Dual-Modal Presentation of Sequential Information

Shuang Xu sxu@cti.depaul.edu
Xiaowen Fang xfang@cti.depaul.edu
Jacek Brzezinski jbrzezinski@cti.depaul.edu
Susy Chan schan@cti.depaul.edu
School of Computer Science, Telecommunications, and Information Systems, DePaul University

ABSTRACT

Based on Baddeley's (1986) working model and research on human attention, this study intends to design a visual-auditory information presentation to: (1) minimize the interference in information processing between visual and auditory channels; and (2) improve the effectiveness of mental integration of information from different modalities. Baddeley suggests that imagery spatial information and verbal information can be concurrently held in different subsystems within human working memory. Accordingly, this research proposes a method to convert sequential textual information into its graphical and verbal representations and hypothesizes that this dual-modal presentation will result in superior comprehension performance and higher satisfaction as compared to pure textual display. Simple T-tests will be used to test the hypothesis. Results of this study will help to address usability problems associated with small-screen computers. Findings may also benefit interface design of generic computer systems by alleviating the overabundance of information output in the visual channel.

Keywords

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

INTRODUCTION

The advancement of wireless technology has promised users mobile communications and information access. But there are many inherent constraints in wireless devices, such as small screens and low-resolution (Chan, Fang, Brzezinski, Zhou, Xu and Lam, 2002). Technologies for speech recognition and synthesis are becoming increasingly sophisticated and provide support for information processing via multi-modal interfaces. The benefit of delivering information across different sensory modalities is often justified by the presumable independence of multi-modal information processing. It is usually assumed that there is no interference between tasks and thus no degradation in performance (Cook, Cranmer, Finan, Sapeluk and Milton, 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 recent applications of voice-based interface (Cook et al., 1997). Therefore, it is imperative to reduce the potential interference between different sensory modalities in order to design an effective multi-modal interface.

The objective of this research is to develop a dual-modal interface that: (1) minimizes the interference in information processing between visual and auditory channels; and (2) improves the effectiveness of mental integration of information from different modalities. This study focuses on the dual-modal presentation of textual information that describes sequential or chronological events. Results of this study will help to address the usability problems associated with small-screen computers and the mobile information access via handheld devices. Findings of this study may also benefit interface design of generic computer systems by alleviating the overabundance of information output in the visual channel.

LITERATURE REVIEW

To develop an effective dual-modal information presentation, we have examined prior research in human attention, working memory, visual and auditory interfaces, and knowledge representation.

Human Attention

The topic of attention has long been of interest to researchers in cognitive psychology. Proctor and Van Zandt (1994) distinguish human attention in three aspects: selective attention that concerns human ability to focus on certain sources of information and ignore others; divided attention that involves human ability to divide attention among multiple tasks; and the amount of mental effort required to perform a task.

Researchers have proposed several models of human attention. Bottleneck models (e.g., Broadbent, 1958) specify a particular stage in the information-processing sequence at which the amount of information that humans can attend to is limited. In contrast, resource models (e.g., Kahneman, 1973) view attention as a limited-capacity resource that can be allocated to one or more tasks, rather than as a fixed bottleneck. Among various attention models, multiple-resource models propose that there are several distinct subsystems, each having their own limited pool of resources.

Wickens (1980, 1984) proposes a three-dimensional system of resources consisting of distinct stages of processing (encoding, central processing, and responding), codes (verbal and spatial), and input (visual and auditory), plus output (manual and vocal) modalities. This model assumes that two tasks can be efficiently performed together to the extent that they require separate pools of resources.

Allocation of attentional resources during complicated time-sharing tasks across multiple modality channels has long been of interest to cognitive psychology researchers. Research shows that introducing auditory channel into prototypes of civil and military cockpits has resulted in degraded performance (Cook et al., 1997). One explanation is that the total amount of attentional resources is limited. When demanded simultaneously by multi-modal information processing tasks, resources allocated to non-dominant channel decrease, as compared to single-modal information processing. Another explanation is that mental integration of different multi-modal information causes a heavy cognitive load in working memory. If this integration is critical to understanding information received from different sensory channels, performance will degrade.

Cook et al (1997) suggest that speech-based interfaces could be used in a restricted, well-defined task to manipulate the demand on central resources by changing the nature of visual discrimination task and the demand on memory. Wickens and Ververs (1998) examined the effects of display location and image intensity on flight path performance. Their findings suggest that attention is modulated by tasks, which are consistent with the limited attentional resources assumption. Faletti and Wellens (1979) explored the seemingly uneven weighing systems for concurrent information processing across different modalities. They believe that approach-avoidance tendencies in response to specific combinations of design elements might be predicted by developing a formula to integrate environmental information. The use of cell phones in automobiles has increased the public concerns for safety issues. Studies on voice-based car-driver interfaces indicate that performing other tasks while driving takes away from a driver's limited attentional resources. An effective multi-modal interface used in automobiles should minimize the driver's investment in attention, and minimize interference and distraction (Starner, 2002; Siewiorek, Smailagic and Hornyak, 2002; Cellario, 2001; Guglielmetti, 2003; Titsworth, 2002).

The above research findings indicate that both the allocation of attentional resources and interactions between information perceived via visual and auditory channels significantly affect a user's comprehension of a dual-modal interface.

Working Memory

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

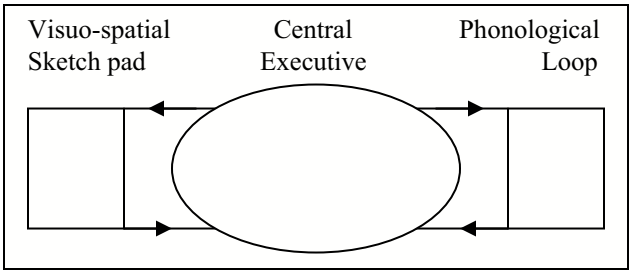

Figure 1. Baddeley's Working Memory Model (1986)

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 model has been confirmed by many studies. For example, Mousavi, Low, and Sweller (1995) show that students' performance was significantly improved when the verbal representation and image representation of a geometry problem were respectively presented in auditory and visual modes. They further suggest that distributing relevant information in visual and auditory modalities might effectively increase working memory.

Visual and Auditory Information Presentation

After comparing visual and auditory information representation, prior research shows that voice is more informal and interactive for handling complex, equivocal and emotional aspects of collaborative tasks (Chalfonte, Fish and Kraut, 1991). As Streeter (1998) indicates, universality and mobile accessibility are major advantages of speech-based interface, whereas its disadvantage is the slow delivery rate of voice information. Archer, Head, Wollersheim, and Yuan (1996) compared the user's preferences and the effectiveness of information delivery in visual, auditory, and visual-auditory modes. They suggest that information should be organized according to its perceived importance to the user, who should also have flexible information access at different levels of abstraction.

Multi-modal interfaces have been widely used to support collaborative work, as well as in teaching systems. Researchers (Nardi, Schwarz, Kuchinsky, Leichner, Whittaker and Sclabassi, 1993) indicate that integration of

video information and other data sources (e.g., aural input, time-based physical data, etc.) helps surgeons choose the correct action and interpretation during remote medical operations. Research on interaction between sound, written words, and the image of objects shows that when different sources of information are integrated, a learner's cognitive overload remains light and does not limit learning (Dubois and Vial, 2000). Stock, Strapparava, and Zancanaro (1997) show that hypertext and digital video sequences help users explore information more effectively. By exploring the integration of captioning, video description, and other access tools for interactive learning, Treviranus and Coombs (2000) demonstrated how to make the learning environment more flexible and engaging for students. Dubois and Vial (2000) suggest that several factors affect the effectiveness of integration of multi-modal information. These factors include not only the presentation mode, the construction of co-references that interrelate to the different components of the learning materials, but also the characteristics of the task.

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

Schema (or script, frame) has been widely used in knowledge representation (Proctor and Van Zandt, 1994; Johnson-Laird, 1983, 1989). Schemas are frameworks that depict conceptual entities, such as objects, situations, events, actions, and the sequences between them. Schemas not only represent the structure of a person's interest and knowledge, but also enable a person to develop the expectancy about what will occur. Sequences and events are salient information to form schemas. After summarizing the experimental studies on the relationship between imagery and text processing, Denis (1988) indicates that narrative texts that strongly elicit visual imagery for characters, scenery, and events are highly imageable. Denis' finding suggests that sequential information contained in texts can be converted to imagery. According to schema theory, imagery of the sequence of events may help users form schemas by reducing the cognitive demand for converting textual information into effective schemas and thus improve their comprehension of the information because the imagery information is processed by the visual-spatial sketchpad (Baddeley, 1986).

Based on the above discussions, we propose a dual-modal information presentation that presents the sequential information contained in texts as flowchart-like diagrams, and presents the remaining textual information as voice message. 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.

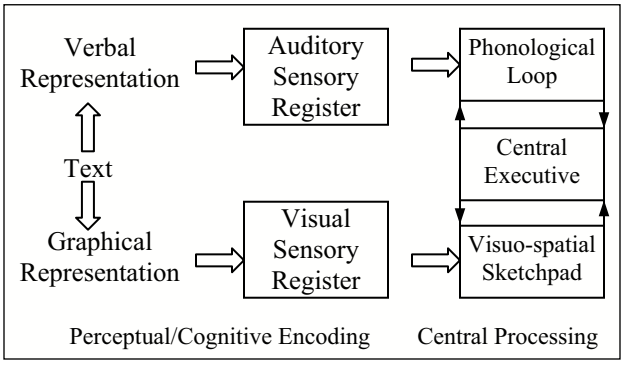

Figure 2. Splitting Textual Information

As suggested in Denis' study (1988), textual description of a series of events is highly imageable. After combining Baddeley's working memory model and Denis' findings (1988), a new dual-modal information presentation is proposed (see Figure 3). In this dual-modal presentation, sequential information contained in texts will be extracted, converted to, and presented as a flowchart. The remaining textual information will be delivered through the auditory channel. The following hypothesis is proposed to test the effectiveness of this new dual-modal information presentation.

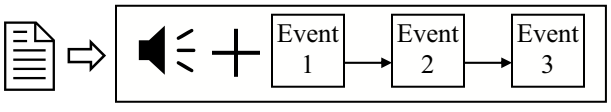

Figure 3. Proposed Dual-modal Presentation

Hypothesis: The dual-modal presentation of sequential information will improve the user's comprehension of information and result in higher user satisfaction as compared to pure textual display.

According to Baddeley's model, pure visual display of textual information will be processed entirely in the the phonological loop. Non-speech verbal input 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. 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 the 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. Research in human attention has shown that many voice-based interfaces caused degraded comprehension performance because of the interference between disparate information perceived from visual and auditory channels. In the proposed dual-modal information presentation, graphic and voice information are derived from the same textual information, and should be highly relevant and complementary to each other. The schema theory (Proctor and Van Zandt, 1994; Johnson-Laird, 1983, 1989) suggests that imagery of sequential information might help users form schemas and thus facilitate the mental integration. Therefore, mental integration of visual and auditory information will be easier during comprehension.

With a reduced cognitive workload and easier mental integration in working memory, the proposed dual-modal information presentation may significantly improve the effectiveness of users' information comprehension.

METHOD

This study will use analytical tests from the Graduate Record Examination (GRE) for the experiment because these tests are designed to measure subjects' analytical comprehension and reasoning skills without assessing specific content knowledge. An experiment Web site will be built to present the GRE analytical tests. The task is to perform GRE analytical tests through the experiment Web site. A GRE analytical test takes 30 minutes. Two tasks will be performed in our experiment, each takes 30 minutes.

The only independent variable is information presentation mode. There are two treatments: Text (T) mode and Graphic+Voice (GV) mode. In the T mode, all information will be visually presented as texts on a Web page. In the GV mode, the original textual information will be split into a flowchart-like diagram and speech output. Figure 4 shows an example of these two display modes. Three faculty members with rich teaching experience will be asked to manually convert the GRE analytical tests into a graph + voice presentation according to the proposed method (see Figure 3). Only sequential information will be converted into graphics.

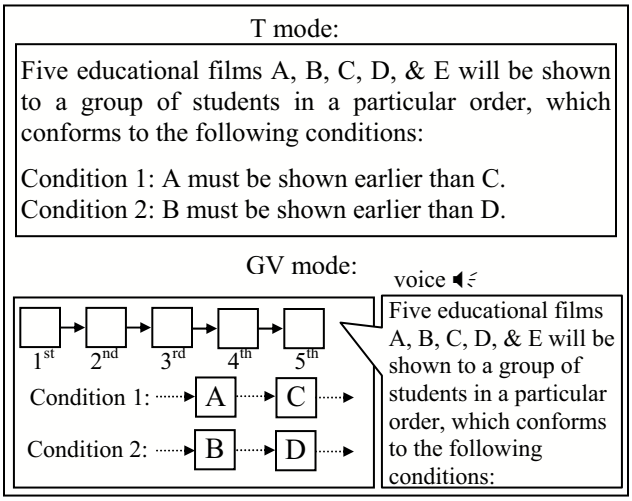

Figure 4. An example of display modes

The two dependent variables are users' performance and satisfaction. Subjects' performance is measured by the number of correctly answered questions within a 30-minute period. An analytical test starts when the first analytical problem is presented on the screen, and ends when time is up. User satisfaction will be measured by a satisfaction questionnaire using a 7-point Likert scale. Based on Technology Acceptance Model (TAM) (Davis, 1989; Koufaris, 2002), this satisfaction questionnaire is designed to measure subjects' perceived usefulness and ease of use of the two interfaces. In addition, one question will be added to measure user's general satisfaction.

Sixty university students will be recruited to participate in this experiment. They will be evenly and randomly distributed into two treatment groups. Their background information will be recorded to ensure a controlled balance in demographic characteristics between groups. Because individual participants' analytical comprehension and reasoning skills may vary greatly and such skills could affect their performance in the experiment, we propose to use an independent GRE analytical test as a pre-test to estimate a participant's skills before the actual experiment task is performed. In the pre-test, all information will be visually presented as texts on a Web page for both groups. The estimate of analytical comprehension and reasoning skills or possibly other test-taking skills from the pre-test will be used as a covariate in the analysis of the experiment task performed later.

The experiment design is a simple t-test. Subjects will perform two GRE analytical tests. The first test serves as the pre-test for estimating each individual's analytical comprehension, reasoning, and test-taking skills. The second test will be presented in T vs. GV mode for comparing the differences of these two presentation modes.

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 problem will be used to explain the interface, browsing rules, time limit, graphic notations (for GV-mode group), and voice control (for GV-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 two 30-minute analytical tests. They will be allowed to browse back and forth within each problem to find or correct their answers. Subjects will click a submit button to move on to the next analytical problem after they finish the current one, but they are not allowed to go back to the previous problem. For the GV-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. During the experiment, subjects will take two 30-minute tests. They are allowed to take a break between Test 1 (the pre-test) and Test 2. 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. Table 1 presents the two tests and the experiment procedure.

	Test 1 (30 minutes)	**Test 2** (30 minutes)	**Satisfaction Survey**
T-mode Group	Solve problems in T-mode presentation	Solve problems in T-mode presentation	Satisfaction questionnaire
GV-mode Group	Solve problems with T-mode presentation	Solve problems with GV-mode presentation	Satisfaction questionnaire

Table 1. Experiment Tasks and Procedure

The following information will be saved into a database for further analysis:

- Subjects' background information.
- Subjects' answers to analytical problems in Tests 1 and 2, and time spent on each problem.
- Subjects' response to the satisfaction survey.
- Subjects' online activities (e.g., manipulating voice messages, changing answers, etc.).

NEXT STAGE

A controlled experiment will be conducted to test the research hypothesis and validate the proposed new dual-modal information presentation. We are currently performing pilot studies and expect to complete the experiment during the next three months. Preliminary results of this study will be presented at the workshop.

SELECTED REFERENCES

1. Baddeley, A. D. (1986) Working memory, New York: Oxford University Press.
2. Cook, M. J., Cranmer, C., Finan, R., Sapeluk, A. and Milton, C. (1997) Memory load and task interference: Hidden usability issues in speech interfaces, *Engineering psychology and cognitive ergonomic*, 3, 141-150.
3. Denis, M. (1988) Imagery and prose processing, In M. Denis, J. Engelkamp and J. T. E. Richardson (Eds.), *Cognitive and neuropsychological approaches to mental imagery*, 121-132, Dordrecht / Boston / Lancaster: Martinus Nijhoff Publishers.
4. Johnson-Laird, P. N. (1983) Mental models, Cambridge, MA: Harvard University Press.
5. Johnson-Laird, P. N. (1989) Mental models. In M. I. Posner (Ed.) *Foundations of cognitive science* (pp.469-499), Cambridge, MA: MIT Press.
6. Mayer, R. E. (1989) Models for understanding, *Review of educational research*, 59, 1, 43-64.
7. Mayer, R. E. and Gallini, J. K. (1990) When is an illustration worth thousand words? *Journal of educational psychology*, 82, 4, 715-726.
8. Mayer, R. E. and Anderson, R. B. (1991) Animations need narrations: an experimental test of the dual-coding hypothesis, *Journal of educational psychology*, 83, 4, 484-490.
9. Mousavi, S. Y., Low, R. and Sweller, J. (1995) Reducing cognitive load by mixing auditory and visual presentation modes, *Journal of education psychology*, 87, 2, 319-334.
10. Proctor, R. and Van Zandt, T. (1994) Human factors in simple and complex systems, Needham Heights, MA: Allyn and Bacon.
11. Wickens, C. (1980) The structure of attentional resource, In R. S. Nickerson (ed.), *Attention and Performance VIII,* 239-257, Hillsdale, NJ: Lawrence Erlbaum.
12. Wickens, C. (1984) Processing resources in attention. In R. Parasuraman and R. Davies (eds), *Varieties of Attention*, 63-102, New York, NY: Academic Press.
13. Wickens, C. D. and Ververs, P. M. (1998) Allocation of attention with head-up displays, *Technical report of aviation research lab, Institute of aviation*, DOT/FAA/AM-98/28.