AUDITORY DISPLAYS ON THE DEPTH OF HYPERTEXT

Ju-Hwan Lee, Myoung-Hoon Jeon, Min-Sun Kim

Cognitive Engineering Lab.
Dept. of Psychology, Yonsei University
Seoul, 120-749 KOREA
leejuhwan@yahoo.com philart77@hotmail.com kimmin0414@hotmail.com

ABSTRACT

This study presents an overview of work on the effects of auditory displays on the navigation of hypertext and the results of 2 new experiments. In the present study, auditory information was pertinently mapped with the depth dimension of structural hierarchy in the hypertext. The general finding of the experiments is that auditory displays have the possibility of the redundant effect facilitating the performance on target search tasks, especially difficult target search situations in website navigation. These findings will be discussed and related to the theoretical framework of multimodality effect in the information processing theories.

1. INTRODUCTION

In everyday lives, we are asked to process overflowing information with visual sense but in fact we unconsciously handle the information with other senses as well as visual one. Imagine that you make out a crucial report in the office room. Knocking at the door gives you information that someone visits you. A crying coffee pot tells that it is coffee break. R-r-ring sends you to answer the telephone. Besides, these events would occur simultaneously. In a simple situation that does not exceed the cognitive capacity, it might be enough with visual sense. However, in a complex and difficult situation that is beyond attentional limitation as above, we have to depend on various senses. Because human perception of visual information is independent from that of other information [1], using multimodality decreases the cognitive burden and improves the performance. We can also overcome cognitive load in systems by applying multimodality to user interface. Such multimodal interface would allow a richer and more natural communication between the computer and the user [2].

In spite of these potential advantages of the auditory display, user interface research has tended to focus on the visual modality due to technical problems. But recently, some researchers try to add the auditory information to user interface. Brown, Newsome and Glinert [3] performed visual search experiments using auditory cues. Their aim was to reduce visual workload by using multiple sensory modalities. The experiments showed that the auditory modality could be partly more effective than the visual one. Gaver [4] and Kramer [5] have pointed out that non-speech auditory cues can supplement visual interfaces by increasing a user's sense of engagement, enhancing monitoring of background processes, and reinforcing visually represented information. Maranda and Celestine [6] also showed that the effects of multimodal display aids on human performance. The result of the experiments indicates that auditory cues positively enhance the detection of critical conditions for certain types of visual indicators. Brewster et al. [2, 7, 8] made broader researches about auditory cues. He conducted many experiments in which he added auditory cues on buttons, scroll bars, menus, alert boxes, windows, and drag and drop. In these experiments, he verified that auditory feedback improves the performance and usability. Moreover, he investigated subjective annoying rate in auditory condition and there is no significant difference from visual one. These preceding studies provide a view that the redundant auditory information is more effective and improving usability.

In addition, Gaver [9] categorized auditory cues into two types, auditory icons and earcons. Auditory icons are everyday sounds we hear usually, while earcons are artificially synthetic sounds changing other properties such as pitch, loudness, timbre and rhythm.

In the aspect of applying the auditory cues, the Web is considered as a niche for utilizing auditory displays for ameliorating the performance and experience in the World Wide Web (WWW), a collection of hypermedia information distributed across the Internet. The reason, as state above, is why the multimodal displays applied in the user interfaces such as a web browser help acquire and process information increasing in geometric progression. Albers and Bergman [10] have mentioned that the highly visual task of scanning through text for links suggested that 'Mosaic' users could benefit from auditory. Their aim was to provide more information modality by using auditory rather than visual enhancements while shifting additional cognitive load to a different. Albers and Bergman have argued that the Audible Web uses auditory cues to enhance the interaction between users and Mosaic in three ways: aiding users' monitoring of data transfer in progress, by providing feedback for user actions, and by providing content feedback to aid navigation.

In the present study, auditory feedback was applied to hypertext on web site. The hypertext would ask much more complex processes of users than simple menu structure. Users need to pay more attention to perform these tasks. In the hypertext navigation, it is important to recognize the depth position where users are located. Yet users are apt to miss the visual cues to their location and lose their location as they rapidly move between pages and pages by using links. If auditory cues were given users proper information for a certain structural position, they could tacitly recognize their present location and plan the next destination. Since auditory information is innately temporal and is fit to detect the changing events, it may be advantageous in the hypertext which implies incessant changes.

Moreover, we expect that the facilitating effect of task-specific sounds would emerge from a certain level on the degree of task-difficulty. It will affect more in a relatively difficult task that make users less understandable about its structure than in an easy task. The present study includes the interaction between the degree of task-difficulty and the redundant effect of the auditory displays.

Following the preceding studies [2, 7, 8, 11, 12], we designed and applied some sounds to aid users to navigate tremendous contents in the website by changing pitch and rhythms according to the structural hierarchy. In the meanwhile we used the sounds as earcons to provide additional information to visual displays.

2. EXPERIMENT 1

In previous researches with auditory display on human-computer interfaces, it has been shown that redundant auditory information leads to improving the usability of human-computer interfaces when compared to a situation with no auditory information. These results support that the multimodality of information display is more effective than the use of a single modality.

In Experiment 1, participants were provided some shopping targets on website made of hypertext and asked to search them as quickly as possible. The targets were divided into two levels according to the degree of difficulty in the website. This experiment hypothesized that task performance would be better in redundant display setting with visual and auditory than visual only.

2.1. Method

2.1.1. Participants

Sixteen university students (10 male, 6 female; mean age 23.4 years) with normal or corrected-to-normal vision and hearing served as participants. None were aware of the hypothesis being tested.

2.1.2. Apparatus and stimuli

Participants sat alone at a table with a Pentium PC, located in a laboratory room. The computer monitor was a 17 inch CRT display set to a resolution of 1024 x 768 pixels, with 32-bit color. The listeners wore Panasonic RP-HS40 headset, adjusted for fit and comfort, and used a standard two-button wheel mouse to respond. The frequency response of the headset was 14 Hz-24 kHz.

Figure 1. *Image of the screen layout seen by participants in experiments. The experiment was written in HTML, JavaScript and Visual Basic 6.0, and run in an Internet Explorer 6.0 window.*

The four sound stimuli were synthesized at 16-bit, 44.1 kHz using ‘Csound’, and saved in WAV format. Each stimulus was composed of a one-beat long pure sine wave tone. The sound stimuli were a sound at pitch C4 (261.63 Hz), C5 (523.25 Hz), C6 (1046.50 Hz) each with the duration of 200ms and a combination of three sounds from C4 to C6 each with the duration of 100ms. The intensity of stimuli was 85dB. Each sound displayed a given position of hierarchical structure in hypertext, that is, four sound stimuli at C6, C5, C4, and a combination of three sounds mapped respectively the first depth level from the main page, the second depth level, the third depth level, and a return to the main page on the experimental website.

When the mouse button was pressed down over a graphical link button a given sound at pitch mapped with the depth dimension of hypertext was played.

2.1.3. Design and procedure

A 2 x 2 (modality by the degree of difficulty) repeated-measure design was used. The modalities were visual display only and redundant display (visual and auditory); the degree of difficulty was easy or difficult. The 5 trials in each of the 4 conditions were ordered randomly for each participant, forming a block. Each participant performed 4 blocks for a total of 20 trials and the order of blocks was counterbalanced.

At the beginning of the experimental session, participants heard a set of instructions that explained the procedure for the target search task. Participants were provided with a written target product below the screen. Participants were seated approximately 60cm from a computer monitor. At the beginning of each trial, the main page of an internet shopping website[1] redesigned for experiment appeared on the computer monitor,

[1] Website address: http://www.interpark.com

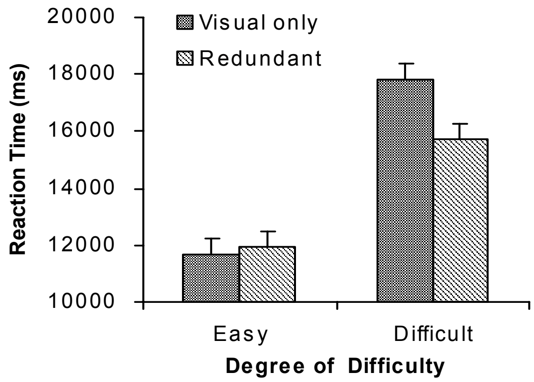

Figure 2. *Results from Experiment 1 showing the mean reaction times for performing easy and difficult target search tasks at the degree of difficulty in the visual only and redundant conditions.*

which served as a ready signal. At this time, if participants were ready, participants pressed the start button, and then searched a presented target as quickly as possible. Each trial was completed by pressing the finish button to measure the reaction time in millisecond.

2.2. Results and Discussion

The results of Experiment 1 are shown in Figure 2. An analysis of variance (ANOVA) revealed a main effect for difficulty level, $F(1, 15) = 34.36$, $MSe = 11503929.57$, $p = .00$. It appeared that two difficulty levels were well divided. There was no similar main effect for modalities, $F(1, 15) = 2.16$, $MSe = 5897870.36$, $p > .05$. We also did not find a significant statistical interaction between modalities and the degree of difficulty, $F(1, 15) = 3.74$, $MSe = 5800939.91$, $p = .07$ but there was a tendency. It means that when target search tasks were easy, there is no difference between modalities, in the meantime when target search tasks were difficult, in redundant display settings the performances were better than visual only. In other words, there was a type of multimodality effect regarding the degree of difficulty.

The results of Experiment 1 demonstrate that redundant display could be better for difficult tasks which demand high attention.

3. EXPERIMENT 2

As the results of Experiment 1, we supposed that the change of pitch was insufficient for mapping the depth dimension of hypertext because sound stimuli served only as a relative index in Experiment 1. In Experiment 2, we wanted to make different sound stimuli as an absolute index.

Stimuli in Experiment 2 were different from those used in Experiment 1. While in Experiment 1 the change of sound stimuli at pitch was mapped with the depth dimension of hypertext, in Experiment 2 the change of sound stimuli on interaction between pitch and the number of sounds was mapped with the depth dimension of hypertext. In other words, the main page was played into a single sound, the second level was played into a combination sound of C6 and C5, the third level was played into a combination sound of C6, C5, and C6, finally the fourth level was played into a combination sound of C6, C5, C6, and C5.

3.1. Method

3.1.1. *Participants*

Sixteen university students (4 male, 12 female; mean age 25 years) with normal or corrected-to-normal vision and hearing served as participants. None had participated in Experiment 1, and none were aware of the hypothesis being tested.

3.1.2. *Apparatus and stimuli*

Apparatus were the same as those used in Experiment 1. Stimuli were changed a little.

The four sound stimuli were synthesized at 16-bit, 44.1 kHz using ‘Csound’, and saved in WAV format. Each stimulus was composed of a one-beat long pure sine wave tone. The sound stimuli were a sound at pitch C6 (1046.50 Hz) with the duration of 200ms, a combination sound of C6 and C5 (523.25 Hz) each with the duration of 100ms, a combination sound of C6, C5, and C6 each with the duration of 100ms, lastly a combination sound of C6, C5, C6, and C5 each with the duration of 100ms. The intensity of stimuli was 85dB. Each sound displayed a given position of hierarchical structure in hypertext, that is, four sound stimuli at a combination of C6 and C5 mapped respectively the main page, the first depth level from the main page, the second depth level, and the third depth level on the experimental website.

When the mouse button was pressed down over a graphical link button a given sound at pitch and the number of sounds mapped with the depth dimension of hypertext was played.

3.1.3. *Design and procedure*

In Experiment 2, only the sound stimuli were different, the experimental design and independent variables were the same as those used in Experiment 1. The modalities were visual display only and redundant display (visual and auditory); the degree of difficulty was easy or difficult. The 5 trials in each of the 4 conditions were ordered randomly for each participant, forming a block. Each participant performed 4 blocks for a total of 20 trials and the order of blocks was counterbalanced.

As in Experiment 1, the task was to search target products as quickly as possible in the website redesigned for experimental purpose.

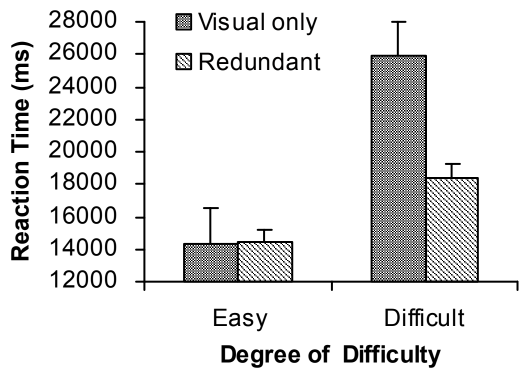

Figure 3. *Results from Experiment 2.*

3.2. Results and Discussion

The results of Experiment 2 are shown in Figure 3. First, the mean of reaction time regarding every condition was shown longer than in Experiment 1. It might be derived from the sexual difference because female participants were slower than male in searching time on the website. An analysis of variance (ANOVA) revealed a main effect for difficulty level, $F(1, 15) = 10.40$, $MSe = 92570691.64$, $p = .00$. It appeared that two difficulty levels were well divided. There was no similar main effect for modalities, $F(1, 15) = 2.75$, $MSe = 79762434.47$, $p = .12$. We also did not find a significant statistical interaction between modalities and the degree of difficulty, $F(1, 15) = 3.06$, $MSe = 73191953.70$, $p = .10$. Like Experiment 1, there was only multimodality effect regarding the degree of task-difficulty.

Although, in Experiment 2, we used the number of sounds as an absolute index to complement a relative index, there was no significant difference between visual and redundant condition. It seems that an absolute index is not a critical factor in this case.

4. GENERAL DISCUSSION

In the present study, auditory display was experimentally tested to discover if it improved usability in website navigation. Auditory information was mapped with the depth dimension of structural hierarchy in the hypertext. The result showed tendency between visual display and multimodality. Especially, in the difficult targets, auditory information played a complementary role. That is, auditory displays have potential as additional feedback in complex situations. Redundant auditory information allows users to allocate visual attention on the other information.

Although experimental tasks were counterbalanced, individual difference in knowledge of target stuffs may affect the results. Since the variance is so large, individual difference might shadow the multimodality effect. In addition, if we informed what the sounds mean before an experimental trial, participants could take advantage of the auditory feedback. Though we manipulated sounds considering as natural mapping, participants may have needed to learn the mapping relation between sounds and structural hierarchy.

In this study, dependent variable was the reaction time. Usability is consisted of two parts, performance and user experience but the reaction time can not reflect the quality of experience like subjective satisfaction. Therefore, in further researches, sensible aspect such as individual preference should be considered as another decisive factor in auditory displays.

5. REFERENCES

[1] A. Baddeley, “The concept of working memory: a view of its current state and probable future development.” *Cognition*, vol. 10, pp. 17-23, 1981.

[2] S. Brewster, “The design of sonically-enhanced widgets,” *Interacting with Computers*, vol. 11, pp. 211-235, 1998.

[3] M. L. Brown, S. L. Newsome, & E. P. Glinert, “An experiment into the use of auditory cues to reduce visual workload.” in *Proceedings of CHI ’89*, Austin, TX, ACM Press, 1989, pp. 339-346.

[4] W. W. Gaver, “Synthesizing Auditory Icons.” In Proceeding of INTERCHI ’93, Amsterdam, Netherlands: ACM, 1993, pp. 228-235.

[5] G. Kramer, “An introduction to auditory display”. In G. Kramer (Eds.), *Auditory display: Sonification, audification, and auditory interface, Santa Fe Institute studies in the sciences of complexity, Proceedings XVIII,* pp. 1-77, Reading, MA: Addison-Wesley, 1994.

[6] E. M. Maranda & A. N. Celestine, “The effects of Multimodal display aids on human performance.” *Computers industrial Engineering*, vol. 33, pp. 197-200, 1997.

[7] S. A. Brewster, P. C. Wright, & A. D. N. Edwards, “The design and evaluation of an auditory-enhanced scrollbar.” in *Proceedings of CHI ’94*, Boston, MA: ACM Press, 1994a, pp. 173-179.

[8] S. A. Brewster, P. C. Wright, A. Dix, & A. D. N. Edwards, “The sonic enhancement of graphical buttons.” in *Proceedings of Interact ’95*, Lillehammer, Norway: Chapman & Hall, 1994b, pp. 43-48.

[9] W. W. Gaver, “The sonicfinder: an interface that uses auditory icon.” Human-Computer Interaction, vol. 4, 67-94, pp. 339-346. Austin, TX, ACM Press. 1989.

[10] M. C. Albers and E. Bergman, “The Audible Web: Auditory enhancements for Mosaic,” in *Proceeding of CHI ’95 Conference on Human Factors in Computer Systems Conference Companion*, Denver, Colorado, May 1995, pp. 318–319.

[11] J. G. Neuhoff, G. Kramer, and J. Wayand, “Pitch and loudness interact in auditory displays: Can the data get lost in the map?”, *Journal of Experimental Psychology: Applied*, vol. 8, no. 1, pp. 17-25, 2002.

[12] B. N. Walker, *Magnitude estimation of conceptual data dimensions for use in sonification.* Unpublished doctoral dissertation, Rice University, Houston, TX, USA, 2000.