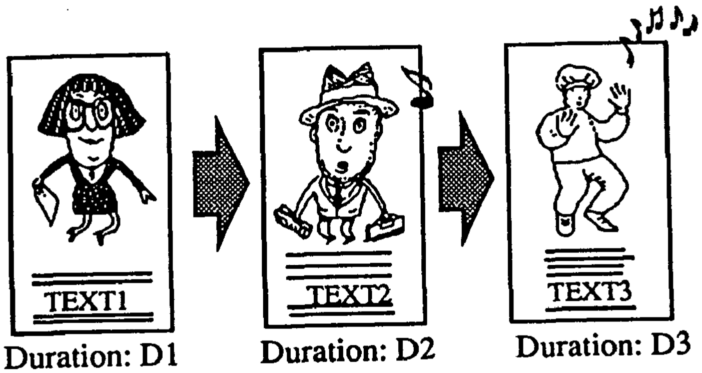

Figure 3: Example of Slide Show

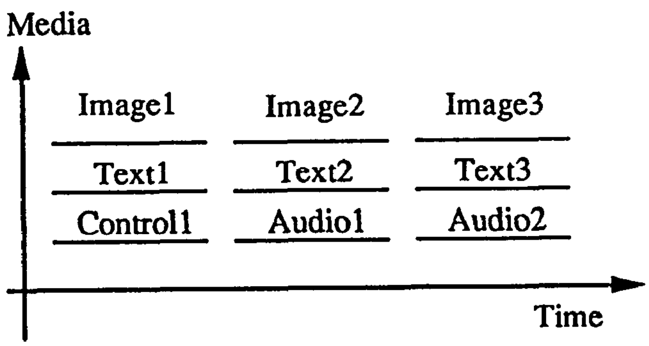

Figure 4: The Temporal Relationships of The Slide Show

presented simultaneously until the timer of the control media object expires.

Next, by three presentation expressions:

$$p_{audio1} = (not\ m_{control})\ and\ m_{audio1}$$
$$p_{image2} = (not\ m_{control})\ and\ m_{audio1}$$
$$p_{text2} = (not\ m_{control})\ and\ m_{audio1},$$

it is described that the values of $p_{text2}, p_{image2}, p_{audio2}$ are set to the result of (not $m_{control}$) and m_{audio1}. Therefore, image2, audio1 and text2 are presented simultaneously from the time when the timer of control1 expires to the time the presentation of audio1 finishes.

Finally, by :

$$p_{audio2} = (not\ m_{audio1})\ and\ m_{audio2}$$
$$p_{image3} = (not\ m_{audio1})\ and\ m_{audio2}$$
$$p_{text3} = (not\ m_{audio1})\ and\ m_{audio2}$$

it is described that, as the values of p_{text3}, p_{image3}, and p_{audio2} are set to the result of (not m_{audio1}) and m_{audio2}. Because m_{audio2} remains 1 until the presentation of audio1 finishes, the result of ((not m_{audio1}) and m_{audio2}) changes from 0 to 1 when m_{audio1} changes from 1 to 0, that is, when the presentation of audio1 is finished. Thus, audio2, image 3, and text3 are simultaneously presented from the time the presentation of audio1 completes to the time audio2 presentation finishes.

6 An Animal Guide System Using Our Model

Multimedia animal guide system based on our model was implemented in which video, audio, image and text can be presented synchronously in order to explain animal life. The user can control the sequence of the presentation.

First, author creates a scenario of what he wants to present. The information presented to the user is created as follows.

1. How to combine each presentation of a media object is determined. At this stage, it is important to select appropriate kinds of media objects.

2. The temporal relationships among them are described according to the scenario. In current system, auto generation of a set of presentation expressions which represents the temporal relationships is not supported; presentation expressions are created manually by the author.

3. The described presentation expressions are verified. The system provides a tool by which the temporal relationships described by the presentation expressions can be represented graphically. By the use of this tool, the author can easily make sure whether or not the set of presentation expressions properly represent the temporal relationships. At this time, if some description errors are found, the author can rewrite the presentation expressions and they are verified again.

4. Data included in the media objects are created. For example, if there exist an audio object, an image object and a text object, corresponding data is created by an audio editor, an image editor and a text editor, respectively.

5. According to the presentation expressions, PM(Presentation Manager),who has the responsibility of presenting each media object, starts presentation of the scenario.

This prototype was implemented on a SUN Sparc Station with a full-color video board that can display real time video image from a laser disk player which can be controlled through RS-232C.

7 Conclusion

In this paper, a multimedia synchronization model was proposed in which multimedia synchronization was formally described by a set of logical expressions.

In order to describe various types of inter-media synchronization, a concept of media object and temporal relationships description method were introduced. Various types of media were classified into two types of media objects: OUTPUT media objects and INPUT media objects. It is shown how temporal relationships among media objects can be represented by a set of presentation expressions consisting of a logical expression including logical operators(and, or, not) and media variables. Furthermore user interaction can be represented by describing temporal relationships between input media objects and output media objects. Multimedia animal guide system based on our model was implemented in which video, audio, image, and text can be presented synchronously to explain animals and the user can control the sequence of the presentation.

References

[1] T. D. C. Little, A. Ghafoor, Synchronization and Storage Models for Multimedia Objects, IEEE Journal on Areas in Communications, Vol. 8, No. 3, April 1990

[2] International Organization for Standardization, ISO Document No. 8613, ISO, Geneva, Switzerland, Mar. 1988.

[3] R. Steinmets, Synchronization Properties in Multimedia Systems, IEEE Journal on Areas in Communications, Vol. 8, No. 3, April 1990

[4] T. Yoneda, Y. Matsushita, A New Communication Tool: Time Dependent Multimedia Document, IEEE Phoenix 92 to be published.

[5] G. Pancaccini, F. Stajano, Media Composition and Synchronization Aspects in an Interactive Multimedia Authoring Environment, Human Aspects in Computing: Design and Use of Interactive Systems and Work with Terminals, Elsevier Science Publishers B.V., 1991

[6] C. Nicolaou, An Architecture for Real-Time Multimedia Communication Systems, IEEE Journal on Selected Areas in Communications, Vol.8, No.3, April 1990

[7] A. Poggio, J. J. Garcia Luna Aceves, E. J. Craighill, D. Moran, L. Aguilar, D. Worthington, J. Hight, CCWS: A computer-based multimedia information system, IEEE Comput., Oct. 1985