EVALUATION OF ANTHROPOMORPHIC USER INTERFACE FEEDBACK IN AN EMAIL CLIENT CONTEXT AND AFFORDANCES

Pietro Murano[1], Amir Malik[1] and Patrik O'Brian Holt[2]
[1]*School of Computing Science and Engineering, University of Salford, Gt. Manchester, M5 4WT, U.K.*

[2]*Interactive Systems Research Group, School of Computing, The Robert Gordon University St. Andrew Street, Aberdeen, AB25 1HG, Scotland*

Keywords: Anthropomorphism, User interface feedback, Evaluation, Affordances.

Abstract: This paper describes an experiment and its results concerning research that has been going on for a number of years in the area of anthropomorphic user interface feedback. The main aims of the research have been to examine the effectiveness and user satisfaction of anthropomorphic feedback. The results are of use to all user interface designers. Currently the work in the area of anthropomorphic feedback does not have any global conclusions concerning its effectiveness and user satisfaction capabilities. This research is investigating finding a way for reaching some global conclusions concerning this type of feedback. This experiment, concerned the context of downloading, installing and configuring an email client which is part of the domain of software for systems usage. Anthropomorphic feedback was compared against an equivalent non-anthropomorphic feedback. The results indicated the anthropomorphic feedback to be more effective and preferred by users. It was also the aim to examine the types of feedback in relation to Affordances. The results obtained can be explained in terms of the Theory of Affordances.

1 INTRODUCTION

The user interface is one of the most important aspects of any software system used by humans. Badly designed user interfaces can lead to decreased productivity, less profits for a company (due to reduced productivity and more errors) and frustration for the end users.

The aim of this research is to improve user interface feedback and discover which methods may be best. The authors are particularly investigating the effectiveness and user satisfaction of anthropomorphic feedback. To achieve this direct comparisons are being made with non-anthropomorphic feedback in an experimental setting. Furthermore, the authors of this paper are also trying to explain the results of conducted experiments in terms of appropriate theories. One such theory that is being investigated in conjunction with the experimental results is the Theory of Affordances.

Anthropomorphism at the user interface usually involves some part of the user interface, taking on some human quality (De Angeli, Johnson, and Coventry, 2001). Some examples include a synthetic character acting as an assistant or a video clip of a human (Bengtsson, Burgoon, Cederberg, Bonito and Lundeberg, 1999).

The area of anthropomorphic feedback has been investigated for several years by various different researchers. One aspect that is clear is that the results available do not reveal an overall consistent pattern. In some cases anthropomorphic feedback is shown to be more effective and preferred by users and in some cases the converse has been shown. This is also shown in the work of Murano and his collaborators (see Murano, 2002a, 2002b, 2003, 2005, Murano, Gee and Holt, 2007 and Murano, Ede and Holt, 2008).

An example can be seen in a study by Moreno Mayer and Lester (2000). The main thrust of their experiment involved comparing anthropomorphic and non-anthropomorphic information presented in the context of tutoring about plant designs. Their results suggested that the anthropomorphic information was better for 'transfer' issues (i.e. using the knowledge to solve new similar problems)

Murano P., Malik A. and O'Brian Holt P. (2009).
EVALUATION OF ANTHROPOMORPHIC USER INTERFACE FEEDBACK IN AN EMAIL CLIENT CONTEXT AND AFFORDANCES.
In *Proceedings of the 11th International Conference on Enterprise Information Systems - Human-Computer Interaction*, pages 15-20
DOI: 10.5220/0001856000150020

and better in terms of users' having a more positive attitude and inclination towards learning more about plant designs.

However a contrasting example can be seen in a study by Moundridou and Virvou (2002). They also tested anthropomorphic information and equivalent non-anthropomorphic information in the context of algebra tutoring. They found no significant differences to do with issues of effectiveness of the feedback types. However they did find significant differences in terms of user attitudes, where the anthropomorphic feedback fostered better user attitudes.

Also in the realm of the author's work (Murano, 2002b, 2005), a study conducted in the context of English as a foreign language pronunciation, anthropomorphic and non-anthropomorphic feedback types were compared. This experiment indicated with significant results that the anthropomorphic feedback was more effective and preferred by users.

However in another experiment conducted by the author (Murano et al, 2008) in the context of PC building, comparing anthropomorphic and non-anthropomorphic feedback, the results showed no difference in terms of effectiveness. However there was a marginal result showing the anthropomorphic feedback to be preferred by users.

This brief review of some of the literature indicates that the study of anthropomorphism as a means of user interface feedback is incomplete. While some of the differences in results could be attributed to experimental design issues, some of the differences could be attributed to issues of affordances at the user interface.

The original Theory of Affordances (Gibson, 1979) has been extended by Hartson (2003) to cover user interface aspects. Hartson identifies cognitive, physical, functional and sensory affordances. He argues that when a user is doing some computer related task, they are using cognitive, physical and sensory actions. Cognitive affordances involve 'a design feature that helps, supports, facilitates, or enables thinking and/or knowing about something' (Hartson, 2003). One example of this aspect concerns giving feedback to a user that is clear and precise. If one labels a button, the label should convey to the user what will happen if the button is clicked. Physical affordances are 'a design feature that helps, aids, supports, facilitates, or enables physically doing something' (Hartson, 2003). According to Hartson a button that can be clicked by a user is a physical object acted on by a human and its size should be large enough to elicit easy clicking. This would therefore be a physical affordance characteristic. Functional affordances concern having some purpose in relation to a physical affordance. One example is that clicking on a button should have some purpose with a goal in mind. The converse is that indiscriminately clicking somewhere on the screen is not purposeful and has no goal in mind. Lastly, sensory affordances concern 'a design feature that helps, aids, supports, facilitates or enables the user in sensing (e.g. seeing, feeling, hearing) something' (Hartson, 2003). Sensory affordances are linked to the earlier cognitive and physical affordances as they complement one another. This means that the users need to be able to 'sense' the cognitive and physical affordances so that these affordances can help the user.

Therefore the remaining sections in this paper will discuss the results of an unpublished experiment and links will be made to the affordances as identified by Hartson (2003).

2 EMAIL CLIENT EXPERIMENT

2.1 Aims

The aim of this experiment was to gather data regarding effectiveness and user satisfaction in the downloading, installing and configuring of an email client context which is part of the domain of software for systems usage. Specifically the aim was to find out if anthropomorphic user interface feedback fostered a better interaction experience with fewer errors and therefore a better task completion rate.

Two identical prototypes were developed with only the feedback methods varying. The first had textual feedback available and the second had anthropomorphic feedback in the form of the MS Agent Merlin character with voice and speech bubbles. The system was built to identically emulate the basic task of downloading, installing and applying a basic configuration to the Kerio (2006) email client.

Further, the authors were also interested to find out if the user interfaces designed were appropriately facilitating the affordances.

2.2 Users

- 20 participants were used in the experiment. These were selected by means of contacts at local colleges and personal acquaintances.
- All the participants taking part in the study were in the 18-40 age groups.

- Although gender was not the main issue of this research, the participants were all adult males and females.
- All the participants were novices in terms of overall computing experience.
- All participants had downloaded software from the Internet in the past, but had not downloaded email clients. However most participants had experienced difficulty with the downloading and installation process.

2.3 Design

A between users design was used. The 20 participants were randomly assigned to one of the two conditions being tested. Random allocation to one of the two experimental groups was achieved by alternately assigning an individual to one of the groups.

2.4 Variables

The independent variable was the type of feedback, (Textual instructions - Non-anthropomorphic and MS Agent synthetic character – Anthropomorphic).

The dependent variables were the participants' performance in carrying out the tasks and their subjective opinions.

The dependent measures were that the performance was measured by counting the number of input errors made, the number of incorrect selections, the number of requests for help, the number of manifested participant hesitations – minor or major in nature (A minor hesitation was of the kind that involved a participant taking longer than ten seconds to choose an option after having obtained some feedback. A major hesitation was when a participant was given some feedback and then proceeded to not take any action at all) and whether the participant completed a task. These factors were then used in a scoring formula in order to achieve a single score per participant (see note below). The formula was devised because it was felt that the factors of errors, hesitations and actually completing the tasks, were related to overall success. These factors were recorded by means of an observation protocol.

The subjective opinions were measured by means of a post-experiment questionnaire, which included questions regarding the user interface and the users' experience etc.

(NOTE – The formula used was as follows:

- Each participant (unknown to them) was started on ten points for each task.
- For every incorrect selection made, one point was deducted. An example of this 'error' was the participant not selecting the 'next' option to begin the installation process.
- For every input error made, one point was deducted. An example of this 'error' was the participant not entering a password for the email account.
- For every obvious minor (>ten seconds) hesitation, e.g. taking longer than ten seconds to make choose an option after having received feedback, half a point was deducted.
- For every obvious major hesitation, e.g. the participant being given some feedback and then not acting on the feedback at all, one point was deducted.
- For every help request made, one point was deducted.
- If the participant completed a task correctly the score was left as described above.
- If the participant did not complete a task a further one and a half points were deducted to give a final score.)

In the actual experiment no major hesitations were observed and all participants completed the task, therefore two elements of the formula described above were not used in practice.

2.5 Apparatus and Materials

The equipment used for the experiment was: A laptop running Windows XP and 448 Mb RAM, the laptop's own speakers and TFT display were used, Microsoft Agent 2.0 ActiveX component and Lernout and Hauspie TruVoice Text-To-Speech engine. Lastly the prototype was engineered with Visual Basic 6.

2.6 Procedure and Tasks

The first step was to recruit a suitable number of participants particularly meeting the requirement of being novices to computers, not having downloaded and installed email clients in the past and if any other software had been downloaded in the past that some degree of difficulty had been experienced on their part (see Users section above). The recruitment process was achieved by the participants completing a pre-experiment questionnaire. The questionnaire had various questions which were mainly designed to elicit prospective participants' experience with computers, the Internet and downloading and installing software (including email clients).

Each participant was briefed with the same information, i.e.: 1. A brief narrative regarding the

purpose of the experiment. 2. The purpose of the experiment was not to test the participant. 3. Participants should do their best to concentrate whilst carrying out the tasks. 4. Each participant was given an explanation regarding the type of feedback they would be using. 5. Most of their interaction would be mouse based. 6. Feedback would be given by the system if the participant made any errors. 7. Participants were asked if the instructions were clear and if not, further (un-biasing) explanations were given. The further instructions did not use leading language which would have given clues on how to complete the task. Also no specific examples were used to further attempt better control on the matter. 8. A post-experiment questionnaire would need to be completed at the end of the experiment. 9. It was verbally made clear that if they wanted to leave at any time, they could do so and if they did not want their data to be used at the end of the experiment, that this was their prerogative. Also data collected as part of the experiment would be kept confidential. 10. Completing the whole experiment would mean each participant would be entered into a prize draw for a £20 Selfridges voucher.

Then the procedure described below was carried out in the same way for all participants using the same environment, equipment and questionnaires/observation protocols. Each participant was treated in the same manner. This was all in an effort to control any confounding variables.

There was one basic global task with several stages. This was to download, install and prepare an email domain with the Kerio email client. The following stages were required to complete the overall global task:

1. Click the appropriate download link. 2. Choose a folder for storing the downloaded file. 3. Await the download process to complete. 4. Initiate the installation process. 5. Choose the appropriate language. 6. Await the file extraction process to complete. 7. Read the welcome message and choose proceed. 8. Read the licence information and choose proceed. 9. Select a folder for the Kerio email client. 10. Select the install type. 11. Await the installation process to complete. 12. Use the Kerio configuration wizard to create an email domain. 13. Enter an email domain. 14. Enter a user name. 15. Enter a password. 16. Complete the installation.

The session was started by the system presenting a small tutorial using the feedback mode of the relevant condition. The tutorial explained what the task was and gave instructions regarding what had to be done if help was required during the carrying out of the task.

Once the tutorial material had been received by the participants the actual task was undertaken with the appropriate feedback condition. The simulation that was built, emulated the actual stages required for the task. Therefore each stage of the interaction was accompanied by either anthropomorphic or non-anthropomorphic instructions (depending on experimental condition) regarding what had to be done to complete the stage and go on to the next stage. The instructions were basically of the kind which instructed the user on what had to be done, e.g. choosing a 'typical' installation and clicking next etc. During the carrying out of the task, each participant was observed and data was recoded on the appropriate observation protocol.

Lastly the participants were asked to complete a post-experiment questionnaire regarding their subjective opinions about the software.

2.7 Results

The data was analysed using a multifactorial analysis of variance (MANOVA) and when significance was found, the particular issues were then subjected to post-hoc testing using in all cases either t-tests or Tukey HSD tests. The following significant results were observed (NB: DF = Degrees of Freedom, SS = Sum of Squares, MSq = Mean Square):

For the variables 'total score' and 'group', there is a significant ($p < 0.05$) difference. The anthropomorphic group scored significantly higher scores than the non-anthropomorphic group, with an F-ratio of 4.87*. Table 1 shows the F Table:

Table 1: MANOVA, total score/group.

Source	DF	SS	MSq	F Ratio
Model	2	5.63	2.81	4.87
Error	17	9.81	0.58	Prob > F
C. Total	19	15.44		0.02

For the variables 'Understandable UI feedback' and 'group', there is a significant ($p < 0.05$) difference. The anthropomorphic group felt that the user interface feedback was significantly more understandable when compared to the non-anthropomorphic group, with an F-ratio of 3.83*. Table 2 shows the F Table:

Table 2: MANOVA, understandable feedback/group.

Source	DF	SS	MSq	F Ratio
Model	2	1.80	0.90	3.83
Error	17	4.00	0.24	Prob > F
C. Total	19	5.80		0.04

For the variables 'Sufficient UI Feedback' and 'group', there is a significant difference. The anthropomorphic group felt that the user interface feedback was significantly ($p < 0.01$) more sufficient compared to the non-anthropomorphic group, with an F-ratio of 10.37**. Table 3 shows the F Table:

Table 3: MANOVA, sufficient feedback/group.

Source	DF	SS	MSq	F Ratio
Model	2	2.50	1.25	10.37
Error	17	2.05	0.12	Prob > F
C. Total	19	4.55		0.001

For the variables 'Friendly UI Feedback' and 'group', there is a significant ($p < 0.01$) difference. The anthropomorphic group felt that the user interface feedback was significantly more friendly compared to the non-anthropomorphic group, with an F-ratio of 20.40***. Table 4 shows the F Table:

Table 4: MANOVA, friendly feedback/group.

Source	DF	SS	MSq	F Ratio
Model	2	7.20	3.60	20.40
Error	17	3.00	0.18	Prob > F
C. Total	19	10.20		<.0001

For the variables 'UI Feedback not Intimidating' and 'group', there is a significant ($p < 0.01$) difference. The anthropomorphic group felt that the user interface feedback was significantly less intimidating compared to the non-anthropomorphic group, with an F-ratio of 41.00***. Table 5 shows the F Table:

Table 5: MANOVA, not intimidating feedback/group.

Source	DF	SS	MSq	F Ratio
Model	2	4.10	2.05	41.00
Error	17	0.85	0.05	Prob > F
C. Total	19	4.95		<.0001

For the help' and 'group', there is a significant ($p < 0.01$) difference. The anthropomorphic group felt that the help given was significantly more friendly compared to the non-anthropomorphic group, with an F-ratio of 18.94***. Table 6 shows the F Table:

Table 6: MANOVA, friendly help/group.

Source	DF	SS	MSq	F Ratio
Model	2	11.70	5.85	18.94
Error	17	5.25	0.31	Prob > F
C. Total	19	16.95		<.0001

For the variables 'friendly error info' and 'group', there is a significant ($p < 0.01$) difference. The anthropomorphic group felt that the error information given was significantly more friendly compared to the non-anthropomorphic group, with an F-ratio of 9.13**. Table 7 shows the F Table:

Table 7: MANOVA, friendly errors/group.

Source	DF	SS	MSq	F Ratio
Model	2	6.50	3.25	9.13
Error	17	6.05	0.36	Prob > F
C. Total	19	12.55		0.0020

2.8 Discussion of Results

The clearest result shows that the anthropomorphic feedback was more effective for the global task of downloading, installing and preparing an email domain with the Kerio email client. The scores achieved in the task were significantly higher in the anthropomorphic condition.

As expected the perceptions of participants in the anthropomorphic condition tended to be more positive about the system. They clearly found the task easier to complete and therefore had more positive perceptions about the system. Specifically they felt that the feedback was more understandable, sufficient, friendly and less intimidating. Also the anthropomorphic group felt that the help was more friendly and that the error information was more friendly in nature. The suggestion is that a higher success rate in a task can elicit more positive perceptions about a system.

Overall the results suggest that in the specific context of downloading, installing and preparing an email domain, the anthropomorphic feedback was more effective and fostered more user satisfaction.

3 THE EXPERIMENT AND AFFORDANCES

This experiment had results where the anthropomorphic Merlin character condition was significantly more effective and significantly more satisfying for the participants. The anthropomorphic condition had the Merlin character utter explanatory speech, which was also displayed in speech bubbles on the screen. The textual non-anthropomorphic condition had the same text displayed in text boxes. The difference in display design was that the text boxes in the non-anthropomorphic condition were not placed close to the area on the screen that they were attempting to explain – compared to the anthropomorphic condition. The result of this could have been that the cognitive affordances would have been negatively affected with respect to the

facilitation of the participant 'knowing' or 'thinking' appropriately about accomplishing the tasks. Further the sensory affordances would have also been affected and not provided the appropriate support for the cognitive affordances. This could have happened because part of the explanations for the download, installation and configuring of the email client involved completing form based aspects as part of an on-screen dialogue. If the text boxes were not close enough to the area requiring the interaction, the sensory affordance concerning 'seeing' could have been also negatively affected and therefore not supported appropriately the cognitive affordance aspect. The physical affordances in this experiment tended to be the fields and buttons of the email client dialogue, which were used by the participants with the keyboard and mouse. These were the same under both conditions and should therefore not have affected matters either way. The functional affordances should therefore not have been affected either, as the experiment aimed to 'explain' or guide the user through the various steps of the field filling and dialogue stages. The actual results of the statistical analysis give some support to this argument because the participants in the non-anthropomorphic condition significantly perceived the feedback to be less understandable, insufficient, less friendly and more intimidating. Lastly this group achieved significantly lower performance scores compared to the anthropomorphic group. These aspects do suggest that due to the textual instructions being laid out onto the screen in the manner described, could have negatively affected the various strands of affordances.

4 CONCLUSIONS

As has been considered in this paper, the study of anthropomorphic feedback is still incomplete. Various researchers have obtained disparate sets of results with unclear reasons for these. However, the authors of this paper, suggest that potentially the issues of whether anthropomorphic feedback is more effective and preferred by users, is strongly linked with how the affordances are dealt with at the user interface. This aspect could also provide a reason regarding why there are so many disparate sets of results in the wider research community, concerning anthropomorphic feedback. Further, the principal author of this paper is continuing to investigate these issues and the affordances in light of other work by the principal author of this paper and work of the wider research community.

REFERENCES

Bengtsson, B, Burgoon, J. K, Cederberg, C, Bonito, J, and Lundeberg, M. (1999) The Impact of Anthropomorphic Interfaces on Influence, Understanding and Credibility. Proceedings of the *32nd Hawaii International Conference on System Sciences*, IEEE.

De Angeli, A, Johnson, G. I. and Coventry, L. (2001) The Unfriendly User: Exploring Social Reactions to Chatterbots, *Proceedings of the International Conference on Affective Human Factors Design*, Asean Academic Press.

Gibson, J. J. (1979) *The Ecological Approach to Visual Perception*, Houghton Mifflin Co.

Hartson, H. R. (2003) Cognitive, Physical, Sensory and Functional Affordances in Interaction Design, Behaviour and Information Technology, Sept-Oct 2003, 22 (5), p.315-338.

Kerio (2006) http://www.kerio.co.uk/ Accessed 2008.

Moundridou, M. and Virvou, M. (2002) Evaluating the Persona Effect of an Interface Agent in a Tutoring System. Journal of Computer Assisted Learning, 18, p. 253-261. Blackwell Science.

Moreno, R., Mayer, R. E. and Lester, J. C. (2000) Life-Like Pedagogical Agents in Constructivist Multimedia Environments: Cognitive Consequences of their Interaction. ED-MEDIA 2000 Proceedings, p. 741-746. AACE Press.

Murano, P, Ede, C. and Holt, P. O. (2008) Effectiveness and Preferences of Anthropomorphic User Interface Feedback in a PC Building Context and Cognitive Load. *10th International Conference on Enterprise Information Systems*, Barcelona, Spain, 12-16 June, 2000 - INSTICC.

Murano, P, Gee, A. and Holt, P. O. (2007) Anthropomorphic Vs Non-Anthropomorphic User Interface Feedback for Online Hotel Bookings, *9th International Conference on Enterprise Information Systems*, Funchal, Madeira, Portugal, 12-16 June 2007 - INSTICC.

Murano, P. (2005) Why Anthropomorphic User Interface Feedback Can be Effective and Preferred by Users, *7th International Conference on enterprise Information Systems*, Miami, USA, 25-28 May 2005. INSTICC.

Murano, P. (2003) Anthropomorphic Vs Non-Anthropomorphic Software Interface Feedback for Online Factual Delivery, *7th International Conference on Information Visualisation*, London, England, 16-18 July 2003, IEEE.

Murano, P. (2002a) 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 October 2002. Published in Lecture Notes in Computer Science - Springer.

Murano, P. (2002b) 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.