INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION, *20*(1), 3–17
Copyright © 2006, Lawrence Erlbaum Associates, Inc.

A Study of the Feasibility and Effectiveness of Dual-Modal Information Presentations

Xiaowen Fang
Shuang Xu
Jacek Brzezinski
Susy S. Chan
School of Computer Science, Telecommunications, and Information Systems
DePaul University

Multimodal interfaces with both visual and auditory output are becoming important, especially for applications using small-screen displays and for user access under mobile conditions. The research presented here investigated the feasibility of simultaneously presenting distinct textual information through both visual and auditory channels by examining two multimodal interfaces with irrelevant or relevant auditory information. These interfaces were intended to study two problems: (a) Can users attend to and process additional information delivered through the auditory channel during a typical Web-browsing process, and (b) what are the effects of information overlap between the visual and auditory channels? Controlled experiments were conducted to evaluate these two questions. The findings suggest that users can attend to auditory information while visually browsing textual information and that information overlap may reduce distraction. These findings have implications for the design of multimodal interfaces for small-screen mobile applications.

1. 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, users of handheld devices may not be able to read much textual information on a small screen, and automobile drivers cannot read texts from a computer screen while driving. Under such circumstances, textual presentation becomes inefficient.

Multimodal interfaces are being used increasingly in computer applications. Previous research has examined the effects of presenting information in both audi-

Correspondence should be addressed to Xiaowen Fang, School of Computer Science, Telecommunication, and Information Systems, DePaul University, 243 S. Wabash Avenue, Chicago, IL 60604–2301. E-mail: xfang@cs.depaul.edu

tory and visual modes. In general, the major advantage of using speech in an interface is its universality, as most people understand spoken language. The main disadvantage is that human beings can process voice output only at a relatively slow speed (Streeter, 1988). Therefore, auditory presentation should be used primarily when the task is performed in continuous motion or when temporal information is involved (Proctor & Van Zandt, 1994). Auditory presentation is suitable if (a) the message is simple, (b) the message is short, (c) the message will not be referred to later, (d) the message deals with events in time, (e) the message calls for immediate action, (f) the visual system of the person is overburdened, (g) the receiving location is too bright, and/or (g) the task performed by the user requires continuous movement (Deatherage, 1972; McCormick & Sanders, 1982; Proctor & Van Zandt, 1994). In contrast, visual presentation should be used if (a) the message is complex, (b) the message is long, (c) the message will be referred to later, (d) the message deals with a location in space, (e) the message does not call for immediate action, (f) the person's auditory system is overburdened, (g) the receiving location is too noisy, and/or (h) the task performed by the user requires the individual to remain stationary (Deatherage, 1972; McCormick & Sanders, 1982; Proctor & Van Zandt, 1994).

Based on Wickens's (1980, 1984) multiple-resource human attention model, two tasks can be performed together more efficiently to the extent that they require separate pools of resources, such as different modalities. In other words, humans can accept information simultaneously from two different channels—visual and auditory—with minimal interference. Therefore, if voice output is integrated with visual presentation, such a multimodal presentation may remedy some of the difficulties in reading textual information, especially in the mobile context and when using small-screen displays. However, prior research on computer interfaces has only investigated the effects of presenting duplicate textual information through the auditory channel (Archer, Wollersheim, & Yuan, 1996). To date, the impact of presenting *extra* information through the auditory channel in addition to regular textual display on computer interfaces remains unknown.

As an initial step toward designing effective multimodal interfaces, we examine in this study the impact of simultaneously presenting distinct verbal information through both visual (i.e., text) and auditory (i.e., speech) channels by investigating the feasibility and effectiveness of two dual-modal information presentations. These dual-modal presentations provide extra information through the auditory channel in addition to regular textual display in a Web-browsing context. This research addresses the following questions: Can users attend to and process additional information delivered through the auditory channel during a typical Web-browsing process, and what are the effects of information overlap between the visual and auditory channels? A few terms used throughout this article are defined as follows:

- Relevant information: Information that can be used to perform the same tasks (i.e., the browsing task).
- Irrelevant information: Information that has no direct relevance to the primary task but can be used to perform a secondary task.

- Information overlap: The same information delivered through both visual and auditory channels.
- Extra information: Information delivered only through auditory channel but not visual display.
- Assistive information: Information that helps users perform the browsing task and answer questions derived from the text-based Web site.

Findings from this research may shed light on future research in multimodal interfaces.

2. BACKGROUND LITERATURE

2.1. Human Attention

The topic of human attention has long been an area of interest for researchers. Proctor and Van Zandt (1994) distinguished human attention in three aspects: selective attention that concerns human ability to focus on certain sources of information while ignoring 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 attention. Bottleneck models (Broadbent, 1958; Treisman, 1964) 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 (Kahneman, 1973; Navon & Gopher, 1979; Wickens, 1980, 1984) 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 is no single attention resource. Rather, several distinct subsystems each have their own limited pool of resources. Wickens (1980, 1984) proposed 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. Wickens's model assumes that two tasks can be performed together more efficiently to the extent that they require separate pools of resources.

2.2. Human Working Memory

Figure 1 shows a diagram of the working memory model proposed by Baddeley (1986). In this model, acoustic or phonological coding is represented by the phonological loop, which plays a role in vocabulary acquisition, learning to read, and language comprehension. The phonological loop is a slave system specialized for the storage of verbal materials. This model also includes visual coding, in the form of the visuo-spatial "sketch pad." This visuo-spatial sketch pad is a slave system specialized for processing and storage of visual and spatial information and of verbal materials that are subsequently encoded in the form of visual imagery. The central

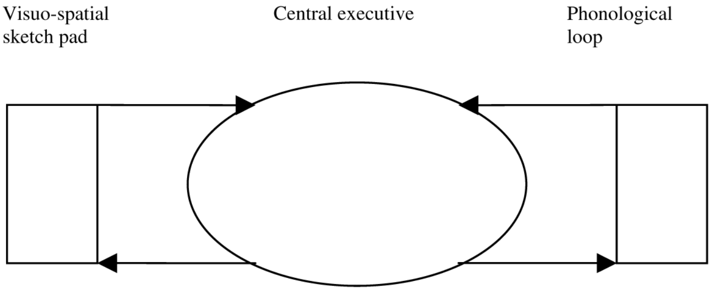


FIGURE 1 Working memory model proposed by Baddeley (1986).

executive is an attention control system that supervises and coordinates the visuo-spatial and phonological subsystems. According to Baddeley, visual imagery information and acoustic verbal information can be held simultaneously in separate cognitive storage systems, which can be further integrated by the central executive with minimum cognitive cost. Therefore, tasks should not interfere with each other if they use different subsystems. Many studies have reported evidence supporting this model. For instance, Mousavi, Low, and Sweller (1995) examined the use of a partially auditory and partially visual mode of presentation for geometry examples. In their study, three presentation modes were tested: Text + Diagrams + Voice messages, Text + Diagrams, and Diagrams + Voice messages. The effects of presentation modality suggest that working memory has partially independent processors for handling visual and verbal materials. Presenting materials in a mixed rather than a unitary mode may increase effective working memory.

2.3. Visual and Auditory Interfaces

The majority of displays encountered in human–machine systems are either visual or auditory. Previous studies on visual and auditory interfaces are summarized in Table 1.

3. EXPERIMENTAL DUAL-MODAL INFORMATION PRESENTATIONS

An extensive literature review has yielded few studies on the effectiveness of a computer interface when different information is presented simultaneously through voice and text modes. To develop an effective multimodal interface, it may be essential to use the auditory channel to deliver extra information in addition to a regular textual display. A multimodal interface may enable users to perform the same task or different tasks more effectively based on the extra information. The feasibility of multimodal interfaces depends on whether users can simultaneously attend to and process different information from different sensory modalities. In this research, experiments were conducted to investigate this issue and whether

Table 1: Summary of Studies on Visual and Auditory Interfaces

Investigators	*Modalities of the Interface*	*Major Findings*
Archer, Head, Wollersheim, and Yuan (1996)	Text and speech	Adding text to voice output improves the perceived acceptability of voice, but adding voice to text does not alter the perceived acceptability of text. When the same information was separately presented in three modes (text, voice, text plus voice), text mode was most efficient in performing information search, followed by voice mode, and text plus voice mode.
Baggett and Ehrenfeucht (1983)	Visual and auditory	Three presentations were studied: visual and narration presented simultaneously, visual followed by narration, and narration followed by visual. Results suggest that there is no competition for resources when related information is presented simultaneously in visual and auditory channels.
Chalfonte, Fish, and Kraut (1991)	Text and speech	Voice was preferred for addressing higher level issues in suggesting document modifications, but text was preferred for more detailed and lower level comments.
Cohen (1992)	Keyboard, screen, pointing device, and auditory channel	Users need to utilize at least two channels, such as auditory and keyboard, to complete a task.
DeHaemer and Wallace (1992)	Visual and auditory	Duplicate instructions were added as voice to a microcomputer workstation for decision support. The visual and audio modes of receiving information appear to be noninterfering.
Nugent (1982)	Media (pictures, audio, and print)	Student learning could be improved when the same content was presented in all three channels (picture, audio, and print). When different information was presented in the visual and audio modes, however, student learning was not affected by the addition of new channels and the presence of visual information did not interfere with processing the audio, and vice versa.
Proctor and Van Zandt (1994)	Visual versus auditory displays	Spatial information is best conveyed through visual displays because spatial discrimination can be made most accurately with vision. Auditory displays work best with temporal information because temporal organization is a primary attribute of auditory perception.
Schlosser, Belfiore, and Nigam (1995)	Speech	The presentation of additional auditory stimuli in the form of synthetic speech is effective in assisting individuals with mental retardation to learn associations between graphic symbols with spoken words.
Sipior and Garrity (1992)	Visual and auditory	Presentation with a mix of audio and visual accompaniments improves receptiveness attributes such as perception, attention, comprehension, and retention.
Streeter (1988)	Speech and text	The main advantage of using speech is that it can be universally accessed by everyone on the move. One notable disadvantage is that voice delivers information at a slower rate than text. Any combination of voice and text is likely to slow information acquisition process.

the level of information overlap between the auditory and visual channels may affect users' task performance.

The following sections describe the two dual-modal information presentations examined in this study.

3.1. Dual-Modal Information Presentation: Visual + Auditory

In this presentation format, Visual + Auditory (VA), a regular Web page was displayed in the normal visual/textual mode while additional irrelevant information was presented as voice output. As discussed earlier, Wickens's (1980, 1984) multiple-resource human attention model suggests that two tasks can be performed together more efficiently to the extent that they require separate pools of resources, such as different modalities. Based on Wickens's model, users may be able to allocate resources to attend to the auditory information while browsing a textual document, because listening and reading can be considered two tasks requiring different modalities. Furthermore, Baddeley's (1986) working memory model indicates that visual imagery information and acoustic verbal information can be held simultaneously in separate storage systems, which can be further integrated by the central executive with minimum cognitive cost. Accordingly, when users receive brief auditory information during a browsing process, information from two different modalities (visual and auditory) may be stored in different subsystems. The visual context of the browsing task may likely be processed by the visuo-spatial sketch pad, and verbal information from both auditory and visual channels may be stored in the phonologic loop. If the time spent on processing the auditory information is relatively short, users might be able to store the visual context of the browsing task in the working memory temporarily and then resume browsing without much disruption after the voice information is processed. Users, therefore, may be able to receive brief information from the auditory channel while they are retrieving information visually. Accordingly, the following hypothesis was postulated concerning this dual-modal information presentation (VA).

> H1: The user can attend to and process additional auditory information presented during a Web-browsing process. Such information will not impede the user's effectiveness in browsing textual information.

This hypothesis proposes that the introduction of additional auditory information does not significantly interfere with the Web-browsing process. Validation of this hypothesis would support the feasibility of multimodal interfaces that present distinct information simultaneously through both visual and auditory channels to optimize information delivery.

3.2. Dual-Modal Information Presentation: Visual + Assistive Auditory Information

In the presentation Visual + Assistive Auditory Information (VAA), a regular Web page was displayed in the normal visual/textual mode while additional information helping users to understand this page was presented as voice cues. According to the multiple-resource human attention model (Wickens, 1980, 1984), users may be able to allocate resources to attend to the auditory information while browsing a textual document. After users receive auditory information, they are able to integrate this information with the primary browsing task because the auditory information helps them to understand the textual information. Hence, users may be able to use this information to improve their browsing performance. The following hypothesis was postulated concerning the dual-modal information presentation (VAA).

H2: VAA presentation will allow the user to receive additional helpful information and thus improve the user's effectiveness in browsing textual information.

4. METHOD

4.1. Participants

Fifty-eight participants were recruited from a university in the U.S. midwest region, which hosts a variety of students representing different age groups, ethnicities, computer experience levels, and knowledge backgrounds. Participants were randomly assigned to one of the following three groups: V (regular visual display), VA, or VAA. As shown in Table 2, participants in the three groups share similar profiles.

4.2. Tasks and the Experiment System Setups

The experimental system included two types of tasks: primary tasks and secondary tasks. Primary tasks were general Web-browsing tasks designed to measure participants' Web-browsing performance. Secondary tasks required participants to listen to voice cues presented in the VA mode and answer questions derived solely from voice cues. These secondary tasks were intended to examine whether participants using the VA presentation could attend to and process information delivered through the auditory channel during the Web-browsing process.

As illustrated in Figure 2, a text-based Web site containing generic curriculum information was developed for this study. Additional curriculum information for speech was developed and prerecorded for auditory presentation. Participants performed tasks on a personal computer. Two sets of questions about the Web site

Table 2: Profile of Participants

				Average Age	
Group	*Total Participants*	*Number of Male Participants*	*Number of Native English Speakers*	*M*	*SD*
V	18	8	10	28.1	5.03
VA	18	8	9	28.6	6.46
VAA	22	11	14	29.9	6.94

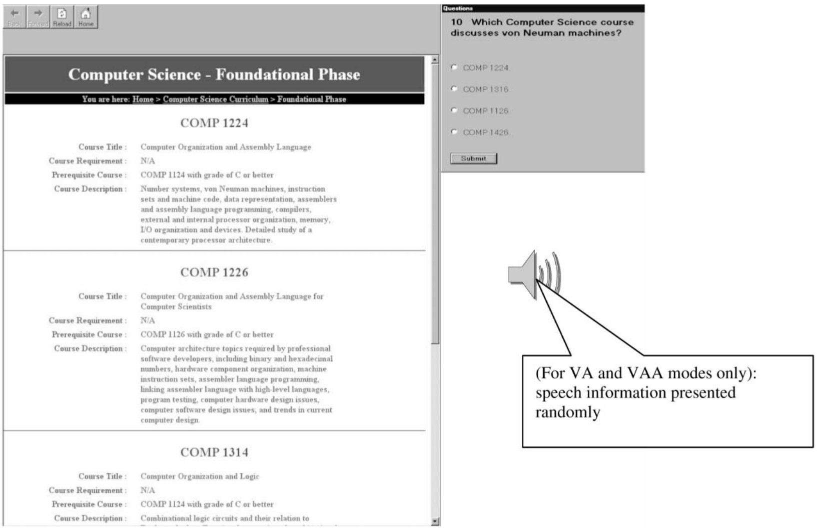
Note. V = regular visual display; VA = Visual + Auditory; VAA = Visual + Assistive Auditory Information.

and the additional speech information were designed to respectively measure participants' performance on both primary and secondary tasks. Prior to the experiment, a pilot study was conducted with 22 participants to ensure the functions of the experiment system and the appropriateness of text-based and voice-based questions. Results from the pilot study contributed to determining the appropriate duration of the experiment and the total number of questions for the experiment. Accordingly, a large number of questions were developed so that no participant could complete all of them. Considering that the fatigue factor might affect participants' capability for reading and listening comprehension, the total duration of the experiment was limited to 30 min. All questions were multiple choice and could be answered by a mouse click.

The following sections explain the setups of V, VA, and VAA modes.

V mode. In the V mode, generic curriculum information was presented visually without auditory cues. This is the control group for testing both hypotheses H1 and H2. The primary tasks for participants were to browse textual information contained in the Web site and to answer text-based questions. No secondary tasks were required.

Participants using the V mode were instructed to browse the text-based Web site, find information relevant to predefined task questions based on textual Web pages, and answer as many questions as they could in 30 min. For example, one of the text-based task questions was, "What is the requirement for taking 'Senior Design Project' in Computer Science?" There was no auditory presentation involved.



Computer Science - Foundational Phase

You are here: Home > Computer Science Curriculum > Foundational Phase

COMP 1224

Course Title : Computer Organization and Assembly Language
Course Requirement : N/A
Prerequisite Course : COMP 1124 with grade of C or better
Course Description : Number systems, von Neuman machines, instruction sets and machine code, data representation, assemblers and assembly language programming, compilers, external and internal processor organization, memory, I/O organization and devices. Detailed study of a contemporary processor architecture.

COMP 1226

Course Title : Computer Organization and Assembly Language for Computer Scientists
Course Requirement : N/A
Prerequisite Course : COMP 1126 with grade of C or better
Course Description : Computer architecture topics required by professional software developers, including binary and hexadecimal numbers, hardware component organization, machine instruction sets, assembler language programming, linking assembler language with high-level languages, program testing, computer hardware design issues, computer software design issues, and trends in current computer design.

COMP 1314

Course Title : Computer Organization and Logic
Course Requirement : N/A
Prerequisite Course : COMP 1124 with grade of C or better
Course Description : Combinational logic circuits and their relation to

10 Which Computer Science course discusses von Neuman machines?

○ COMP 1224
○ COMP 1316
○ COMP 1126
○ COMP 1426

(For VA and VAA modes only): speech information presented randomly

FIGURE 2 Screen shot of the experimental system.

VA mode. In the VA mode, additional curriculum information, irrelevant to the text questions, was randomly presented through the auditory channel while participants were browsing the textual content of the Web site. In addition to the primary browsing tasks, participants in this group had to listen to auditory output that might be presented at any time during the experiment and answer questions derived from auditory information. Voice cues were presented randomly during the browsing session. Immediately after each voice message was played, a question derived from this voice message would pop up on the screen in text format. The participant had to answer the voice-based question (as a secondary task) before moving on. The time that the participant spent on answering voice-based questions was not counted toward the 30-min period for Web-browsing tasks.

VAA mode. In the VAA mode, assistive information helping participants to understand the textual information was presented randomly through the auditory channel while participants browsed the textual content of the Web site. In other words, in addition to the primary browsing task, participants in this group needed to listen to the auditory output that might be presented at any time during the experiment. The assistive speech information was synchronized with randomly selected text-based questions so that participants could not anticipate and wait for voice cues. The assistive information would help participants find information needed to answer a text-based question. For example, the assistive information for the text-based question, "What is the requirement for taking 'Senior Design Project' in Computer Science?" was "The 'Senior Design Project' course can be found in the course requirements page at Department of Computer Science." Participants could use the auditory information they received to facilitate the primary browsing tasks. No secondary tasks were used in the VAA mode.

4.3. Independent and Dependent Variables

The only independent variable was the information presentation mode. There were three modes in total: V, VA, and VAA.

The dependent variable was user performance. User performance was measured by the number of correctly answered questions related to the text-based Web site, the number of correctly answered questions related to the voice cues (for the VA group only), and accuracy. Accuracy was defined as the number of correctly answered questions divided by the total number of questions that the participant attempted to answer. Accuracy was included as one performance measure for precaution, because some participants might have answered additional questions by guessing, which did not necessarily indicate improved performance.

Satisfaction was not a dependent variable used to test the two hypotheses but it was measured as a precaution, because a user's subjective perception is vital to measuring the effectiveness of an interface. User satisfaction was measured by a questionnaire derived from the technology acceptance model (Davis, 1989). According to this model, perceived ease of use and perceived usefulness are predictors of a user's intention to adopt a new technology. If a user feels more satisfied with a technology's ease of use and usefulness, it is more likely that he or she will adopt the technology. Therefore, it is reasonable to use perceived ease of use and perceived usefulness as surrogates for satisfaction. In the satisfaction questionnaire, five questions measure the perceived ease of use, five questions measure the perceived usefulness, and one question measures the user's general satisfaction. These questions were drawn from Davis's study. The question on the user's general satisfaction states "In general, I am satisfied with the Curriculum Planning Application." The total scores of the items measuring perceived ease of use and perceived usefulness were calculated and included in the analysis.

4.4. Procedure

Each participant was asked to sign a consent form and complete a pre-experiment questionnaire before participating in the experiment. Upon completion of the questionnaire, each participant received instructions, online and in hard copies, for using the experimental browser and performing the required tasks. Each participant then began a training session to browse a sample Web site with visual (and auditory) information similar to the version the participant would later encounter in the experiment. Upon successful completion of the training test, the participant began the experimental tasks. Before beginning the experiment, participants were informed of the time limit and asked to focus their attention on the tasks at hand. The browsing process was programmed to automatically terminate at the 30-min mark. All participants were asked to correctly answer as many questions as they could and as quickly as possible. A large number of questions were developed so that no participant could finish all of them. Upon completing the experimental tasks, participants were asked to complete a questionnaire about their satisfaction regarding the information display mechanism. Participants in the VA and VAA groups were then debriefed to provide additional insights about their experience in processing and using the dual-modal interface. The debriefing sessions were recorded for data analysis.

5. RESULTS AND DISCUSSIONS

5.1. Test of Hypothesis H1

The intention of hypothesis H1 was to test the effectiveness and feasibility of presenting additional information through the auditory channel during the Web-browsing process. H1 postulates that additional auditory information presented during a Web-browsing process can be received by users and will not negatively impact their performance on browsing tasks. The dependent variables were the number of correctly answered questions related to the text-based Web site, the number of correctly answered questions related to the voice cues, and accuracy. A simple *t* test between the V and VA groups was used to test H1. If the average number of correctly answered questions that are related to the voice cues heard by the

VA group was significantly greater than zero, then users can attend to additional auditory information during a Web-browsing process.

Table 3 presents the mean values, standard deviations, and *t*-test results of the dependent variables measuring user performance for both V and VA groups. No significant differences were found between the two experimental groups in the number of correctly answered questions related to the text-based Web site, *t*(34) = –0.10, *p* = .92, and accuracy for text-based questions, *t*(34) = –1.34, *p* = .19, at α = .05 level. Power analyses suggest that the power of detecting 10% difference in number of correctly answered text questions was 0.11 and the power for accuracy was 0.50. An additional *t* test indicates that the average number of correctly answered questions related to the voice cues in the VA group was significantly greater than zero (i.e., no speech information was perceived), *t*(17) = 16.27, *p* = .0001. The mean of accuracy for voice-based questions was also significantly greater than 0.25 (*M* = 0.959), *t*(17) = 47.85, *p* = .0001. Because there were four options for each voice-based question, the average probability of guessing the correct answers would be 0.25. An accuracy rate significantly greater than 0.25 implies participants may have derived the correct answers from information presented through the auditory channel instead of guessing. These results indicate that participants in the VA group did successfully receive some information delivered through the auditory channel.

The results of this experiment fully support hypothesis H1. Furthermore, results agree with prior research findings. Based on the multiple-resource theory model (Wickens, 1980, 1984), two tasks can be performed together more efficiently to the extent that they require separate pools of resources, such as different modalities. According to Baddeley's (1986) working memory model, tasks using different working memory subsystems (visuo-spatial sketch pad and phonologic loop) should not interfere. While users are visually browsing information, they may be able to receive brief information from the auditory channel.

Table 4 presents the Cronbach's alpha value, mean, standard deviation, and *t*-test results for perceived ease of use, perceived usefulness, and general satisfaction. The high Cronbach's alpha values for perceived ease of use and perceived usefulness suggest that the questionnaire was reliable and valid. No significant differences between the V and VA modes were found in perceived ease of use, *t*(34) =

Table 3: Comparison of User's Performance Between V and VA Groups

	V Mode[a]		*VA Mode*[b]			
Variables	*M*	*SD*	*M*	*SD*	*t*	*Pr* > | *t* |
Number of correctly answered questions related to the text-based Web site	21.6	8.99	21.9	7.31	–0.10	0.92
Accuracy for text-based questions	0.714	0.1245	0.767	0.1114	–1.34	0.19
Number of correctly answered questions related to the voice cues	NA	NA	10.1	2.62	16.27	0.0001
Accuracy for voice-based questions	NA	NA	0.959	0.0628	47.85	0.0001

Note. V = regular visual display; VA = Visual + Auditory. For the VA group only, the number of correctly answered questions related to the voice cues was tested against a null hypothesis H0 = 0, and accuracy for voice-based questions was tested against a null hypothesis H0 = 0.25.

[a]*n* = 18. [b]*n* = 18.

–1.89, $p = .067$; perceived usefulness, $t(34) = –0.61$, $p = .54$; and general satisfaction, $t(34) = –0.95$, $p = .35$, at $\alpha = .05$ level. Results as presented in Table 4 indicate that the introduction of auditory information in a browsing process did not significantly reduce participants' satisfaction.

- Debriefing interviews conducted for the VA group at the end of the experiment reveal the following points:
- Participants could recognize and remember key words and phrases contained in voice messages.
- The first few words in the voice messages were distracting but caught the participant's attention.
- There seemed to be a filtering process when a voice message started. During this filtering process, a participant would decide whether to listen to the voice information. If the voice information appeared to be relevant to the primary browsing task, the participant would more likely pay attention to it. This observation is consistent with the filter theory proposed by Broadbent (1958).
- Participants felt that information irrelevant to the primary browsing task would cause more distraction than relevant information and thus might hinder the browsing task.

5.2. Test of Hypothesis H2

The intention of hypothesis H2 was to test the effectiveness and feasibility of presenting helpful information through the auditory channel during the Web-browsing process. H2 postulates that helpful information presented through the auditory channel during the Web-browsing process can be received by users and will improve their performance on browsing tasks. The dependent variables were the number of correctly answered questions and accuracy.

Table 5 presents the descriptive statistics and t-test results of the dependent variables. T tests between the two groups suggest that the differences were not significant at the $\alpha = .05$ level—the number of correctly answered questions, $t(38) = –0.99$, $p = .33$; accuracy, $t(38) = –1.74$, $p = .090$. A power analysis was conducted, and the results suggest that with the current sample size and standard deviation, the power of number of correctly answered questions was 0.125 and the power of accuracy

Table 4: Comparison of User's Satisfaction Between V and VA Groups

	V Mode[a]			*VA Mode*[b]				
Variables	*Cronbach's* α	*M*	*SD*	*Cronbach's* α	*M*	*SD*	*t*	*Pr > |t|*
Perceived ease of use	0.91	19.8	8.13	0.88	23.8	3.78	–1.89	0.067
Perceived usefulness	0.90	22.6	7.73	0.88	22.9	5.71	–0.61	0.54
General satisfaction	NA	4.7	1.80	NA	4.9	1.32	–0.95	0.35

[a]$n = 18$. [b]$n = 18$.

was 0.398. The low statistic power might have contributed to the failure of detecting differences between the two groups.

Because only some of the text-based questions were provided with assistive voice cues, the benefits gained by participants in the VAA group were limited and might be too small to be reflected in the overall performance. Therefore, it was imperative to look into user's performance associated specifically with the questions with assistive voice cues. In Table 6, the number of correctly answered questions with voice hints, accuracy of these questions, and time spent on each question was analyzed. Results suggest that users in the VAA group answered the questions with voice cues with a significantly higher accuracy, $t(38) = –2.25$, $p = .030$, and spent significantly less time on each question, $t(38) = 2.10$, $p = .043$, than did users in the V group. Two questions with voice hints that were answered by all participants were also analyzed. Results consistently show that users in the VAA group spent less time on both questions—Question 1, $t(38) = 2.05$, $p = .048$; Question 9, $t(38) = 3.90$, $p = .0004$—than did users in the V group. These analyses supported that users in the VAA group spent significantly less time and correctly answered more questions with voice hints than did users in the V group.

Table 5: Comparison of User's Performance and Satisfaction Between V and VAA Groups

	V Mode[a]		*VAA Mode*[b]			
Variables	*M*	*SD*	*M*	*SD*	*t*	*Pr* > | *t* |
Number of correctly answered questions related to the text-based Web site	21.6	8.99	23.7	7.14	–0.99	0.33
Accuracy for text-based questions	0.714	0.1245	0.778	0.1061	–1.74	0.090
Perceived ease of use	19.8[c]	8.13	22.5[d]	5.51	–1.26	0.22
Perceived usefulness	22.6[e]	7.73	23.4[d]	5.82	–0.98	0.33
General satisfaction	4.7	1.80	4.1	1.34	–0.63	0.53

Note. V = regular visual display; VAA = Visual + Assistive Auditory Information.
[a]*n* = 18. [b]*n* = 22. [c]Cronbach's α = 0.91. [d]Cronbach's α = 0.93. [e]Cronbach's α = 0.90.

Table 6: User's Performance in Questions With Voice Hints

	V Mode[a]		*VAA Mode*[b]			
Variables	*M*	*SD*	*M*	*SD*	*t*	*Pr* > | *t* |
Number of correctly answered questions with voice hints	7.2	3.85	8.6	–1.48	0.147	
Accuracy of all questions with voice hints	0.677	0.2412	0.801	0.1214	–2.25	0.030
Average time spent on each question with voice hints (sec)	80.39	52.719	55.22	18.365	2.10	0.043
Time spent on Question 1 (with voice hints; sec)	150.76	87.245	101.58	64.597	2.05	0.048
Time spent on Question 9 (with voice hints; sec)	93.99	64.237	39.11	14.644	3.90	0.0004

Note. V = regular visual display; VAA = Visual + Assistive Auditory Information.
[a]*n* = 18. [b]*n* = 22.

No significant difference was found in the satisfaction between V and VAA groups: perceived ease of use, *t*(38) = –1.26, *p* = .22; perceived usefulness, *t*(38) = –0.98, *p* = .33; general satisfaction, *t*(38) = –0.63, *p* = .53.

Debriefing interviews conducted for the VAA group at the end of the experiment reveal that the voice information helped participants perform the primary tasks in addition to findings from interviews of VA users:

Based on the previously mentioned analyses, hypothesis H2 was supported by results from this experiment. The results agree with prior research findings (Wickens, 1980, 1984). Based on the multiple-resource human attention model (Wickens, 1980, 1984), two tasks can be performed together more efficiently to the extent that they require separate pools of resources, such as different modalities. While users are visually browsing information, they may be able to receive brief information from the auditory channel and use this information in their primary browsing tasks.

6. CONCLUSIONS AND LIMITATIONS

This study investigated the feasibility and effectiveness of presenting information using multiple modalities. Two dual-modal information presentations were proposed and tested through a controlled experiment. Findings from this study suggest the following: (a) Users may be able to attend to and use information presented through the auditory channel while visually browsing textual information, (b) relevant speech information may facilitate the efficiency of user performance, and (c) relevant speech information seems less distracting based on participants' observations. Therefore, it is possible to use multimodalities (visual plus auditory modes) to present more information than in a single modality. These findings have profound implications to future research in multimodal interface design. Multimodal interfaces are especially promising for mobile applications due to the nature of wireless technology. Mobile devices have two main constraints: small screen size and their mobile usage (Chan et al., 2002). Compared to desktop or laptop computers, mobile devices typically have a very small screen on which only a very limited amount of visual information can be presented. When the device is used on the move, it makes the reading of textual information even more difficult. Multimodal interfaces may help to address these constraints by delivering information through multiple sensory modalities such as visual and auditory channels.

This study presents a first step toward the research on multimodal interface design. There are still many issues about multimodal interfaces to be explored and investigated. Future study should focus on addressing what information and how much information should be presented in the visual and auditory modes, respectively.

REFERENCES

Archer, N., Head, M., Wollersheim, J., & Yuan, Y. (1996). Investigation of voice and text output modes with abstraction in a computer interface. *Interacting With Computers, 8,* 323–345.

Baddeley, A. (1986). *Working memory.* New York: Oxford University Press.

Baggett, P., & Ehrenfeucht, A. (1983). Encoding and retaining information in the visuals and verbals of an educational movie. *Educational Communication and Technology Journal, 31*(1), 23–32.

Broadbent, D. E. (1958). *Perception and communication*. Elmsford, NY: Pergamon.

Chalfonte, B., Fish, R., & Kraut, R. (1991). Expressive richness: A comparison of speech and text as media for revision. In *Proceedings of CHI'91* (pp. 21–26). New York: ACM Press.

Chan, S., Fang, X., Brzezinski, J., Zhou, Y., Xu, S., & Lam, L. (2002). Usability for mobile commerce across multiple form factors. *Journal of Electronic Commerce Research, 3,* 187–199.

Cohen, P. R. (1992). The role of natural language in a multimodal interface. In *Proceedings of the ACM Symposium on User Interface Software and Technology* (pp. 143–149). New York: ACM Press.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13,* 319–340.

Deatherage, B. H. (1972). Auditory and other sensory forms of information presentation. In H. P. Van Cott & R. G. Kinkade (Eds.), *Human engineering guide to equipment design* (Rev. ed., pp. 123–160). Washington, DC: U.S. Government Printing Office.

DeHaemer, M., & Wallace, W. (1992). The effects on decision task performance of computer synthetic voice output. *International Journal of Man–Machine Studies, 36,* 65–80.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.

McCormick, E. J., & Sanders, M. S. (1982). *Human factors in engineering and design*. New York: McGraw-Hill.

Mousavi, S. Y., Low R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology, 87,* 319–334.

Navon, D., & Gopher, D. (1979). On the economy of the human information processing system. *Psychological Review, 86,* 214–255.

Nugent, G. (1982). Pictures, audio, and print: Symbolic representation and effect on learning. *Educational Communication and Technology, 30,* 163–174.

Proctor, R., & Van Zandt, T. (1994). *Human factors in simple and complex systems*. Needham Heights, MA: Allyn & Bacon.

Schlosser, R., Belfiore, P., & Nigam, R. (1995). The effects of speech output technology in the learning of graphic symbols. *Journal of Applied Behavior Analysis, 28,* 537–549.

Sipior, J., & Garrity, E. (1992). Merging expert systems with multimedia technology. *Data Base, 23*(1), 45–49.

Streeter, L. (1988). Applying speech synthesis to user interfaces. In M. Helander (Ed.), *Handbook of human–computer interaction* (pp. 321–343). New York: Elsevier Science.

Treisman, A. M. (1964). Verbal cues, language, and meaning in selective attention. *American Journal of Psychology, 77,* 206–219.

Wickens, C. (1980). The structure of attentional resource. In R. S. Nickerson (Ed.), *Attention and performance VIII* (pp. 239–257). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Wickens, C. (1984). Processing resources in attention. In R. Parasuraman & R. Davies (Eds.), *Varieties of attention* (pp. 63–102). New York: Academic.