Natural Language and Direct Manipulation Search Tools in a Multimodal Information System

Sheryl L. Miller
Human Factors and Applied Cognition
ARCH Lab - George Mason University
Fairfax, VA 22030-4444 USA
703 993 4667
smillerb@osf1.gmu.edu

J. Gregory Trafton
Naval Research Laboratory, Code 5513
Washington, D.C. 20375-5337 USA
202 767 3479
trafton@itd.nrl.navy.mil

ABSTRACT This study evaluated the usability of an existing, information-dense workspace, which places burdensome search requirements on its users. Direct manipulation principles were contrasted with natural language and multimodal interaction. The results indicated that natural language was an effective and usable search tool in a complex information workspace

Keywords Natural language, direct manipulation, graphical user interface, multimodal system

INTRODUCTION The quality of human-computer interactions is limited by the ease with which people manipulate information in the world. This limitation is increasingly important as people become more reliant on information obtained and stored on personal computers. Information that was organized in file cabinets, calendars, and bookshelves is now constrained by the size of a screen. Clutter makes the location of screen objects more difficult, increasing the need for effective search strategies. Natural language and direct manipulation are frequently contrasted as effective search tools (Cohen, 1991; Shneiderman, 1998; Oviatt, 1996; Oviatt, DeAngeli, & Kuhn, 1997). Does natural language provide the user with search functionality beyond that found in direct manipulation interfaces (i.e., the graphical user interface (GUI))? And if different modalities do afford greater success in information-dense environments, are users sensitive to these capabilities?

It is proposed here that natural language is actually a successful tool for on-screen object search, despite its characterization as a textual query tool in a "machine-centered artificial intelligence (Star Trek) scenario" (p.294, Shneiderman, 1998). However, direct manipulation is the standard mode of computer interaction (most users interact with computers via a GUI). It is proposed that people will be sensitive to task differences that make one modality more appropriate than another (natural language or direct manipulation) for interacting with a complex information space and will transition between easily.

METHOD

Participants and Materials Sixteen undergraduates were each randomly assigned to use one of two interfaces: Graphical User Interface only (GUI) or both Graphical User Interface and Natural Language interface (Combined). The natural language system was designed to offer users comparable functionality to the GUI, allowing them to type using everyday language. A demonstration and all experimental trials were run on a Sun SPARC Computer Workstation.

Interface, Task, and Procedure InterLACE is an existing geographical system equipped with the natural language processor, NAUTILUS (Wauchope, 1996), in which users navigate a map of Germany and gather environmental information. This system is representative of many complex information systems. Multiple objects are displayed on-screen and are symbolized in some pictorial way (e.g., rivers as blue lines). However, only some objects are labeled by name. The ambiguity imposed by the lack of object labels makes selecting among objects a difficult search task. It is predicted that object search will be more efficient when students are able to use natural language, rather than only direct manipulation.

Students were evaluated by the number of steps that they use to complete each task instruction. An example of a DM step is that in which a student highlights an object (X) and clicks on the "Identify" or "Distance" button, to get information about the particular object. A comparable NL step requires the student to type "Where is X?" or "Show X" to find the particular object or "How far away is X" to find out distance. While a student in the GUI condition may use only DM steps, a student in the Combined condition may use DM steps, NL steps, or a combination of steps (both DM and NL steps to complete an instruction). Although students in the Combined conditions have these options, they are not given explicit instruction as to when they should

implement action in a certain modality. After training, each student completed a series of 26 experimental instructions and a questionnaire that assessed the usability of the map application. In an effort to be conservative in our evaluation of the multimodal interface, natural language errors were counted against a user's performance

RESULTS The results of a MANCOVA indicated that there were no significant effects of the covariates (age and spatial and verbal ability pre-test scores, for all $p>0.28$)). Overall, students used more steps to complete Instructions that contained Unlabeled Objects than contained Labeled Objects ($p<0.001$), indicating that Unlabeled Objects create a more complex search task than Labeled Objects. Students in the Combined condition used less steps than students in the GUI condition to complete instructions only when the instructions contained Unlabeled Objects ($F(1,14)=36.98$, $p<0.001$). How were they able to improve their performance over GUI students?

Students in both conditions used the same number of DM steps to complete instructions containing Labeled Objects. However, students in the Combined condition used significantly fewer DM steps for instructions containing Unlabeled Objects than did students in the GUI condition ($p<0.001$). The students in the Combined condition also utilized NL steps to complete instructions, and they used significantly more NL steps for those instructions that contained Unlabeled objects than Labeled objects. These findings indicate students using a multimodal interface dealt with search tasks (looking for Unlabeled Objects) by using NL steps. This is an option that students in the GUI condition did not have, leading them to use significantly more (DM) steps to complete instructions requiring search.

A closer look at student actions in the Combined condition shows how these students alternatively used the DM and NL modalities. Users were expected to utilize NL to search through Unlabeled information more than they used DM, evidence that natural language is frequently used to search the display. Only 9 percent of the time, did students in the Combined condition use DM alone to handle an instruction containing an Unlabeled object; 91 percent of the time, they used NL or some combination of DM and NL. In contrast, they used DM alone to search for Labeled objects 46 percent of the time. The typical pattern of interaction for combining DM and NL involved the use of natural language to find an object, followed by a subsequent DM step (or set of DM steps) to complete the instruction (e.g., to move, to identify size, or to calculate distance). Take, for example, the instruction, "What size is airstrip Lager Hammelburg?" One participant in this study (#001) used two steps to complete the instruction. First, the student typed, "show me lager-hammelburg" (NL). Second, the student clicked the Identify button to determine the size of the airstrip (DM). This pattern was seen for 86 percent of all the instructions on which modalities were combined. By doing so they were able to reduce the number of steps needed to perform complex search tasks, compared to students who used the GUI alone.

DISCUSSION Almost with out exception, students used natural language to locate an object on the map. Not only was this method efficient, but it was self-generated; students were not taught to perform this way. Critics might argue that the GUI here was simply a poor interface. Yet, the behavior of the study participants does not indicate this. First, they all were familiar with traditional GUI design in that they all had experience with personal computers, and InterLACE's interface was traditional, allowing drag and drop, mouse click selection, etc. Second, the study was designed such that the task could be completed by using only the GUI. Third, the participants in the Combined condition frequently used the GUI, and they used it regularly to complete certain tasks. Finally, the students in the GUI and Combined conditions rated the interface equally on the exit questionnaire for issues of usability (all questions non-significant, $p > 0.15$). The lowest mean rating for these questions was above average (4.667 on a Likert Scale of 1 to 7). Therefore, the behavior of the users may be linked not to the quality of this particular interface, but to the users' perceptions of the task and system functionality.

The inclusion of natural language in a multimodal interface is an effective way to reduce the number of steps that users implement to complete a search task. Users appear to be sensitive to certain strengths and weaknesses in different modalities and use them accordingly. The possible incorporation of language into systems is prevalent in the literature and in design anecdotes (e.g., software agents, in-vehicle highway systems, intelligent tutoring systems, home automation, and games (Moore, 1996; Grasso, Ebert, & Finin, 1997; Lester, Barlow, Converse, Stone, Kahler, & Bhogal, 1997; Shneiderman, 1998). This prevalence suggests a need for consideration of the variety of task characteristics that can influence human-computer interaction. One solution for dealing with different task characteristics is attempting to interlace different modalities.

ACKNOWLEDGMENTS This research was supported in part by a grant number 55-7249-A8 from the Office of Naval Research to the Naval Research Laboratory. The Authors wish to thank Erik Altmann, Deborah Boehm-Davis, Wayne Gray, Irvin Katz, Audrey Lipps, Paula Raymond, Marilyn Salzman, Christian Schunn, Lelyn Saner, Susan Trickett, Adam Tucker, & Ken Wauchope for their remarks.