An Interaction Model for Experimental Data Analysis Using Spatial Positioning

Yasuhiro Yamamoto [1] Kumiyo Nakakoji [1,2,3] Atsushi Aoki [2]

[1] Graduate School of Information Science, Nara Institute of Science and Technology
[2] Software Research Associates, Inc.
[3] PRESTO, JST

8916-5, Takayama-cho, Ikoma, Nara, 630-0101, Japan
{yasuhi-y, kumiyo}@is.aist-nara.ac.jp aoki@sra.co.jp
http://ccc.aist-nara.ac.jp/
Tel: +81 743-72-5381; Fax: +81 743-72-5383

ABSTRACT

In software usability testing, empirical software engineering, or user evaluation, experimenters collect various types of data. Understanding such data requires cognitively intensive qualitative analysis based on the cycle of discovery and validation processes. We need computational tools that are suitable for supporting experimenters to discover important aspects of the data, to collect them, to store them and to share them with peers. Our approach to support these processes is to provide the tool, Time-ART, which uses spatial positioning as an interactive schema. This paper discusses the nature of experimental video analysis tasks and identifies requirements for supporting such tasks. An interaction model for the Time-ART system uses spatial positioning. The interaction model is based on our previous experience of developing the ART system, which supports writing by using two-dimensional positioning.

KEYWORDS

cognitive tools, video-analysis support, human-computer interaction, two-dimensional positioning as representations, interaction models

INTRODUCTION

In software usability testing, empirical software engineering, or user evaluation, experimenters* collect various types of data on subjects before, during, and after an experimental task. Examples of such data include pre-task questionnaire, pre-interviews about a subject profile, video-recorded screen display of a subject's computer during the experiment, video-recorded screen display of an experimenter's computer during the experiment, protocols made by a subject during the experiment, a video and voice record of a subject's behavior, a subject's input data such as mouse-trace or keystrokes during the experiment, biometrical data collected during the task, and video-taped post-interviews with the subject.

Among such a wide variety of data, one of the most important and critical sources of information is the data collected during the experiment. Different from data collected before or after the experiment, data collected during the experiment are all time-stamped, and multiple sources of data are collected in parallel. The multiple data sources allow the experimenter to analyze behavior, thoughts, or mental situations of a subject during the experiment from a variety of perspectives.

Those time-stamped data collected during the task are recorded independently, but they are essentially interdependent on each other. In analyzing such data, some can be analyzed by simply applying statistical analysis techniques, but some are more difficult to analyze requiring cognitively intensive qualitative analysis based on the cycle of discovery and validation processes.

Most of such data is video-taped data recorded during the task. Faster CPUs, larger hard-disks and better data-compression techniques have made it possible to analyze video data on a personal computer. However, most of existing tools which deal with video data are video authoring tools, which aims at producing a movie or a video-clip, or automated video summarization tools that help users to understand the contents of the video or quickly browse video clips. They are not suitable for experimental video-analysis tasks helping a user discover interesting phenomena and understand what is going on in a video-taped session.

Video-analysis tasks for empirical studies are to understand what was happening during the experiment. Experimenters need to understand and discover which portions are going to be important, where to look at to construct empirical hypothesis, or to validate the hypothesis. We need computational tools that are suitable for supporting experimenters and practitioners to discover important aspects of the data, to collect them, to store them and to share them among peer researchers.

---

* We use the term experimenter referring to a person who designs an experiment, conducts the experiment, and analyzes the result, for the sake of simplicity.

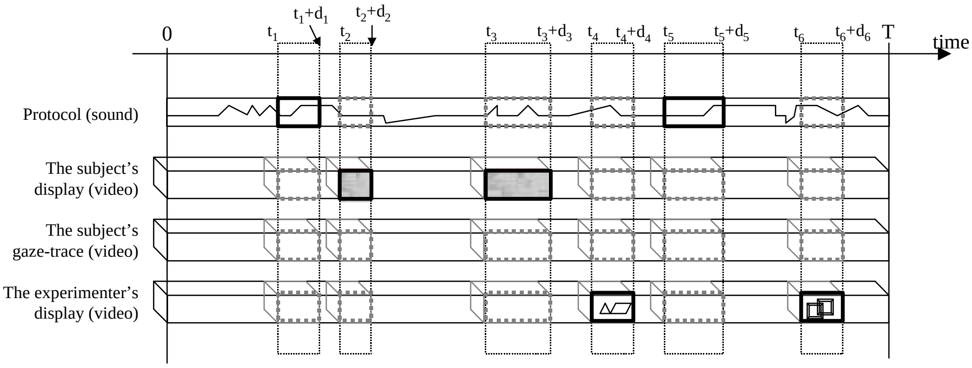


Figure 1: Experimental Data Analysis Tasks: Experimental data includes multiple resources of data. Each data is time-stamped between *0* and *T*, the duration of the experimental session.

Our approach is to support such a video-analysis process with a computer tool using two-dimensional positioning as an interactive schema. The tool, Time-ART, allows a user to segment any portions of a video clip and position it in a free two-dimensional space. This positioning allows the user to annotate each portion by representing things such as how this portion is related to other portions, which portions are supporting a hypothesis, or which portions need more careful analysis. In addition, Time-ART allows a user to automatically obtain other types of data which share the same time-stamp with the video portion. A user may add textual annotation to each segmented video.

In what follows, we first describe the nature of experimental video analysis tasks and identify requirements for supporting such tasks. Then, we describe an interaction model for the Time-ART system using spatial positioning. The interaction model is based on our previous experience of developing the ART system [13,14], which supports writing by using two-dimensional positioning. We discuss related work and conclude the paper.

VIDEO ANALYSIS TASKS

At Nara Institute of Science and Technology, we have been developing a CAESE (Computer-Aided Empirical Software Engineering), which is an integrated environment for empirical software engineering supporting various aspects of the empirical study processes, such as collecting and analyzing data [11]. The basic architecture consists of a data transmitting component, a data recording component, a data play-back component, a data displaying component, and a data analysis component. The type of data that the CAESE can currently handle includes mouse and keyboard strokes, eye traces, three-dimensional movement, and video-tapes. The tool we describe in this paper is a part of the data analysis component of the CAESE. Our approach is based on an assumption that video and other multimedia data that experimental data analysis tasks analyze are all synchronized with timestamps.

In usability testing or empirical software engineering, video is an effective medium for recording a subject's behavior as well as a computer screen. Such video can be used to record:

- a screen of a subject's computer system,
- protocol, face impressions and body movement of a subject, and/or
- eye-movement (gaze) traces superimposed on an experimenter's monitoring screen.

These video data are usually accompanied with other types of data, such as command sequences, keystroke records, three-dimensional motions captured, or biometrical data including brain-wave or skin-resistance level (Figure 1).

Such experimental video-taped data have unique characteristics different from other types of video, such as family movies or instructional movies. The characteristics include:

1. experimental video-taped data can be all time-stamped and synchronized with other video-taped data or other types of data, such as keystrokes;

2. experimental video-taped data will not be edited or added at a later time. The taped session is a result of an experiment to be understood as it was. It is not supposed to change or add anything to the content of the video once the experiment is over;

3. experimental video-taped data do not have any intended purpose for video viewers. It is created in such a way that helps a viewer (experimenter) discover and understand what was hapenning during the experiment. Nobody knows which part would be important or what interesting stories would be recorded in the video (as opposed to an educational movie which has an embedded goal in the movie); and

4. the purpose of experimental video-taped data is in the end to produce *findings*. Such findings should be made explicit and recorded so that they can be shared among other peer researchers and practitioners (as opposed to

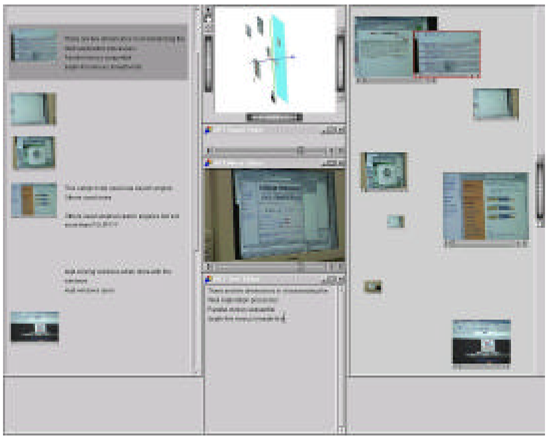


Figure 2: The Time-ART System

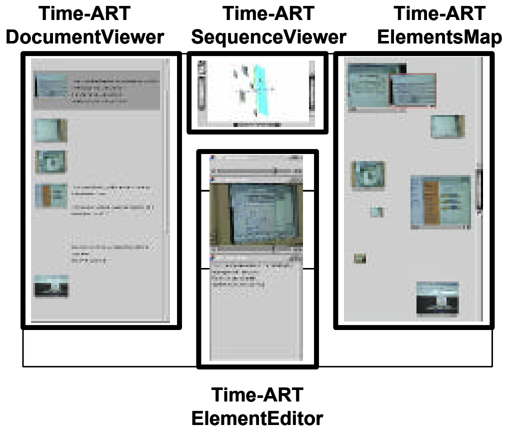


Figure 3: Components of Time-ART

an artistic movie the viewers of which do not have to produce any report).

Tools that support such video analysis tasks, therefore, must allow a user to perform intensive intellectual exercise using videos as multimedia materials. At the same time, the tool must support a user to express and record what has been found and understood.

During the analysis task, an experimenter plays a lot of what-if games. By looking at a certain portion of a video, the experimenter creates a hypothesis; then the experimenter needs to validate the hypothesis by looking at other types of data with the same time-stamp, or other parts of the video. This process is iterated rather randomly. It does not proceed in a sequential manner such as finding a hypothesis and validate it one-by-one. Instead, all these complicated cognitive processes proceed in parallel.

Figure 1 illustrates the process. The experimenter finds interesting portions within multiple sources of information; for instance, a sound stream for protocols, a video stream for the subject's display, a video stream for the subject's gaze-trace, or another video stream for the experimenter's display. Each of those data sources are time-stamped and synchronized, starting from the beginning of the experiment and ending at the time T, which is the duration of the experiment. When an experimenter finds a certain portion of a protocol interesting, the portion can be identified with the starting time of the protocol t_1, and the duration of the partial protocol d_1. Then, the experimenter may need to examine what was happening during the time period (between t_1 and $t_1+ d_1$) and analyzes other sources of data, such as the video data storing the subject's display. Figure 1 illustrates a situation when an experimenter finds six portions of experimental data interesting during an experimental analysis task.

In sum, for supporting experimental video analysis tasks, a user must be able to:

1. focus on a certain portion of a single source of data (e.g., sound data for a protocol generated between t_1 and $t_1+ d_1$);
2. compare the focused portion with different sources of multimedia data (e.g., sound data for protocol with video data for the gaze-trace);
3. obtain other types of data that has the same time-stamp with the focused portion (e.g., video data for the experimenter's display between t_1 and t_1+d_1);
4. overview multiple portions of multiple sources of data to understand overall phenomena indicated by the experimental data;
5. annotate the portion when necessary; and
6. integrate what have been found to be shared with peer experimenters.

The next section describes the Time-ART system, which is designed based on the above requirements.

THE TIME-ART SYSTEM

The interaction model of Time-ART that we propose in this paper is based on the approach of the ART system. The ART system [13] [14] supports document construction as a design task. The system allows users to position segmented text chunks as "elements" in a two-dimensional space. An element is any unit that writers choose to think of as one, such as a phrase, a sentence, a paragraph, or a longer piece of text. Using the eye-tracking analysis technique and through a series of user studies of the ART system, we have identified that two-dimensional positioning is a powerful interactive scheme for annotating a chunk of text, expressing things such as the level of completeness, the level of commitment, or the degree of relatedness with other chunks [10].

We have adapted the interaction model of ART to help experimenters analyze experimental video and designed

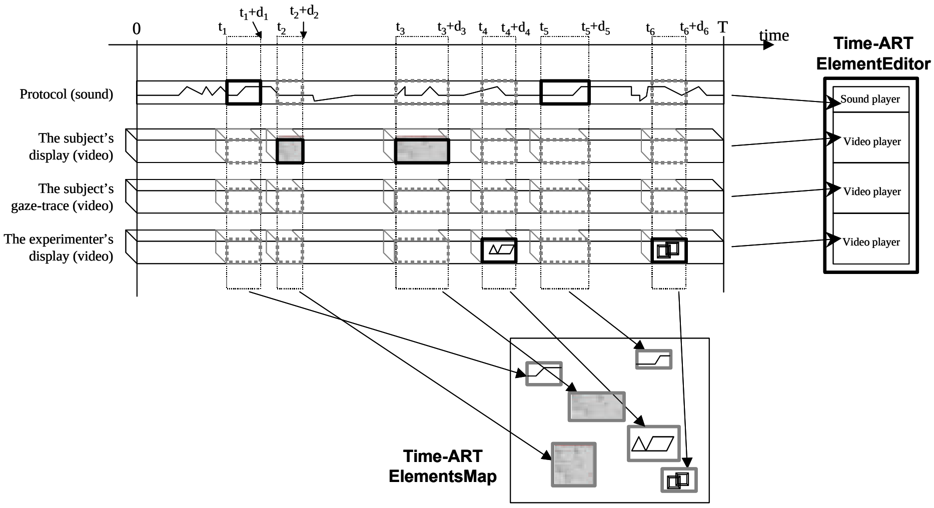


Figure 4: The Roles of Time-ART ElementEditor and Time-ART ElementsMap

Time-ART (Figure 2). The Time-ART system consists of four parts: ElementsMap, ElementEditor, DocumentViewer, and SequnceViewer (Figure 3). A user of Time-ART can annotate any portions of experimental video data by clipping any part of the original video data (with ElementEditor), and position it in a two-dimensional space (ElementsMap). DocumentViewer of Time-ART displays automatically constructed hypermedia document based on the segmented multimedia data with annotations.

Figure 4 illustrates the role of ElementEditor of Time-ART. Using the Time-ART ElementEditor, the user can browse all time-stamped data in a synchronized manner. When the user finds a portion of some of the data interesting, the user can clip the portion by specifying the starting and ending point. The segmented/clipped portion can then be placed in the ElementsMap (Figure 4).

Figure 4 illustrates how segmented multimedia data, called elements, are represented in the ElementsMap. The ElementsMap consist of layers each of which corresponds to different media recorded during an experiment. When a user selects one of the elements displayed in ElementsMap, all of the corresponding data are displayed in ElementEditor (Figure 4).

In addition to multiple information sources, a user can annotate each segmented clip with textual descriptions in ElementEditor. Each annotated data-clips are integrated and displayed in Time-ART DocumentViewer.

Now we briefly describe functions of each component of Time-ART.

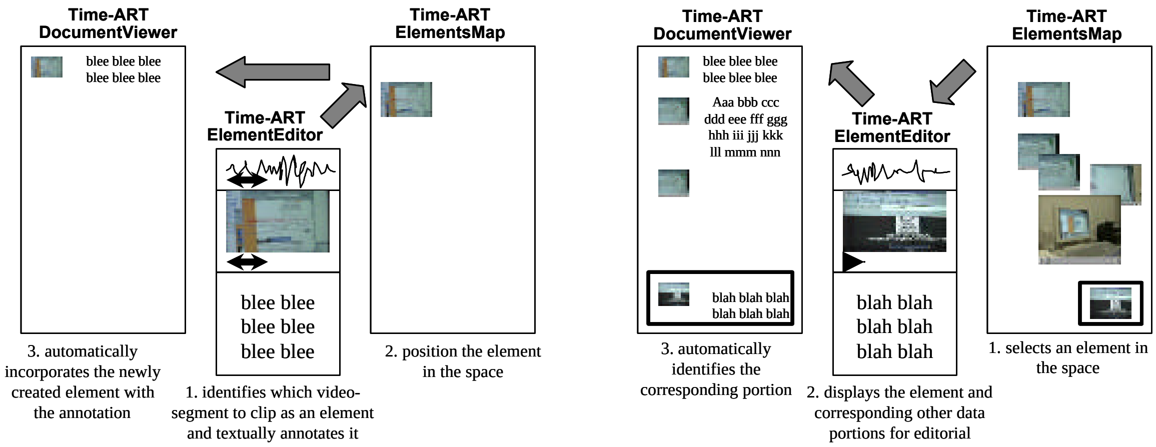
ElementsMap
ElementsMap is a two-dimensional space where a user can position segmented video, sound or data clips. Each segmented portion is called an element, and represented as a thumbnail image in the ElementsMap. Each element represents multiple data sources that has the same time-stamp.

A user can position each element anywhere in the ElementsMap and freely change the size. As we have observed through user studies of the ART system that supports writing by freely positioning chunks of text in a two-dimensional space [10], a user of Time-ART gradually develops understanding about experimental video data by understanding the role of each segmented video, its relationship to other elements, or its relationship with data represented in other media with the same time-stamp, through interacting with the two-dimensional spatial positioning.

ElementEditor
When no elements are selected in the ElementsMap, ElementEditor displays the entire experimental video, sound and other data of the whole experimental session. A user can segment a portion of the data by specifying the starting time and the ending time. The segmented data becomes an element as stated above.

When one element is selected in the ElementsMap, ElementEditor displays the content of the element and a user can play the video, sound and other data in the duration of the specified timestamp. A user can also change the starting time or the ending time of the element. In the text editing part at the bottom, a user can write annotation for the element representing comments including what has been understood about the element, findings from the element, and hypotheses made about the element.

SequenceViewer
SequenceViewer displays segmented elements displayed in the ElementsMap in the three-dimensional space with the time sequence as the Z axis. As stated above, a user can



(a) Creation of a new element

(b) Display contents of a selected element

Figure 5: The Use of Time-ART

freely change the position of each element in the ElementsMap. While SequenceViewer dynamically reflects the changes in the ElementsMap, it maintains the time-sequence of each element according to the starting time of the timestamp of each element; the newer the element is in terms of its timestamp, the closer it is displayed to a user in the SequenceViewer.

DocumentViewer

DocumentViewer integrates segmented elements together with textual annotations producing a hypermedia document consisting of each segmented thumbnailed data element with its associated annotation. Looking at the information displayed in the DocumentViewer allows users to understand what has been found and understood about the experimental video data. This information is automatically produced by simply combining elements with their annotations in the order of time sequence as represented in SequenceViewer. DocumentViewer allows users of Time-ART to share and understand results and findings as well as processes and current status of experimental analysis.

Figure 5 illustrates how a user interacts with the Time-ART system in an experimental video analysis task. First, a user examines the entire video-clip, sound data or other data recorded during the experiment. Whenever the user finds some portion of one of the multimedia data interesting, the user can clip the portion in Time-ART ElementEditor and position it in the ElementsMap window of Time-ART (see Figure 5-(a)). When clipping the element, the user may add textual annotations to the element. This segmented data portion together with textual annotation is automatically displayed in the Time-ART DocumentViewer window.

As the analysis task proceeds, the ElementsMap window of Time-ART includes many portions of multimedia data (see Figure 5-(b)). To analyze each element, the user simply selects one of the displayed elements in the ElementsMap window (see the bottom of ElementsMap in Figure 5-(b)). The content of the element as well as other data portions that has the same time-stamp with the original element are automatically presented in the ElementEditor window in the system, where the user can examine the content or further clips smaller portion out of the displayed multimedia data portion. When selecting an element in ElementsMap, the corresponding portion is automatically highlighted in the DocumentViewer window of Time-ART.

A prototype system based on the above design has been developed on VisualWorks (Smalltalk) and 3-D graphics library Jun [7].

RELATED WORK

Existing computer tools for dealing with video can be categorized into three types: video analysis tools, video summary tools, and video annotation tools.

Many of video analysis tools support automatically processing video data with the purpose of more quickly find a certain portion of the video or more quickly understand the whole [3]. For instance, VIS (Visual Information Seeking) is a framework for information exploration where users filter data through direct manipulation of dynamic query filters [6]. Uchihashi et al. [12] presents methods for automatically creating pictorial video summaries. In their system, image and audio analysis is used to automatically detect and emphasize meaningful events. DIVA is an exploratory data analysis tool with multimedia streams [9]. The system enables users to visualize, explore and evaluate patterns in data that change over time. The system is intended to help researchers annotate and visualize patterns and relationships among time-based multimedia data, and not to help authoring multimedia documents. CEVA [4] is a synchronous groupware tool that provides an animated direct manipulation WYSIWIS interface for multithreaded collaborative video analysis.

Video summary tools support users to more quickly find important scenes in a shorter time. Many video summary tools support users to browse a video-taped meeting session, or a video-taped lecture or seminar [8]. Boreczky et al. describes their approach to summarize videos and navigate the summaries by using an interactive comic book presentation metaphor [2].

Video annotation tools are mostly aimed at supporting training or learning. A typical video annotation tools allow a viewer to annotate a certain portion of a video clip as they view the whole instructional movie. For instance, Synopsus focuses on helping make video into a viable and useful learning medium for novice computer users by enabling them the ability to watch and control a video movie while at the same time helping them create an interactive hyperlinked personal summary of the movie they are watching [5]. MRAS [1] is an asynchronous educational environment featuring on-demand streaming video. The system is Web-based for annotating multimedia Web contents and students can share annotations with each other and instructors.

As we discuss in Section 2, videos that these tools deal with are produced with a certain goal and intention in mind. For instance, if a movie is about an instructional movie about the Universe, then there are certain themes, topics, and goals that need to be conveyed to viewers (students) by looking at the video. Experimental video analysis, on the other hand, is not produced with any intention in mind. It is totally a viewer's responsibility to find out what was there. Trial-and-error type understanding is more critical with experimental video analysis tasks.

CONCLUSION

In this paper, we described the nature of experimental multimedia data analysis tasks, and identified requirements for supporting tools for such tasks. We described an interaction model for Time-ART that supports the tasks. The system uses spatial positioning as an interaction metaphor allowing experimenters playing what-if games in building hypothesis and understanding findings. We are currently implementing a prototype system based on the interaction model described in this paper .

ACKNOWLEDGEMENTS

This study was supported by the Proposal-based New Industry Creative Type Technology R&D Promotion Program from the New Energy and Industrial Technology Development Organization (NEDO) of Japan.

REFERENCES

1. Bargeron, D., Gupta, A., Grudin, J. and Sanocki, E.: "Annotations for Streaming Video on the Web: System Design and Usage Studies," Computer Networks (Netherlands), Vol. 31, No. 11-16, Elsevier Science, pp.1139-1153, 1999.

2. Boreczky, J., Girgensohn, A., Golovchinsky, G., and Uchihashi, S.: "An Interactive Comic Book Presentation for Exploring Video," in Proc. of CHI2000, ACM Press, pp.185-192, 2000.

3. Cheyer, A. and Julia, L.: "MVIEWS: Multimodal Tools for the Video Analyst," in Proc. of IUI 98, pp.55-62, 1998.

4. Cockburn, A. and Dale, T.: "CEVA: A Tool for Collaborative Video Analysis," in Proc. of the International Conference on Supporting Group Work (ACM Group'97), ACM Press, pp.47-55, 1997.

5. Dekel, A. and Bergman, O.: "Synopsus: A Personal Summary Tool for Video," in Extended Abstracts of CHI2000, ACM Press, pp.4-5, 2000.

6. Hibino, S. and Rundensteiner, E.A.: "MMVIS: A Multimedia Visual Information Seeking Environment for Video Analysis," in Proc. of the CHI '96 conference companion on Human factors in computing systems: common ground, ACM Press, pp.15-16, 1996

7. http://www.sra.co.jp/people/aoki/Jun/Main_e.htm

8. Li, F.C., Gupta, A., Sanocki, E., He, L.-w., and Rui, Y.: "Browsing Digial Video," in Proc. of CHI2000, ACM Press, pp.169-176, 2000.

9. Mackay, W.E. and Beaudouin-Lafon, M.: "DIVA: Exploratory Data Analysis with Multimedia Streams," in Proc. of CHI'98, ACM Press, pp.416-423

10. Nakakoji, N., Yamamoto, Y., Reeves, B.N., and Takada, S.: "Two-Dimensional Positioning as a Means for Reflection in Design," Design of Interactive Systems (DIS 2000), ACM, New York, NY, August, 2000 (to appear).

11. Torii, K., Matsumoto, K., Nakakoji, K, Takada, Y., Takada, S., Shima, S.: "Ginger2: An Environment for CAESE (Computer-Aided Empirical Software Engineering)", IEEE Transactions on Software Engineering, Vol.25., No.4, July/August, 1999, pp.474-492.

12. Uchihashi, S., Foote, J., Girgensohn, A., and Boreczky, J.: "Video Manga: Generating Semantically Meaningful Video Summaries," in Proc. of ACM Multimedia 99, ACM Press, 1999, pp.383-392.

13. Yamamoto, Y., Takada, S., and Nakakoji, K.: "Representational Talkback: An Approach to Support Writing as Design," in Proc. of 3rd Asia Pacific Computer Human Interaction Conference, IEEE Computer Society, 1998, pp125-131.

14. Yamamoto, Y., Nakakoji, K., Takada, S.: "Hands-on Representations in a Two-Dimensional Space for Early Stages of Design," Knowledge-Based Systems Journal, Elsevier Science, 2000 (to appear).