

## Investigating Effective ECAs: an Experiment on Modality and Initiative

Alistair Sutcliffe<sup>1</sup>, Faisal Al-Qaed<sup>2</sup>

Centre for HCI Design, School of Informatics, University of Manchester, UK

<sup>1</sup> [alistair.g.sutcliffe@manchester.ac.uk](mailto:alistair.g.sutcliffe@manchester.ac.uk)

<sup>2</sup> [faisal.al-qaed@postgrad.manchester.ac.uk](mailto:faisal.al-qaed@postgrad.manchester.ac.uk)

**Abstract.** This paper investigates the effectiveness of conversational agent-based delivery of task strategy and operational help for an interactive search tool. The study tested three modalities of advice (text-only, text-and-audio, and text-audio-and-agent) in addition to a control group with no advice. User- and system- initiated advice modes were also compared. Subjects in the text-only group outperformed other modality groups in usability errors, search performance, advice uptake and in their positive comments in the debriefing interview and post-test questionnaire. User-initiated advice was preferred and was more effective. Users criticized speech advice for being too long and difficult to control. The results suggested that the computer as social actor paradigm might not be effective for advisory applications.

**Keywords:** Multi Modal Interfaces, Advisor Agents, User initiative

### 1 Introduction

Embodied Conversational Agents (ECAs) have attracted considerable attention as a means of enhancing user experience in e-commerce [1], and in social interaction on the Internet [10]. Following Fogg's [12] advocacy of computers as a persuasive technology, ECAs are being used in e-commerce as store guides to advise users [1, 31]. Although the computer as social actor theory [28] predicts that the use of human image and voice will invoke a powerful response by users, treating computers as if they were human agents, it is not clear how far these experimental findings generalize to interactive settings. Conversational agents or avatars can improve user engagement and motivation [3, 7]; however, evidence of task-related performance improvements is less clear.

ECAs using TTS (Text-To-Speech) voice have positively influenced users' preferences and purchasing behavior [8]; however, in another web-based study, the presence of animated agents that actively monitored users' behavior was found to decrease users' performance and increase their anxiety level [29]. Generally, there are few convincing demonstrations that ECAs or animated characters significantly influence users' behavior or improve task-related performance [11, 21]; indeed, Berry et al. [2] found that text-based advice was better than an ECA for persuading users to adopt healthy life styles, although the agent version with more natural dialogue and facial expressions was preferred. This paper extends previous studies by investigating

the effectiveness of the CSA paradigm for task-based advice. Most previous evaluations have assumed system initiative and control of the conversation by the ECA.; however, given the well known user aversion to system initiated help, for instance the Microsoft Clippie agent; and politeness rules posited in [28] user initiated ECA advice might receive a more positive reaction. To complement previous studies we also investigated differences in user- and system- initiated interaction.

The context for the study is providing decision making advice based on strategies from the Adaptive Decision Making (ADM) theory [24] and online help for users using a visual browsing/search tool based on Shneiderman's Spotfire display [32]. Since users tend to access help and explanation systems when they perceive the task to be difficult, this scenario investigates the difference between task-based advice for decision making which may be useful but not essential, and online help which will be necessary if problems are encountered.

The paper is organized as follows: the next section provides a brief background of related research. This is followed by the discussion of the methods and results of the experiment. The discussion reviews the experimental findings and concludes with plans for future research.

## **2 Related Work**

ECAs can emulate human behaviors and influence user behavior and judgment if they portray a believable human presence [7]. Substantial attempts [6, 7, 21] have been made to improve the design and responsiveness of ECAs with emotional reactions and appropriate facial expressions to provide an experience approaching the quality of natural human-to-human interaction. Preliminary evaluations of Cassel's RHEA showed that agent-based advisors for health topics can be effective. Pelachaud and Bilvi [25] presented a computational model for the creation of non-verbal behaviors associated with speech for natural ECAs that can emulate human behaviors, facial expressions, body gesture and speech. Informal evaluations of their GRETA agent also suggest the computer as social actor can be an effective advisor.

De Angeli et al. [9, 10] have shown that even limited capability ECAs or chatterbots designed with social anthropomorphic reactions can establish social relationships with users. In a Wizard of Oz study on help for online banking, they demonstrated that text-based dialogues with chatterbots closely resemble human-human conversations, and that people prefer ECA agents to text-based advice. Similar limited capability ECAs based on Microsoft agents have been implemented as advisor agents in e-commerce applications [31] in which two or more agents converse with the user to advise on different aspects of a product. Once again informal evaluations seem to show that users respond well to agent-based advisors.

Commercial use of ECA-style shopping assistants is becoming widespread on the Internet. In a study of e-commerce websites, Keeling et al. [15] investigated the appropriateness of different ECAs (e.g. cartoon-like agents and human-like agents) for a variety of the websites and users. They concluded that great care should be taken in matching ECAs with websites, considering both the physical characteristics of the ECA and the goals and motivations of the users.

Riegelsberger et al. [30] illustrated how users respond more effectively and more naturally to rich media representations (video, audio, avatar, photos) compared to

textual messages, and showed that video followed by audio gives the user the most detailed insight into expertise and promotes trust with agents. Oviatt et al. [22, 23] investigated natural modality integration patterns and proposed a predictive model for modality integration. This suggests that ECAs using speech with complementary non-verbal communication (facial expression and gesture) should be more effective than other single modalities. Overall it appears that, apart from Berry et al.'s study, most of the evidence points towards agent-based advisors as being more effective than standard text-based advice.

ECAs have tended to use speech to deliver advice; in contrast, chatterbots and avatars on the Internet tend to be text based. Guidelines for modality combination suggest that speech may be more effective than text when advice also involves inspecting a text-based resource, for example user interface menus, prompts and displays. The modality design guidelines and the Split-Attention principle developed by Moreno and Mayer [17] recommend avoiding using the same modality for two competing tasks, and that material delivered on different modalities should be complementary rather than dividing the user's attention. In contrast, multimedia design guidelines [14, 34] advocate presenting detailed material in text rather than speech. Interpretation may depend on how detailed the advice is and whether users can split their attention between two tasks (e.g. operating the interface and comprehending advice) and information presented on one modality.

### **3 Experimental Design**

Forty subjects (25 males and 15 females, aged 20 to 39) from the University of Manchester participated in the experiment. Most subjects were postgraduate students with intermediate to expert Internet search experience. The subjects were paid £10 for their participation. The interview sessions were video-taped and lasted for 45 minutes, with experimental task durations ranging from 10 to 15 minutes.

The study tested three modalities of advice (text-only, text-and-audio, and text-audio-and-agent); in addition there was a control group with no advice, resulting in four between-subject groups with 10 subjects each. Following Moreno and Mayer's [17] multimedia design principle (Split-Attention Effect), the audio and text were not played concurrently, to prevent the audio from interfering with text advice.

In each modality group, two advice initiative conditions were used. In the user-initiative condition the subjects accessed the advice by clicking on appropriate buttons of the agent. In the system initiative condition the advisor delivered the advice automatically, although the speech/text delivery could be cancelled. The subjects were able to close the advisor window or get advice any number of times in both initiative conditions. The initiative conditions were counter-balanced for each modality group.

Prior to starting the experimental tasks, the subjects were given a brief tutorial on how to control the advisor. They were then given a scenario of finding an apartment to rent in Manchester using a student accommodation database with two tasks:

Use the Scatterplot filtering tool to:

- Find the best apartment that has 2 rooms and a weekly rental less than £200.

- Find the number of apartments that are 3-4 miles from the city centre and with popularity rating of at least 8. (Hint: You may need to re-configure the slider settings).

The hint directed subjects' attention to the need to change the sliders on the tool before the second search task could be completed. Thus the first task was relatively simple and could be achieved without help whereas the second task should be perceived as being more difficult; and motivate access to the agent's help. The Scatterplot filtering tool (see Figure 1), inspired by Shneiderman's [32] Filmfinder tool using the concept of dynamic query filters, was developed using Java applets that communicated with Java Servlet programs at the server-side to dynamically retrieve data from the database. Operation of the tool was reasonably intuitive although users had to guess that both ends of the slider could be manipulated to set a range. The advisor agent was implemented using Microsoft Agents toolkit, using a male agent with limited facial expressions and gestures. Text was presented in a separate window partition once the speech was completed. The two tasks required subjects to use different features of the tool. Two identical types of advice were given in each modality group, Strategy advice and Tool advice. Strategy advice provided decision strategies that could be employed with the selected tool, whereas Tool advice contained specific explanation about how to use different features of the tool. The control group subjects carried out the tasks without any advice. Finally, all subjects were presented with a post-test questionnaire eliciting their subjective assessment of the advisors, and a debriefing interview in which they were asked for their advisor modalities and initiative preferences; and to provide suggestions for improvements with explanations of any particular problems they had experienced.

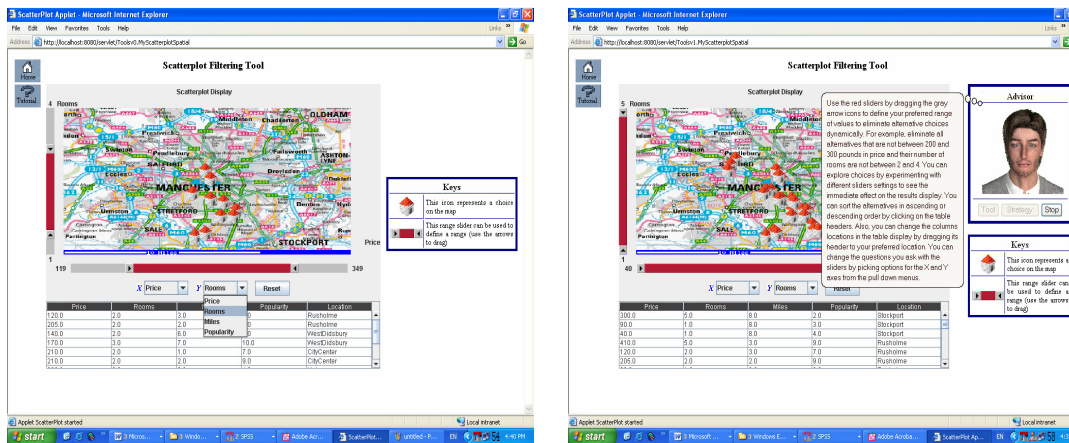


Figure 1: Scatterplot Filtering Tool without the Advisor (left), and with the Advisor (right) showing the text appearing after the speech advice.

## 4. Results

### 4.1 Task Times and Errors

The subjects' search performance times differed significantly between the groups (see Figure 2) in Task 1 (ANOVA,  $F=13.876$ ,  $df=3,39$ ,  $p<0.001$ ) and for Task 2 ( $F=12.431$ ,  $df=3,39$ ,  $p<0.001$ ). There were no differences in search times between tasks since users learned search tool operation in Task 1 (mean 507.6s) which compensated for the additional configuration actions required in Task 2 (mean 496.5s). All modality groups had better search performance times than the control group in both tasks ( $p<0.001$ , Tukey HSD post-hoc test). In Task 1, the control group was significantly worse than the text-only, text-and-audio, and text-audio-and-agent groups (henceforth agent group) ( $p<0.001$ ,  $p<0.05$  and  $p<0.01$  respectively), while the text-only condition was better than text-and-audio and agent groups ( $p<0.05$  and  $p<0.05$  respectively). In Task 2, similar significant differences were found between the control and all other groups ( $p<0.001$ ,  $p<0.05$  and  $p<0.05$  respectively). In addition, the text-only condition was better than the text-and-audio and agent modalities ( $p<0.05$  and  $p<0.05$  respectively). Similar inter-modality differences were found when times were adjusted to allow for the imposed latency in the agent and text conditions when subjects had to listen to the speech before the text advice appeared.

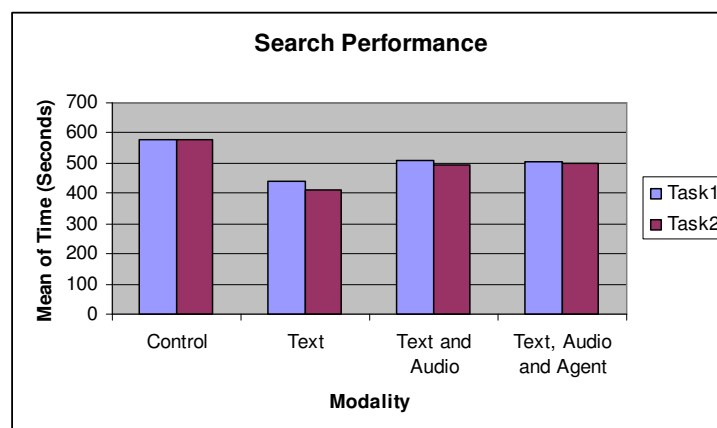


Figure 2: Mean Search Times (not adjusted for imposed latency)

A one-way between-group ANOVA with observed usability errors as the dependent variable showed a significant difference between groups ( $F=6.063$ ,  $df=3,39$ ,  $p<0.01$ ). The control group experienced significantly more errors than the advice text-only and agent groups (Tukey HSD post-hoc tests  $p<0.001$  and  $p<0.05$  respectively); however, the difference between the control group and text-and-audio group was not significant (see Figure 3). More errors occurred in Task 1 than in Task 2 ( $p<0.05$  Binomial test on totals for all conditions) even though Task 2 was more difficult and involved configuration so the learning effect was minor. The agent group had fewer usability errors than the text-and-audio group even though the same advice was present in both conditions, although this difference was non-significant. Most of the usability errors were accounted for by four design problems: users couldn't find the sliders, they did not realize that both ends of the slider bar could be moved to create a

range query, the sort results function was not clear as it required the user to click on the header bar to sort by the selected attribute (location, price), and the configuration cue was poor since users had to guess that clicking on the x or y axis box was necessary to change the settings. The tool advice enabled all these problems to be solved.

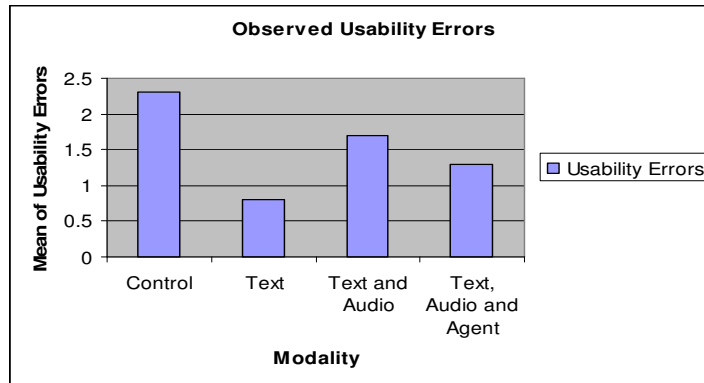


Figure 3: Observed Usability Errors Means

Figure 4 shows that text was the preferred modality for delivering advice. Total advice access was significantly different between modalities ( $F=8.694$ ,  $df=2,29$   $p<0.01$ ), as was Tool advice ( $F=6.927$ ,  $df=2,29$   $p<0.01$ ); however, Strategy advice access did not differ significantly between modalities. In pairwise comparisons text-only had a higher frequency of total advice access compared to text-and-audio, and agent modalities (both  $p<0.01$  Tukey HSD). For Tool advice, the text-only showed significantly more frequent access than the text-and-audio group ( $p < .01$ ) but no differences in strategy advice access were apparent. Total advice access differed between tasks ( $p<0.01$ ,  $T1>T2$ ); so it appears that Task 1 received more advice but also caused more errors, even though it was simpler than Task 2. This apparent contradiction is analyzed further in section 4.4.

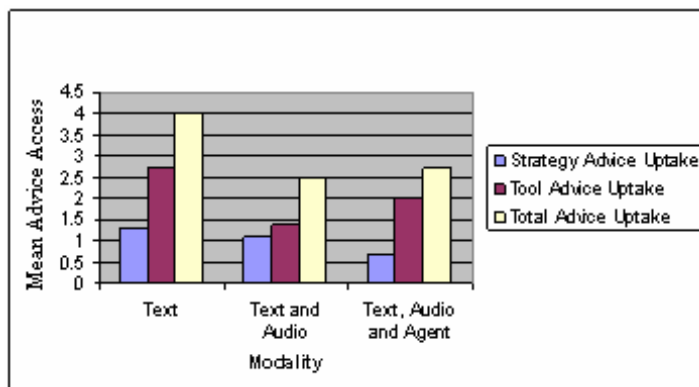


Figure 4: Advice Access frequencies by Modality

Tool advice was accessed more frequently than Strategy advice (see Figure 5) probably because it was perceived to be more useful, from debriefing interview comments. In the debriefing interview most subjects commented that the Strategy advice agreed with their normal approach to search tasks hence it had limited utility and they ignored it. Many subjects (56.7%) complained that the speech advice was too long, which may have caused them to access it less frequently than text presentations.

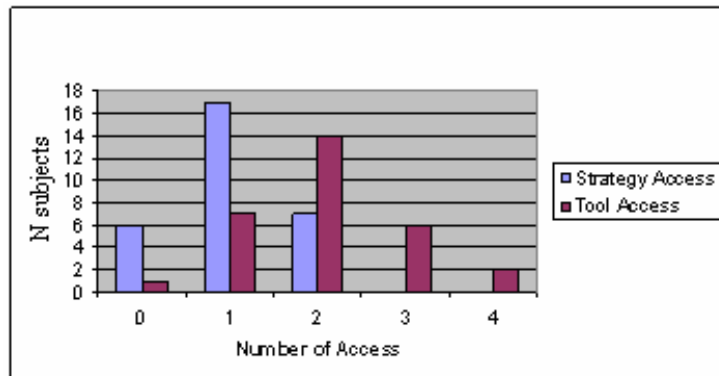
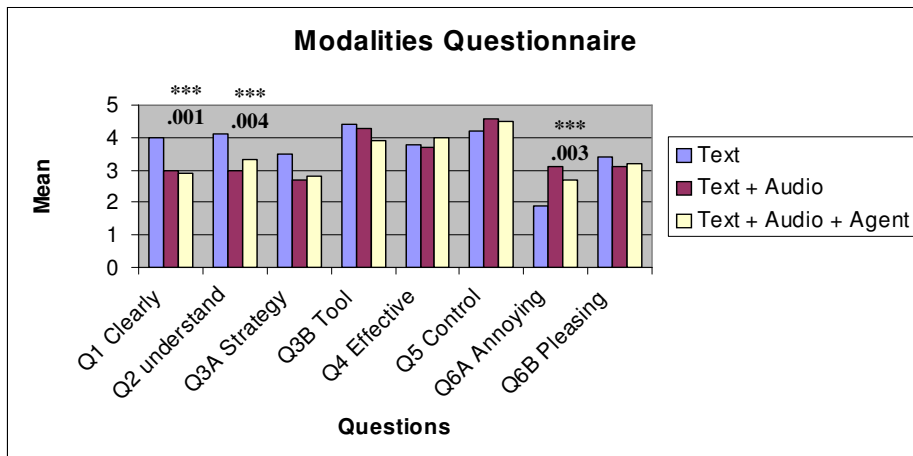


Figure 5: Number of subjects accessing each type of advice

#### 4.2 Questionnaire Ratings



Note: Stars indicate significant differences between advisor modalities (Sig. level <0.05) and the numbers below them indicate the Sig. Value. Q represents question number explained within text.

Figure 6: Post-test Questionnaires ratings

Figure 6 shows users' ratings captured in the post-test questionnaire. In Q1 (The Advisor delivers the advice clearly) differences between modalities were significant

( $F=9.165$ ,  $df=2,39$   $p<0.001$ ) with the text-only advisor outperforming the other modalities in delivering clear advice in pairwise comparisons ( $p<0.005$ ). Q2 (I was able to follow and understand the system advice) showed a similar pattern between modalities ( $F=6.715$ ,  $df=2$ ,  $p<0.01$ ); text-only outperformed the other modalities  $P<0.01$  with text-and-audio and  $p<0.05$  with the agent). In Q6A (System delivered advice in a way that was annoying), the difference between modalities ( $F=7.252$ ,  $df=2$ ,  $p<0.01$ ), favored text-only presentation as being less annoying than the other modalities ( $p<0.01$  with text-and-audio and  $p<0.05$  with the agent). However, Q3A (The system was useful for Strategy advice), Q3B (The system was useful for Tool advice), Q4 (The advisor's way of delivering advice is effective), Q5 (User controls for getting advice are easy to use) and Q6B (System delivered advice in a pleasing way) showed no differences between the modalities. From the users' subjective opinion ratings the agent does not appear to have been treated as a social actor in this advisor context.

#### 4.3 Debriefing Interviews

In debriefing interviews, 63.3% of the subjects stated that they preferred the agent advisor modality, followed by the text-only advice (16.7%); while 80% of the subjects preferred the text advice compared to audio and text. Their preference for the advisor modality was significant (Chi-Square = 37.3,  $df=4$ ,  $p<0.001$ ). Although they perceived the agent interface to be more attractive, most subjects (80%) suggested giving the advisor more personality to improve its social presence

The audio was disliked by 80% of the subjects, who commented that it was annoying to have to listen to a long speech clip especially when they needed to access the advice more than once (the duration for the audio of Tool advice was 52 seconds and for Strategy advice was 24 seconds). In contrast, the subjects preferred the text advice because they could read it at their own pace. 55% of the subjects from audio groups disliked the artificial voice and recommended using more natural recorded speech. A third of the subjects (30%) recommended better matching of voice, accent and tone to the user group, e.g. speaker gender, while two thirds (70%) suggested better controls for the pace and content of the speech and use of smaller speech segments (65%).

The subjects were asked to state which advice initiative (system or user) they preferred. As expected, user initiative was preferred in both tasks (86.7%, Chi-Square = 16.1,  $df=1$ ,  $pP<0.001$ ). Several subjects commented that the system initiative was annoying, especially the audio conditions. Advice access frequency was controlled by the experimental condition, so only repeated access (second access after cancelling speech) could be compared; however, no significant differences were apparent between system and user initiative.

#### 4.4. Behavior Analysis

In this section we explore the effects of advice initiative in more depth by analyzing the users' behavior patterns. Using the video recordings the subjects' search behavior was categorized into actions: using sliders to search, browsing and sorting the results table or getting advice. Usability errors were classified as attempting an action, e.g. attempt use sliders. Sequences of behavior were analyzed by casting transition frequencies for all combinations of behavior categories in a matrix (i.e. frequencies



where A was followed by B, B was followed by C, etc.) and then constructing behavior network graphs for all subjects. To test for significant sequence transitions, an expected value was calculated for each cell in the matrix by formula (1):

$$Exp_{A, B} = \frac{Frequency_A \times Frequency_B}{NT} \quad (1)$$

where NT=total transition frequencies in the whole matrix divided by number of cells.

The expected value was then used in the Binomial test for sample sizes where  $N > 25$  in formula (2), where  $x$  = the observed transition frequency between behavior (A,B) and  $N$  was the expected value:

$$z = \frac{(x \pm 0.5) - 0.5N}{0.5\sqrt{N}} \quad (2)$$

In the network diagrams, only frequencies above 1% of the overall total are reported, to make the diagrams more readable. This may cause inconsistencies between the total input and output transitions for some behaviors.

Figure 7 illustrates the behavior patterns for Task 1, user initiative condition. Most subjects either attempted to use the sliders or browsed, then accessed the Tool advice. Seven subjects cancelled the advice (all in the speech condition), while eight accessed the advice twice. After accessing the advice most users operated the tool successfully and completed the task, although few used the sort function. In the system initiative condition (see Figure 8) all users cancelled the strategy advice and most cancelled the tool advice before it ended. Four used the tool sliders successfully then browsed results, while four were unsuccessful even after accessing advice. Many users accessed the advice on their own initiative after experiencing difficulties. This explains the apparent anomaly of frequent advice access and more errors in Task 1; users were following a trial and error strategy and accessing advice after experiencing problems.

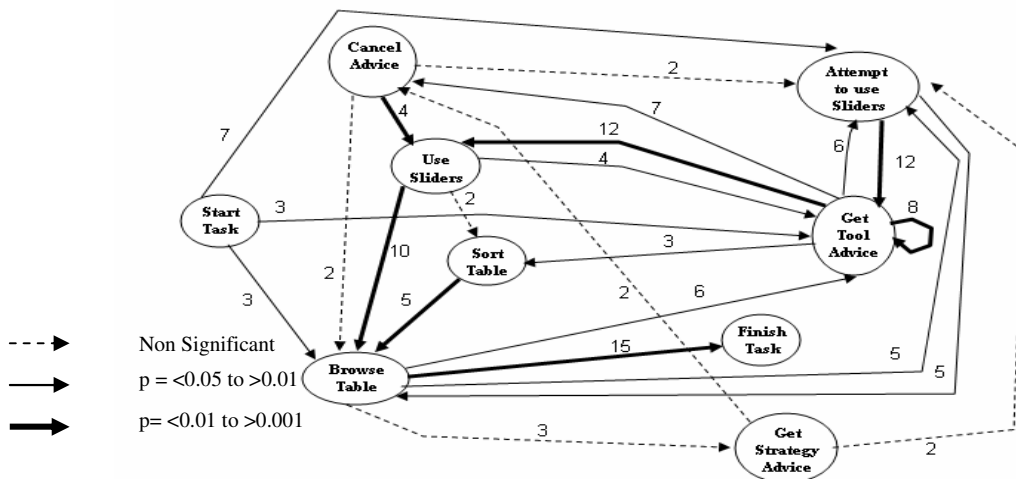
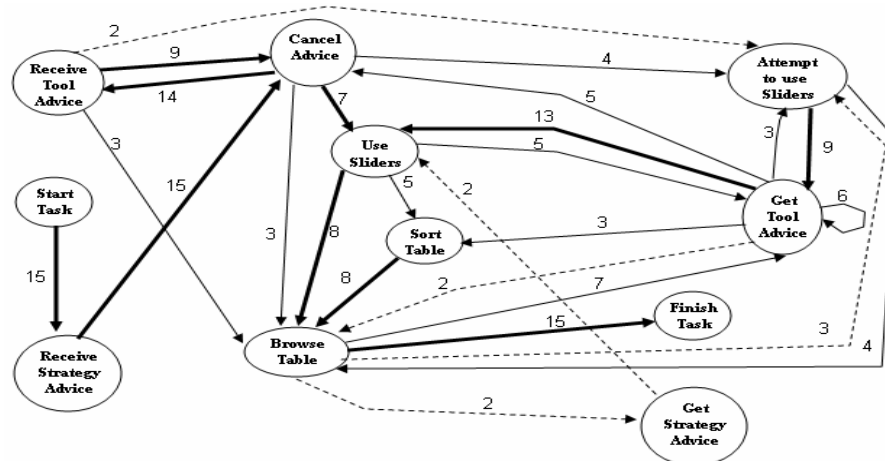
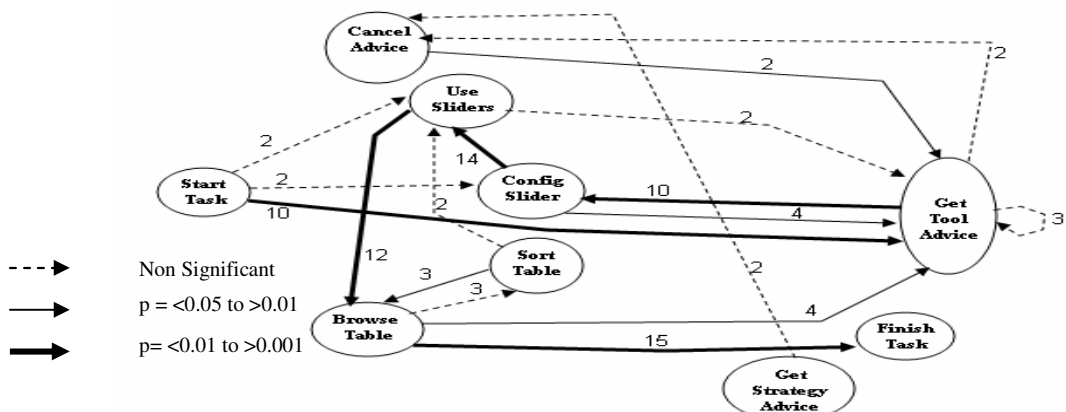


Figure 7: Behavior Network for Subjects in Task 1 – User Initiative



**Figure 8: Behavior Network for Subjects in Task 1 – System Initiative**

In Task 2 the subjects were required to learn how to configure the slider settings to effectively finish the task. In the user initiative condition, depicted in Figure 9, most subjects accessed the advice before attempting to configure the tool. This may be due to a learning effect after Task 1, coupled with the greater perceived difficulty of configuring the sliders compared with simply operating the Scatterplot tool. Few errors or advice cancels were observed. In the system initiative condition (see Figure 10) all users cancelled the Strategy advice, but few cancelled the Tool advice, possibly reflecting a more positive perception of the utility of the advice for the more difficult configuration task. They then followed a variety of pathways before re-accessing the advice and proceeding with the configuration-search task.



**Figure 9: Behavior Network for Subjects in Task 2 – User Initiative**

Most subjects cancelled the speech advice before it ended in both tasks (75%). Advice accesses were more frequent in the text-only condition, while cancels were more frequent in the speech conditions ( $p < 0.001$  Binomial test on totals for all

subjects). System initiative was bound to have more cancellations due to the experimental design; however, repeated access did not differ between initiative conditions.

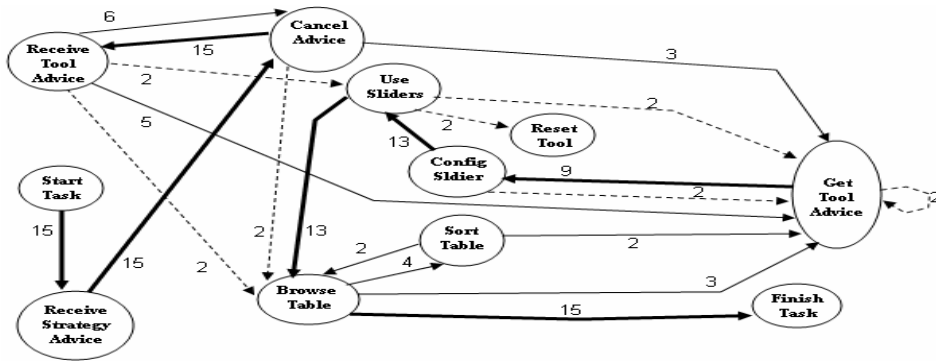


Figure 10: Behavior Network for Subjects in Task 2 – System Initiative

Overall the behavior analysis shows that user initiative was more effective, as might be expected; however, when the task is more difficult the difference between the patterns is less clear cut. Users appear to be more tolerant of system control if they perceive the advice may be important. Interestingly, when we examined the patterns by modality, it appeared that the agent condition had little impact on the pattern of user behaviour.

## 5 Discussion

Previous studies on using ECAs to improve the effectiveness of interaction have produced mixed results, with some showing positive effects [2, 4, 7, 11] while others showed no improvement over text-based advice [2, 27]. This study has added evidence to the contention that human-like agents are not always effective, suggesting that ‘the computer as a social actor’ paradigm [28] may not be influential in certain conditions when usability errors may override the persuasive potential of human-like avatars. Whereas previous evaluations of ECAs have measured users’ attitudes and perceptions such as social presence and flow [27], our study on task-related advice demonstrated that the agent was worse than text-only presentation in both performance and satisfaction measures. Berry et al. [2] also found no advantage for an ECA in a persuasion task, and that text-only was more effective for comprehensibility of advice. In our study, the subjects’ comments and preference for the ECA condition indicated that the agent may have had some positive or ‘social actor’ effect; but when usability concerns became critical, as in the agent-speech condition, we conjecture that the positive social presence of the agent is over-ridden by the users’ adverse reaction to the speech dialogue.

One reason for the ECA’s poor performance may lie in design of the speech dialogue, since users can have an adverse reaction to inappropriate artificial voices [18]. However, experimental studies and speech dialogue principles [18, 19] suggest

that users are tolerant of poor quality TTS speech [20]) and that male voices carry more authority for delivering advice [18, 19], which supports our choice of avatar. We suspect that lack of user control over speech proved to be the major cause of dissatisfaction. This suggests that presence of ECAs alone may be insufficient for delivering task-related help and that careful design of an appropriate dialogue is critical. However, the dialogue in our study was limited, and studies with scripted ECAs have shown that addition of social chat to task-oriented dialogue can improve trust in agents [3]; furthermore, Reeves and Nass [28] demonstrated that an agent's social presence is a function of its on-screen size. The agent in our study was in a small window.

Most users cancelled the advice, especially when it was spoken, and preferred the text advice because they could read it at their own pace. This agrees with the design principle of user control over the pace of media delivery [34] and ISO standards [14, 34]; however, guidelines [16] are less forthcoming about the acceptable duration of speech segments. Our results indicate that simple stop controls, which we implemented, may not prevent user dissatisfaction, and the duration of speech segments should be short, possibly less than 10 seconds. The default assumption of using speaking ECAs to deliver task-related advice [7, 25] may have to be reconsidered. Speech may be used more profitably for giving background information and attracting the users' attention with the ECA, while detailed advice is delivered by text. This follows previous recommendations for partitioning topics in explanatory dialogues and matching the delivery modality to different types of information [26].

The decision strategy advice tended to be ignored by most subjects, probably because it was not relevant to the task in hand. However, the Tool advice was perceived as effective and it was accessed particularly after users had experienced difficulties. Furthermore, the Tool advice was accessed more frequently in Task 2, indicating that users' predisposition to help may be susceptible to experience and perceived difficulty of the task. This agrees with previous findings that users tend to access help and explanation only when they perceive the task to be difficult; otherwise, trial and error is the usual strategy [5, 13]. In our future work we will improve the advisor's presence and the quality of the advice itself to make the ECA more conversational since dialogue and feedback is known to improve the acceptability of advisor systems [33].

## References

- [1] J. Aberg and N. Shahmeri, "An Empirical Study of Human Web Assistants: Implications for User Support in Web Information Systems," *Conference on Human Factors in Computing Systems*, 2001, New York: ACM Press.
- [2] D.C. Berry, L.T. Butler and F. DeRossis, "Evaluating a Realistic Agent in an Advice Giving Task," *International Journal of Human-Computer Studies*, vol. 63, 304-327, 2005.
- [3] T. Bickmore and J. Cassell, "Relational Agents: A Model and Implementation of Building User Trust," *Conference on Human Factors in Computing Systems*, 2001, 396-403, New York: ACM Press.
- [4] T. Bickmore and J. Cassell, "Social Dialogue with Embodied Conversational Agents," in *Natural, Intelligent and Effective Interaction with Multimodal*

- Dialogue Systems*, New York: Kluwer Academic, 2004.
- [5] J.M. Carroll and M.B. Rosson, "Paradox of the Active User," in *Interfacing Thought: Cognitive Aspects of Human-Computer Interaction*, Cambridge MA: MIT Press, 1987.
  - [6] J. Cassell, C. Pelachaud, N.I. Badler, M. Steedman, B. Achorn, T. Becket, B. Douville, S. Prevost and M. Stone, "Animated Conversation: Rule-Based Generation of Facial Expression, Gesture and Spoken Intonation for Multiple Conversational Agents," *Computer Graphics Annual Conference Series (SIGGRAPH-94)*, 1994, 413-420, New York: ACM Press.
  - [7] J. Cassell, J. Sullivan, S. Prevost and E. Churchill, *Embodied Conversational Agents*, Cambridge MA: MIT Press, 2000.
  - [8] C. Darves, S. Oviatt and R. Coulston, "The Impact of Auditory Embodiment on Animated Character Design," *International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2002)*, 2002, Embodied Agents Workshop.
  - [9] A. De Angeli, G.I. Johnson and L. Coventry, "The Unfriendly User: Exploring Social Reactions to Chatterbots," *International Conference on Affective Human Factors Design*, 2001, London: Asean Academic Press.
  - [10] A. DeAngeli, "To the Rescue of a Lost Identity: Social Perception in Human Chatterbot Interaction," *Virtual Agents Symposium*, 2005, 7-14, .
  - [11] D.M. Dehn and S. VanMulken, "The Impact of Animated Interface Agents: A Review of Empirical Research," *International Journal of Human-Computer Studies*, vol. 52, 1-22, 2000.
  - [12] B.J. Fogg, *Persuasive Technology: Using Computers to Change What We Think and Do*, San Francisco: Morgan Kaufmann, 2003.
  - [13] S. Gregor, "Explanation From Knowledge-Based Systems and Cooperative Problem Solving: An Empirical Study," *International Journal of Human-Computer Interaction*, vol. 54, 81-106, 2003.
  - [14] ISO, *ISO 14915 Multimedia User Interface Design Software Ergonomic Requirements, Part 1: Introduction and Framework; Part 3: Media Combination and Selection*: International Standards Organisation, 1998.
  - [15] K. Keeling, S. Beatty, P. McGoldrick and L. Macaulay, "Face Value? Customer Views of Appropriate Formats for Embedded Conversational Agents (ECAs) in Online Retailing," *37th Hawaii International Conference on System Sciences*, 2004, Los Alamitos CA: IEEE Computer Society Press.
  - [16] S. Luz and N.O. Bernsen, "A Tool for Interactive Advice on the Use of Speech in Multimodal Systems," *Journal of VLSI Signal Processing*, vol. 29, 129-137, 2001.
  - [17] R. Moreno and R.E. Mayer, "A Learner-Centered Approach to Multimedia Explanations: Deriving Instructional Design Principles From Cognitive Theory," *Interactive Multimedia Electronic Journal of Computer Enhanced Learning*, 2000.
  - [18] C. Nass and S. Brave, *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*, Cambridge MA: MIT Press, 2005.
  - [19] C. Nass and L. Gong, "Speech Interfaces From an Evolutionary Perspective," *Communications of the ACM*, vol. 43, 37-43, 2000.
  - [20] C. Nass and K.M. Lee, "Does Computer-Generated Manifest Personality? An

- Experimental Test of Similarity-Attraction,” *Conference on Human Factors in Computing Systems*, 2000, 49-57, New York: ACM Press.
- [21] S.L. Oviatt and B. Adams, “Designing and Evaluating Conversational Interfaces with Animated Characters,” in *Embodied Conversational Agents*, Cambridge MA: MIT Press, 2000.
- [22] S.L. Oviatt, R. Coulston and R. Lunsford, “When Do We Act Multimodally? Cognitive Load and Multimodal Communication Patterns,” *Sixth International Conference on Multimodal Interfaces (ICMI 2004)*, 2004, New York: ACM Press.
- [23] S.L. Oviatt, R. Coulston, S. Tomko, B. Xiao and R. Lunsford, “Toward a Theory of Organized Multimodal Integration Patterns During Human-Computer Interaction,” *International Conference on Multimodal Interfaces*, 2003, New York: ACM Press.
- [24] J.W. Payne, J.R. Bettman and E.J. Johnson, *The Adaptive Decision Maker*, Cambridge: Cambridge University Press, 1993.
- [25] C. Pelachaud and M. Bilvi, “Computational Model of Believable Conversational Agents,” in *Communication in MAS: Background, Current Trends and Future*, Berlin: Springer-Verlag, 2003.
- [26] H.C. Purchase and J. Worrill, “An Empirical Study of on-Line Help Design: Features and Principles,” *International Journal of Human-Computer Studies*, vol. 56, 539-567, 2002.
- [27] L. Qiu and I. Benbasat, “The Effects of Text-to-Speech Voice and 3D Avatars on Consumer Trust in the Design of 'Live Help' Interface of Electronic Commerce,” *International Journal of Human-Computer Interaction*, vol. 19, 75-94, 2005.
- [28] B. Reeves and C. Nass, *The Media Equation: How People Treat Computers, Television and New Media Like Real People and Places*, Stanford CA/Cambridge: CLSI/Cambridge University Press, 1996.
- [29] R. Rickenberg and B. Reeves, “The Effects of Animated Characters on Anxiety, Task Performance, and Evaluations of User Interfaces,” *Conference on Human Factors in Computing Systems*, 2000, 49-56, New York: ACM Press.
- [30] J. Riegelsberger, M.A. Sasse and J.D. McCarthy, “Rich Media, Poor Judgement? A Study of Media Effects on Users' Trust in Expertise,” *HCI 2005*, 2005, Berlin: Springer.
- [31] T. Rist, E. André and S. Baldes, “A Flexible Platform for Building Applications with Life-Like Characters,” *International Conference on Intelligent User Interfaces*, 2003, New York: ACM Press.
- [32] B. Shneiderman, “Dynamic Queries for Visual Information Seeking,” in *Readings in Information Visualization: Using Vision to Think*, San Francisco: Morgan Kaufmann, 1999.
- [33] Silveira M.S, de Souza, C.S. & Barbosa S.D.J, Semiotic Engineering Contributions for Designing on line Help Systems. In *Proceedings of HCI International*, San Francisco CA, Eds Smith M., Salvendy G. and Koubek R, pp31-38, Elsevier, 1997
- [34] A.G. Sutcliffe, S. Kurniawan and J. Shin, ., “A Method and Advisor Tool for Multimedia User Interface Design,” *International Journal of Human-Computer Studies*, vol. 64, 375-392, 2006.