PAPER

The Efficiency of Various Multimodal Input Interfaces Evaluated in Two Empirical Studies

Xiangshi REN[†], ***Regular Member***, **Gao ZHANG**[††], ***and*** **Guozhong DAI**[†††], ***Nonmembers***

SUMMARY Although research into multimodal interfaces has been around for a long time, we believe that some basic issues have not been studied yet, e.g. the choice of modalities and their combinations is usually made without any quantitative evaluation. This study seeks to identify the best combinations of modalities through usability testing. How do users choose different interaction modes when they work on a particular application? Two experimental evaluations were conducted to compare interaction modes on a CAD system and a map system respectively. For the CAD system, the results show that, in terms of total manipulation time (drawing and modification time) and subjective preferences, the "pen + speech + mouse" combination was the best of the seven interaction modes tested. On the map system, the results show that the "pen + speech" combination mode is the best of fourteen interaction modes tested. The experiments also provide information on how users adapt to each interaction mode and the ease with which they are able to use these modes.

key words: *multimodal interface, mode combination, interface evaluation, interface efficiency*

1. Introduction

Multimodal user interfaces (MMIs) have been proposed to make full use of the user's sense and motion channels (or modalities), e.g. gesture, speech, pen, touch and so on [3]. Multimodal interfaces have long been considered as alternatives and as potentially superior. Many studies of multimodal user interfaces have been reported for tasks such as text processing [14], map-based tasks [8], [12], and in the communication environment [10]. Studies on speech and keyboard input [4], [6], mouse and speech input [2], [9], speech and gesture input [1], [7] have also been conducted. These studies did not compare all reasonable combination modes, such as uni-modal and tri-modal combinations.

Researchers generally believe that a multimodal interface is more effective than a unimodal interface, e.g. Hauptmann et al. (1989) [7] who observed a surprising uniformity and simplicity in the user's gestures and speech, and Oviatt et al. (1997) [12] who reported that users overwhelmingly preferred to interact multimodally rather than single-modally. However, little quantitative research and evaluation has been reported on multimodal combinations which would certify that one combined input mode is more natural and efficient than another in particular environments, e.g. a map system, a CAD system etc.. Suhm et al. (1999) [13] asked the question "which modality do users prefer?," however, the authors only answer with reference to single modes. They did not state which is the most effective mode, nor did they report on combined modes.

This study evaluates the differences in modalities and their combinations through usability testing on a CAD system and a map system respectively. We look at how interaction modes are adapted to different applications. We are interested in what is the most effective modality that users prefer in a given application. We also seek to provide information on how users choose different interaction modes when they work on an application.

2. Experiment One: A CAD System

A CAD system is a complex interactive system, with which users can draw graphic objects with a mouse, choose and drag objects and tools, select colors and other properties from menus or dialogue boxes, and manage their plans through the file system. We consider that a CAD system is an application-oriented system, and the study of the usability of its multimodal interface ought to be based on an applied CAD system. Furthermore, we give consideration to all manipulations including drawing, location, modification and property selection (thickness/color etc.).

Driven by this, we set up a multimodal user interface environment on an AutoCAD system, where users could use pen, speech, and/or mouse to draw an engineering plan. Based on this environment, we designed an evaluation experiment to investigate the differences in these modalities and their combinations.

Besides the traditional mouse, users can use pen and speech to draw a plan in this multimodal user interface environment. They can draw graphic objects with a pen in a more intuitive way than with a mouse. They can also simultaneously select color and line width properties through the speech modality without manu-

Manuscript received July 24, 2000.
Manuscript revised April 17, 2001.

[†]The author is with the Department of Information Systems Engineering, Kochi University of Technology, Kochi-shi, 782–8502 Japan.
[††]The author is with Microsoft Research, China, 5F, Beijing Sigma Center, Zhichun Road, Beijing 100080, China.
[†††]The author is with the IM&D, Software Institute, Chinese Academy of Sciences, P.O. Box 8718, Beijing 100080, China.

ally selecting complex menus or dialogue boxes. Thus the drawing procedure can continue uninterrupted.

2.1 Method

2.1.1 Participants

Twenty-four subjects (20 male, 4 female, all right handed; 14 students, 10 workers) were tested for the experiment. Their ages ranged from twenty to thirty five years. Ten of them had had previous experience with AutoCAD systems, the other had no experience, but they could all use the mouse and keyboard proficiently. Only five of them had had previous experience with the pen used in the experiment.

2.1.2 Apparatus

The hardware used in the experiment was a pen-input tablet (WACOM), a stylus pen, a microphone, and a personal computer (P166, IBM Corp.). The software used in the experiment was Windows 95, AutoCAD 12.0 for Windows, a drawing recognition system which we developed and a speech recognition system (CREATIVE Corp.).

2.1.3 Design

We did not use a keyboard in the experiment because the task was only drawing. There are seven possible interface combinations for a mouse, a pen, and speech: mouse, pen, speech used individually, mouse + pen, mouse + speech, pen + speech, pen + speech + mouse. Obviously, speech-only cannot accomplish the drawing tasks efficiently.

In order to simplify the experiment, we performed a preliminary experiment to compare the difference between the use of the mouse and the use of the pen in the CAD system. The procedure in this pre-experiment was the same as the one described in Sect. 2.1.4. The result showed the pen was suitable for drawing the outline of the plan. The pen's efficiency and subject preference rating were better than the mouse's but the pen was not as accurate as the mouse. We inferred from this result that the pen + speech mode was better than the mouse + speech mode for outlining and we therefore omitted tests for the mouse + speech mode. Furthermore, the frequent change between mouse and pen takes a lot of time and is not convenient. We therefore assigned the pen to drawing tasks and the mouse to modification tasks and we made speech simultaneously available to both pen and mouse operations.

Thus, in order to investigate the differences between different input modes and their combined use, each subject tested four modes: the mouse, pen, pen + speech, and pen + speech + mouse modes. The use of pen and mouse in the pen + mouse + speech mode interface eliminated frequent changing between mouse and pen. The mouse was used as a supplemental device to the pen at the modification stage to ensure accuracy.

2.1.4 Task and Procedure

First the experiment was explained to each of the subjects, who were each given 30 minutes to learn how to use the pen and the speech input equipment. The CAD system chose one of the four interface modes randomly and showed the corresponding instruction information on the title bar of the AutoCAD system. After receiving the mode information, a dialogue box with three buttons appeared: "beginning to draw," "beginning to modify," and "finishing drawing." The subject chose "beginning to draw" to begin the test.

The AutoCAD system was opened and a sample plan appeared on the screen. This plan was selected as the test object and appeared as a sample, which could not be altered by the users. In the test, the subject tried his/her best to match the sample plan. In order to establish the drawing time and modification time, the subject was not allowed to modify the drawing before he/she had finished all of the drawing. After finishing all of the drawing, the subject could choose "beginning to modify" to begin the modification stage. After finishing all the modifications, the subject could choose "finishing drawing" to finish the current test.

The subject was asked to do six tests for each interface mode. The first was a practice and the results of the other five were recorded as formal tests. Each subject had to test all four modes. Whenever they finished a test, they were allowed to have a rest.

Data for each interface mode was recorded automatically as follows: (1) The time taken to draw the plan: This is the time lapsed from the selection of the "beginning to draw" button to the selection of the "beginning to modify" button. (2) The time taken to modify the plan: The time lapsed between selection of the "beginning to modify" button and the selection of the "finishing drawing" button. (3) Accuracy of drawing: At the beginning of each test, the system provided each subject with a background paper. During the test, the subjects were told to trace the background drawing. When the drawing was finished, the system calculated the matching percentage between the background paper and drawing paper. We reckoned the matching percentage to be the degree of accuracy. (4) Subject preference: The subjects were questioned about their preferences after they finished testing each interface mode. They were asked to rank (on a scale of 1–10) the mode just tested according to their satisfaction with the mode and their desire to use that mode.

2.2 Results

We performed an ANOVA (analysis of variance)† with repeated measures on the within-subject factors on the interface modes used, with drawing time, modification time, total time (drawing time + modification time), accuracy, and subjective preference as dependent measures. Accuracy was calculated according to the matching rate between the sample plan and the user's drawing plan.

2.2.1 Pen + Speech + Mouse Mode

The mean drawing time for each of the four interface modes is shown in Fig. 1. The pen + speech mode was faster than the other three in drawing time (mean = 9.9 minutes), $F(3,92) = 185.97$, $p < 0.0001$, however, the pen + speech + mouse mode was faster than the other three in modification time (mean = 5.0 minutes), $F(3,92) = 145.06$, $p < 0.0001$, in total time (mean = 15.0 minutes), $F(3,92) = 35.44$, $p < 0.0001$.

The mouse-based interface was the most accurate (mean accuracy = 91.6%), $F(3,92) = 136.88$, $p < 0.0001$.

The pen + speech + mouse mode also had the highest satisfaction rating by the subjects (mean = 7.8, see Fig. 2), $F(3,92) = 7.18$, $p < 0.0001$.

2.2.2 Location and Modification Issues

The results show that the pen-based interface was slower than the mouse interface for modification, $F(1,46) = 240.47$, $p < 0.0001$. The pen-based interface was less accurate than the mouse interface, $F(1,46) = 515.8$, $p < 0.0001$. The subjective rating results also show that the pen-based interface was not as satisfactory as previously thought.

We noted that users took a lot of time and energy learning to use the pen with the complex menus, especially the 19 novices (out of 24 subjects) who had no experience with the pen. However, the pen + speech mode was faster than the pen-based mode in total time, $F(1,46) = 91.46$, $p < 0.0001$.

Regarding accuracy, a significant difference was found between the pen + speech mode and the pen-based mode, $F(1,46) = 86.14$, $p < 0.0001$, the results show that the combination of pen + speech was more accurate than the pen on its own. Subjective preferences show that the pen + speech mode had higher ratings (mean = 7.29) than the pen-only mode (man = 6.54), $F(1,46) = 8.41$, $p < 0.005$.

2.3 Discussion

We presented an evaluation experiment based on the applied CAD system. The experimental results show that the pen + speech + mouse mode was the best of the seven interaction modes on the CAD system, in terms of modification time, total time, and subjective evaluation. In particular the pen + speech + mouse mode was faster than the pen + mouse mode in total drawing and modification time, $F(1,46) = 96.77$, $p <$

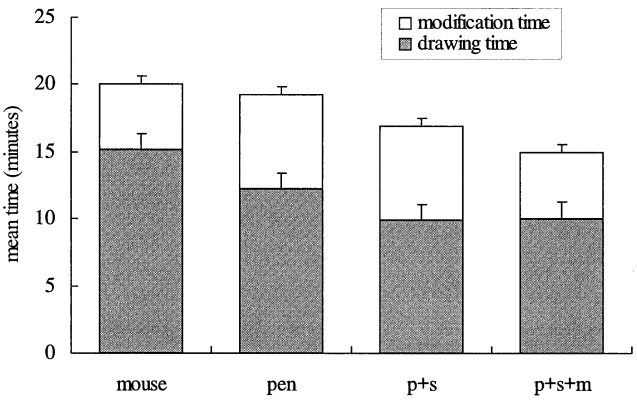


Fig. 1 Means (with standard error bars) for drawing time and modification time for the four interaction modes (p: pen; s: speech; m: mouse).

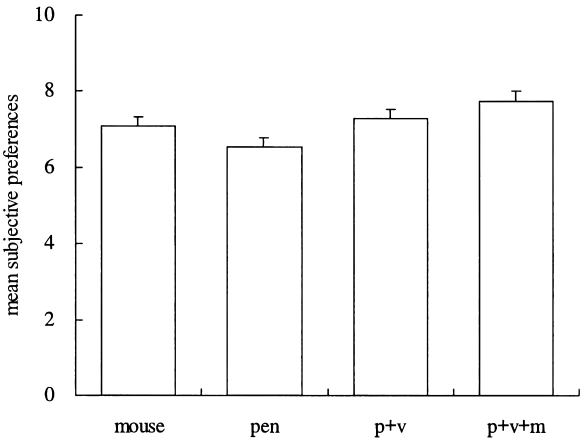


Fig. 2 The subjective preferences on a scale of 1-to-10 for the four interaction modes (1 = lowest preference, 10 = highest preference).

†For readers who are not familiar with the notation, ANOVA, ANnalysis Of VAriance, is a method for analyzing variations between groups especially three or more groups. It is thus a more meaningful way to evaluate data than a comparison of simple averages. It tells us whether the variation of multiple groups is significant by putting all the data into one number "F ratio" and providing one "p" for null hypothesis. $F(m-1, n-m)$ = (found variation of the group averages) ÷ (expected variation of the group averages). "m" is the total number of groups (4 in Experiment One and 6 in Experiment Two). "n" is the total number of leaves (96 in Experiment One and 138 in Experiment Two). "p" reports the significance level, which indicates the probability that these differences could have occurred by random chance alone.

0.0001. We do not only show that the multimodal interface is better than the unimodal interface for CAD systems but we also show the results of comparisons between combined modes.

The results also show that a proper combination of input modes can improve the interactive efficiency and user satisfaction of CAD systems. The pen was suitable for drawing the outline of the plan. The mouse was useful for accurate modification procedures. Speech was suitable for inputting the descriptive properties of graphic objects and selecting menus.

Many studies suggest the use of new interactive modes such as the pen or speech to replace the traditional mouse and keyboard. However, based on our experiment, the mouse is still useful for modification and location procedures because location and modification with the pen may not be accurate enough. We recommend the pen for outlining, initial location and layout procedures and the mouse for modification because it is more accurate. We should pay more attention to location and modification technology. In the meantime, we suggest that the traditional mouse continue to be used for modification procedures in CAD systems.

3. Experiment Two: A Map System

Map systems are usually used in public places. They require a more convenient user interface. Some multimodal interactive methods, such as spoken natural language, gesture, handwriting, etc. have been introduced into map systems to produce more natural and intuitive user interfaces [5], [11].

We set up a prototype multimodal map system where users can use pen, speech (spoken natural language, Chinese), handwriting (handwriting of Chinese characters), pen-based gestures (drawing graphics with a pen), as well as mouse and keyboard modes (typing, selecting, and dragging), to accomplish a number of trip plan tasks, e.g. to get the necessary information to plan their travel routes. For this environment, we designed an experiment to investigate which is the best of the different combination modes.

3.1 Method

3.1.1 Participants

Twenty-four subjects (12 male, 12 female, all right handed; 12 students, 12 workers) were tested for the experiment. Their ages ranged from twenty to thirty-five. None of them had had any experience in using this kind of trip plan system.

3.1.2 Apparatus

The hardware was the same as used in Experiment One (see Sect. 2.1.2). The software was Windows 95, a drawing recognition system (developed by us) and a speech recognition system (CREATIVE Corp.).

3.1.3 Design

We used the keyboard (as well as the mouse, pen and speech) because we considered that the keyboard was useful for information retrieval. Thus the possible combinations for keyboard, mouse, pen and speech are: mouse, keyboard, speech, pen used individually, mouse + keyboard, mouse + speech, mouse + pen, keyboard + speech, keyboard + pen, speech + pen, m + k + s, m + k + p, k + s + p, s + p + m.

We designed the experiment in two steps. Step One was to exclude modes and combinations seldom-used by the subjects. The mean success rate was $1/14 = 7\%$, we treated those combinations with success rates above 7% as useful combination modes. Step Two was to more accurately compare the differences between useful multi-modal combinations using combination modes above 7%.

3.1.4 Task and Procedure

There were four classes of task in the map system: distance calculation; object location; filtering; information retrieval. All these tasks could be accomplished by multiple modal combination modes.

In Step One, each of the subjects had 30 minutes to learn how to use the input equipment (mouse, keyboard, pen, and speech) to accomplish trip plan tasks. After they were familiar with the experimental environment, the experiment began. The subjects were asked to accomplish four tasks for each mode combination selected. The system allowed 10 minutes for each class of task. All possible mode combinations were given randomly to the subjects to perform. The subjects were asked to perform the task as soon as possible by use of the appointed mode combination. If the subjects accomplished the appointed task in a given time (20 seconds), an automatic program running in the background recorded the performance time and procedure. If the task took longer than twenty seconds the testing system assumed that the user had failed to perform this task.

In Step Two, the subjects were asked to accomplish 24 tasks for each mode combination selected. The order of the 24 tasks was randomly assigned in each combination mode. Data for each mode combination was recorded automatically as follows. The program running in the background automatically recorded the performance time. This was the time lapsed from the beginning of the first task to the end of the last task. The subjects were questioned about their preferences after they finished testing each mode combination. They were asked to rank (on a scale of 1–10) the mode just

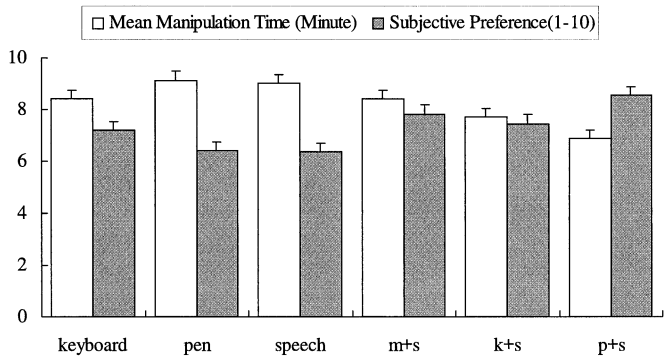


Fig. 3 Means (with standard error bars) for manipulation time and subjective preference (1 = lowest preference, 10 = highest preference) for the six interaction modes (p: pen; s: speech; m: mouse, k: keyboard).

tested according to their satisfaction and their desire to use it.

3.2 Results

An ANOVA (analysis of variance) was also conducted on the results of the experiment. The manipulation efficiency and subjective preference differences of these multimodal interactive modes were compared.

3.2.1 Useful Combination Modes

From Step One, we analyzed useful combination modes for accomplishing the trip plan tasks. The statistical results showed that in single-modality mode, the success rate for the mouse was 1%, for the keyboard it was 10%, speech 15% and pen 12%.

In bi-modality mode, the success rate for mouse + speech was 13%,, keyboard + speech 11% and speech + pen 18%. The success rate for mouse + keyboard was 5%, mouse + pen 3%, and keyboard + pen 5%. The tri-modality modes were seldom used with success rates of less than 3%.

We therefore chose the following six modes for further testing in Step Two: keyboard, speech, pen, mouse + speech, keyboard + speech, pen + speech.

3.2.2 Pen + Speech Mode

Significant differences in the six modes were found in mean manipulation time, $F(5,138) = 105.6$, $p < 0.0001$. The results revealed the pen + speech combination was faster than the other five modes in total time (mean manipulation time of pen + speech was 6.8 minutes, see Fig. 3). On the other hand, the pen-only interface was the slowest among the six modes with a mean manipulation time of 9.1 minutes.

There was also a significant difference in the six modes for subject preferences, $F(5,138) = 105.6$, $p < 0.0001$. The pen + speech mode had the highest satisfaction rating (mean = 8.5, see Fig. 3). The speech-only mode had the lowest satisfaction rating (mean = 6.3).

Based on the analyses, the pen + speech combination was the best of the six interaction modes.

3.3 Discussion

Regarding individual modes, the success rate for the mouse was 1%, the keyboard was 10%, the speech was 15%, the pen was 12%. This reveals that the mouse (though it is accurate) is not suitable for trip plan tasks. In combination modes (pairs), speech plays an important supplemental role in trip plan tasks (mouse + speech 13%, keyboard + speech 11%, pen + speech 18%). All tri-modality modes rated less than 3%. Tri-modality modes were seldom used.

Overall, the pen + speech combination was the best of the fourteen interaction modes.

On the other hand, in Experiment One, we show that the mouse is still useful for modification and location procedures in CAD systems, however, map systems do not need a mouse because these kinds of systems do not call for accurate fine tuning.

This experiment shows that modes used in pairs were better than tri-modality modes. The results also show that more modes may not necessarily be better than less modes, e.g. the pen + speech + mouse mode was better than the pen + speech mode on CAD systems, however, the pen + speech mode was better than the tri-modality modes on map systems. The optimal number of combined modes for each environment should be investigated.

4. General Discussions and Conclusions

This paper provides information on how users choose different interaction modes that are available currently when they work on a tangible application. Obviously, an interaction mode which is designed for a particular environment will take into account the most suitable software applications for that environment, and future developments in software applications. One voice pack-

age may be suitable in one situation while another will be more suitable in a different environment. This is a matter of specific design for specific environments. We have attempted to establish an evaluation method which will help designers determine the best combination of modes for their particular environment. By substituting various (e.g. voice) applications when testing interaction modes using our method the designer will more accurately be able to choose the best application in each species (voice, mouse, pen etc.). We have also established a general rule that one or other combination of modes is more or less efficient than another.

First, the pen + speech + mouse mode was the best of seven interaction modes tested on CAD systems. The pen + speech mode was the best of fourteen interaction modes tested on map systems. Second, we also show that more combination modes may not necessarily be better than less modes. Third, our tests for the first time give statistical support to the view that the mouse is still useful for accurate modification and location procedures especially in multimodal interfaces for CAD systems. Finally, we contributed to the body of information about how users adapt to each interaction mode, and the ease with which they are able to use them.

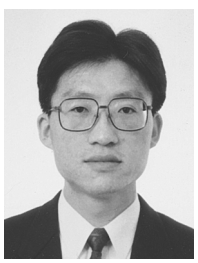
These tests included keyboard, mouse, pen, and speech. Other interaction modes such as gaze input and touch, should also be tested in combination modes.

Acknowledgements

This study is supported by the key project of National Natural Science Foundation of China (No.69433020) and the key project of China 863 Advanced Technology Plan (No. 863-306-ZD-11-5).

References

[1] H. Ando, H. Kikuchi, and N. Hataoka, "Agent-typed multimodal interface using speech, pointing gestures and CG," Symbiosis of Human and Artifact, pp.29–34, 1995.
[2] M.M. Bekker, F.L.van Nes, and J.F. Juola, "A comparison of mouse and speech input control of a text-annotation system," Behaviour & Information Technology, vol.14, no.1, pp.14–22, 1995.
[3] R.A. Bolt, "The integrated multi-modal interface," IEICE Trans., vol.E70, no.11, pp.2017–2025, Nov. 1987.
[4] R.I. Damper and S.D. Wood, "Speech versus keying in command and control applications," International Journal of Human-Computer Studies, no.42, pp.289–305, 1995.
[5] A. Cheyer and L. Julia, "Multimodal maps: An agent-based approach," SRI International, 1996, http://www.ai.sri.com/čheyer/papers /mmap/ mmap.html.
[6] M. Fukui, Y. Shibazaki, K. Sasaki, and Y. Takebayashi, "Multimodal personal information provider using natural language and emotion understanding form speech and keyboard input," The Special Interest Group Notes of Information Processing Society of Japan, vol.64, no.8, pp.43–48, 1996.
[7] A.G. Hauptmann, "Speech and gestures for graphic image manipulation," Proc. CHI '89 Conference on Human Factors in Computing Systems, pp.241–245, 1989.
[8] L. Julia and A. Cheyer, "A multimodal computer-augmented interface for distributed applications," Symbiosis of Human and Artifact, Elsevier Science B.V., pp.237–240, 1995.
[9] J.J. Mariani, "Speech in the context of human-machine communication," ISSN-93, pp.91–94, 1993.
[10] T. Ohashi, T. Yamanouchi, A. Matsunaga, and T. Ejima, "Multimodal interface with speech and motion of stick: CoSMoS, symbiosis of human and artifact," Elsevier B.V. pp.207–212, 1995.
[11] S. Oviatt, "Toward empirically-based design of multimodal dialogue system," Proc. AAAI 1994-IM4S, pp.30–36, Stanford, 1994.
[12] S. Oviatt, A. DeAngeli, and K. Kuhn, "Integration and synchronization of input modes during multimodal human-computer interaction," Proc. CHI '97 Conference on Human Factors in Computing Systems, pp.415–422, 1997.
[13] B. Suhm, B. Myers, and A. Waibel, "Model-based and empirical evaluation of multimodal interactive error correction," Proc. CHI '99 Conference on Human Factors in Computing Systems, pp.584–591, 1999.
[14] S. Whittaker, P. Hyland, and M. Wiley, "Flochat: Handwritten notes provide access to recorded conversations," Proc. CHI '94 Conference on Human Factors in Computing Systems, pp.271–277, 1994.



Xiangshi Ren is currently an assistant professor in the Department of Information Systems Engineering at Kochi University of Technology. He received a B.E. degree in electrical and communication engineering, M.E. and Ph.D. degrees in information and communication engineering from Tokyo Denki University, Japan, in 1991, 1993 and 1996 respectively. He was an instructor in the Department of Information and Communication Engineering at Tokyo Denki University during 1996–1999. His research interests include all aspects of human-computer interaction, in particular, multimodal interactions and usability. He is a member of the IPSJ and the Human Interface Society, both in Japan, the ACM, the ACM SIGCHI, the IEEE Computer Society, and the British HCI Group.

Gao Zhang is currently an Associate Researcher at Microsoft Research China. He received B.E. and M.E. degrees in Software at HuaZhong University of Science and Technology, Ph.D. degree in software at Software Institute, Chinese Academy of Sciences. His research interest is Software User Interface.

Guozhong Dai is a professor of the Institute of Software, Chinese Academy of Sciences. He received a B.S. degree in mathematics from University of Science & Technology of China. His current research areas include computer graphics and human-computer interaction.