

Exploring Persuasive Potential of Embodied Conversational Agents Utilizing *Synthetic Embodied Conversational Agents*

John Shearer, Patrick Olivier, Marco De Boni, and Robert Hurling

Culture Lab – Newcastle, Grand Assembly Rooms, King’s Walk, Newcastle University,
Newcastle upon Tyne, NE 1 7RU, UK

john.shearer@ncl.ac.uk, p.l.olivier@ncl.ac.uk

Unilever Corporate Research, Unilever Colworth Sharnbrook, Bedford, MK44 1LQ, UK
marco.de-boni@unilever.com, bob.hurling@unilever.com

Abstract. This study presents *synthetic embodied conversational agents*, and how they can be used to explore the persuasive potential of *real* embodied conversational agents. Utilizing a novel Wizard-of-Oz style approach and a direct measure of behavior change we explore whether ‘ideal’ embodied conversational agents have a similar persuasive impact as real people, and demonstrate the importance of visually perceiving for embodied conversational agents to be persuasive.

Keywords: persuasion, embodied conversational agents, virtual characters.

1 Introduction

The behavior of present-day embodied conversational agents (ECAs) is limited compared to real humans, especially with respect to non-verbal behavior. Previous research indicates that ECAs have social influence [1], with computer interfaces (such as ECAs) treated as social actors [2]. A *synthetic* ECA appears to be a real (computer generated) ECA, but in fact is simply video and audio transmission of a real human transformed giving the appearance of an ECA –the behavior is that of a real human, resolving the behavioral limitations of present-day ECAs. We consider ‘persuasion’ to mean the change of interactant B’s behavior caused by interactant A. Significant previous work exists on persuasion and social influence [1, 3, 4] using self-reports of attitudes/beliefs, but little work has measured behavior change directly, which is the focus of this study. Present-day ECAs have demonstrated social influence [1, 2, 5].

2 Implementation of Synthetic Embodied Conversational Agent

It is not presently possible to evaluate the persuasive effect of ECAs with behavioral quality approaching that of real humans, due to them having only limited behavioral fidelity. Synthetic ECAs bypass the behavioral quality limitation of present ECAs. Synthetic ECAs use a real human (a wizard) for the behavioral functionality

implemented either by driving a real ECA from motion capture and speech recognition of the wizard, or by transforming video and audio of the wizard. We use the video/audio transformation approach using only commodity hardware.

Present day ECAs appear to be computer generated. This is not due the lack of technology, but merely that it is not presently used in these agents. Photo-realistic ECAs have a natural advantage in terms of persuasive potential, but a photo-realistic synthetic ECA driven by a real person would be indistinguishable from a real human and would therefore be inappropriate for this study. The synthetic ECA must appear to be computer generated to support the belief that it is a real ECA. Furthermore, it was important to verify that people believed the synthetic ECA to be a real ECA, not a synthesized one, so as a precursor to the study on persuasion we ran a study using eye-tracking and questionnaires to determine differences in perception of the synthetic ECA and beliefs about it. The results (yet to be published) conclude that subjects believed the synthetic ECA to be a real ECA.



Audio was transformed using commercial voice transformation software MorphVox [6] and was synchronized with the video which was transformed by ‘cartoonising’ each frame using a custom real-time algorithm based on cartoonising filter in the GIMP[7], implementing in EyesWeb [8]. Previous work cartoonising video streams has been for augmented reality[9] and for creating differing video styles[10].

Fig. 1. Synthetic Embodied Conversational Agent

3 Persuasive Effect of Synthetic Embodied Conversational Agents

Persuasive effect was measured directly by giving each subject the opportunity to donate money (from zero to £10 in 50p increments) from their payment to a single specific charity after interacting with the synthetic ECA presenting information on that charity. Measuring behavior change for *each* subject is impossible. Asking subjects to donate before hand or asking how much they *would* donate would inevitably influence the later donation. But, behavior change can be measured over a group of subjects. Subjects were randomly assigned to one of four conditions:

1. *human* – subjects saw real video and audio of wizard; wizard saw and heard the subject – control condition.
2. *synthetic ECA with video* – subjects saw transformed video and audio of the wizard; wizard saw and heard the subject – ECA with vision condition.
3. *synthetic ECA without video* – subjects saw transformed video and audio of the wizard; wizard only heard the subject – ECA without vision condition.
4. *audio only* – subjects only heard real audio of the wizard; wizard could only hear the subject – telephone style condition.

The human wizard was a female in all conditions, appeared the same throughout the study and believed she was engaged only in a video chat. Both male and female wizards would allow measuring gender impact, but the focus was not on gender differences, and it requires more subjects. A variety of previous studies [3, 4] have investigated gender effects and our conclusions cannot be validly applied to males wizards. Each experiment was self-guided with no interaction with the experimenter.

We predicted that subjects would donate most under the human condition, with reduced donations under the other three conditions, also that the synthetic ECA *with* video condition would have more donated than *without* video – reflecting the believed importance of seeing the other interactant. Pre-interaction and post-interaction questionnaires were included for completeness and verification, but did not include measures directly relevant to the study – merely concerning the nature of the interaction; the subjects’ beliefs as to the computer-generated or human nature of the character; and a personality test (Myer’s Briggs).

4 Results

Statistical analysis showed non-normal distribution of donations making both ANOVA and t-Tests invalid, so a non-parametric Kruskal-Wallis test was performed. The test statistic (Chi squared) was 7.754, equating to a probability of 0.051. Strictly this should be <0.05 to go on to further comparisons of the means – but 0.051 is near enough. A Wilcoxon test was run to compare non-normally distributed means, finding the probability for the difference between synthetic ECA *with* and *without* video to be $p=0.003$. A Bonferroni correction was applied due to multiple paired comparisons giving a significance criteria as $0.05/6 = 0.0083$.

Table 1. Amount donated to charity vs condition – 76 subjects (44 female; 32 male)

		Mean	N	Std. Dev.
condition	synthetic ECA <i>with</i> video	£3.50	21	£3.17
	audio only	£2.94	18	£3.67
	Human condition	£2.47	19	£3.58
	synthetic ECA <i>without</i> video	£1.36	18	£1.62
Total		£2.61	76	£3.17

No significant difference was found between any other pairs of conditions, so we cannot support the hypothesis that a synthetic ECA is less persuasive than a real human, and the large variances preclude concluding that they are equally persuasive. Additionally, no significant genders differences were found, though previous studies [3, 4] suggests there may be significant gender differences. With the limited subject numbers these may have been missed

In summary we can conclude that there is a reasonably robust significant difference between the female synthetic ECA *with* and *without* video, and therefore that female ECAs with vision have a greater persuasive potential than female ECAs without vision.

5 Discussion

Results indicate similarity between how persuasive female *synthetic* ECAs are as compared to humans, but we cannot draw hard conclusions about that similarity. Results show that it is important for female synthetic ECAs to be able see their interactants. As subjects were unaware that the synthetic ECA was not a real ECA, we conclude that seeing their interactants is important for persuasion in *real* ECAs.

Results are limited the interactions within a webcam interface context and may not generalize to more realistic environments and do not apply to agents that are more pro-actively persuasive. The quantization of monies given to subjects and exact denominations may have affected the amounts donated. The large variances require a larger numbers of subjects for more conclusive results.

It might be expected that with audio only condition would have the lowest donation, but this wasn't the case. A variety of theories could be presented as to why the audio condition had a similar persuasive effect as the conditions with video transmission in both directions. Perhaps, when unable to see an agent the subjects no longer expects appropriate non-verbal behavior, putting the whole interaction under different assumptions. In contrast, in the synthetic ECA *without* video condition we could perhaps presume that subjects expect non-verbal behavior in response to their own non-verbal behavior, but it cannot be provided.

Overall, we conclude that it is important for female ECAs to be able to see their interactants to effect behavior change. This demonstrates the utility of future work using simple vision methods (and other modalities) to inform the behavior of ECAs.

References

1. Bailenson, J.N., Yee, N.: Digital Chameleons. Psychological Science (2005)
2. Reeves, B., Nass, C.I.: The media equation: how people treat computers, television, and new media like real people and places. Cambridge University Press, New York (1996)
3. Baylor, A.L.: Interface Agents as Social Models: The Impact of Appearance on Females' Attitude Toward Engineering. In: CHI, ACM Press, Montreal, Canada (2006)
4. Blascovich, J.: A theoretical model of social influence for increasing the utility of collaborative virtual environments. In: Collaborative virtual environments, pp. 25–30. ACM Press, Bonn, Germany (2002)
5. Fogg, B.J.: Persuasive Computers: Perspectives and Research Directions. CHI 98 , 225–232 (1998)
6. Screaming Bee: MorphVox. Screaming Bee LLC (2006)
7. Kimball, S., Mattis, P.: GNU Image Manipulation Program (2006)
8. Camurri, A., Coletta, P., Massari, A., Mazarino, B., Peri, M., Ricchetti, M., Ricci, A., Volpe, G.: Toward real-time multimodal processing: EyesWeb 4.0. AISB 2004 Convention: Motion, Emotion and Cognition, Leeds, UK (2004)
9. Fischer, J., Bartz, D.: Real-time Cartoon-like Stylization of AR Video Streams on the GPU. In: Wilhelm Schickard Institute for Computer Science, University of Tübingen (2005)
10. Wang, J., Xu, Y., Shum, H.-Y., Cohen, M.F.: Video Tooning. In: SIGGRAPH, 23th edn., pp. 574–583. ACM, Los Angeles, CA, USA