WHY ANTHROPOMORPHIC USER INTERFACE FEEDBACK CAN BE EFFECTIVE AND PREFERRED BY USERS

Pietro Murano

University of Salford, School of Computing, Science and Engineering, Computer Science Research Centre, Interactive Systems and Media Laboratory, Gt. Manchester, M5 4WT, England, UK

Keywords: User interface feedback, Anthropomorphism, Human behaviour.

Abstract: This paper addresses and resolves an interesting question concerning the reason for anthropomorphic user interface feedback being more effective (in two of three contexts) and preferred by users compared to an equivalent non-anthropomorphic feedback. Firstly the paper will summarise the author's three internationally published experiments and results. These will show statistically significant results indicating that in two of the three contexts anthropomorphic user interface feedback is more effective and preferred by users. Secondly some of the famous work by Reeves and Nass will be introduced. This basically shows that humans behave in a social manner towards computers through a user interface. Thirdly the reasons for the obtained results by the author are inextricably linked to the work of Reeves and Nass. It can be seen that the performance results and preferences are due to the subconscious social behaviour of humans towards computers through a user interface. The conclusions reported in this paper are of significance to user interface designers as they allow one to design interfaces which match more closely our human characteristics. These in turn would enhance the profits of a software house

1 INTRODUCTION

User interface feedback in software systems is being improved as time passes and developers dedicate more time to the feedback and realise that feedback to the user is just as important as the rest of an application.

In line with the goal of constant improvement and better understanding of user interface feedback this research has looked at the effectiveness and user approval of anthropomorphic feedback. This was compared to an equivalent non-anthropomorphic feedback.

Anthropomorphism at the user interface usually involves assigning human characteristics or qualities or both to something which is not human, e.g. a talking dog or a cube with a face that can talk etc. A well known example is the Microsoft Office Paper Clip. It could also be the actual manifestation of a real human such as a video of a human (Bengtsson et al, 1999).

This issue has been considered because there was a division between computer scientists where certain computer scientists are against (e.g. chapter by Shneiderman in ((Bradshaw, 1997) and (Shneiderman, 1992)) anthropomorphism at the user interface and others are in favour (e.g. Agarwal (1999), Cole et al. (1999), Dertouzos (1999), Guttag (1999), Koda and Maes (1996a), (1996b), Maes (1994) and Zue (1999)) of using anthropomorphism at the user interface. However there has not been concrete enough evidence to show which opinion may be correct.

Experiments (summarised below and detailed in Murano (2001a), (2001b), (2002a), (2002b), (2003)) have been conducted where it has been shown with statistical significance that in certain contexts anthropomorphic user interface feedback is more effective and preferred by users. However these experiments concentrated on 'what' type of feedback was better (i.e. anthropomorphic or non-anthropomorphic) and not on 'why' a particular type of feedback was better over the other.

This issue of 'why' was raised as an interesting question at various international conferences attended by the author. Hence firstly this paper aims to address this question and provide an answer by means of the body of evidence produced by Reeves and Nass. It is believed by the author that no other researchers outside of Reeves and Nass' influence have used and validated some of their results in such a detailed manner. Secondly, the experiments conducted and summarised below, are innovative in that while they follow the guidelines of Reeves and

Table 1: Comparison of Video Vs Diagrams and text

	Comparison of Video Vs. Diagrams and Text
t-Observed	2.14
t-Critical (5%)	1.74

Table 2: Overall User Preferences

	Overall User Preferences	
	Mean	Standard Deviation
Video	8.17	1.10
Diagrams and Text	7.11	2.17

Nass, this is the first time that the guidelines have been applied to a more realistic context. The experiments by Reeves and Nass were more artificial in nature.

This is because despite many computers and applications being in homes and businesses, there are still many prospective users who are afraid of computers. These prospective users could become actual enthusiastic users, thus potentially increasing business profits for a software house. This could be achieved by the improvement of user interface feedback by using the findings of this paper.

2 SUMMARY OF EXPERIMENTS

In the next three sections below a brief summary is presented of the three experiments. Full details for repeatability can be found in Murano (2001a), (2001b), (2002a), (2002b), (2003). However for each of the three experiments within users' designs were used. This meant that in each experiment all the subjects tried all tasks and had the opportunity to use all relevant kinds of feedback. Considerable efforts were made to maintain laboratory conditions constant for each subject. Also efforts were made to control possible confounding variables.

2.1 Experiment One

The first experiment Murano (2002a) was in the context of software for in-depth understanding. This was specifically English as a foreign language (EFL) pronunciation. The language group used was Italian native speakers who did not have 'perfect' English. Software was specifically designed to automatically handle user speech via an automatic speech recognition (ASR) engine. Further, in line with EFL literature by Kenworthy (1992) and Ur (1996) exercises were designed and incorporated as part of

the software to test problem areas that Italian speakers have when pronouncing English.

Anthropomorphic feedback in the form of a video of a real EFL tutor giving feedback was designed. This in effect was a set of dynamically loaded video clips which were activated based on the software's decision concerning the potential error a user had done (if no errors were made no pronunciation corrections were made by the software). This type of feedback was compared against a non-anthropomorphic equivalent. In this case two-dimensional diagrams with guiding text were used. The diagrams were facial cross-sections aiming to assist a user in the positioning of their mouth and tongue etc. for the relevant pronunciation of a given exercise. This type of feedback was based on EFL principles found in Baker (1981) and Baker (1998). No feedback type was ever tied to the same exercise, i.e. feedback was randomly assigned to an exercise.

The results for 18 Italian users (with imperfect English pronunciation) taking part in a tightly controlled experiment, going through a series of exercises were statistically significant. Users were scored (scores used in hypothesis testing statistical analysis) according to the number of attempts they had to make to complete an exercise successfully.

The statistical results suggested the anthropomorphic feedback to be more effective. Users were able to self-correct their pronunciation errors more effectively with the anthropomorphic feedback. The scores obtained were approximately normally distributed. These were then used in a t-test. The results are in the Table 1.

Furthermore it was clear that users preferred the anthropomorphic feedback. The actual scores obtained from the questionnaires using a Likert scale, where 1 was a negative response and 9 was a positive response, are detailed in Table 2.

Hence it was concluded that the statistically significant results suggested the anthropomorphic

Table 3: Comparison of Video Vs Text

	Comparison of Video Vs Text
t-Observed	10.21
t-Critical (5%)	1.67

Table 4: Overall User Preferences

	Overall User Preferences	
	Mean	Standard Deviation
Video	7.53	1.40
Text	6.35	1.84

feedback to be more effective and preferred by users.

2.2 Experiment Two

The second experiment by Murano (2001a), (2001b), (2002b) was in the context of software for online systems usage.

This was specifically concerned with the using of UNIX commands. This was an interesting area as typically novice users of UNIX commands can find it difficult to master the concepts of the command structure and to remember relevant commands in the first place. Software was designed to emulate a small session at the UNIX shell covering a sub-set of UNIX commands. As in the first session, an ASR engine was used which allowed the users to 'query' the system verbally. The users which were recruited for the experiment were complete novices to UNIX commands.

In this experiment anthropomorphic feedback was compared with a non-anthropomorphic equivalent. In this case the anthropomorphic feedback consisted of dynamically loaded video clips of a person giving the command verbally for the current context the user was in. The feedback was prompted by the user requesting the feedback from the system (through the ASR engine). The non-anthropomorphic feedback was a textual equivalent (based on the structure used in Gilly (1994)) appearing in a supplementary window next to the main X-Window. A small set of typical tasks a beginner might engage in, involving UNIX commands, were designed. Since the users had no knowledge of UNIX commands, they were obliged to make use of the feedback if they wished to complete the tasks. The two types of feedback were randomly assigned to the tasks so that one task was not tied to one type of feedback.

The results for this tightly controlled experiment, which involved 55 users who were novices to UNIX commands, were statistically significant. The users were scored (scores used in hypothesis testing statistical analysis) as they attempted a set of tasks using UNIX commands. Scores were devised according to the number of errors, hesitations and completions/non-completions a user was able to carry out. Further, scores were obtained via a questionnaire for users' opinions on the system feedback given them.

The statistical results suggested the anthropomorphic feedback to be more effective. The scores obtained were approximately normally distributed. These were then used in a t-test. The results are in the Table 3.

Furthermore it was clear that users preferred the anthropomorphic feedback. The actual scores from the questionnaires using a Likert scale, where 1 was a negative response and 9 was a positive response, are detailed in Table 4.

Hence it was concluded that the statistically significant results suggested the anthropomorphic feedback to be more effective and preferred by users. The users were able to carry out the tasks more effectively with the anthropomorphic feedback, whilst indicating a preference for the anthropomorphic feedback.

2.3 Experiment Three

The third experiment by Murano (2003) was in the context of software for online factual delivery. Specifically the context for this area was direction finding. Software was developed to give directions to two different but equivalent locations (equivalence was concerned with approximately equal distances and difficulty), where the aim was for test subjects to physically find their way to the given locations. The subjects were to use the directions given to them by the system. Hence it was a prerequisite that the subjects should not have known where the locations were before taking part in the experiment (this was determined as part of a questionnaire).

In this experiment anthropomorphic feedback was compared with a non-anthropomorphic equivalent. In this case the anthropomorphic feedback consisted of dynamically loaded video

Table 5: F-test Results - Diagram Vs. Video

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

Table 6: Overall user preferences

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

clips of a person giving directions to a location. This was compared with an equivalent non-anthropomorphic feedback consisting of a map with guiding text based on the principles found in Southworth and Southworth (1982). One type of feedback was not tied to one particular location in the experiment. The feedback was rotated so that each location had either type of feedback at some point in the experiment.

The results for this tightly controlled experiment, which involved 53 users, were statistically significant (for effectiveness in favour of the map). The users in the experiment were scored (scores used in hypothesis testing statistical analysis) according to the amount of mistakes they made (i.e. wrong turnings taken), visible hesitations and if the subjects actually reached the prescribed location. Users were also given a questionnaire which allowed them to express their opinions concerning the feedbacks. The results were statistically significant in favour of the map with guiding text feedback. The users overall performed the tasks of direction finding more effectively with the map and guiding text.

The actual data collected was found to be approximately normally distributed and was used in an F-test. The results are in Table 5, suggesting the map to be more effective.

The opinions of the users concerning their preferences were much less clear. The actual scores from the questionnaires using a Likert scale, where 1 was a negative response and 9 was a positive response, are detailed in Table 6.

As can be seen from the above table, the scores for the opinions (overall user preferences) showed the map to be only slightly better than the anthropomorphic feedback. Many of the users liked very much the idea of having 'someone' give them directions rather than the map. This resulted in them scoring the anthropomorphic feedback much higher than expected.

2.4 Overall Discussion of Experiments

The experiments summarised in the last three sections show clearly that anthropomorphic feedback is generally liked by users in most situations. However the effectiveness of such feedback is dependant on the domain of concern. Hence certain domains appear to not be suited to anthropomorphic feedback, such as the domain for online factual delivery, particularly the direction finding context. This is also confirmed by the suggestion based on other research discussed in Dehn and van Mulken (2000). However as the third experiment showed, users still like seeing and interacting with anthropomorphic feedback even if it is not the best mode of feedback for them to achieve their tasks.

These results suggest the conclusion that it would be better for designers of feedback to include anthropomorphic feedback in the domains shown to be better suited to such a style. For the domains not suited to anthropomorphism it clearly needs stating that a suitable non-anthropomorphic feedback should be used instead. However based on what users like, it may be suitable to combine non-anthropomorphic feedback with some form of anthropomorphic feedback. An example based on the third experiment described above is to have the map with guiding text (which was more effective), and to perhaps have a video or synthetic character of a person giving some 'external' (not the actual directions) information. 'External' information could simply be to introduce the user to study the map being presented to them. This would give the user the benefit of anthropomorphism and the effectiveness of the map with guiding text. This suggestion may seem simple. However many software packages have failed due to bad user interfaces and feedback. Sometimes the problems could have been resolved by fairly simple means. Hence this suggestion is in line with the idea that sometimes minor adjustments can dramatically improve the usability of a system.

3 THE WORK OF REEVES AND NASS APPLIED TO USER INTERFACE FEEDBACK

As stated in the introduction, these issues really only deal with the 'how'. This means that it has been discovered how we should give feedback in certain domains, i.e. certain domains are better suited to anthropomorphic feedback. However the issue of 'why' has not been addressed by the experiments, i.e. why is it that in certain domains anthropomorphic feedback is more effective and preferred by users? The answer lies in us as humans.

Reeves and Nass (1996) have for many years conducted research very compatible with the research summarised in this paper. In their book 'The Media Equation' (1996), they have discussed empirical findings which give us the answer to the question posed in the previous paragraph. In this large body of research they have found that people in general (including computer scientists) tend to interact with a computer in a social manner and in a very similar manner to the way one interacts away from a computer, i.e. with other people etc. Also they have found that people apply the basic social rules of every day life to their interaction with computers. This is done automatically and intuitively by humans. In fact they do not even realise they are behaving in this manner.

These points are crucial to the findings of the three experiments summarised above. The suggestion here is that the subjects concerned were subconsciously applying human social rules whilst interacting with the feedbacks and because one of the types of feedback (anthropomorphic) was more compatible with the applying of social rules, the results showed more effectiveness (in two of the experiments) and very importantly high user approval (in all three experiments).

This suggestion does not explain why the results did not apply to one of the experiments (the direction finding experiment). One explanation is that whilst users would have been applying social rules in all circumstances, the direction finding experiment was much better suited to the map with guiding text feedback. However the issue of user approval for the anthropomorphic feedback in the direction finding experiment does support the findings that users will apply social rules whilst interacting with computers.

This issue leads to other aspects of the research by Reeves and Nass – the unique characteristics of anthropomorphic feedback. There are various characteristics which easily occur in anthropomorphic feedback that are compatible with the subconscious use of social rules. It is these characteristics that account for the 'why' or the reason for anthropomorphic feedback being more effective in certain domains and being mostly preferred by users.

One aspect concerns the fact that human-to-human communication involves eye contact and it has been shown in Ekman (1973) and Ekman et al (1972) that if a person is looking at a 'face', about half of the time used in this activity is used to look at the eyes. Also if one matches modalities this usually incurs a better response, e.g. if one sends an email to a friend, usually the reply will be sent by email and not by a telephone call, as stated by Reeves and Nass (1996). Reeves and Nass (1996)argue that this human phenomenon could work at the user interface if one could overcome the obvious barriers to this, e.g. if an electronic voice issues advice/information to the user it would be better to have the system accept input verbally from the user, via an ASR engine. This is because communication modalities are being matched and are closer to the human-to-human social rules.

These two aspects are very important with respect to the anthropomorphic feedback used in the experiments. This is because as stated above the anthropomorphic feedback consisted of dynamically loaded video clips of a person. These clips showed the face of the person clearly and within the limits of the video one could see the eyes. This is very important based on the material of Ekman (1973) and Ekman et al (1972) because the users would have been subconsciously looking at the face and spending a good proportion of the time looking at the eyes of the person in the video clips (whilst listening to the help given). Furthermore the communication modalities were well matched as the user communicated with the system via the ASR engine and clearly the anthropomorphic feedback (video of human talking) was also communicating verbally.

Another aspect that should be considered is the way the experiments were conducted. Reeves and Nass argue that when one is testing a product presented by a computer the computer should not ask the user for evaluations. If the evaluations are to be done electronically it would be better to use a different computer. Alternatively these would be better conducted by some paper based means. However caution would still need to be deployed because if the person conducting the experiment was also the person helping the user in some way, then Reeves and Nass (1996) state that a subject may subconsciously look for the most diplomatic responses so as not to upset anyone. They suggest that the best way to overcome these problems is to test two 'products' against each other. In this way

the subjects do not feel obliged to respond in some socially acceptable manner.

The experiments were conducted in the manner suggested by Reeves and Nass. In each case two types of feedback were the basis of each experiment, being tested against each other. The subjects were asked for their evaluations on these for usability etc. Furthermore the evaluations were carried out by the subjects not in an electronic manner, but away from the computer by means of carefully designed paper-based questionnaires. Thus it is suggested that any bias on the evaluation concerning the applying of certain social rules should have been dramatically reduced if not eliminated completely.

Another aspect requiring consideration is that Reeves and Nass (1996) and Reeves at al (1992) discovered that people tend to have similar reactions with a picture of a person as they do with a real person in front of them. When a person sees another person that is near them, the human subconscious result is that people will evaluate that person more intensely, pay more attention to them and remember them better. They found that these principles still applied if one looked at a picture of a person. This is important as it affects the way a person could view some anthropomorphic feedback, either of a person or some synthetic character.

The experiments conducted by the author had as stated above anthropomorphic feedback consisting of video clips of a person. These were filmed in such a manner so as to follow the principles found by Reeves and Nass. The person in the clips was seen to be near to the person using the feedback. This was achieved simply by filming the person from not too far a distance. Also there were no large 'open spaces' around the person being filmed. This resulted in the person appearing quite close to the user viewing the feedback. This filming strategy would have resulted in users feeling a more intense evaluation of the feedback along with enhanced memory results and actually paying more attention to the feedback during the session. It is suggested that this would have resulted in the users basically performing their tasks better and also evaluating the feedback very positively.

Reeves and Nass (1996) also discuss the effects of having 'unnecessary peripheral motion'. They say that having this in an interaction leads the user to be distracted from their current attention giving activity to the 'item' moving at some other position in the screen or window. This clearly results in something being ignored from the primary interaction.

This suggestion was put into practice for the feedbacks. Simply no 'peripheral motion' was used so that all the attention could be put onto the feedback and the help being given for achieving the tasks.

4 CONCLUSION

The issues discussed above provide a reasonable explanation based on empirical findings concerning the reasons for the anthropomorphic feedback being more effective (in two of the three contexts tested) and liked by users in all cases.

One aspect is that humans, whether they admit it or not, behave in a social manner with computers, applying various social rules as Reeves and Nass found. This has been crucial to the effects observed in the experiments summarised above.

The next important aspect as discussed above is the fact that anthropomorphic feedback and the way it is presented to users can very naturally provide humans with the appropriate 'cues' for them to behave in a more social manner towards the feedback. This in turn results in better task completions and a higher satisfaction rate in certain contexts.

This work has also shown the validity of the work by Reeves and Nass in this area as the findings of the reported experiments corroborate some of the findings by Reeves and Nass. This work also takes the work of Reeves and Nass further as this work has been conducted in a much more realistic set of contexts compared to the contexts used by Reeves and Nass.

Concerning interface designers it is suggested that they should take seriously the use of anthropomorphic feedback, as the suggestion is that it leads to better more productive interactions and more usable interfaces. This is significant as more and more people who are not 'professionals' are using computer systems and their software. This in turn brings the requirement of developing better user interfaces by using results such as the ones discussed in this paper. This in turn could have a beneficial effect to the profits of a software house. Clearly if they can apply these principles they could potentially attract a new market, particularly composed of those who may be afraid of computers.

However despite the potential for applying these findings immediately in a business context, there is more work required. One of the experiments showed that the anthropomorphic feedback was not as effective as the non-anthropomorphic equivalent. This leads to the requirement of investigating other areas of user interface feedback to try and find other domains not suited to this type of feedback. Ultimately, a taxonomy of possible domains and suitable types of feedback could be devised over time. If this was available based on empirical findings, user interface designers could be helped when they are faced with the many decisions they have to take when designing user interfaces.

A further area that would need investigating is feedback in a virtual reality setting and also the very different setting of performing two or more tasks at once. An example would be a car driver driving a car and trying to find directions to some location (currently this is done with navigation software and any future study should take into account issues such as cognitive load and divided attention etc.). Alternatively, investigating this type of feedback for pilots could lead to some interesting findings and perhaps a more comprehensive taxonomy of the kind suggested in the previous paragraph.

ACKNOWLEDGMENTS

The School of Computing, Science and Engineering at the University of Salford, Prof. Sunil Vadera, Prof. Tim Ritchings and Prof. Yacine Rezgui are thanked for their support.

REFERENCES

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

Baker, A. (1981) *Ship or Sheep? An Intermediate Pronunciation Course*, Cambridge University Press.

Baker, A. (1998) *Tree or Three? An Elementary Pronunciation Course*, Cambridge University Press.

Bengtsson, B., Burgoon, J.K. et al.(1999) The Impact of Anthropomorphic Interfaces on Influence, Understanding and Credibility. *Proc of the 32nd Hawaii International Conference on System Sciences,*. IEEE

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

Brennan, S.E and Ohaeri, J.O. (1994) Effects of Message Style on Users' Attributions Toward Agents. *CHI '94 Human Factors in Computing System*

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,*

Dehn, D. M. and van Mulken, S. (2000) The Impact of Animated Interface Agents: A Review of Empirical Research. *International Journal of Human-Computer Studies 52: 1-22*

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

Ekman, P. (ed.) (1973) *Darwin and Facial Expression: A Century of Research in Review*, Academic Press. New York.

Ekman, P., Friesen, W.V., Ellsworth, P. (1972) *Emotion in the Human Face: Guidelines for Research and an Integration of Findings*, Pergamon Press. New York.

Gilly, D. (1994) *UNIX In a Nutshell*, O'Reilly and Associates

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

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

Kenworthy, J.(1992) *Teaching English Pronunciation*, Longman.

Koda, T. and Maes, P. (1996a) Agents With Faces: The Effect of Personification. *Proc of the 5th IEEE International Workshop on Robot and Human Communication,*, IEEE.

Koda, T. and Maes, P. (1996b) Agents With Faces: The Effects of Personification of Agents. *Proc of HCI '96*, British HCI Group.

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

Murano, P. (2001a) A New Software Agent 'Learning' Algorithm. *People in Control An International Conference on Human Interfaces in ControlRooms, Cockpits and Command Centres*, IEE.

Murano, P. (2001b) 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*, IEE.

Murano, P. (2002a) Effectiveness of Mapping Human-Oriented Information to Feedback From a Software Interface. *24th International Conference* Information Technology Interfaces.

Murano, P. (2002b) 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,. Published in Lecture Notes in Computer Science (C) - Springer.

Murano, P. (2003) Anthropomorphic Vs Non-Anthropomorphic Software Interface Feedback for

Online Factual Delivery. *7th International Conference on Information Visualisation (IV 2003) An International Conference on ComputerVisualisation and Graphics Applications*, (c) – IEEE.

Nass, C., Steuer, J. et al. (1994) Computers are Social Actors. *CHI '94 Human Factors in Computing Systems – 'Celebrating Interdependence'*, ACM.

Reeves, B., Lombard, M. and Melwani, G. (1992) Faces on the Screen: Pictures or Natural Experience, *International Communication Association*.

Reeves, B. and Nass, C. (1996) *The Media Equation How People Treat Computers, Television, and New Media Like Real People and Places*, Cambridge University Press.

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

Southworth, M. and Southworth, S. (1982) *Maps a Visual Survey and Design Guide*. Little, Brown and Co.

Ur, P. (1996) *A Course in Language Teaching - Practice and Theory*, Cambridge University Press

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