Appeared in “Designing Effective and Usable Multimedia Systems,” A. Sutcliffe, J Ziegler, P. Johnson (Eds.), Kluwer Academic Publishers, The Netherlands, pp.91-102, September, 1998.

Finding the “Right” Image: Visualizing Relationships among Persons, Images and Impressions

K. Nakakoji[1,2,3], Y. Yamamoto[3], K. Sugiyama[3], S. Takada[3]
[1]PRESTO, Japan Science and Technology Corporation (JST),
[2]Software Engineering Laboratory, SRA Inc.,
[3]Graduate School of Information Science
Nara Institute of Science and Technology
8916-5, Takayama-cho, Ikoma, Nara, 630-0101, Japan
tel: +81-743-72-5381; fax: +81-743-72-5319
email: kumiyo@is.aist-nara.ac.jp

Abstract

In designing multimedia information, authors need to use images that best represent the design intention. The use of images, however, can be misleading. Images can be interpreted in many ways, and the same image may be thought of as pretty, cute, childish, or immature depending on the context. Due to this situation, searching for images from an image library becomes a challenging task because authors cannot specify what they are looking for without understanding what impression people have with each image. The goal of the research presented here is to support casual authors in finding images for their authoring task by understanding (1) how their task is best represented using affective words, and (2) what impressions people have with images. We propose a model to visualize individual differences in impression of images allowing users to explore the space of relationships. The model consists of a set of triplets of a person (P), an image (I), and an affective word that represents impression (W). We have built a computational environment EVIDII to visualize the relationship among P, I and W using two types of spaces. The paper concludes with a discussion of effectiveness of the approach.

Keywords
image search, impressions, Kansei information, visualization

1 INTRODUCTION

The goal of multimedia information design is that meanings behind the information are well-communicated - the meaning is shared among audience as the author has intended (Nakakoji 1996). As more and more casual users have an opportunity to author multimedia information, we have found more cases where authors are misusing multimedia (Forsythe 1998). Casual users do not have domain knowledge about the use of color, for example, and know little about what effects or consequences colored representations have on humans (Murch 1984). This results in multimedia information that miscommunicates the author's intention only because the author used colors in a wrong manner.

An image (computer graphics and pictures) is a multimedia domain that requires domain knowledge to be appropriately used in an authoring task. Images can be more ambiguous than text. For example, one image may be interpreted as a cute picture or a silly picture depending on the context. Factors that determine the context include topics of the information, purposes of the information, and most importantly, who receives the information. Professional multimedia authors have domain knowledge to understand or predict what effects images have. Casual users do not have such knowledge, and therefore, may keep looking for an image which is inappropriate for their authoring task.

Our goal is to support casual multimedia authors to find images for their authoring task that will be interpreted in the same manner by the audience as intended by the authors. There are two challenges in pursuing this goal. First, images can be interpreted in many ways, and we need to help users understand what impressions other people have for images. When an author wants to design a 'refreshing' homepage, an image selected for the page needs to be interpreted as 'refreshing' or as a similar impression; otherwise, the use of the image is 'wrong' in the authoring task.

Second, the goal of an authoring task can be vague and ambiguous, and fluctuate (Nakakoji 1996). Authors are not able to completely articulate what the goal is at the beginning of the authoring task as authoring tasks are ill-defined design tasks (Simon 1981, Fischer 1991). An approach to simply retrieve images based on what authors have specified as the initial requirements will not work.

Existing research in image retrieval based on 'feelings' or 'meanings' that represent an authoring task aims at developing a best mapping between words and physical properties of images (Kurita 1992, Inoue 1996, Hasegawa 1997, Isomoto 1996). This approach involves three issues. First, it is impossible to map everybody's impression in a generic manner. It may be possible to have multiple associations for different groups of people, but it will not solve the problem. We have found in one of our user studies that even the same person has different impressions at different times for the same image. Second, words themselves give different impressions to different people. Different people have different connotations for the same textual expression; for example, one might think that the term 'gorgeous' reminds him of a beautiful woman, whereas others might associate 'gorgeous' with a more negative meaning, such as being overly expensive. Third, the approach does not take into account the fact that authors' requirements fluctuate. Although authors may look for 'refreshing' images as they think their task is to develop a 'refreshing' homepage, they may later find that their task is better represented with the term 'natural' than 'refreshing.' Retrieving images based on a specific requirement as a one-shot affair will not suffice to support casual authors in finding images for an authoring task.

Instead of trying to train a machine to find the right mapping between words and images so that users can retrieve images based on a task specification, our approach is to use the machine to visualize the differences of impressions of images, and to allow people to find associations by exploring the visualized information spaces. The power of multimedia allows us to use the human perception skill on externalized representations instead of simply presenting numbers calculated as correlation factors among impressions and images (Zhang 1997).

We have developed a model that deals with three elements: people, images, and words. The (P,I,W) triplets are then viewed from each element: the person-based, image-based, and word-based perspectives. Each perspective is then mapped to a two-dimensional or three-dimensional space, such as the HBS-space for the image-based perspective. The EVIDII (Environment for Visualizing Differences in Individual Impressions) system has been developed based on this model. The system provides four different types of views to explore the relationships among persons, images and words by looking at how the triplets are distributed. In exploring the relationships that are visually represented, authors have a better understanding of what words better represent the goal of their authoring task, and what images better serve their tasks by looking at what others think of the images.

In what follows, we first present the EVIDII model. Section 3 describes the EVIDII system, the rationale for the design of the system, and also presents a brief scenario of how users interact with EVIDII. We conclude the paper with a discussion of implications of our approach and future directions.

2 THE EVIDII MODEL

The EVIDII (Environment for Visualizing Differences in Individual Impressions) model uses the three elements of persons, images and words to denote and visualize relationships among them. We first give an overview of the model, then describe the three spaces that are used in the model, as well as the views which are used by the users.

2.1 Overview of EVIDII

The EVIDII model uses the three elements of persons, images and words to represent the space of association. S(P,I,W) represents a set of triplets {(p, i, w)} where p is a person identifier, i is an image identifier, and w is an affective word, such as ‘clear’, ‘soft’, or ‘cute’. For example,

(Jack, Image#31, refreshing)

represents ‘Jack thought Image#31 as refreshing’. Such data is collected using a questionnaire to ask which image is associated with which words.

The goal of the model is to allow people to explore how the three elements are related to each other. For example, authors should be able to ask questions such as:

- *“In addition to Jack, who else found Image#31 to be refreshing?”*
- *“Are there any other images that Jack thinks refreshing?”*
- *“Does Jack find the image just ‘refreshing’?”*

- *“What words best describe my task?”*

Thus, EVIDII is an environment that visualizes relationships among the three elements so that the users of the system would be able to answer such questions by interacting with the system.

2.2 Basic Spaces

The EVIDII model provides three basic kinds of spaces using each of the three elements in the model as the base element: image-based space, word-based space, and person-based space. This approach of focusing on one element at a time was taken because none of people, images, or words can be linearly ordered in any understandable way.

When visualizing differences in relationships among persons, images and words, the actual representation of each basic space is also important, as it will affect the user’s interpretation (Zhang 1997). Each of the basic space can be visualized in the following way:

Image-based space. There is a number of ways to map images onto a two- or three-dimensional space. For example, a system can compute the most frequently used color in each image or compute the RGB or HBS color coordination values.

Word-based space. To determine the physical location of the impression words, one way is to perform a survey asking people how ‘close’ pairs of words are, and compute the relative distance between each pair of words.

Person-based space. A personality test can be used to map people onto a two- or three-dimensional space.

As described in the next section, the EVIDII system based on this model has taken the HBS representation as one type of visualization using image as the base element. The NCDR-Word space, which is designed by the Nippon Color and Design Research Institute, is chosen for visualizing the word-based space. The person-based space has not been implemented in the current EVIDII system. In what follows, we focus only on the former two basic spaces in accordance with the current implementation of EVIDII.

2.3 Views

Two views each are provided for the image-based and word-based basic spaces. Table 1 summarizes the spaces and views, while Figures 1 and 2 represent the relationships among the basic spaces and views, and how the relationships among the three elements, persons, images and words, are visualized in the EVIDII model.

The image-based space (Figure 1) offers the ‘image-based word view’ and the ‘image-based person view’. In the word view (Figure 1-(a)), one can focus on a certain word, and examine the persons who associated that word to each image. For example, we can examine how everyone associated the word ‘refreshing’ to each image. In the person view (Figure 1-(b)), one can focus on a certain person, and examine the impressions that the person associated to each image. For example, we can examine how Jane associated the words to each image.

In the same way, the word-based space (Figure 2) offers the 'word-based image view' and the 'word-based person view'. In the image view (Figure 2-(a)), one can focus on a certain image, and examine the persons who associated that image to each word. For example, we can examine how everyone associated impressions to Image#10. In the person view (Figure 2-(b)), one can focus on a certain person, and examine the images which are associated to each word by that person. For example, we can examine how Jane associated images to each word.

Table 1 Summary of Spaces and Views in EVIDII

Basic Space	View	Meaning
Image Based	word	who selects which images for a word w?
	person	what words are associated with each image by a person p?
Word Based	image	who selects which words for an image i?
	person	which images are associated with each word by a person p?

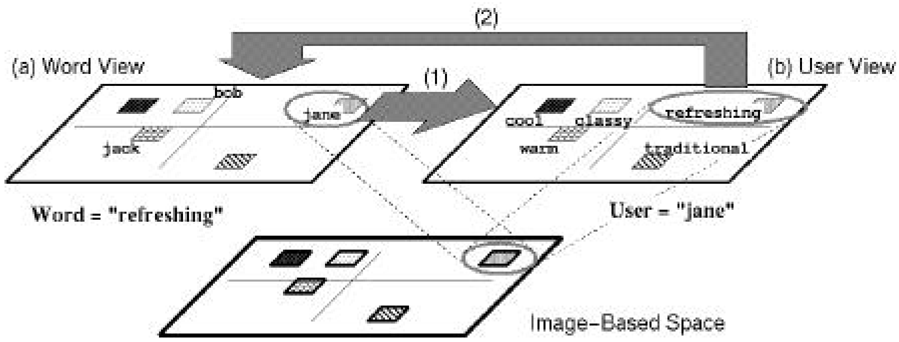


Figure 1 The Image-Based Space and Word- and User-Views. The figure represents the transitions by (1) focusing on the user Jane, and (2) focusing on the word 'refreshing'.

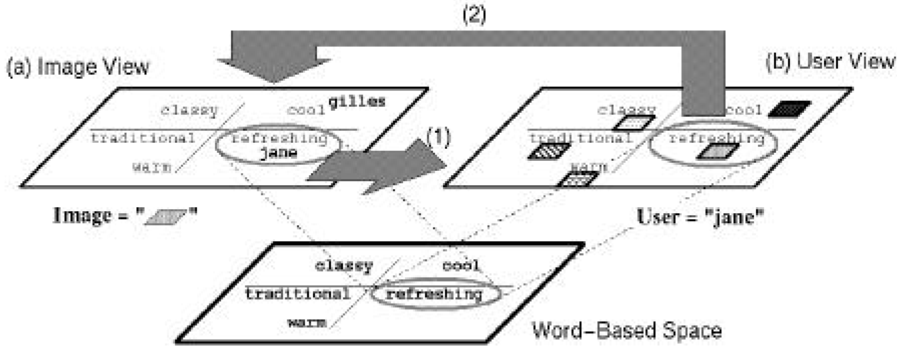


Figure 2 The Word-Based Space and Image- and User-Views. The figure represents the transitions by (1) focusing on the user Jane, and (2) focusing on the image that Jane thought to be ‘refreshing’.

3 THE EVIDII SYSTEM

This section describes the EVIDII system which is based on the EVIDII model. In this section, EVIDII will refer to the EVIDII system rather than the EVIDII model.

3.1 Overview of the EVIDII System

The current implementation of EVIDII supports two basic spaces: image-based space and the word-based space. EVIDII consists of the following components:

- Data collection interface
- HBS space (image-based space)
- Word space (word-based space)

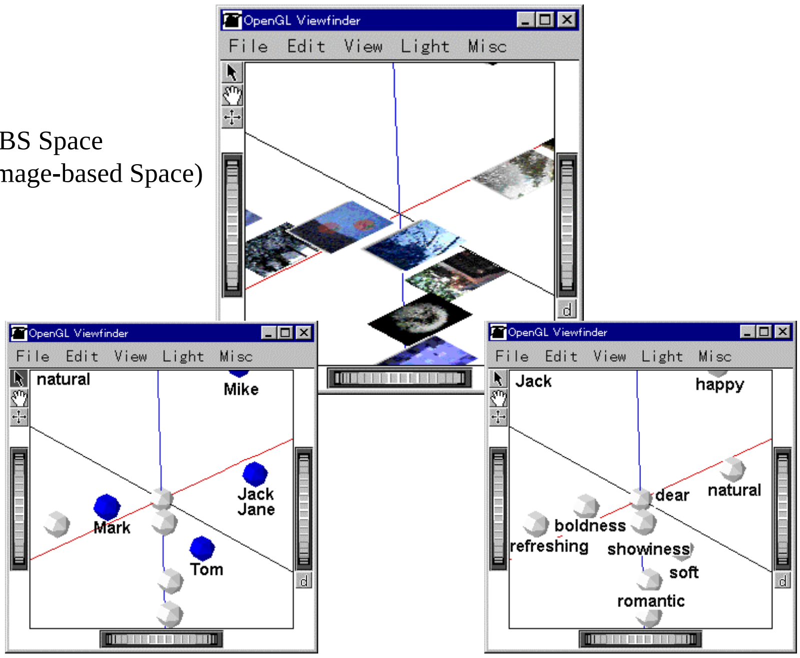
The data collection interface of EVIDII is used to input data that are used in the HBS space and Word space. Given a set of images in GIF format and a set of impression words, EVIDII asks a user to associate words with each image, or vice versa.

The HBS space (Figure 3) and NCDR-Word space (Figure 4) allows a user to explore the space of relationships among persons, images and words. Thumb wheels on the left, right, and bottom of the windows can be used to rotate and move the space to better see the parts of the space that the user wants to see.

Figure 3 presents the HBS space. In the figure, there are twenty images allocated to positions in the three-dimensional space according to the values of Hue, Brightness and Saturation of the color used most in each image (Figure 3-(a)). By selecting one of the affective words in the space, the HBS space displays who associated that word to each image (Figure 3-(b)). Users can go back and forth between the two views. By selecting one of the users, the person view of the HBS-space shows the impressions that the user associated to each image. For example, in Figure 3-(c), Jack’s impressions are displayed in the HBS space.

Figure 4 presents the NCDR-Word-space. 174 words are allocated in the two dimensional space according to the word-space defined by the Nippon Color and Design Research Institute (Figure 4-(a)). The two dimensional space is represented in the cool-warm (x-axis) and soft-hard (y-axis). By selecting an image, the system displays the image-view, which shows who associated which word to this particular image (Figure 4-(b)). Users can go back and forth between these two views as well. By selecting a person (person-view), the Word-space displays what images this person associated with each word (Figure 4-(c)).

(a) HBS Space
(Image-based Space)

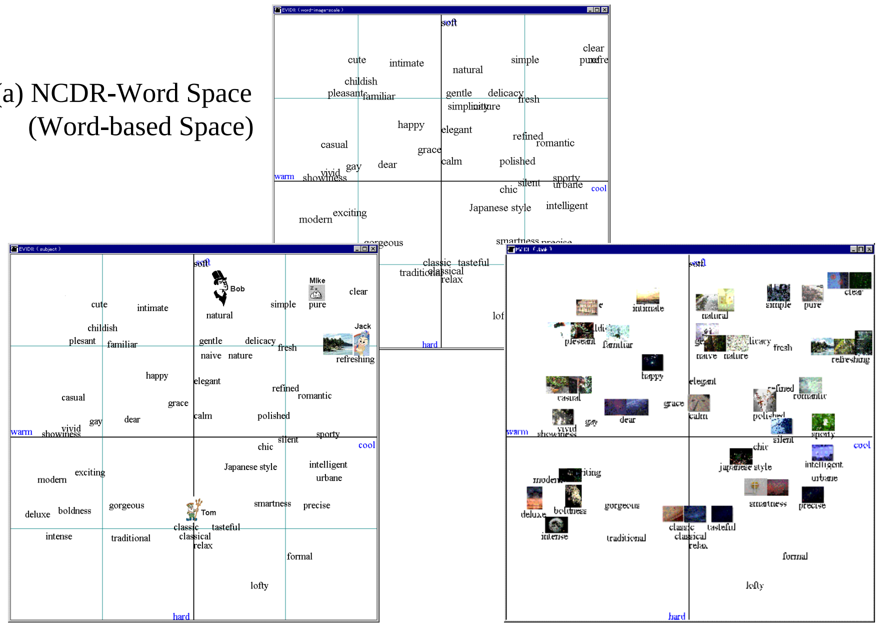


(b) Image-based Word View

(c) Image-based Person View

Figure 3 HBS Space and its Person- and Word-Views of EVIDII.

(a) NCDR-Word Space
(Word-based Space)



(b) Word-based Image View

(c) Word-based Person View

Figure 4: Word Space and its Person- and Image-Views of EVIDII.

3.2 Example Usage Scenario

This subsection presents a typical usage scenario using EVIDII.

Suppose Jane, a Biology student at a university needs to author a home page for her research group retreat. This home page should encourage people to participate in this retreat. Since the retreat is held in a mountain, she first wants to use an image related to nature.

Her research group members had already filled in EVIDII questionnaires to associate images with impressions so that students can use EVIDII to find appropriate images for their home page authoring task. Jane now starts using EVIDII to explore images that are ‘natural’.

First she uses the word view of the HBS space to see what images are thought of as ‘natural’ by whom (Figure 3-(b)). The system shows that Jane and Jack, her senior researcher, selected the same image of a stream in a woods as ‘natural’. She becomes interested in how Jack thinks of other images. She selects the person view of the HBS space for Jack (Figure 3-(c)). Now, EVIDII displays the affective words that are associated with each image by Jack. In the display, she finds an image of a river with a green bank adjacent to the one for ‘natural’, with which Jack associated the word ‘refreshing.’

This reminds her that ‘refreshing’ might be a more appropriate term to represent her intent for her authoring task. She looks at the person view of the NCDR-Word space to see what other images that Jack thought of as ‘refreshing’ (Figure 4-(b)). Jack had selected a couple of other pictures as ‘refreshing’. She likes an image of a meadow with blue sky. She changes to the image view of the NCDR-Word space to see how other people think of this image (Figure 4-(c)). There, she finds that her advisor Bob associated the term ‘natural’ for this image. She becomes quite sure about the image and decides to use that image for her authoring task.

3.3 Discussion

The scenario presented above indicates quite interesting possibilities for the EVIDII system.

- One can know how other people think of an image one is interested in.
- One can explore what image is thought of in a certain way by how many people.
- One can know what tendency images have in terms of the physical characteristics of the image.
- One can find other persons who have a similar or opposite tendency from oneself.

EVIDII supports people to understand what people think of an image. The system provides visual representations through which people can compare differences of impressions rather than using symbolic computation to represent similarity indices using algorithms such as the fuzzy logic or statistics.

Not only allowing people to understand what other people think of images, EVIDII supports users in understanding what they really need. As illustrated with the scenario, an author may not have a clear understanding for a requirement, as Jane first thought that she was looking for a ‘natural’ picture, and later changed her search to a ‘refreshing’ picture. Such situations are not supported in existing Kansei information systems and other types of image retrieval. Systems may be able to retrieve ‘natural’ images for Jane, but cannot support the refinement of her requirements. By allowing users to visually explore relationships among persons, images and words, EVIDII gives users an opportunity to see ‘what other people think of this image.’

Integration of multiple views with multiple element basis and the visual spaces allow people to explore the relationships among persons, images and words through related elements.

4 FUTURE WORK

This paper presented the EVIDII model and the EVIDII system as our approach to support casual users in selecting images in their authoring tasks by visualizing relationships among persons, images and impressions.

Ongoing and future work include:

Distributed EVIDII. We are currently working on implementing EVIDII on Internet using VisualWave. This will enable us to collect through the network data from a large number of people from a variety of cultures, and will allow such people to use EVIDII for their authoring tasks. We may be able to identify cultural characteristics of people-image-impression relationships. We may even be able to visualize culture (not necessarily geographic culture, but age groups, professions, religions, etc.); people from a certain culture think blue images as noble, for example.

Survey-setup Interfaces. When EVIDII is distributed and accessible over the network, users should be able to set up a set of images and affective words to make a survey to produce the person-image-word set. This might be more appropriate for a small group of people working together, for example, to collaboratively design a homepage. One can suggest tens of images and set up an environment for a discussion to develop a shared understanding about the image selection process.

Changeable Scales. Current implementation of EVIDII uses the HBS space and the NCDR-word space suggested by the Nippon Color and Design Research Institute. This, however, does not mean that these spaces are the best ones to visualize the relationships among persons, images and words. Rather, they should be thought of as an instance of mapping schemes. Users should be able to select a space of their choice, for example, users can choose for the image-based space the HBS of the most frequently used color, or that of the RGB, etc. This will allow people to explore which space best represents a certain tendency.

5 CONCLUSION

EVIDII helps people understand what other people might think of a certain image through exploring image-based and word-based spaces. This addresses the issue that people associate different meanings to the same multimedia representations. EVIDII also helps people refine their understanding of the task itself. By exploring words and images associated with each other through other people's eyes, the system supports authors by having them become aware of new words that better describe their task. This word, then can be used to search for images that better match their authoring task.

We depend on the power of multimedia representations by visualizing the relationships. Much of what EVIDII offers do not necessarily have to be visualized. For example, if we wanted to find what other people think about a certain image, a simple list would suffice. Our approach here, however, is to make the best use of multimedia representations by using human perception skill (Norman 1986, Zhang 1997). The power of external representations have been underutilized (Yamamoto 1998). The work presented here is an attempt to use the power of visualization, which is not offered by simply using numbers to represent interdependence between elements. When possible, representation that enhance the understanding should be used. In EVIDII, for example, the image-based space enhances which images are similar in terms of physical attributes, while the word-based space enhances which images are similar in terms of impressions.

More and more people have an opportunity to author multimedia information, which can be seen as a metaphor of the 'computation era' of the past and the 'representation era' of now and the future. Computers are used not just for scientific computation, but also for showing many types of ideas and thoughts. By being faced with the shift from the 'computation era' to the 'representation era', we have to seriously consider what multimedia representations best represent the task at hand. This paper presents our approach to address the challenge not on how to design effective multimedia information, but on how to understand what multimedia representation effectively reflects the intention.

6 REFERENCES

Forsythe, C., Grose, E. and Ratner, J. (1998) Human Factors and Web Development, Lawrence Erlbaum Associates, Inc., Publishers, Mahwah, NJ.

Fischer, G. and Nakakoji, K. (1991) Empowering Designers with Integrated Design Environments, *Artificial Intellegence in Design '91*, 191-209.

Hasegawa, T. and Kitahara Y. (1997) Basic Concept of Multimedia Kansei Synthesis Method and Evaluation of Its Experimental System, *Trans. of Information Processing Society of Japan*, **38-8**, 1517-1530 (in Japanese).

Inoue, M., Tanaka, S., Ishiwaka, M. and Inoue, S. (1996) Influence of Color-filtering on Impressions from Natural Picture, *Technical Report of IEICE*, **PRMU96-62**, 25-30 (in Japanese).

Isomoto, Y. and Nozaki, H. (1996) Application Method of Fuzzy Thesaurus to Express Sensitive Heart, *Technical Report of IEICE*, **ET96-63**, 25-32 (in Japanese).

Kurita, T., Kato, T., Fukuda, I. and Sakakura, A. (1992) Sense Retrieval on a Image Database of Full Color Paintings, *Trans. of Information Processing Society of Japan*, **33-11**, 1373-1383 (in Japanese).

Murch, G.M. (1984) Physiological Principles for the Effective Use of Color, *IEEE Computer Graphics and Arts,* 49-54.

Nakakoji, K., Aoki, A. and Reeves, B.N. (1996) Knowledge-Based Cognitive Support for Multimedia Information Design, *Information and Software Technology*, **38-3**, 191-200.

Simon, H.A. (1981) The Sciences of the Artificial, MIT Press.

Yamamoto, Y., Takada, S., Gross, M.D. and Nakakoji, K. (1998) Representational talkback: An approach to support writing as design, *Proceedings of the APCHI '98 Conference* (to appear).

Zhang, J. (1997) The nature of external representations in problem solving, *Cognitive Science*, **21-2**, 179-217.

7 BIOGRAPHY

Kumiyo Nakakoji, a research fellow at PRESTO Japan, is an Adjunct Associate Professor for the Cognitive Science Laboratory at Nara Institute of Science and Technology. Her research interests include considerations of culture, communication and creativity in systems and design. She has worked for Software Research Associates, Inc. (Tokyo, Japan) for the last twelve years as a senior researcher at the Software Engineering Laboratory. She received her BS from Osaka University (1986), and MS (1990) and PhD (1993) in Computer Science from University of Colorado at Boulder.

Yasuhiro Yamamoto is a PhD student at Nara Institute of Science and Technology. His research interests include computer support for writing as design. He received his BS from Kyoto University (1996) and MS from Nara Institute of Science and Technology (1998).

Kimihiko Sugiyama currently works for Mitsubishi Electric Corporation. His research interests include Kansei information systems. He received his BS from Seikei University (1996) and MS from Nara Institute of Science and Technology (1998).

Shingo Takada is a Research Associate for the Cognitive Science Laboratory at Nara Institute of Science and Technology. His research interests include information search and object-oriented technology. He received his BS (1990), MS (1992), and PhD (1995) in Computer Science from Keio University.