Users' Performance and Preferences for Online Graphic, Text and Auditory Presentation of Instructions

Ellen Ayoob[a], Kerry Bodine[a], Dan Bohus[b], Alexander Rudnicky[b] and Jane Siegel[a]
[a]Human-Computer Interaction Institute and [b]Department of Computer Science
Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213 USA
[ema, kbodine, dbohus, air, jals] @cs.cmu.edu

ABSTRACT
Traditional technical manuals consist primarily of text supplemented by tabular and graphic presentation of information. In the past decade technical information systems have increasingly been authored for presentation on computers instead of on paper; however a stable set of standards for such manuals has yet to evolve. There are strong beliefs but little empirical evidence to guide standards development within companies producing Interactive Electronic Technical Manuals (IETMs).

The current study compares three different modes of instruction presentation for mechanical assembly tasks (graphic, text, and auditory), using a Wizard of Oz paradigm. Study participants preferred graphically-presented information and they completed the tasks fastest using this presentation mode. We found no significant difference in performance or preference between text and audio conditions. Nevertheless users indicated a clear desire that graphic presentation be supplemented by other modes. Study results will be useful for designers of multi-modal interfaces for online instruction systems.

Keywords
IETMs, technical manuals, mode preferences, online documentation systems.

INTRODUCTION
Little is known about the effective presentation of technical information, such as IETMs, or users' modality preferences for purposes of supporting tasks that require mobility or involve users whose eyes and/or hands are busy. Past studies that have compared online presentation of instructions to paper have found that online systems do not improve users' performance [1, 4]. Currently deployed technologies are computer notebook-based and support keyboard or pen-based interaction [5]. Both of these options are problematic when users need their hands to take apart, repair or reassemble equipment such as for aircraft, ship or automotive systems. Online technical manual systems that enable users to invoke speech input and audio output along with high quality visual information are potentially highly useful. Understanding how these modes are used online is therefore of great interest to designers and users of online instructional products.

We report an initial experiment that studies, under controlled conditions, the relative effectiveness of graphic, text and audio presentation of technical instructions. The purpose of the experiment is to inform design choices for a wearable system presenting rich multi-media technical information for maintenance personnel in the field using a heads-up display and speech input and output. The technical materials will include step-by-step procedure presentation, searchable IETMs, and animated simulations of mechanical and electronic systems.

The LEGO® MindStorms Robotics Invention System™ 2.0 (model 3804) is an advanced LEGO kit, directed at children 12 years or older, that enables assembly of complex, programmable robots. It has been used in college-level Mechanical Engineering and Robotics courses and has spawned a diverse and sophisticated user community[1]. Based on these considerations we believe that MindStorms is a good analogue for our target domain (maintenance and repair of mechanical and electronic systems [4]) while providing the benefit of tractability (conducive to laboratory experimentation using a student population).

METHOD
Assembly Tasks
To identify a set of comparable LEGO tasks, we asked twenty pilot participants to each put together three different assemblies selected from the standard MindStorms instruction manual, for a total of five trials for each of twelve assemblies. All instructions were presented in their original form (participants worked from the manual). The completion times and errors, as well as number of parts and unique types of parts, are shown in Table 1.

[1] For a course see http://www.ceeo.tufts.edu/al/en10/syllabus.html. For a resource guide see: http://portal.acm.org/citation.cfm?id=374699 &jmp=indexterms&dl= portal &dl=ACM

Task	*Task Time (median)*	*Final Error (mean)*	*Parts (types)*
Roverbot (RB) Base	**12:27**	**1.20**	**83 (21)**
RB Legs	7:01	0.40	38(9)
RB Tracks	18:14	0.60	46(19)
RB Single Bumper	**9:30**	**1.20**	**32 (18)**
RB Double Bumper	15:09	1.00	38(19)
Acrobot Base	16:23	0.00	93(21)
Inventorbot (IB) Base	20:10	0.40	71(30)
IB Body	18:21	0.20	66(27)
IB Slap Arm	10:20	0.80	30(20)
IB Squeeze Arm	6:23	0.20	29(19)
IB Hat Arm	**11:02**	**1.40**	**43 (22)**
IB Thrower Arm	**3:27**	**0.00**	**17 (11)**

Table 1: Standard LEGO assembly procedure (tasks in bold are the ones selected for use in the experiment)

The criterion for choosing assemblies to use in the main experiment was a combination of median assembly time and error rate. A practice assembly task (Inventorbot Thrower Arm) was chosen for its short completion time and low error rate. Three main assembly tasks (Roverbot Base, Roverbot Single Bumper, and Inventorbot Hat Arm) were chosen for their moderate completion times (approximately 9 to 12 minutes) and high error rates. Table 2 shows the number of screens, steps and average number of steps per screen for each of the chosen assemblies.

Task	*Screens*	*Total Steps*	*Average Steps/Screen*
Base	22	61	2.8
Single Bumper	19	52	2.7
Hat Arm	20	66	3.3

Table 2: Experimental task characteristics**.**

Design

There were three experimental conditions based on the presentation mode of the assembly instructions: text, graphic, or audio. The order of modes and assembly tasks was randomized according to a Græco-Latin scheme.

The *Graphic* condition presented unaltered part selection and assembly diagrams from the LEGO MindStorms instruction manual. (See Figures 1a & 1b.) The *Text* instructions were written by the experimenters, who iteratively tested them for comprehensibility with pilot participants over a period of two months. Text was presented as browser-generated HTML in 12 point Verdana. *Audio* instructions were spoken versions of the text instructions recorded by a female speaker.

Each of the three main assembly tasks consisted of approximately twenty screens. For the diagram condition, one diagram was displayed on the screen at a time. In order to maintain parallelism among the three main conditions, the text and audio instructions were broken down into sets of one or more steps corresponding to the actions depicted in each diagram. (See Table 2) For the text condition, one set

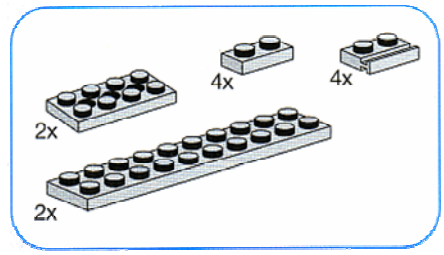

Figure 1a. Example diagram: part selection

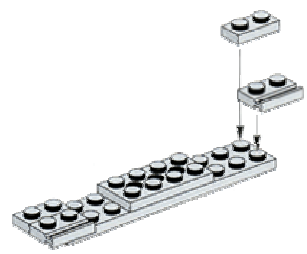

Figure 1b. Example diagram: assembly

of analogous steps was displayed on the screen at a time. (See Figure 2.) For the audio condition, one set of steps was also associated with each screen, but the content area of screen was blank and the steps were played one at a time through the speakers. The audio paused at the end of each step until the participant issued a command to go to the next or previous step, or to repeat the current step.

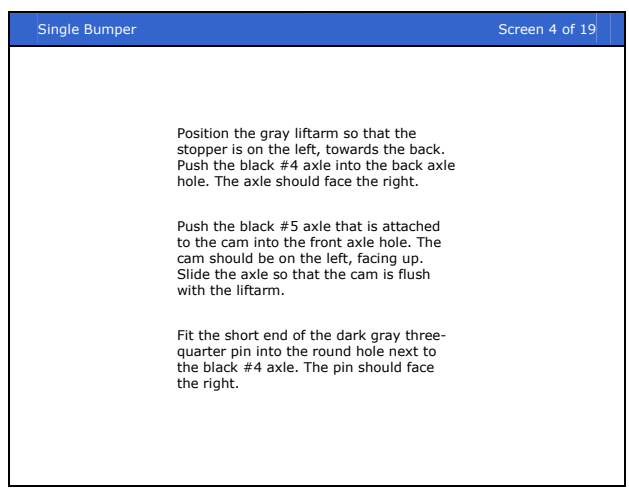

Figure 2. Text layout for a single screen with 3 steps.

For all conditions, the screen number (e.g., "Screen 4 of 19") was displayed in the upper right corner. Several additional cues, used only in conjunction with the audio condition, helped to orient participants to their relative location within the instructions. At the beginning of each screen, a message stated the new screen number (e.g., "Screen

four"). At the end of each step, a single "ding" sound played. At the end of the last step for each screen, a double "ding" sound played.

For all conditions, a diagram of the completed assembly was displayed following the final screen of instructions so that participants could make any necessary adjustments.

Participants were able to use spoken commands for navigating through the instructions (e.g., "next", "previous", "repeat") so that they could keep their hands on the assembly tasks. They were told that there was a speech recognition system; however a second experimenter (the Wizard) actually manipulated the system based on the participants' spoken commands.

Completion times and the number of final errors in each assembly were recorded. A post-experiment survey probed participants' reactions to the tasks and to the different modes, asking them to note, for each mode, its positive and the negative qualities, and to make specific suggestions for improvement.

Equipment

Figure 3 shows the participant's work space. Text and diagram instructions were displayed on the screen, while audio instructions were presented through speakers to the left of the monitor. Participants assembled the pieces in a work area, approximately 12" x 18". Individual LEGO parts were sorted by type and placed in a plastic tray with labeled bins between the monitor and the work area. A speech command reference key was to the left of the monitor. A video record of the session captured an image of the user and the workspace. Two desk-top microphones were used to capture participant speech and to convey it to the Wizard. The primary experimenter was present at all times in the room, sitting behind and to the side of the participant.

Figure 3. Lab configuration: Participant view

The Wizard sat behind a partition and was not visible to the participant. The Wizard used an HTML and JavaScript-based control application developed for this experiment, consisting of a user-visible window plus additional windows for control and feedback generation. The Wizard was instructed to respond only to legal user commands and to issue one of several pre-recorded error messages otherwise.

Participants

Eighteen undergraduate and graduate students (aged 18-26, median 21; 10 male, 8 female) were recruited. The age range of our participants corresponds to the age range of mechanics in our target population. Prospective participants were prescreened for fluency in English and unfamiliarity with the specific experimental assemblies. Participants also received a standardized spatial reasoning test to insure reasonable proficiency (this excluded one low-scoring participant from the study).

Procedure

Upon arrival, users were read a standard set of training materials and instructions and had the opportunity to ask clarifying questions. They then assembled the simple Thrower Arm, which familiarized them with the LEGO parts and nomenclature, three presentation modes (each mode was used for a third of the instructions), and speech commands. They then proceeded to the main part of the experiment, where they were asked to build the three main assemblies (see Figure 4 for an example), each of which used a different presentation modality. Upon completion, they filled out a personal demographic information form and the post-experimental questionnaire. Finally, all participants were debriefed.

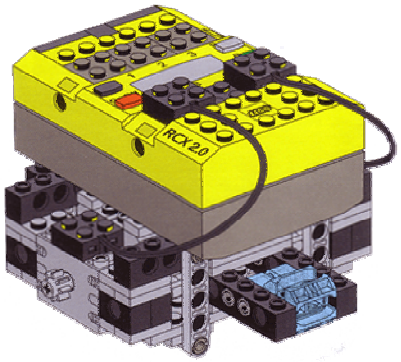

Figure 4. Roverbot Base, one of four assemblies used

RESULTS

The usefulness of a presentation mode depends both on its effectiveness in supporting task performance and in its acceptability to the user. We accordingly examined both task-based and subjective evaluations.

An ANOVA comparing Completion Times, using Task, Mode and Order as factors found significant Task ($F_{2,24}$=18.1, p<0.001) and Mode ($F_{2,24}$=49.5, p<0.001) effects with no Order effect. Individual comparisons (using a Tukey test) show that the Graphic mode is significantly faster than Text and Audio (p<0.001), which in turn are not different from each other (p>0.1).

An ANOVA comparing Mode preferences is significant ($F_{2,24}$=61, p<0.001) with only Graphic appearing signifi-

cantly better in individual comparisons; we find a similar pattern for instruction understandability ($F_{2,24}$=18, p<0.001). Results are shown in Table 3.

Task	*Time (median)*	*Comple-tion Rate*	*Error (mean)*
Roverbot (RB) Base	20:05	83%	0.27
RB Single Bumper	11:07	94%	0.12
Inventorbot Hat Arm	16:31	89%	0.13

Mode	*Time*	*Compl*	*Error*	*Pref*	*Unders*
Graphic	9:44	100%	0.11	1.00	1.72
Text	19:33	83%	0.13	2.44	3.06
Audio	20:56	83%	0.27	2.5	3.22

Table 3: Performance by Task and by Mode. *Compl*: task completion rate; *Pref*: participant mode preference (mean); *Unders*: participant-rated understandability of instructions (mean).

Participant questionnaire responses provided useful insight into functional requirements for design of online instructions. Each observation reported below is supported by comments from at least 5 different participants.

While we fully expected that the original LEGO graphic instructions would very effectively communicate information, we were surprised to find that participants noted several shortcomings: the inability to manipulate the graphic (in 3-D) in order to better understand the target assembly, the lack of verbal annotations that could explain difficult operations, and the packing of too many individual steps into a single drawing. Text and Audio were both faulted for providing poor visual orientation, while Audio presentation was additionally faulted for lack of control over presentation pace and for poor navigation functionality. At the same time, Audio was positively noted for its ability to support eyes-busy assembly. Participants had no negative comments about speech input, but expressed a desire for better navigation functionality and faster reaction speed. A true speech recognition-based interface would of course introduce errors and potentially lead to a different judgment.

DISCUSSION

Taken together the results of this study suggest the primacy of graphic information in assembly tasks, and in that respect are consistent with results from other studies [2]. In itself this should not be a surprising finding, particularly as the graphic materials we selected have benefited from long-term improvement (LEGO first published graphic assembly instructions in 1964).

Of particular interest, therefore, is the strong suggestion that participants in this study were not entirely satisfied with the graphic mode by itself, particularly in a setting in which they could reflect on the potential contributions of other modes (text and audio). The picture that emerges is one in which users have a desire to work with multiple, complementary modes of presentation. As well, they seek to have active control over the presentation (such as an easily manipulable graphic display and an interface that allows efficient access to related text and audio information). The challenge at this point is to understand how users organize their use of multiple presentation modes over the course of a task and how an interface design can best support these strategies.

The current study nevertheless suggests some useful design guidelines: Diagrams were particularly useful to participants because they provided ongoing verification of correctness. This implies that task monitoring and feedback are a key property of a good online documentation system.

Participants' comments from their post-experimental questionnaires pointed to several functional requirements for design of online instructions, including:

- Importance of providing graphic/diagrammatic information in conjunction with text-based instructions;
- Providing ability for participants to adjust the orientation of diagrams to see the view they desire; and
- Enabling control of presentation pace of instructions by participants.

ACKNOWLEDGMENTS

This study was conducted with support from the U.S. Navy under contract N61339-02-C-0043. We would like to thank Brian Gollum for his technical contributions to the experiment, and Ilene Harris, Timothy Driscoll, and Erin Li for their assistance in conducting this experiment.

REFERENCES

1. Chen, D. and Rada, R. (1996) Interacting with Hypertext: A Meta-Analysis of Experimental Studies, Human-Computer Interaction, Lawrence Erlbaum Associates, Inc., V11, 125-156.

2. Kraut, R.E., Miller, M.D., and Siegel, J. (1996) Collaboration in Performance of physical tasks: Effects on outcomes and communication, Proceedings of CSCW'96 (57-66). NY: ACM.

3. Oviatt, S., Darrell, T., and Flickner, M. (2004) Multimodal Interfaces that Flex, Adapt, and Persist, Communications of the ACM Vol. 47, No. 1, 30-33.

4. Siegel, J., Hyder, E., Moffett, J, and Nawrocki, E., (2001) IETM Usability: Using Empirical Studies to Improve Performance Aiding, Carnegie Mellon University Technical Report: CMU-CS-01-131.

5. Siegel, J. and Hyder, E. (1999) Enabling Collaborative Work: Work in Progress, Personal Technologies 3:65-71, Springer-Verlag, London.