

A Model and a Tool for Active Watching: Knowledge Construction through Interacting with Video

Akio Takashima, Yasuhiro Yamamoto, Kumiyo Nakakoji

Research Center for Advanced Science and Technology, University of Tokyo

We view knowledge construction as a dynamic process by which people interact with external representations. This paper focuses on the knowledge construction process when interacting with video, such as in analyzing user study videos or surveillance videos. In such tasks, people actively interact with the video while skimming the entire data, examining details of a particular frame, making and validating a number of hypotheses, seeking specific characteristics, or making sense of trends. Existing video browsers do not serve this purpose because they are primarily designed for people to watch movies or TV programs to appreciate their contents on an as-is basis, and users do not actively interact with the video.

To support the knowledge construction process using video data, our approach employs the notion of "active watching." This notion is based on "active reading" (Adler and Doren 1972), in which readers actively interact with text media (such as by highlighting sentences in a scientific paper) while constructing knowledge. We argue that the ability to manipulate temporal and visual properties of video in various manners through direct manipulation is fundamental for active watching.

This paper first provides a theoretical foundation for our approach, and then proposes the Time-based Visual Presentation (TbVP) model, which is a framework for active watching. The TbVP Browser has been developed based on this model. Our user study illustrates how the tool helps people actively interact with video data in constructing knowledge.

The process of information understanding can be differentiated roughly into two types, *passive understanding* and *active understanding*.

In *passive understanding*, people receive information and act as mere passive recipients. Information senders are responsible for the representations of information by taking into consideration how the information should be sent and received by the receivers. That is why publishing companies or producers of a book go to great expense to clarify who the receivers are and how the layout ought to be. We regard this *passive understanding* as a static process.

In contrast, in *active understanding*, people are more engaged in changing representations and receiving methods to understand information better in accordance with their interests. For example, in reading scientific papers, readers change the appearance of the text by highlighting or underlining sentences or by scribbling comments on the text. In this situation, readers are not simply and passively receiving the text; they also are actively working on the text to constructively understand its content. Adler and Doren call this type of emerging understanding "active reading," which combines reading with critical thinking and learning (Adler and Doren 1972). Active reading not only involves interactively changing the appearance of the medium, but also changing reading styles. In a reading process, the likelihood is that readers will riffle through pages to overview its content, layout, atmosphere, and so on, and will pore over the specific parts of a book in which they have an interest. Thus, to change reading styles, such as riffling pages or iterative reading on a specific part, it is important to acquire information and construct knowledge.

What people understand individually differs even if the information resource is the same. Understanding depends not only on the experiences or the knowledge that each reader originally has, but also on the dynamically emerging relation between the reader and the information resource. The activity of constructing knowledge through both acquisition and understanding of information should be regarded as

a dynamic process that includes how receivers interact with the media. In other words, how people interact with media in everyday life involves not mere naive information receiving processes but complex knowledge construction processes.

In this research, the processes of acquiring and understanding information by interacting with media are referred to as knowledge construction. This paper focuses on interaction with video data for knowledge construction.

In what follows, we first discuss knowledge construction processes through text and video data illustrated with related work. Then we introduce the TbVP (Time-based Visual Presentation) model, which enables users to interact with video data by regarding it as temporal visual presentations. Section 3 describes the TbVP Browser, which is an application based on the model, and then Section 4 presents the user study about knowledge construction.

1. Knowledge Construction by Interaction

We view knowledge construction by interacting with media, as a process by which a media receiver (i.e., a reader or a viewer) interactively changes the property values of the media into arbitrary values according to the receiver's needs. In considering this, we use the properties of Media Data (MD) and User Experience (UE) to distinguish between what the data originally offers and what a user actually experiences. Interacting with media is viewed as transforming MD values into UE values. For example, zooming into a picture is regarded as changing a given value of a visual property (e.g., size, resolution) into different values of the properties (larger size, higher resolution). Changing the visual appearances of representations is explained as a transformation from MDV (Media Data Visualization) into UEV (User Experience Visualization).

This section shows the related work of active reading for constructing knowledge from documents as examples of such a transformation. It

then describes knowledge construction from video data, which is the focus of this paper.

1.1. Active Reading of Documents

Schilit et al. developed a touch panel device named XLibris that enables users be engaged in the active reading process with electric documents (Schilit et al. 1998). They describe that active reading with XLibris is useful for constructing knowledge, such as understanding the text, finding information within the text, and summarizing the text, in contrast to passive reading, in which a user receives information from a text with minimum effort.

The existing approaches to active reading mainly stress changing appearance of documents, such as annotating or highlighting. These approaches can be regarded as transformations from Media Data Visualization to User Experience Visualization.

Reading documents on a web browser using Speed-dependent Automatic Zooming (Igarashi and Hinckley 2000) is a notable approach that combines changing the visual appearance of the document with a reader's reading style of the document. The view automatically zooms out when the user scrolls rapidly so that the perceptual scrolling speed in screen space remains constant. The resulting effect is the same as knowledge construction by riffling a book to get an overview of it. Cockburn and Savage compared this technique with the traditional scroll, pan, and zoom method, and the results showed that scrolling tasks are done significantly faster with automatic zooming in both text document and map browsing tasks (Cockburn and Savage 2003).

Because the experience of reading a book differs among readers, depending on the representation or what kind of interaction is possible (Hornbæk and Frøkjær 2003), we consider interaction with media as being deeply concerned with knowledge construction.

1.2. Active Watching of Video

In the same way as with reading, we argue that in watching video data, such as a movie or animated visualization, there are two ways to watch: *active watching* and *passive watching*.

A few studies have supported active watching, even though they may not have used this phrase to describe their work. They have mainly addressed changing appearances, such as transforming MDV to UEV. For example, adding text as annotations on frames in which a user gets interested (Correia and Chambel 1999), or using 2D-spatial positioning to represent relationships among segmented frames (Yamamoto et al. 2001).

In contrast, little is reported that addresses changing the temporal aspects of video data to help knowledge construction. Visual and temporal aspects are the two fundamental properties of video data. Manipulation of time has had little study compared to the field of visualization. Our research goal addresses this challenge in designing methods for interaction with temporal properties.

We consider two types of time to express the playing speed of video: MDT (Media Data Time) and UET (User Experience Time). Changing the playing speed of a video affects the user's impression of it (Kawasaki and Ideguchi 2002). For example, in a football game video, to understand a positioning of one player or a formation shift of the whole team, the sequences may be better presented at a faster speed. Conversely, to understand a ball rotation or the shot motion of a player, a slower playing speed would be more suitable. That is to say, experiencing various playing speeds allows users to construct knowledge from various viewpoints. In this example, MDT refers to the time taken in playing the whole video at the original speed, and UET is the time taken in playing at a user-meaningful playing speed matched to a particular set of needs (Figure 1). Changing playing speed is described as a transformation from MDT into UET. One of techniques of active watching is thus achieved by changing MDT to UET. Passive watching can be viewed as simply watching, in which UET has the same value as

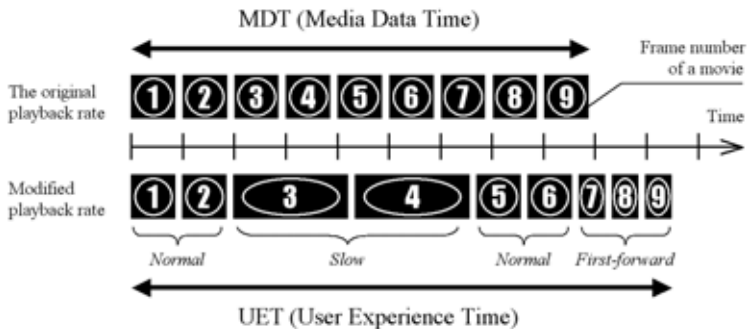


Figure 1 – Media Data Time and User Experience Time

MDT. In addition to the changes from MDT to UET, active watching includes changes of visual appearance such as pictorial size or resolution. The changes from MDV to UEV are similar to the changes of active reading.

Next, we introduce the TbVP (Time-based Visual Presentation) model for actively watching video data to construct knowledge.

2. Designing Systems for Active Watching

Numerous systems have been developed to summarize and visualize video data. For example, Video Manga identifies key frames of a video, then adjusts their sizes to pack them on the page in a style reminiscent of a comic book (Uchihashi et al. 1999). Nam and Tewfik introduced the Dynamic Video Summarization technique for summarizing video data, a system that modifies the local sampling rate to make it directly proportional to the amount of visual activity in localized sub-shot units of the video (Nam and Tewfik 1999). In each case, these systems employ techniques that automatically transform the raw video data into alternative temporal, visual representations. Although these automatically generated summarizations can be consumed faster than by watching the original video, they may not always constitute a perfect fit for the user's particular needs. For example, a system may summarize the video data by identifying and displaying all the scene

transitions, even though the user is mostly interested in a visualization of all the footage in which a particular person is present.

As described in the previous section, we view knowledge construction as a dynamic process by which people interact with external representations. Because no system will be able to generate appropriate visualizations for every conceivable need and situation, we wish to create a system by which users can actively watch video data.

2.1. Design Requirements of Active Watching

The existing interface styles provided to watch video data have not been changed much since the invention of the analog videocassette recorder. The current widely prevailing interface for a videocassette recorder, which is designed for passively watching movies and recorded events, does not support users in freely interacting with temporal representations for knowledge construction.

We have identified the following design requirements to support knowledge creation by active watching:

- the user must transform MD values to UE values; and
- the user must be able to easily interact with the video.

We developed the TbVP (Time-based Visual Presentation) model based on these requirements. The TbVP model is a conceptual framework that illustrates how a user can interact with, and modify, temporal and visual property values of video data through the process of transforming temporal and visual MD values into UE values.

We have taken into account the principles shown below based on the requirements and features of video data. The principles related to temporal properties are that the system should:

- allow a user to modify the playback rate and the direction of video,
- allow a user to interact without losing the temporal continuity of video, and
- allow a user to understand the relationship between the duration (MDT) of the video and the particular scene the user is viewing.

The principles related to visual properties are that the system should:

- allow a user to create multiple video images to compare different UE values,
- allow a user to change the display position and transparency of a video image for easy comparison, and
- allow a user to change the size of video image to focus the spatial representation.

We next explain the TbVP model and then address temporal and visual interaction based on these principles.

2.2. Time-based Visual Presentation

The TbVP model is a conceptual framework that illustrates how a user can interact with and modify the temporal and visual aspects of video data (Figure 2). We refer to these two types of transformation as Time-

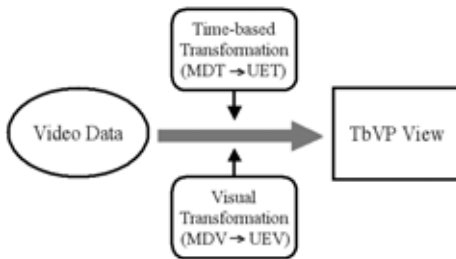


Figure 2 – TbVP model

based Transformation and Visual Transformation. The result through these two transformations is called a TbVP View.

Time-based Transformation enables users to change the playing speed of a video, whereas Visual

Transformation enables users to modify the visual appearance, which is unrelated to time.

By separating temporal aspects from other perceptual aspects, we not only can consider interaction design simply, but we can combine time-based transformations to other transformations in specialized target media. In listening to music, for example, Auditory Transformation that changes the volume of sound or the source direction could be combined with another transformation.

The details of each transformation of the TbVP model, and the interaction that is enabled by these transformations, follow.

2.2.1. Time-based Transformation

Time-based Transformation, as we have just described, is what changes MDT (Media Data Time), which is needed to play the whole video at normal speed, which indicates one meaningful to the user. This transformation enables users to explore video data by interactively changing speed.

One of the challenges of offering tools for active watching of video data lies in providing fluid interaction mechanisms to manipulate the temporal qualities of the video views. For example, it may be desirable to change the speed or direction of the video playback, as well as to provide smooth acceleration between different playback speeds. To address this problem, we developed the *Rate Controller (graph)*.

The *Rate Controller (graph)* uses a graph interface to represent a rate transition (Figure 3). In this paper, the term "rate" is used to describe playing speed and direction of the video. The horizontal axis represents the original frames of the video (the left indicates the start frame, the

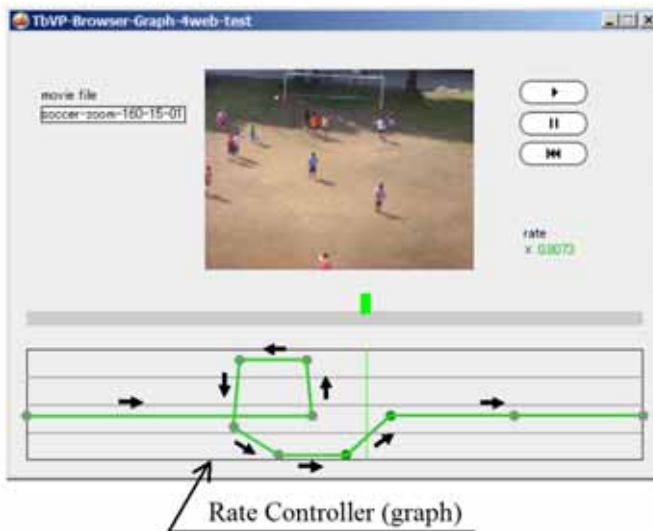


Figure 3 – A graph interface

right, the last frame), while the vertical axis represents playback speed. Users draw line segments through the graph to indicate which portions of the video should be played at what rates. For example, in Figure 3, the video will first progress at a normal rate, then move backward quickly, then progress forward again, at first at a slow rate that is gradually accelerated back to a normal playback rate. The small arrows on Figure 3 were added by the authors to illustrate the flow of browsing. The width of this graph illustrates the entire video duration (MDT) and the indicators (the handle on the time slider and the vertical thin line on the *Rate Controller*) show which part of the MDT is displayed. We designed this graph in one stroke for keeping the temporal continuity of video.

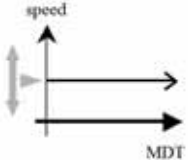
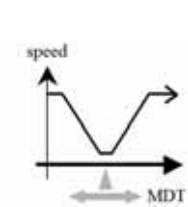
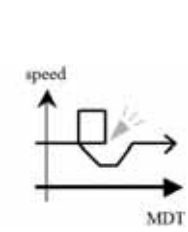
In an informal user observation using the *Rate Controller (graph)*, the users could interact with video data through a fine adjustment of the playback rate. However, in some situations, this highly controllable interface made the interactions more difficult.

Based on an informal observation result, we provide three rate change patterns for more practical exploring. Table 1 shows the graph and features of each pattern. These graphs are described in the same manner as for Figure 3. After creating these rate change patterns, we designed the interactions based on each type. The user interactions are illustrated as gray arrows in Table 1.

The first rate change pattern in Table 1 allows a user to keep the playback speed constant. We named this pattern “entire speed changer (Type1).” In using this pattern, a user dynamically controls the entire speed for the whole video while watching it. This means that the user interacts with the video in terms of UET.

The second pattern is for looking at a particular part at a slower speed and skimming the other part at a faster speed. We named this pattern “frame focus (Type2).” In using this second pattern, a user indicates where the focal point is, which is a scene of interest to the user. This means that the user interacts with the video in terms of the MDT.

Table 1 – Examples of rate change patterns.

	<p>Entire speed changer (Type1):</p> <p>A user could grasp the overview of entire video data at a higher speed or observe details at a lower speed. It works like a “zooming magnifier” for speed manipulation.</p>
	<p>Frame focus (Type2):</p> <p>A user could observe a detail of a particular focal point of video data. Coming toward the focal point, the video plays at a slower speed, and then becomes faster as the displayed frame moves away from the point of view. It is like changing the position of a magnifying glass used on a map.</p>
	<p>Re-examiner (Type3):</p> <p>A user could re-examine and focus on a frame that has just passed during browsing. At the moment when a user finds something interesting in the video, the player first moves backward slightly to replay the just-passed part. Then the player progresses like Type2. It is like putting a magnifier on a map at the precise moment that a user wanted to do this.</p>

The third pattern allows a user to re-examine the part that has “just passed” at a slower speed. We named this pattern “re-examiner (Type3).” By using this pattern, a user should be able to specify the focal point while watching the video. With this pattern, the user interacts with the video in terms of UET.

All of these patterns are designed to satisfy the user needs to skim the overview of a video or observe the detail of a part of the video. These patterns were often produced by the users in informal observation, but patterns are not restricted to these three types. Our future plan includes

implementing functionality to allow a user to add varieties of patterns while interacting with a video.

2.2.2. Visual Transformation

Visual Transformation is what changes Media Data Visualization into User Experience Visualization. Visual aspects of data that users can control include the on-screen location, size, resolution, or transparency of the picture.

The ability to freely modify these characteristics of the video allows a user to create custom visualizations meaningful to the user, especially for comparing two or more TbVP views. For example, a user can use a variety of positioning patterns to represent how views are related to one other, to convey which are more important than others, or to draw attention to particular views for the sake of comparison. The size and transparency of the views can also be changed for comparison purposes or to give attention to the views.

3. TbVP Browser

The TbVP Browser is an application based on the TbVP model that allows users to actively watch videos for knowledge construction (Figure 4). The TbVP Browser provides the capability to create multiple views of a video, and then selectively to alter the size, position, transparency, and time base of each view. The Browser is developed in macromedia Director and works on a web browser with shockwave plug-in. This section describes the functions and usages of the TbVP Browser.

3.1. Design Decisions

We designed the TbVP Browser based on the principles described in section 2.1.

In this application, Time-based Transformation (mapping from MDT to UET) produces the three types of rate transitions shown in Table 1, so the TbVP Browser includes them as a default set of rate change patterns.



Figure 4 – A screen shot of the TbVP Browser

To compare multiple TbVP Views, we have constructed a two-dimensional space in which users can freely place multiple TbVP Views. In this space, a user can change the size and opacity of each TbVP View and place them or pile up two or more TbVP Views.

Each TbVP View has an inherent color to distinguish it, and the components of the view, such as the handle on the sliders and the button and the border of the picture, are drawn in the same color.

The duration (MDT) of a video is mapped with the length of the slider bars.

3.2. General Operation

The engine of the movie player on the TbVP Browser is the same as the QuickTime Player, and has several of the functions of an ordinary movie player, such as play, stop, go back to start, and select a frame with the *Presentation Slider*. The *presentation slider* and indicators on it help the user to understand the MDT of the video and the parts in which each TbVP View displays.

A TbVP View has the same appearances as an ordinary video image, and its content is selected by *Movie file selector*. A TbVP View is displayed on the *pile space* when the *Add TbVP View Button* is pushed. The number of TbVP views displayed simultaneously is currently restricted to five for the sake of the implementation constraints.

3.3. Temporal Operation

The TbVP Browser has three types of predefined rate change patterns as Time-based Transformations (mapping from MDT to UET). To use each pattern, a user drags a *Pattern Icon* and then drops it on a TbVP View. If the user wants to watch a TbVP View in default rate, the “normal” *pattern icon* should be dragged and dropped.

When each function is assigned, the controller that specialized in each function is displayed; *Rate Controller (slider)* for “entire speed changer (Type1),” *Target Selector (slider)* for “frame focus (Type2),” and *Focus Selector (button)* for “re-examiner (Type3).” As argued above, the interfaces for each pattern enable users to interact with video data by specifying a moment on MDT or UET. In using the *Rate Controller (slider)* for a Type1 transition, a user manipulates the slider handle and controls its speed in real time by specifying UET. To set a focal point for Type2, a user manipulates the indicator on the *Target Selector (slider)*. The slider corresponds to the time length of a video, so a user can indicate a focal point on the MDT. In using the *Focus Selector (button)* for Type3, a user specifies a focal point by pushing the button. This means that the user controls the focal point on UET.

Additional functions can be useful for active watching, such as adjusting the details of these patterns and combining these patterns, but they are not available yet.

If video data include sounds, they will play at the same rate as the video image, and will be silent when the playback rate is set to zero.

3.4. Visual Operation

Visual Transformation (mapping from MDV to UEV) functions to modify the picture size and opacity of a TbVP View. These functions allow a user to change the focus or importance of each TbVP View.

To change the opacity of a TbVP View, a user manipulates the *Opacity Slider*. To change the picture size of a TbVP View, a user drags and drops an edge of a TbVP View with the Ctrl button.

In addition to its interaction with visual properties, the TbVP Browser has several spatial interactions. Users could give a meaning freely to the *Layout Space* and then place each TbVP View with some meaning into the space by dragging and dropping it. The *Pile Space* is for arranging TbVP views into a pile correctly.

These visual and spatial interactions could support a user's active watching for knowledge construction.

4. User Study



In order to observe how the TbVP Browser helps people actively interact with video data in constructing knowledge, we conducted a user study to compare user interactions using two types of video browsers.

4.1. Method and Video Data

We have observed users' interaction processes with video data by using two browsers, the TbVP Browser and QuickTime Player as a regular video browser.

The study comprised solving two tasks, as summarized in Table 2. The video image of Task1 is an actual screenshot, whereas the image of Task2 is a redrawn image due to copyright constraints. The video data used in Task1 was a record of a basketball game for one quarter. In that game, the white team predominantly advanced the ball and defeated the blue team by a score of 23 to 7. The video data used in Task2 was a record of a usability test of an online shopping web site using an eye

Table 2 – Task details.

	video image	content	questions
Task1		A record of a basket ball game (19min. 11sec.)	1) Which team won? 2) How was the overview of the game? 3) Are there any key plays?
Task2		A record of a usability test of an online shopping web site using an eye tracking system. (3min. 0sec.)	1) Suggest better web page designs.

tracking system. The subject user recorded in the video kept watching online shopping web pages to select an item to buy.

In our user study, both tasks were not simple search tasks but cognitively demanding problem-solving tasks. In the study, the subjects had to first identify what kinds of scenes to look for to answer the respective questions, and then to understand the contents of the video based on the identified scenes. The difference between the two tasks, Task1 and Task2, is that whereas the questions asked in the former task make it easier for the subjects to formulate what kind of scenes to look for, the questions in the latter task make it harder for the subjects to understand what to look for as scenes. We intentionally chose a video with the long duration in the first task and one with the short duration in the second task. This would allow us to compare the effects of time length (MDT).

In this study, we observed two subjects. SubjectT used TbVP Browser and SubjectQ used QuickTime Player; both were doctoral students. SubjectT had not used the TbVP Browser before, and he was instructed on the usage of TbVP Browser before starting the tasks for five minutes. SubjectQ had known the basic usage of QuickTime Player, and no instructions were given. Each subject was encouraged to speak aloud while searching to obtain think-aloud protocols (Ericsson and Simon

1984), and was instructed to finish browsing in about 20 minutes in each task.

4.2. Overview of the Results

The subjects' answers to the questions in each task are described in Tables 3 and 4. Comparing the answers of both user studies, there is

Table 3 – The answers to the questions in Task1.

Subject	Question	Answer
Subject T TbVP Browser	(1)	The winner was the white team.
	(2)	The white had an advantage because of a lot of shots.
	(3)	The blue had many mistake shots. The white made many goals with some three pointers.
Subject Q QuickTime Player	(1)	The winner was the white team.
	(2)	The white was dominant. The blue was defensive, although the goals percentage seemed better than the white.
	(3)	The rebounds of the white team were much better.

Table 4 – The answers to the questions in Task2.

Subject	Question	Answer
Subject T TbVP Browser	(1)	Making pages to show only the parts where users might be interested in. Making pages small in order to exclude a scrollbar.
Subject Q QuickTime Player	(1)	Arranging uniformly the amount of information described on a page. Making pages small in order to exclude a scrollbar. Showing the history of which pages a user visited. Making pages which users can look through all items.

only one clear difference about impressions of the goal percentage of the blue team in the second question in Task1.

We have not found any quantitatively significant differences in the two subjects' performances. However, the process leading to finding out the answers was clearly different.

The process of browsing by SubjectT with TbVP Browser was as follows:

1. Grasping the overview of the entire video with Type1 at a high speed.
2. Planning a strategy to search scenes to check.
3. Seeking a scene to check with Type1 at a high speed.
4. Getting the scene by iterating use of Type3 after overreaching the scene, then watching the detail.
5. Placing the TbVP View that displays the scene after stopping.
6. Making a new TbVP View to continue watching.
7. Returning to step 3 to repeat the process.

Figure 5 shows the flow of browsing by SubjectT illustrating the use of *Layout Space*.

The process of browsing by SubjectQ with QuickTime Player was as



Figure5 – An example of TbVP Browser use (SubjectT, Task1)

follows:

1. Fast-forwarding while looking for some characteristic scenes.
2. Playing at a normal speed, when the scene was found.
3. Replaying with rewinding or time slider bar if needed.
4. Returning to step 1 to repeat the process.

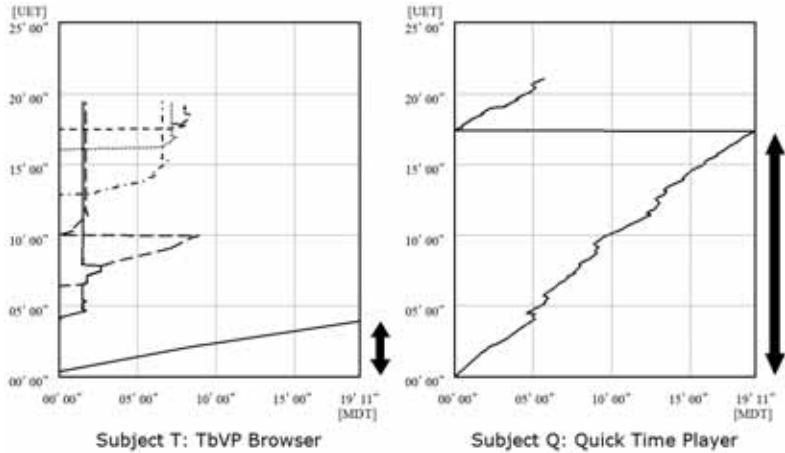


Figure 6 – The browsing processes in Task1

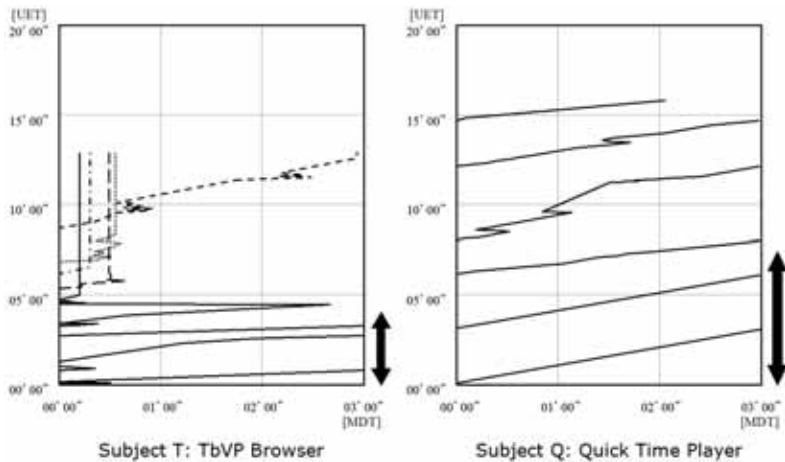


Figure 7 – The browsing processes in Task2

The following subsection describes more details about such differences.

Figures 6 and 7 show how each subject actually experienced MDT of the video as UET. Several types of dotted line on the left graph show a respectively different TbVP View. Observing these graphs, we could understand that SubjectT with TbVP Browser took a shorter time to grasp the overview of the video data than did SubjectQ with QuickTime Player. This is indicated by arrows on the graphs and described in the next subsection.

4.3. Observed Active Watching; A Qualitative Analysis

Through the user study, we could observe several distinctive behaviours of Active Watching, explained here with the actual situations.

Enlarging Video Image for Spatial Detail

In the each task, each subject started browsing by enlarging the TbVP View (Figure 5, upper left) or Quick Time Player. They did this because the default sizes of the video image were not large enough, and each subject wanted to watch the video image in detail spatially.

Skimming the Overview

As shown in Figures 6 and 7, both subjects browsed through the whole video at the beginning of each user study. The number of times of run-through by subjects was once in Task1, and three or four in Task2. The required time for skimming with TbVP Browser was clearly shorter than with QuickTime Player.

SubjectT browsed at the max speed (6X) set by Type1 in order to *“get a sort of an overview of the game quickly.”* In Task1, SubjectT finished the initial browsing from the start to the end of video in about three and a half minutes (Figure 6, left), and then argued that the winner seemed to be the white team. In Task2, SubjectT also browsed the video data in the same manner. The initial browsing thorough the whole video took about 30 seconds. Immediately after that, SubjectT argued that the user on the video was taking a long time to read detailed descriptions of the

size or color of the items on the web site rather than the large images of items on the site.

In Task1, SubjectQ's strategy involved taking an overview of video by fast-forwarding because the MDT of the video data (19 minutes, 11 seconds) was almost the same as the time limit of this study (20 minutes). Actually, SubjectQ quit fast-forwarding and then played at normal speed when an interesting scene, such as shooting or playing under the basket, appeared. Occasionally, the subject used the rewind button and time slider to check whether the ball had gone into the basket. Using this browsing method, it took 17 minutes to finish the initial browsing of the whole video (Figure 6, right). During the initial browsing, SubjectQ said, "I don't understand what points each team has scored" and "I'm not sure whether this goal was got or not." In Task2, SubjectQ described a plan to browse in normal speed because the MDT of the video data (3 minutes) was short.

Examining Details

In the study, the both subjects tried to examine particular scenes in detail. The scenes were, for instance, shooting scenes in Task1, and window-scrolling scenes in Task2. In order to watch a video to re-examine the scenes, SubjectT mainly used the Type3 rate transition of the TbVP Browser, and SubjectQ used rewinding and selecting by the time slider of QuickTime Player. SubjectQ could not operate these functions and said repeatedly that the contents could not be checked exactly, and then replayed several times. In contrast, SubjectT could operate the function comfortably. This is due to the playing speed of the part where users want to focus. Using QuickTime Player, SubjectQ had to replay at normal speed. On the other hand, the lowest speed of Type3 rate transition of TbVP Browser is 0.4X, so that SubjectT could re-examine scenes easily by watching details of the scenes at a slower speed.

We have described above that the impressions of the goal percentage of two subjects were somehow different. The actual goal percentage of the

blue team was 18%, whereas the white team made 45%. We could say that SubjectQ was not able to investigate scenes in detail.

Selecting a Specific Moment

To examine focal points of the video, SubjectT began to search some characteristic key plays. The scenes the subject wanted to view were goal scenes by the white team and failed shot scenes by the blue team. To find these scenes, the subject sought them at a high speed with Type1. When a scene was found, he pushed *Focus Selector (Button) for Type3* several times to rewind to the beginning of the scene, which had just passed by, and then re-watch the scene at a slower speed with Type3. SubjectT preferred this interaction rather than using the *Presentation Slider*.

SubjectQ tried to investigate details in each task as described above, and mentioned difficulty in using the time slider for selecting a moment of the video data. It seems much more difficult, especially in Task1 in which the MDT of the video date is longer. It is because, the temporal continuity of the video image was lost by playing without smoothness when SubjectQ has to operate the time slider.

Expressing Findings

The *Layout Space* of the TbVP Browser had originally been designed for comparing several TbVP Views. However, the space was actually used for expressing the meanings or relationships of findings.

SubjectT made a TbVP View stop at an interesting scene, made it smaller, and then positioned it in the Layout Space, repeating this procedure several times in the session. The subject used the positioned scenes as “visual bookmarks” (Figure 5, upper right). After storing several scenes, SubjectT relocated them to two sides: the left side for successful goal scenes made by the white team, and the right side for failed shot scenes made by the blue team (Figure 5, lower right). In Task2, the scenes SubjectT wanted to pick up were that the user on the video was reading details of items, and the user was scrolling a page. The subject placed the reading detail scenes on the left side, and the scrolling scenes to the right side on the *Layout Space*.

Users could organize or structure objects by positioning objects in space (Marshall and Shipman 1995), and it might be effective in this situation.

5. Discussion

Through the user study, we have observed the following activities of the subjects constituting active watching:

- enlarging video image for watching spatial detail
- skimming the overview
- examining details
- selecting a specific moment
- expressing findings

Although this user study had only two subjects, we could observe a number of active watching processes. We found that users attempt to interact with video data for constructing knowledge.

This section discusses future development of systems that support active watching based on the results of the user study.

5.1. Adjusting Comfortable Speed

Li et al. argued that the behaviors in browsing a digital video depend on not only the user's personality but also on the content type of a video (Li et al. 2000). Moreover, through the user study, we have found difficulty in predefining comfortable speeds for browsing.

The results show that the subjects were not satisfied with speed control. In using QuickTime Player, SubjectQ complained repeatedly because the predefined speeds of QuickTime Player were not comfortable for browsing. Even when TbVP Browser was used, SubjectT often browsed a video at the maximum speed (6x) set by Type1. This leads us to the opinion that faster speeds should be selectable by users.

The *Rate Controller (graph)* shown in Figure 3 could set the speed and rate change pattern freely, but it makes interactions complex. There is a trade-off problem between changing speed and easily interacting with speed. One solution is that systems provide predefined rate change

patterns and allow users to modify the patterns relatively and absolutely along their individual needs.

5.2. Interaction for Selecting a Moment

In designing a system related to temporal data, we must consider interactions carefully.

TbVP Browser has three types of rate change patterns as the default set. The interactions presented by these patterns are distinguished in two types, two interactions specifying UET, and one specifying MDT.

“Entire speed changer (Type1)” and “re-examiner (Type3)” provide interactions to specify a moment on UET. That is, the user interacts with a video while watching the video in real time. The results from the user study show that interactions specifying a moment on UET were comfortable for active watching because these interactions were performed frequently.

“Frame focus (Type2)” provides interactions specifying a moment on MDT. Compared with Type1 and Type3, Type2 was not used by SubjectT. The interface for Type2 is a slider that corresponds with the MDT of a video. It was difficult for SubjectT to map the MDT onto the slider length, which is why Type2 was not used.

One possible way to interact with video data using Type2 was proposed by subjectT. The idea was that a system regarding several TbVP Views can be stopped by a user as multiple focal points. Then another TbVP View would play at a slower speed at each moment indicated by the focal points. This approach changes the browsing style based on past interactions by a user.

Another possibility for Type2 is shown in Figure 8. To help users map the MDT of a video onto the *Target Selector (Slider)*'s length, the thumbnail images captured in a fixed interval from the video could be displayed on the upper side of the slider bar. The set of aligned thumbnails illustrates a static overview of the video, and it allows users to select a focal point more easily.

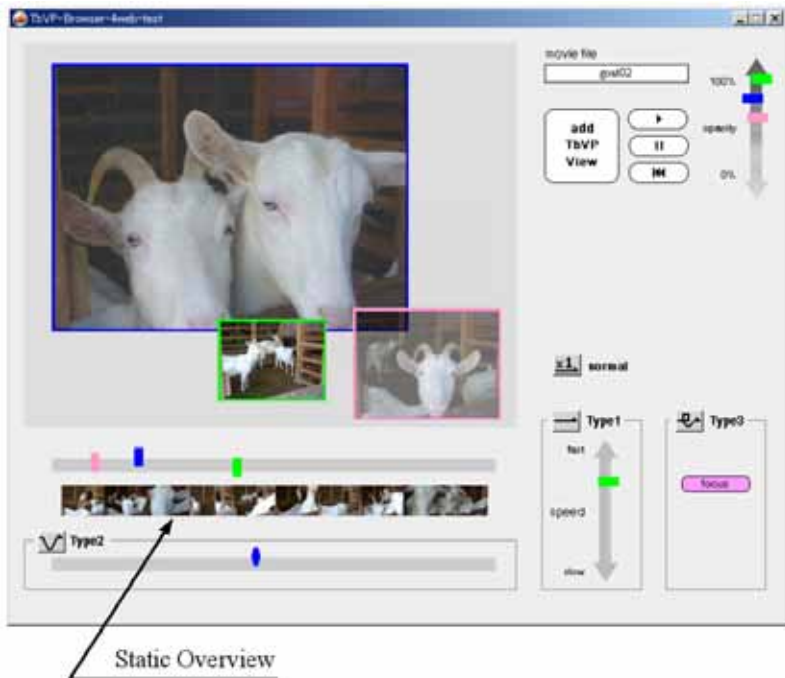


Figure 8 – TbVP Browser with static overview

6. Future Work

This paper describes the knowledge construction process when interacting with external representations, in particular video data, and describes TbVP Browser, which enables users to interact with video data. Through user studies, we have confirmed that TbVP Browser supports active watching by the user.

Since this user study had only two subjects, it is thought that the influence of individual characteristics exists, so additional experiments are needed.

Representations with temporal changes have been shown to affect a recipient's understanding in exploratory data analysis (Nakakoji et al. 2001), so we view that providing active interactions with temporal

properties is useful for knowledge construction. Although TbVP Browser was designed and developed for exploring video data, Time-based Transformation and temporal interactions are applicable to other temporal data, for example, combining Time-based Transformation with Auditory Transformation for musical data, as mentioned in section 2.2. In using speech data, users could listen to words easily by a time stretch function (changing speed with constant pitch), or users could understand what temporal operation is being done by listening to the speech as an auditory display that is played at a modified speed. In these cases, it is important to make recipients experience the MD values as the UE values, depending on the situations. In other words, concerning mapping MD values onto UE values means not only expressing how a user operates media representations, but also expressing dynamically emerging relationships between recipients and the representations.

Our view is that what MD values and UE values are, and how they ought to be, is important, and this plays a crucial role in designing interfaces and interactions between recipients and media.

Acknowledgement

This research was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Scientific Research (A), 16200008, 2004--2007.

References

Adler, M. J. and Doren, C. V. (1972) *How to Read a Book*, Simon and Schuster, New York.

Cockburn, A. and Savage, J. (2003) Comparing Speed-Dependent Automatic Zooming with Traditional Scroll, Pan, and Zoom Methods, in *People and Computers XVII: British Computer Society Conference on Human Computer Interaction*, pp. 87-102.

Correia, N. and Chambel, T. (1999) Active video watching using annotation, in *Multimedia'99 Proceedings (Part 2)*, pp. 151–154, ACM Press.

Ericsson, K. A. and Simon, H. A. (1984) *Protocol Analysis: Verbal Reports as Data*, MIT Press, Cambridge, MA.

Hornbæk, K. and Frøkjær, E. (2003) Reading Patterns and Usability in Visualizations of Electronic Documents, *ACM Transactions on Computer-Human Interaction (TOCHI)*, Vol. 10, No. 2, pp. 119–149.

Igarashi, T. and Hinckley, K. (2000) Speed-dependent Automatic Zooming for Browsing Large Documents, in *Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology*, pp. 139–148, ACM Press.

Kawasaki, T. and Ideguchi, T. (2002) Factor Analysis of Video Picture Impression and Influence of Video Transcribing Speed on Each Factor, *The Institute of Electronics, Information and Communication Engineers, Journal A*, Vol. J85-A, No.9, pp. 1022-1025.

Li, F. C., Gupta, A., Sanocki, E., He, L. and Rui, Y. (2000) Browsing digital video. In *Proceedings of CHI 2000 (April, The Hague, Netherlands)*, ACM, pp. 169-176.

Marshall, C. C. and Shipman, F. M. (1995) Spatial Hypertext: Designing for Change, in *Communications of the ACM*, Volume 38, Issue 8, pp. 88-97.

Nakakoji, K., Takashima, A., and Yamamoto, Y. (2001) Cognitive Effects of Animated Visualization in Exploratory Visual Data Analysis, in *IV-2001 Proceedings*, pp. 77–84, IEEE.

Nam, J. and Tewfik, A. H. (1999) Dynamic video summarization and visualization, in *Proceedings of the Seventh ACM International Conference on Multimedia (Part 2)*, pp. 53-56.

Schilit, B. N., Golovchinsky, G., and Price, M. N. (1998) Beyond paper: Supporting active reading with free form digital ink annotations, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 249–256, ACM Press / Addison-Wesley Publishing Co.

Uchihashi, S., Foote, J., Girgensohn, A., and Boreczky, J. (1999) Video Manga: Generating semantically meaningful video summaries, in Proceedings of the ACM Multimedia 99, pp. 383-392.

Yamamoto, Y. Nakakoji, K. Aoki, A. (2001) ARTWare: A Component Library for Building Domain-Oriented Authoring Environments, International Conference on Future Software Technology 2001, Software Engineers Associates, ZhengZhou, China, pp. 246-251.