

A Study of Commenting Agents as Design Support

Mikael Ericsson¹, Magnus Baurén², Jonas Löwgren¹ and Yvonne Wærn²

Dept. of Computer and Information Science
Linköping University,
581 83 Linköping, Sweden
Email: {miker,jlo}@ida.liu.se

Department of Communication Studies
Linköping University,
581 83 Linköping, Sweden
Email: {magba,yvowa}@tema.liu.se

ABSTRACT

Sixteen subjects were observed using a simulated (Wizard-of-Oz) commenting agent in a design support system. Different commenting behavior was tested, and the overall usefulness evaluated. The interaction was logged and recorded on video, and the subjects rated the agent with respect to usefulness, understandability, system competence, disturbance and perceived stress. Perceived mental workload was measured using RTLX. The results show that a commenting tool is seen as disturbing but useful, that the comments from an active tool risk being overlooked, and that comments pointing out ways of overcoming identified design problems are the easiest to understand.

Keywords

design support, commenting system, Wizard-of-Oz

INTRODUCTION

Much research and development work has been aimed at supporting the provision of design knowledge. One of the main themes in the work to date has been to create support tools to provide the user-interface designer with some of the knowledge needed, on levels ranging from domain-independent, lexical and syntactic design considerations to semantic and task-related knowledge. Another theme of interest is how the support tool interacts with the designer in providing the knowledge. Passive tools that allow the designer to browse relevant information have been investigated as well as active tools that intervene in the user-interface design work or automate parts of it.

In a current research project [3], we try to investigate the needs for design support among systems developers (focusing on the detailed interface design work), and to explore the added value of *commenting* tools. For this, we combine results from previous feasibility studies, explorative interviews, empirical studies of different variations on commenting and, finally, cooperative design and evaluation of new support tools. In this paper, we describe the results from the empirical studies, in which we used a dedicated Wizard-of-Oz system [2]: the WUDS system (Wizard-of Oz for Design Support) [3].

The general case in favor of using commenting tools as knowledge support in a design environment has been addressed elsewhere [4,5,6]. We focus on the specific questions:

- Are comments from a support tool useful in a user-interface design setting?
- What is the most appropriate form of the comments?
- What is the most appropriate behavior of the commenting tool?

In addition to these main questions, we also explore issues related to the commenting timing strategy, i.e., the mechanisms that decide when the system should deliver a comment if an active tool is preferred. Hence, we are interested in investigating the user's (designer's) perceived mental workload (MWL) when using a commenting tool.

EXPERIMENT DESIGN AND PROCEDURE

In the experiment, the participants used a User Interface Builder (UIB) to design the static user interface for a hotel search system. The participants were divided into five groups. One group (control) used only the UIB, and the other four also had a simulated support tool evaluating and commenting upon the design. A crossed design was used, where one variation consisted in active/passive *mode* and the other in declarative/imperative *mood*.

The participants were given a description of a design task. In the description, a hotel search system was outlined, and the participant was instructed to work on the visual design part of the systems development. A total of 16 interface designers participated in the experiment. All of them currently studied at the university. Fifteen were male and one was female; 11 of the participants came from Computer Science and 5 from other study programmes.

The Wizard, simulating the commenting agent, used a limited number (33) of guidelines. Two mood versions were used—a declarative and an imperative—with the intent to keep their semantic content similar.

Likert scales were used for collecting participant ratings of the complete design environment, the commenting support tool and specific comments. The rating questionnaires used included questions about usefulness, understandability, system competence, disturbance, perceived stress and ability to find referred interface

objects. The scales ranged from 1 to 8. Perceived mental workload was measured using RTLX [1], which is a simplified version of the subjective task load measurement technique NASA TLX (Task Load index).

Procedure

The participants were asked to design the user interface, and were (in the active condition) given advice as soon as the Wizard noticed some flaw in their design. In the passive condition, the Wizard provided advice when the participants demanded such. The control group did not receive any comments. During the experiment, all interaction was recorded in a file, and the participants screen was recorded on video.

After the session, the participants were asked to rate the system and to describe their impression of the total system. The participants were then shown three short video clips from their work on the design. For the non-control group participants, one or two of the clips showed situations where the participants received comments, starting one minute before the comment was given. They were asked to rate the particular situation pertaining to this clip in two different ways. First, the participant was given an RTLX-form with which perceived MWL was measured. Secondly, they answered a questionnaire with rating scales similar to those used for the total system.

RESULTS

The control group rated the system as very little disturbing (mean=1.7) and the other groups' answers indicate some disturbance (3.7). At the same time, the participants getting comments rated how much they would like to have a system like this in their work environment very high (7.1) This indicates that they found the system to be useful, despite the disturbance.

The linguistic variation in the comments showed that the comments with imperative form were easier to understand than the comments with declarative form (6.3 vs. 2.8, $p < 0.05$), and resulted in lower rated system competence (4.7 vs. 5.8, $p < 0.1$). The mode variation showed that the system was perceived as being more supportive in the active condition than in the passive one (6.6 vs. 5.0, $p < 0.05$). Furthermore, it was considered easier to find objects referred to in the comments in the active mode, than in the passive one (6.9 vs. 3.9, $p < 0.01$).

When scrutinizing the results from the video-clips it was found that the participants in the active condition sometimes did not notice the comments at all. This mostly happened when MWL was rated as high ($RTLX > 50$).

CONCLUSIONS AND DISCUSSION

First, the commenting tool was perceived as useful in the support of user-interface design. However, it was also perceived as disturbing. This apparent contradiction does

not in fact need to be problematic if we also consider the finding that the comments provided the designers with previously unknown information. A plausible interpretation is that the comments interrupt, and thus disturb, the flow of design but that the value of the comments—in terms of their contents—justify the interruption.

It may be proposed that the comment must be "situated" in the particular design moment. The importance of such a "situatedness" is supported by the result concerning the linguistic variation. A comment phrased in a specific form (imperative) may be easier to relate to the particular design moment than a comment phrased in a general form (declarative).

At the same time, it is interesting to note that providing an easy to understand comment lead to a perception of a less competent system than the providing of declarative ones. This result is similar to others, showing that systems, which are easier to understand, are considered to be less competent (cf. Wærn [7]).

ACKNOWLEDGMENTS

We wish to thank the participants of our experiment for their time and efforts. This work was financially supported by the Swedish Research Council for Social Sciences and the Humanities (HSFR).

REFERENCES

1. Byers, J. C. et al. (1989) Traditional and raw task load index (TLX) correlations: Are paired comparisons necessary? In Mital, A., editor, *Advances in Industrial Ergonomics and safety*. Taylor & Francis.
2. Dahlbäck, N., Jönsson, A., and Ahrenberg, L. (1993) Wizard-of-Oz-studies—Why and how. In *Workshop on Intelligent User Interfaces*, Orlando, FL.
3. Ericsson, M. (1996) Commenting tools as design support – a Wizard-of-Oz study. Licentiate thesis, Linköping Studies in Science and Technology. No. 576. Linköpings universitet, Sweden.
4. Fischer, G. et al. (1993) Embedding critics in design environments. In *Proceedings of INTERCHI'93*, pages 157-164. New York: ACM Press.
5. Löwgren, J. and Laurén, U. (1993) Supporting the use of guidelines and style guides in professional user-interface design. *Interacting with Computers*, 5(4):385-396.
6. Silverman, B. and Mezher, T. (1992) Expert critics in engineering design: Lessons learned and research needs. *AI Magazine*, (Spring):45-62.
7. Wærn, Y. & Ramberg, R. (1995) People's perception of human and computer advice. *Computers in Human Behavior*.