Anthropomorphic Vs non-anthropomorphic software interface feedback for online factual delivery

Murano, P

http://dx.doi.org/10.1109/IV.2003.1217970

Anthropomorphic Vs Non-Anthropomorphic Software Interface Feedback for Online Factual Delivery

Pietro Murano
University of Salford, Computer Science, Intelligent Visualisation Laboratory, Gt. Manchester, M5 4WT, UK
p.murano@salford.ac.uk

Abstract

This paper follows on from a group of internationally published papers in the area of investigating anthropomorphic user interface feedback. In this paper an experiment is described which aims to examine the effectiveness and user approval issues of anthropomorphic versus non-anthropomorphic user interface feedback. This experiment is highly relevant to the research community as Computer Scientists are not in agreement concerning the effectiveness and user approval of anthropomorphic user interfaces. The experiment described in this paper is in the context of Online Factual Delivery. Specifically the area of direction finding was used to test anthropomorphic feedback against non-anthropomorphic feedback. Statistically significant results are presented for the benefit of the research community, where it is clear that in this context, the non-anthropomorphic user interface feedback was more effective and users tended to prefer this feedback.

1. Introduction

Multimedia has permeated much of society to the benefit of many individuals and businesses. Most multimedia systems have user interfaces and it is desirable that these user interfaces be as usable as possible. This research contributes to the goal of improving user interfaces. The work described in this paper has involved the comparison of anthropomorphic and non-anthropomorphic user interface feedback by means of an experiment. The issues of effectiveness and user approval of the feedbacks were under investigation.

This research is very relevant as there is disagreement within the Computer Science research community concerning the effectiveness and user approval of interfaces having anthropomorphic feedback. Some of those in favour of anthropomorphism at the user interface are Agarwal [1], Guttag [3], Koda and Maes [5], Maes [6] and Zue [13]. The most famous person against anthropomorphism is Shneiderman in [2] and [11]. However each side of the research argument does not have strong evidence to show their correctness.

Therefore this research is helping to resolve this 'open' issue and this experiment is part of wider research being undertaken ([7], [8], [9], [10]) in this area. In [7], [8] and [10] it has been suggested that in the contexts of software for in-depth understanding and on-line systems usage, an anthropomorphic feedback is more effective and preferred by users. These conclusions are based on statistically significant results. Hence the experiment described in the remaining sections was set in the context of Online Factual Delivery. Direction finding was the specific area used to test the user interface feedbacks.

2. Hypotheses

Answers to the following questions were the aim of this experiment.

- Is a direct mapping (using video as the direct mapping) of human-oriented information to software interface feedback effective and do users like it? (effectiveness for this experiment was defined as the user finding the location, the user taking as few wrong turnings as possible and the user hesitating or faltering as few times as possible)

Proceedings of the Seventh International Conference on Information Visualization (IV'03)
1093-9547/03 $17.00 © 2003 IEEE

- Is an indirect mapping (using a two-dimensional map with guiding text as the indirect mapping) of human-oriented information to software interface feedback effective and do users like it?
- What guideline(s) can user interface designers receive?

Furthermore, two null hypotheses (H_0) linked to the first two questions above were tested:

- There will be no difference between the 2 conditions (video and two-dimensional map with guiding text) - for effectiveness.
- There will be no difference between the 2 conditions (video and two-dimensional map with guiding text) - for user preference.

Two alternative hypotheses (H_1) were also considered:

- The diagram (map) feedback will be more effective than the video feedback.
- Users will prefer the diagram (map) feedback to the video feedback.

Furthermore, this experiment, where direction finding was the 'foundation', rested within the broader area or domain of software for online factual delivery. The reason for this was to try if possible, to make a generalisation based on the results of this experiment, to cover the broader area of similar software.

2.1. Users

- All the users taking part in the study were of varied age groups.
- Males and females took part.
- All the subjects had differing backgrounds. This was taken to be individuals studying different courses, having varied birthplaces and having varied hobbies/work experience. The information was elicited by means of a pre-designed questionnaire.
- Subjects were found through the university population.
- 53 users took part in the study.

2.2. Experimental design

A within users design was used for this experiment. All the users attempted to reach two destinations, which were approximately equivalent to each other. Each destination was a real location within walking distance of the laboratory, but not within sight of the laboratory. Furthermore the actual distances of the locations were balanced with each other, as were the complexities of the various turnings involved for reaching the particular destinations. Each subject was given both types of feedback (video and two dimensional map with guiding text) during the experiment, i.e. video would be given for one destination and the map would be given for the remaining destination

Further, the feedback was rotated, i.e. a type of feedback was not permanently linked to one destination. This was to avoid the possibility of one of the destinations perhaps being more suited to one type of feedback. The result was a randomisation of feedback for the destinations concerned.

2.3. Variables

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

- Two dimensional map with guiding text.
- Video.

To measure feedback effectiveness the dependent variables consisted of recording on an observation protocol the number of wrong turnings users took, the number of user hesitations, the number of times the map was read and the number of times the video was replayed.

To measure user preference of the feedback the dependent variables consisted of users giving a score (as part of the questionnaire) to show their opinions towards each type of feedback. Furthermore each user was hypothetically asked which of the two types of feedback they would choose if they had to make a choice.

The dependent measures used were by observation, i.e. bservations were recorded on an observation protocol. Further a post-experiment questionnaire containing various subjective opinion questions was used for each user.

2.4. 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 ASR engine (including text-to-speech), trained with a male English accented profile. A full training was the reading of 496 English phrases, predefined in ViaVoice. An English female profile was also obtained for use with female subjects (in practice the author obtained several profiles for having a better chance with voice matching issues).

- Head mounted microphone supplied with the ViaVoice kit.

The prototype was engineered with C++ Builder 3 and the ViaVoice Software Development Kit (SDK).

Running the prototype presented a screen with a text message asking the user for the location they wished to find. At the same time the same request was made audibly to the user via the text-to-speech engine. At this point the user asked for directions to the assigned location, via the microphone, using verbal commands given to the user as part of their task sheet. Upon a successful ASR, the prototype would display in a new screen relevant interface feedback (this was randomised so that one location was not tied to one type of feedback).

If it was the map feedback (non-anthropomorphic) that was displayed in response to the user's request, a text-to-speech message asking the user to 'read' the map would be issued. At this point the screen would also contain a button for ending the program.

If it was the video feedback (anthropomorphic) that was played to the user, a video clip of a person giving relevant directions to the user would be played in the centre of the screen. At this point the screen would also contain a button for ending the program, and a button for replaying the video.

2.5. Procedure

The procedure described below 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 as a reward, which they signed for, for their participation.

Each subject was booked an appointment during the day, at a time suitable to both the author 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. Further they would be asked questions concerning their background (see section 2.1 above) before beginning the experiment. At this point the subject was issued with a sheet of paper containing the introductory bullet points concerning the experiment and the details of Tasks 1 and 2 (the details concerned asking the user to find their way to the two pre-defined locations). Each subject was then asked if they knew where the locations detailed on their sheet were (2 individuals knew these locations and were therefore not used - otherwise 55 subjects would have taken part).

A verbal introduction to the system itself was given, to help the subject overcome any false notions about the system. This included aspects of how to use the ASR module, e.g. to speak clearly and 'normally'. Each person was given an indication of the type of feedback that was being tested. Subjects were also briefed on the system behaviour, e.g. the sequencing of the screens involved in the interaction. Furthermore the subjects were assured that the aims of the experiment were to test the software and hypotheses concerning the software, and not to test the person.

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 author physically adjusted the position of the microphone to the correct distance from the mouth, as outlined in [4]. Upon running the program the subject would input (via the microphone) the appropriate command for Task 1. Upon a successful ASR, the appropriate 'directions' would be issued to the user, where the user would have the option of reading the map as many times as necessary or playing the video as many times as necessary (depending on what feedback was issued). When the user was ready, and asked to 'mentally' take note of any hesitations/falterings they had about where to go while on the street, they would physically attempt to make their way to the location, on foot. The author would then follow the subject at a discreet distance. This would give the opportunity to observe any wrong turning the subject may take. Final results, e.g. subject found location etc. would be recorded on the appropriate observation protocol. Having arrived at the location, the author would immediately ask if they had felt any hesitation/faltering while on route. Responses were immediately recorded. The subject and author would then walk back to the laboratory, where when ready, the subject would attempt Task 2, with events proceeding as described above.

The way errors were categorised were that if a user found the location this was recorded accordingly. However, at times a user found the location but still took, while on the way, one or more wrong turnings (a wrong turning was any deviation from the prescribed route). This would be recorded as the user finding the location, but with one or more wrong turnings having been taken. At times the participant took one or more wrong turnings that were serious enough to ensure they did not find the location. This was also recorded to reflect this occurrence. If it was seen that the subject was deviating significantly from the route given in the feedback (e.g. going completely in the wrong direction for a few minutes and showing no sign of corrective

action), the subject would be stopped and asked where they were planning to go from that point. If the subject indicated that they intended to continue going on a completely erroneous route it was deemed at this point that they had not found the location. If this was the case they were then told where they should have gone. However, if it seemed that they were planning some corrective action in the correct direction, they were allowed to proceed.

The number of hesitations were taken from the information given by the participant. Some participants who were at the stage of hesitating strongly, stopped on the street and would turn around and say they were unsure where to go next. At this point the author ensured no extra help was given, but encouraged the participant to try and use the information seen online a few minutes earlier.

Each completed task in relation to a particular type of feedback was converted to a score for statistical purposes, where the successful/unsuccessful completion of the task, any wrong turnings and any user hesitations were used in the scoring formula. The formula used was that each user (unknown to them) was started on 10 points for each task. For every incorrect turning taken, 1 point was deducted. For every hesitation reported by the user 0.5 points were deducted. If the subject found the location the score was left as above. However if the subject did not find the location a further 1.5 points were deducted to give a final score.

Having completed all of the experiment, the subject would then proceed to answer a set of questions on a questionnaire which elicited their opinions of the feedback and their opinions of the general usability of the user interface.

2.6. Results

The data collected for this direction finding experiment was firstly concerned with the effectiveness of the interface feedback, and secondly with the user approval of the feedbacks. The scores of all subjects were plotted on a Normal Probability Plot which showed the scores to be approximately normally distributed. These were used in an F test for the determination of feedback effectiveness. For 53 subjects, the F observed was 1.85, and the F critical (5 %) was 1.67, these are shown in Table 1 below:

Table 1 – F test results

	Comparison of Video Vs. Diagram(Map)
F-Observed	1.85
F-Critical (5%)	1.67

The Overall User Preferences were determined by the scores allocated by the subjects in the elicited post-experiment questions. The mean of the scores was calculated along with the relevant standard deviation. The mean for the diagram (map) was 6.74 (standard deviation = 1.62). The mean for the video was 6.42 (standard deviation = 1.68). These are shown in Table 2:

Table 2 – Overall user preferences

	Overall User Preferences	
	Mean	Standard Deviation
Video	6.42	1.68
Diagram (Map)	6.74	1.62

Also, users were asked to make a choice regarding preference between one of the types of feedback. The results were that 41.51% favoured video and 58.49% favoured the map.

Furthermore subjects were asked questions (via the questionnaire) concerning what they thought about other general interface/system usability issues. The issues covered are in the table below where subjects categorised their opinions on a Likert scale. The scores available to the users ranged from 1 to 9 – 9 being the most positive score one could allocate (i.e. showing a preference or positive attitude towards an interface element). Hence users gave the following scores shown in Table 3:

Table 3 – System usability

	System Usability	
	Mean	Standard Deviation
Comfort of Colours Used	7.36	1.49
Text Readability	8.26	1.02
Text Understandability	8.36	1.21
Buttons Quality of Labelling	8.19	0.98
Buttons Layout Clarity	8.26	1.00
Buttons Consistency	8.30	1.01
Overall Ease of Use of the System	8.11	0.97

2.7. Conclusions

From the results of the F test, significance in favour of the diagrammatical (map) (non-anthropomorphic) feedback can be confidently concluded. Clearly in this context, the non-anthropomorphic feedback is more effective, as overall success in the direction finding tasks was higher when the map was used.

Users' ratings concerning their thoughts on the helpfulness issues of each type of feedback do not show such clear evidence as one can see for the effectiveness

issue. Both the video and map were rated in the 6 range, the map receiving a high 6 (6.74) and the video a middle 6 (6.42). The standard deviation for the map shows a little more consistency in the scoring (but not much more than the video). These are not extremely high scores, but are tending towards the higher end of positive categorising. However when asked to make a choice concerning what feedback the user preferred, the percentages show that almost 60% would choose the map over the video feedback. One can imply from this that overall individuals still liked the video even though it was not as helpful to them, but that most would take the map if given a choice. This can be seen by the fact that various subjects while discussing their thoughts informally after the experiment, said that they liked very much the idea of 'someone' giving them directions. Certain subjects commented that they felt secure with the video. This however does not explain why the map did not receive overall higher approval scores. One explanation concerns what was observed during the experiment. Certain subjects upon seeing the map gave the impression that the task was perhaps 'very easy' in their minds (this was not the case), and therefore perhaps did not 'study' the map well enough. However most subjects when asked how many times they read the map before leaving the laboratory, most admitted to reading the map several times. This again is borne out by the author's observations. Subjects appeared to make an effort at reading the map properly and using the information to find the location. An alternative reason could be one that correlates with certain subjects' feelings towards the video. If they felt secure with the video, perhaps the map (as all maps are) was more of a 'cold' impersonal type of feedback, not instilling a feeling of security as the video gave, thus lowering the actual scores given by subjects.

The scores allocated by subjects concerning the general interface/usability of the prototype are encouraging. The scores are all of a high value with low standard deviations, showing a good degree of consistency in the scoring of the various interface elements. This gives confidence that the interface or system did not adversely affect the results discussed above. It gives confidence that the prototype was a good user friendly environment for the testing of the feedbacks and the hypotheses.

Therefore from the results the first null hypothesis (H_0) (There will be no difference between the 2 conditions (video and two-dimensional map with guiding text) - for effectiveness) can be confidently rejected, as there is statistically significant evidence to suggest that there was a difference in the effectiveness stakes. Therefore the second alternative hypothesis (H_1) (The diagram (map) feedback will be more effective than the Video feedback) can be confidently accepted based on the clear statistical evidence.

However the evidence for the user approval issues of the respective types of feedback is not as clear. Therefore, the null hypothesis (H_0) (There will be no difference between the 2 conditions (video and two-dimensional map with guiding text) - for user preference) concerning preference cannot be rejected. One though has to reject the second hypothesis (H_1) (Users will prefer the diagram (map) feedback). This stance has to be taken as the feedback approval scores are too close to each other to make a reliable conclusion concerning user preference.

This result for this particular software domain and particular context within this domain has been missing in the current world knowledge and will therefore add to and hopefully modify the current world knowledge.

The generalisation that can be made with these results is that in the software domain of online factual delivery, an indirect mapping of human-oriented information to software interface feedback, such as some relevant diagrammatical method, is highly desirable. The suggestion is that this would be more effective for users. Since the results did not show a significant difference in user preferences, it would be advisable for this software domain to also include some anthropomorphic element either as a complement of some diagrammatical method or to have some option for 'switching on' the anthropomorphic element. If adding this type of secondary feedback helps a user to feel more secure, then this would be worthwhile as the long term result would be a system that is liked by users and in turn used more by users. Clearly the effects and user approval of using a combination of the two types of feedback would ideally need to be investigated in another experiment. Software for online factual delivery already exists in various different forms, e.g. car navigation, air traffic control, train command centres, satellite tracking and utility (e.g. electricity boards) systems. These all give specialised information on the state of one or more 'situations' or sets of 'items'. To extend the car navigation example, it is likely that a diagrammatical method (i.e. a map that is usually dynamic) is more effective than an anthropomorphic feedback. In [12] a diagrammatical feedback is already used with the option of having a 'human' voice give directions, thus allowing one to drive at the same time. However this research presents for the first time evidence in favour of the effectiveness of this configuration. Furthermore, if one extends the example of a control room for a utility power station, clearly a diagrammatical (non-anthropomorphic) method displaying the state of certain processes should be more effective than an anthropomorphic system. These conclusions also match with the observation in

[11] by Shneiderman, where bank automated telling machines which were anthropomorphic became failures. This is because this type of artefact is within the domain of online factual delivery and hence anthropomorphism would not be suitable as Shneiderman has found and as has been found by this experiment in direction finding.

The overall recommendation, based on the results, is for software interface feedback designers to consider as their main mode of feedback in software for online factual delivery a diagrammatical non-anthropomorphic type of feedback. Where suitable it is also recommended to have some option for users to choose whether they wish to 'switch on' an anthropomorphic equivalent or have an anthropomorphic equivalent in conjunction with the proven effective diagrammatical feedback.

This research is therefore making a contribution to multimedia systems, particularly user interface feedback, where the overall goal is to make more usable systems that are pleasurable to use by humans.

3. References

[1] Agarwal, A. Raw Computation. Scientific American. 1999, 281: 44-47.

[2] Bradshaw, J. M. Software Agents, AAAI Press, MIT Press. 1997.

[3] Guttag, J. V. Communications Chameleons. Scientific American. 1999, 281: 42,43.

[4] IBM, IBM ViaVoice 98 User Guide, IBM, 1998.

[5] Koda, T. and Maes, P. Agents With Faces. The Effects of Personification of Agents. Proceedings of HCI '96, London, 1996, British HCI Group.

[6] Maes, P. Agents That Reduce Work and Information Overload. Communications of the ACM. 1994, 37(7): 31-40, 146.

[7] Murano, P. Anthropomorphic Vs Non-Anthropomorphic Software Interface Feedback for Online Systems Usage, 7th European Research Consortium for Informatics and Mathematics (ERCIM) Workshop – 'User Interfaces for All' – Special Theme: 'Universal Access' Paris (Chantilly), France, 24,25 Oct. 2002. Published in Lecture Notes in Computer Science - Springer

[8] Murano, P. Effectiveness of Mapping Human-Oriented Information to Feedback From a Software Interface, Proceedings of the 24th International Conference on Information Technology Interfaces, Cavtat, Croatia, 24-27 June 2002

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

[10] Murano, P. Mapping Human-Oriented Information to Software Agents For Online Systems Usage. People in Control An International Conference on Human Interfaces In Control Rooms, Cockpits and Command Centres, UMIST, UK, 2001, IEE.

[11] Shneiderman, B. Designing the User Interface Strategies for Effective Human Computer Interaction, Addison-Wesley, 1992.

[12] VDO Dayton. Car Multimedia Systems, VDO Car Communication UK Ltd., Holford Drive, Birmingham, B6 7UG, 2000

[13] Zue, V. Talking With Your Computer Scientific American. 1999, 281: 40,41