Mapping Human Oriented Information to Software Agents for Online Systems Usage

Pietro Murano

INTRODUCTION

This paper describes an experiment and the preliminary results of the experiment involving software agents. The overall aim of the experiment (one of a suite of experiments) is to determine the effectiveness and user approval of mapping human-oriented information to software agents. The preliminary results presented in this paper involve user approval issues.

This subject area, involving anthropomorphism at the user interface as part of a feedback agent(s), is currently surrounded with controversy. The Computer Science community is divided on this issue, where they are either for (e.g. (Agarwal 1999; Cole, Massaro et al. 1999; Dertouzos 1999; Guttag 1999; Zue 1999)) or against, (e.g. Shneiderman in (Shneiderman 1992; Bradshaw 1997)) anthropomorphism at the user interface. However, they do not provide concrete enough evidence for their stance. The experiment described in this paper aims to begin to resolve this issue, particularly with respect to effectiveness and user approval. The experiment (involving UNIX commands) falls within one broad area or 'strand' of software. This is software for online systems usage.

If interesting and meaningful results can be obtained, these will potentially affect future user interfaces. This will have implications for software in different domains, including software that could be defined as 'non-standard' such as the software sometimes used in command centres and control rooms. These could lead to new types of more usable interfaces. This would be particularly so with respect to training novices to use 'non-standard' user interfaces, or where a command centre or control room in an organisation is migrating to a virtual control room and staff require an update to their training.

EXPERIMENT - UNIX COMMANDS

Hypothesis

Obtaining the preliminary results of the experiment involved asking the following questions :

 Is a direct mapping (using video feedback as the direct mapping) of human-oriented information to software agents liked by users?
 Is an indirect mapping (using guiding text feedback as the indirect mapping) of human-oriented information to software agents liked by users?

Furthermore, a hypothesis linked to the first two questions above was tested :

 Users will prefer the agent text feedback to the agent video feedback.

Users

 All of the users taking part in the study were adults.
 Males and females took part.

 All the subjects had differing personal backgrounds.
 Subjects were students studying various courses, found through the University of Central Lancashire, where the number of students exceeds 10000.
 33 users have taken part in the study so far.

Experimental Design

A Within Users design was used for this experiment. For this experiment the users tried a set of designed tasks (see Appendix 1) and were able to use both types of feedback, i.e. the video and the guiding text condition. The tasks were four in total, where two tasks would be given video as the attached feedback and the remaining two would receive the guiding text. For randomisation purposes the feedback conditions were rotated, so that one type of feedback was not always linked to a particular task.

Variables

The independent variables were the types of feedback, i.e. :

 Guiding text.
 Video.

To measure user preference to the feedback the dependent variables in Appendix 1, Observation Protocol and Debriefing Issues for UNIX Commands Experiment (Nos. 5 - 8) were used. The dependent measures used were by observation and a post-experiment questionnaire (See Appendix 1). (NB : Numbers 1- 4 are being used for the effectiveness issues of the types of agent feedback, currently ongoing and not the subject of this paper).

Apparatus and Materials

The equipment used for the experiment was :

 A PC running Windows 95, 400 MHz and 128 Mb RAM.
 External speakers.
 IBM ViaVoice Executive Automatic Speech Recognition (ASR) engine (including text-to-speech), trained with a male English accented profile. A full training comprised the reading of 496 English phrases, predefined in ViaVoice. An English female profile was also obtained for use with female subjects (in practice the researcher obtained several profiles for having a better chance with voice matching issues).
 Head mounted microphone supplied with the IBM ViaVoice kit.

The prototype was engineered with C++ Builder 3 and the ViaVoice Software Development Kit (SDK), and it was made to 'look like' a UNIX environment.

Running the prototype would present to the user a screen with a single X-Window containing a 'classic' UNIX type shell prompt. At all times an 'end program' button would be available at the top left part of the screen.

Allowing the new 'learning' algorithm (Murano 2001) to take effect, would mean the software agent would infer that the current user was a beginner and therefore present a smaller window to the left of the main X-Window asking the user what they wanted to do. The smaller

window always contained a 'close' button in case the user wished to not have any feedback. The user would then input via the microphone the appropriate command for obtaining the required information for accomplishing the given tasks (see Appendix 1, Tasks 1- 4).

Upon a successful ASR, the prototype would display in the smaller window, relevant agent feedback (this was randomised so that one command was not tied to a particular video or text feedback).

If the guiding text (non-anthropomorphic) was presented, the appropriate command would be displayed in the smaller window.

If the video (anthropomorphic) feedback was presented, a video clip of a person verbally uttering the appropriate command would be played in the smaller window.

Procedure

This procedure was carried out in the same way for all subjects using the same equipment and questionnaires/observation protocols. Each subject was treated in the same manner. This was all in an effort to control any confounding variables.

The experiment took about 30 minutes to complete per volunteer. Subjects were given £3 in cash, which they signed for, for their participation.

Each subject was booked an appointment during the day, at a time suitable to both the researcher and subject. Upon meeting the subject a few pleasantries were exchanged to help the subject relax and then they were given a brief overview of the purpose of the research. Furthermore, they would be asked questions 9 - 18 (Appendix 1 - Observation Protocol and Debriefing Issues for UNIX Commands Experiment) before beginning the experiment.

A verbal introduction to the system itself was given, to help the subject overcome any false notions about the system. Furthermore the subjects were assured that the aims of the experiment were to test the software and research questions concerning the software, and not to test the person. Each participant was also given a brief explanation as to what UNIX is, and the concepts surrounding UNIX commands. A description of what they would see was also given, for the reason that a UNIX interface does not look like a regular PC interface.

When the subject felt they were ready to start the experiment, they were given the head mounted microphone to put on. If required and with the volunteer's permission, the researcher physically adjusted the position of the microphone to the correct distance from the mouth, as outlined in (IBM 1998). Upon running the program the subject would input (via the microphone) the appropriate command for Task 1 (see Appendix 1). Upon a successful ASR, the appropriate command would be issued to the user, where the user would read the text command or view the video (depending on what feedback was issued). The user would then try to input, via the keyboard, the command the system advised the subject to use. Assuming the subject entered the command in the correct manner, they could then proceed to try the remaining tasks (Appendix 1). If they failed/made mistakes in entering the command, they had the option of calling the appropriate help to view the advised command again. This would go on until they achieved the correct command.

Having completed all of the experiment, the subject would then proceed to answer numbers 5 - 8 (Appendix 1 - Observation Protocol and Debriefing Issues for UNIX Commands Experiment).

Results

The data analysed so far concerns the user approval of the video feedback and text feedback by the agent. User approval was rated in terms of how helpful users felt the various forms of feedback to be, where users could allocate a score of 1 (Terrible) through to 9 (Excellent) - (numbers 5 and 6 in Appendix 1). The mean scores and accompanying standard deviations can be seen in table 1:

Table 1 - Means for Video and Text

	Mean	Standard Deviation
Video	7.30	1.49
Text	6.39	1.89

Also users were asked in addition to make a choice between the two agent feedbacks with respect to their preferences (number 7 in Appendix 1). The answers to this question were that 63.64% of subjects would have chosen video if a choice had to be made and 36.36% of subjects would have chosen the text feedback if a choice had to be made.

Experiment Conclusions

From these figures it can be clearly seen that users rated the video feedback with higher scores compared to the text feedback, on average. Hence users on the whole preferred the video agent feedback compared to the text agent feedback. Furthermore it can be seen, from the standard deviations, that the scores for the video were more consistent than the scores for the text. It can be concluded that there was more disparity in the way users scored the text feedback.

The scoring is further confirmed by the fact that 63.64% of the subjects said that they would choose video over text if a choice between the two had to be made. This is consistent with the scoring with respect to helpfulness of each type of feedback. From these statistics the hypothesis raised at the beginning of the experiment can be rejected.

Although the experiment is still running, it is hoped to have results for more subjects (i.e. more than 33) in the near future. Of utmost importance will be the results concerning the effectiveness of the feedback tested.

APPLICATION OF FINDINGS TO OTHER DOMAINS

The results presented are very interesting, as the initial expectation was for users to prefer a more established form of feedback, the text in this case. However as has been seen most users preferred the video feedback. This has implications for the design of user interfaces, particularly concerning issues of agent feedback to a user.

In (Dicken 1999), the developing of effective alarm systems for soft control desks is discussed. Significant efforts have been made to help the user, by using various strategies, e.g. using flashing colours and different colours linked to an alarm condition as feedback to an operator. Furthermore, Dicken describes information being made available via various 'lists' and 'items' of text. The use of anthropomorphic feedback (e.g. video as used in the experiment), could be of help in a training environment. This would be with the aim of helping a new operator become proficient in dealing with alarms, by having anthropomorphic feedback acting as an assistant, giving guidance on where to find information and on how to act for a particular situation. The experiment described above shows that potentially operators could have a higher level of approval for such a type of feedback, as has been the case for UNIX command novices. As the operator moves on to being 'a trained' operator, reliance on this feedback would be less. However, having the availability of this type of feedback in every day situations could also be of use in certain circumstances, e.g. alarm situations not seen very often or the alarm floods mentioned by Dicken. Operators may like this as they could 'feel' that another person is 'with them', despite probably not being alone in a control room. This is confirmed by informal comments received from participants in the experiment described above.

Hence it would be interesting to set up experiments in a control room environment with the aim of fully determining the usefulness of such a scheme. Clearly soft desks are not perfect, as in (Dicken 1999) it is acknowledged that the alarm system was criticised due to excessive amounts of alarm overloads. It is generally accepted by Human Computer Interaction (HCI) experts that users should be satisfied when using a system, even if it can be said that control room operators are usually more motivated (Shneiderman 1992) than say a home PC user. Thus giving operators more satisfaction could lead directly to better efficiency in the control room.

The potential benefits of using the proposed anthropomorphic feedback in training can be implied from (Hvelplund 1999), where it is clearly indicated that much time needs to be dedicated to teaching operators all the facilities in a sophisticated control system. The statements made by Hvelplund indicate that training periods for operators should ideally be longer. Hence the practice of exploring the lesser used aspects of the system by the operators, during quiet periods of working. This paper is suggesting that perhaps having anthropomorphic feedback or guidance as shown by the experiment in using UNIX commands, could be liked by operators who are perhaps exploring the system. Such a feedback could provide a more structured means of exploration. It could be used as a sophisticated self-use and self-contained tutorial. This was of benefit to the users attempting various UNIX tasks. However as mentioned above the requirement for such a level of feedback would diminish as the operators became more proficient and experienced.

CONCLUDING REMARKS

The results presented in this paper, show that users' opinions should always be valued and taken into account when developing software. If one does not take this approach, the result could be functionally sound software that is either not used by users or not used to its full potential.

As the experiment is not yet completed with respect to the effectiveness issues, it will be very interesting to see which type of agent feedback is shown to be more effective. It will be

interesting to see if the most effective type of agent feedback will match the user preferred agent feedback.

REFERENCES

Agarwal, A. (1999). Raw Computation. Scientific American. 281: 44-47.

Bradshaw, J. M. (1997). Software Agents, AAAI Press, MIT Press.

Cole, R., D. W. Massaro, et al. (1999). New Tools for Interactive Speech and Language Training : Using Animated Conversational Agents in the Classrooms of Profoundly Deaf Children. Method and Tool Innovations for Speech Science Education, London, Dept. of Phonetics and Linguistics, University College London.

Dertouzos, M. L. (1999). The Future of Computing. Scientific American. 281: 36-39.

Dicken, C. R. (1999). Soft Control Desks - Producing an Effective Alarm System. People in Control An International Conference on Human Interfaces in Control Rooms, Cockpits and Command Centres, University of Bath, UK.

Guttag, J. V. (1999). Communications Chameleons. Scientific American. 281: 42,43.

Hvelplund, R. (1999). Control Room For High-Efficiency Power Plant. People in Control An International Conference on Human Interfaces in Control Rooms, Cockpits and Command Centres, University of Bath, UK.

IBM (1998). IBM ViaVoice 98 User Guide, IBM.

Murano, P. (2001). A New Software Agent 'Learning' Algorithm. People in Control An International Conference on Human Interfaces in Control Rooms, Cockpits and Command Centres, UMIST, UK, IEE. (To Appear)

Shneiderman, B. (1992). Designing the User Interface - Strategies for Effective Human Computer Interaction, Addison-Wesley.

Zue, V. (1999). Talking With Your Computer. Scientific American. 281: 40,41.

APPENDIX 1 - UNIX COMMANDS EXPERIMENT

 Thank you for taking part in this experiment, which concerns testing two different types of interface feedbacks/agents for their effectiveness and user acceptability.
 The data collected will be kept confidential.
 You will be given (below) 4 tasks concerning UNIX X-Windows commands on an emulation of the UNIX X-Windows interface. (NB : UNIX X-Windows commands are case sensitive)

Task 1:
Use the appropriate command to display all files in the current directory in long format. If

you do not know the command you may ask the system via the microphone. The verbal command to use via the microphone is 'showing files in the current directory in long format'.

Task 2:
Use the appropriate command to display the full path name of the current directory. If you do not know the command you may ask the system via the microphone. The verbal command to use via the microphone is 'displaying full path name of the current directory'.

Task 3:
Use the appropriate command to compress the file called 'letter' and display resulting percentage of reduction for files. If you do not know the command you may ask the system via the microphone. The verbal command to use via the microphone is 'compressing files'.

Task 4:
Use the appropriate command to delete the file called 'letter.Z'. If you do not know the command you may ask the system via the microphone. The verbal command to use via the microphone is 'deleting files'.

Observation Protocol and Debriefing Issues For UNIX Commands Experiment

(1) Does the user display all files in the current directory (text - video)?
How many attempts?
Did user falter? YES/NO If yes how many times?

(2) Does the user display the full path name of the current directory (text - video)?
How many attempts?
Did user falter? YES/NO If yes how many times?

(3) Does the user compress the file (text - video)?
How many attempts?
Did user falter? YES/NO If yes how many times?

(4) Does the user delete the file (text - video)?
How many attempts?
Did user falter? YES/NO If yes how many times?

(5) Helpfulness of Video Feedback/Agent

Terrible Excellent
1 2 3 4 5 6 7 8 9

(6) Helpfulness of Text Feedback/Agent

Terrible Excellent
1 2 3 4 5 6 7 8 9

(7) Which of the two feedbacks/agents would the user prefer if a choice had to be made between them?

(8) Any other comments:
(9) Age group : 17 - 25, 26 - 35, 36 - 45, 46 +
(10) Gender : Male/Female
(11) Status :
(12) Occupation and/or course studied :
(13) Birthplace :
(14) Hobbies :
(15) Does the user have any reading, sight or hearing problems/conditions?
(16) Does the user have any knowledge of UNIX commands?
(17) Computing Experience.
(17.1) Do you use a computer for hobby/work/study purposes? YES/NO
(17.2) If answer to (17.1) is yes - Are you Beginner/Intermediate/Experienced?
(17.3) If answer to (17.1) is yes - How long have you been using a computer?
(18) Would you like to see software developers improving current software? YES/NO