Association for Information Systems
AIS Electronic Library (AISeL)

AMCIS 2006 Proceedings

Americas Conference on Information Systems (AMCIS)

December 2006

Development of a Dual-Modal Presentation of Hierarchical Relationships: Effects of Graphics and Voice

Xiao Fang
DePaul CTI

Priti Kulkarni
DePaul CTI

Jacek Brzezinski
DePaul CTI

Susy Chan
DePaul CTI

Follow this and additional works at: http://aisel.aisnet.org/amcis2006

Recommended Citation
Fang, Xiao; Kulkarni, Priti; Brzezinski, Jacek; and Chan, Susy, "Development of a Dual-Modal Presentation of Hierarchical Relationships: Effects of Graphics and Voice" (2006). *AMCIS 2006 Proceedings*. 355.
http://aisel.aisnet.org/amcis2006/355

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2006 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Development of a Dual-Modal Presentation of Hierarchical Relationships: Effects of Graphics and Voice

Xiaowen Fang
DePaul CTI
xfang@cti.depaul.edu

Priti Kulkarni
DePaul CTI
kpriti@yahoo.com

Jacek Brzezinski
DePaul CTI
JBrzezinski@cti.depaul.edu

Susy Chan
DePaul CTI
schan@cti.depaul.edu

ABSTRACT

Baddeley's (1986) working memory model suggests that imagery spatial information and verbal information can be concurrently held in different subsystems within human working memory. Based on this model and research on human attention, a new dual-modal presentation depicting the hierarchical relationships in texts as a tree structure and delivering remaining texts as voice messages is proposed. This dual-modal presentation attempts to: (1) minimize the interference in information processing between the visual and auditory channels; and (2) improve the effectiveness of mental integration of information from different modalities. The objectives of this study are to test the effectiveness of this dual-modal presentation and to investigate effects of visual imagery and auditory presentations. An experiment will be conducted to test the theory-based hypotheses. An experiment website is built for the experiment. Details of the experiment system will be discussed. About 60 participants will be recruited to test this new interface.

Keywords

Multi-modal interfaces, information presentation, human attention, working memory, interface design.

INTRODUCTION

At present, most information on computers is presented in a textual format. However, textual display may not always be the best way to present information. For example, people may have difficulty reading and comprehending complex texts, and users of handheld devices may not be able to read much textual information on a small screen. Under such circumstances, textual presentation becomes inefficient and multi-modal interfaces are required. Use of multiple sensory modalities is often supported by the independent nature of multi-modal information processing, which assumes that there is no interference between tasks and thus no degradation in performance (Cook et al., 1997). However, research in cognitive psychology shows that visual and auditory perceptual processing is closely linked (Eimer, 1999). Problems related to memory and cognitive workload are found in current applications with voice-based interface (Cook et al., 1997).

The objectives of this study are to develop an effective dual-modal presentation and to investigate effects of visual imagery and auditory presentations. This study focuses on the dual-modal presentation of textual information describing hierarchical relationships. Results of this study will help to address the usability problems associated with small-screen computers and display of textual information.

LITERATURE REVIEW

To develop an effective dual-modal information presentation, we have examined prior research in working memory and graphical representation of texts.

Working Memory

Baddeley (1986) has proposed a working memory model that depicts three components: central executive, visuo-spatial sketchpad, and phonological loop (see Figure 1).

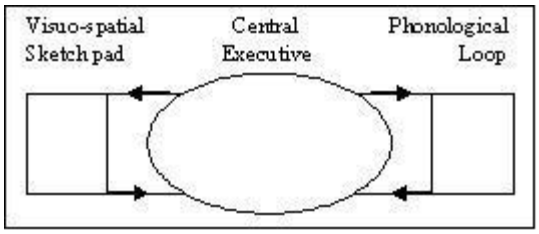

Figure 1. Baddeley's Working Memory Model (1986)

According to this model, human working memory contains two subsystems for storage: phonological loop and visuo-spatial sketchpad. Acoustic or phonological coding is represented by the phonological loop, which plays an important role in reading, vocabulary acquisition, and language comprehension. The visuo-spatial sketchpad is responsible for visual coding and handling spatial imagery information in analog forms. The phonological loop and visuo-spatial sketchpad are able to simultaneously hold verbal and imagery information without interference. Central executive is the control system that supervises and coordinates the information retrieved from the two storage subsystems for further integration.

Baddeley's working memory model has been confirmed by a considerable number of experiments. Mousavi, Low, and Sweller (1995) show that students' performance is significantly improved when the verbal representation and image representation of a geometry problem are respectively presented in auditory and visual mode. These researchers further suggest that, because the phonological loop and visuo-spatial sketchpad can hold distinct information independently, working memory might be effectively increased by distributing information in visual and auditory modalities. An earlier research study by Woodhead and Baddeley (1981) indicates that people who are good at recognizing faces also show better performance in recognizing paintings, but do not differ in recognizing words. Their finding implies image memory is separate from verbal memory.

Graphical Representation of Texts

To design an effective dual-modal information presentation based on Baddeley's working memory model, it is important to understand how textual information should be converted to imagery/graphical and verbal representations.

Researchers in cognitive psychology have been interested in the knowledge representation via mental images for decades. Studies (Minsky, 1975; Kosslyn & Shwartz, 1977) suggest that mental imagery information is generated from a hierarchically presented structure from long-term memory. This hierarchical knowledge representation includes the skeletal shape of the object and details or other components that are attached to the skeleton. The association between the new information and the components in this tree structure significantly affects the effectiveness of people's comprehension. Other researchers indicate that it is useful to be able to see the entire hierarchy while focusing on a particular part so that the relationship of parts-whole can be seen and the focus can be guided to other parts in a smooth and continuous way (Lamping, Rao & Piroli, 1995). For example, Cone Tree display in lucid forms complicated data hierarchies that might be otherwise invisible to the user (Robertson, Mackinlay & Card, 1991).

Mayer's empirical studies (1989 & 1990) indicate that an effective illustration model should use images or diagrams to reorganize and integrate the acquired information. The illustration must be able to guide the user's selective attention towards key items in the presented information. These key items include not only the major entities (such as objects, states, and actions.), but also the relationship among them. Mayer and Anderson (1991) further indicate that an explanatory illustration can be most effective when the (visual) animation and (auditory) narration are presented concurrently.

Based on the above discussions, we propose a dual-modal information presentation that presents the hierarchical relationship depicted in texts as tree-structured diagrams, and the remaining textual information as voice messages. The following section discusses this dual-modal presentation in greater details.

PROPOSED DUAL-MODAL INFORMATION PRESENTATION

Based on Baddeley's working memory model, it is assumed that the effectiveness of human information processing can be improved if the verbal representation and the imagery/graphical representation of certain textual information are presented via auditory and visual output, respectively. As shown in Figure 2, if the verbal presentation of the original textual information is presented via auditory channel, the verbal information will be temporarily stored in the auditory sensory

register, then sent to and processed in the phonological loop in working memory. Meanwhile, information perceived from the graphical presentation will be stored in the visual sensory register and then transferred to visuo-spatial sketchpad. Verbal and graphical information that are concurrently stored in working memory could be respectively retrieved from the phonological loop and visuo-spatial sketchpad, and then integrated by the central executive for comprehension.

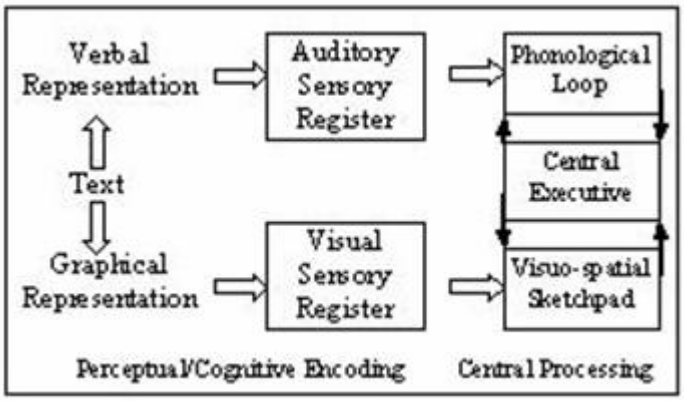

Figure 2. Splitting Textual Information

In this proposed dual-modal presentation (see Figure 3), hierarchical relationship contained in texts will be extracted and presented as a tree structure, with related entities indicated as nodes of this tree. Additional textual information will be delivered through the auditory channel. The following hypotheses are proposed to test the effectiveness of this dual-modal information presentation and examine effects of visual imagery and auditory presentations.

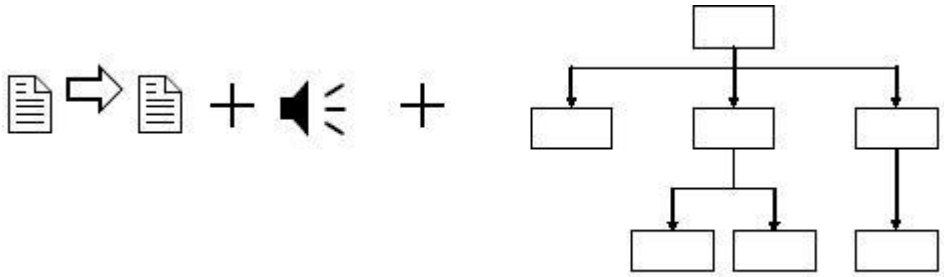

Figure 3. Proposed Dual-modal Presentation

According to Baddeley's (1986) working memory model, the graphical information might be perceived and held in the visuo-spatial sketchpad. However, visually presented texts will be processed entirely in the phonological loop. The graphic tree presentation of hierarchical relationships will likely make the visuo-spatial sketchpad in working memory available for information processing in addition to the phonological loop. Therefore, the graphic tree presentation might improve user comprehension of information by increasing working memory capacity.

H1: The graphic tree presentation will improve user comprehension of the hierarchical relationships depicted in texts.

Based on Baddeley's (1986) working memory model, non-speech verbal input such as visually presented texts must go through a sub-vocal rehearsal to be converted to speech input and temporarily saved in the phonological loop of working memory before further processing. On the contrary to the textual display, the dual-modal presentation delivers verbal information as speech and doesn't likely involve a sub-vocal rehearsal in information processing. By avoiding a sub-vocal rehearsal, the dual-modal presentation could improve user comprehension of the texts.

H2: The dual-modal presentation will improve user comprehension of the texts delivered as speech as compared to the textual display.

In the proposed dual-modal presentation, the graphical information might be perceived and held in the visuo-spatial sketchpad while the speech input is perceived and directly stored in phonological loop. Therefore, by concurrently utilizing the two subsystems in working memory to process the same amount of information, a reduced cognitive workload is expected during information processing. The multiple resource pool model in human attention suggests that two or more tasks can be performed together efficiently as long as they require separate pools of resources (Wickens, 1980). This model

provides another theoretical basis for the proposed dual-modal presentation. In the proposed dual-modal information presentation, the graphic and voice information are derived from the same textual information. The information presented in the two modalities should be highly relevant and complementary to each other. Therefore, the mental integration of the visual and auditory information might be easier than that of disparate information. With a reduced cognitive workload and easier mental integration in working memory, the proposed dual-modal information presentation may significantly improve user satisfaction as compared to single-modal visual display.

H3: The dual-modal presentation will result in superior user satisfaction as compared to single-modal visual display.

METHOD

Subjects

About sixty university students will be recruited to participate in this experiment. They will be evenly and randomly assigned to three treatment groups.

Independent and Dependent Variables

The only independent variable is information presentation mode. There are three treatments: Text (T) mode, Text + Graphic (TG) and Graphic + Voice + Text (TGV) mode. In the T mode, all information will be visually presented as texts on a Web page. In the TG mode, hierarchical relationships contained in the original texts will be presented as tree-structured diagrams. In the TGV mode, hierarchical relationships contained in the original texts will be presented as a tree-structured diagram and selected texts will be delivered as speech. These three treatments allow the authors to investigate effects of tree-structured diagrams and speech display.

The two dependent variables are user performance and intention to adopt the technology. User performance is measured by the number of correctly answered questions within a 30-minute period. Questions are designed to measure user comprehension of the hierarchical relationships and texts delivered as speech in the TGV mode respectively. User intention to adopt will be measured by a questionnaire using a 7-point Likert scale. Based on Technology Acceptance Model (TAM) (Davis, 1989; Koufaris, 2002), this questionnaire is designed to measure subjects' perceived usefulness and ease of use of the three interfaces. In addition to effectiveness, user's intention to adopt is also critical to the success of a new technology.

Experiment System

Encyclopedia articles in biology are selected as the experiment materials. An experiment website that delivers the three versions (T, TG, and TGV) of these articles has been developed. Figures 4, 5, and 6 show the sample screen shots of the three versions.

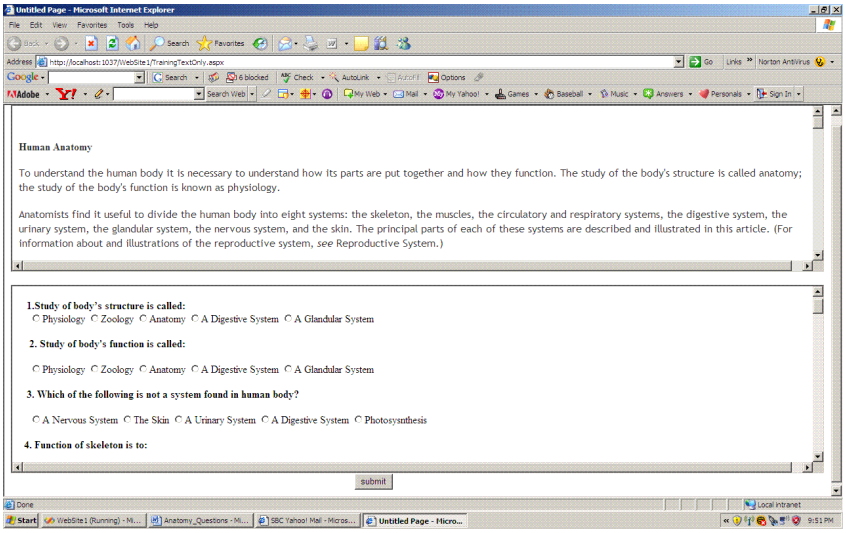

Figure 4. Sample Screen Shot of Text (T) Version

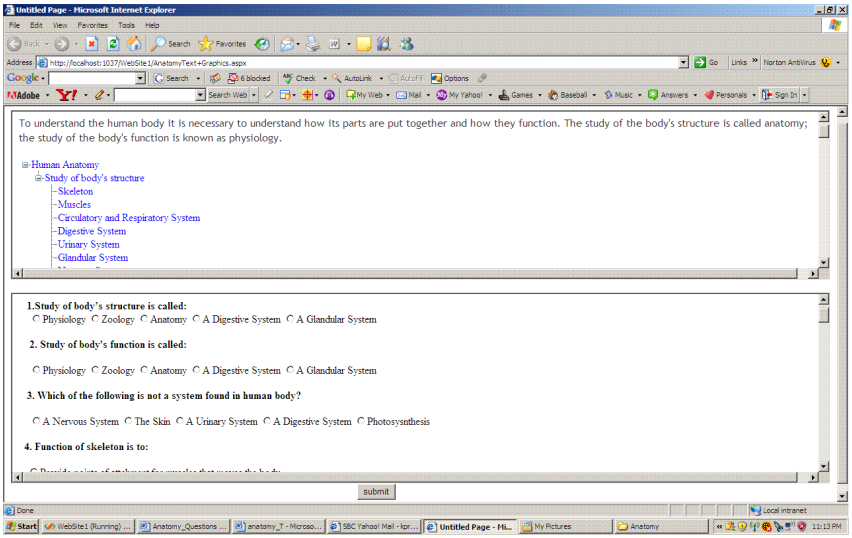

Figure 5. Sample Screen Shot of Text + Graphic (TG) Version

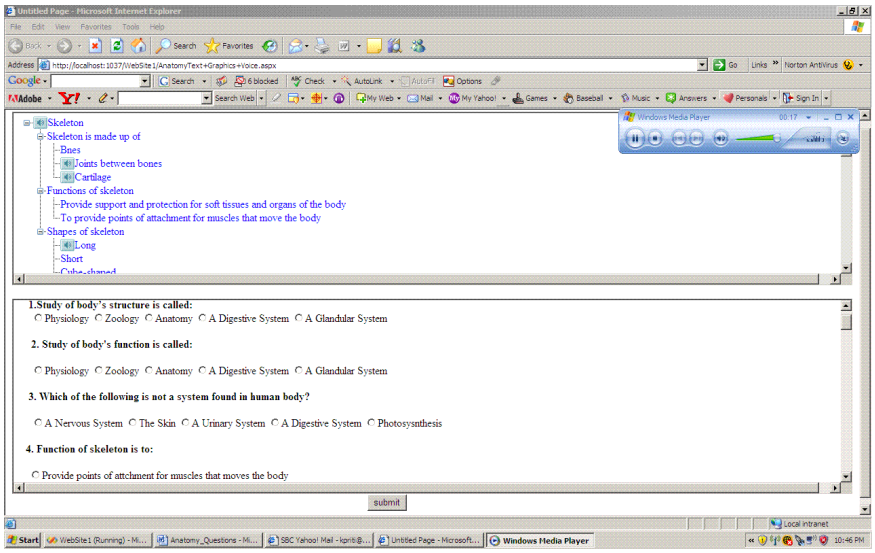

Figure 6. Sample Screen Shot of Text + Graphic + Voice (TGV) Version

Procedure

Each subject will be asked to sign a consent form before participation. During the training session, each subject will fill out a background questionnaire and the experimenter will describe the tasks included in different groups. A sample article will be used to explain the interface, browsing rules, time limit, graphic notations (for TG and TGV groups only), and voice control (for TGV-mode group). Subjects are allowed to ask questions during the training period. They can spend as much time as they need in the training session. Subjects will be encouraged to answer as many questions as they can during the experiment. They will be allowed to browse back and forth within each article to find or correct their answers. Subjects can click a submit button to move on to the next article after they finish the current one, but they cannot go back to the previous article. For the TGV-mode presentation, pre-recorded voice information will be automatically played when the Web page is loaded on the screen. Subjects can use controls on the screen to replay voice messages. Upon completion of these two tests, the subject will be asked to fill out a satisfaction questionnaire. There is no time limit for this satisfaction survey.

NEXT STAGE

This project is still in progress. The authors are finishing the development of the experiment system. A controlled experiment will be conducted to test the research hypotheses and validate the proposed new dual-modal information presentation. Preliminary results of this study will be presented at the conference.

REFERENCES

1. Baddeley, A. D. (1986). *Working memory*. New York: Oxford University Press.
2. Cook, M. J., Cranmer, C., Finan, R., Sapeluk, A., & Milton, C. (1997). Memory load and task interference: Hidden usability issues in speech interfaces. *Engineering psychology and cognitive ergonomic*, 3, 141-150.
3. Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, Sep. 1989, 319-340.
4. Eimer, M. (1999). Can attention be directed to opposite locations in different modalities? An ERP study. *Clinical neurophysiology*, 110, 1252-1259.
5. Kosslyn, S. M. & Shwartz, S. P. (1977). A simulation of visual imagery. *Cognitive science*, 1, 265-295.
6. Koufaris, M. (2002). Applying the technology acceptance model and flow theory to onine consumer behavior. *Information systems research*, 13(2), 205-233.
7. Lamping, J., Rao, R., & Piroli, P. (1995). A focus + context technique based on hyperbolic geometry for visualizing large hierarchies. In *Proceedings the SIGCHI conference on Human factors in computing systems* (pp.401-408). Denver, Colorado, United States Pages: 401 - 408
8. Mayer, R. E. (1989). Models for understanding. *Review of educational research*, 59(1), 43-64.
9. Mayer, R. E. & Gallini, J. K. (1990). When is an illustration worth thousand words? *Journal of educational psychology*, 82(4), 715-726.
10. Mayer, R. E. & Anderson, R. B. (1991). Animations need narrations: an experimental test of the dual-coding hypothesis. *Journal of educational psychology*, 83(4), 484-490.
11. Minsky, M. (1975). A framework for representing knowledge. In P.H. Winston (Ed.), *The psychology of computer vision* (pp.211-277), New York: McGraw-Hill.
12. Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of education psychology*, 87(2), 319-334.
13. Robertson, G. G., Mackinlay, J. D., & Card, S. K. (1991). Cone trees: Animated 3D visualizations of hierarchical information. In *Proceedings of the CHI'91 conference on human factors in computing systems* (pp.189-194).
14. Woodhead, M. M. & Baddeley, A. D. (1981). Individual differences and memory for faces, pictures, and words. *Memory & cognitive*, 9(4), 368-370.
15. Wickens, C. (1980). The structure of attentional resource. In R. S. Nickerson (ed.), *Attention and Performance VIII*, 239-257, Hillsdale, NJ: Lawrence Erlbaum.