

Wizard of Oz Experiments and Companion Dialogues

Jay Bradley

David Benyon
Edinburgh Napier University
School of Computing, Edinburgh, UK

Oli Mival

Nick Webb
ILS Institute
SUNY Albany, Albany, USA
nwebb@albany.edu

j.bradley@napier.ac.uk

d.benyon@napier.ac.uk

o.mival@napier.ac.uk

Novel speech systems such as the conversational agents being developed by the Companions Project (www.companions-project.org) can be simulated using the Wizard of Oz methodology. In this approach technologies that are not yet ready for testing by people are replaced by a human, both for prototyping and collecting additional dialogue data. In the case of Companions we want to observe what it would be like for people interact with a fully functional embodied conversational agent (ECA) and to collect samples of typical dialogue in order to explore, evaluate and model dialogue strategies. One controversial aspect of the Wizard of Oz approach is whether people should be aware that they are interacting with a simulation or whether they should be “fooled” into thinking they are interacting with a real system. Clearly there are ethical issues involved in fooling people, but some argue that unless the participant believes the simulation to be real, the results of any experimentation will not be applicable to the real situation. Over the course of several previous Wizard of Oz experiments our observations suggest that the dialogues produced do not significantly differ whether the participants know that the technology is faked or not. This hypothesis was investigated by collecting dialogues from two groups of participants. One group of participants believed that the Wizard of Oz speech system was in fact a fully computerised prototype and the other group knew that they would be talking through the interface to a hidden person (the wizard). The dialogues were analysed for differences attributable to the participants’ beliefs about the system. This analysis was undertaken by an independent “blind” reviewer, a dialogue expert who attempted to allocate participants to one group or the other. His guess was wrong for four out of the six participants. Thus it appears that whether people believe they are interacting with a real system or not does not effect the dialogues and other factors, for example the personality of the person engaged in a dialogue with an ECA, are more important.

Wizard of Oz, Companions, Dialogue Systems.

1. INTRODUCTION

Wizard of Oz experiments [7] have been used over the past few years by the authors to inform the design of novel conversational agents [1], to provide some idea of a baseline performance for such agents [2] and their evaluation [13]. In a Wizard of Oz experiment participants interact with a computer system which they believe to be fully autonomous but which has some part of its computation replaced by an unseen human. Wizard of Oz methods are used to test novel technologies that either do not yet exist or are costly in time or money to run. The unique aspect of Wizard of Oz experiments is that they should provoke realistic behaviours from the participant. This is as opposed to other forms of prototype testing such as video prototyping or paper based walk-throughs. The participant in an experiment can be made aware that the system is being controlled by a human, in which case it is more similar to role-playing. This can be useful when adopting a human centred design approach where the participant is invited in to be one of the design team [9], often

when the participant is a stakeholder in the design and needs to be kept aware of the design process. Otherwise, Wizard of Oz experiments can be run strictly. That is when the participant believes that the system under test is a functioning autonomous prototype. In this case one can rely more on the fact that the participant is demonstrating genuine behaviour during interaction, genuine in the sense that this is how they would act were the system real. In such circumstances this genuine behaviour is the principle reason to take the Wizard of Oz approach. However, running participants through experiments without full disclosure puts researchers in a difficult ethical situation. The ethics can be overcome by providing full disclosure at the end of the experiment, but people may still object to being duped. It may be enough to say that full disclosure was given and to use clever descriptions when explaining the experimental set-up to the participants, but this still involves an element of deception. The researcher can gloss over the exact workings of the prototype leaving the participant to make assumptions about the underlying technology and rely on omission

being a lesser evil than lying. However, this does still not remove the ethical issues.

In order to explore the role of Wizard of Oz further in this context we undertook a study of six participants using a Wizard of Oz implementation of a Companion, three of whom were told they were interacting with a simulation and three of them were not.

We wanted to focus on the nature of the dialogues that ensued. We know that people are happy talking to computers. This was demonstrated as long ago as the 1970s in Weizenbaum's famous experiments with Eliza. What we do not know is what good dialogues between people and ECAs should be like. From the perspective of much work in dialogue systems and natural language processing, good dialogues are effective when they achieve some task in the shortest time, or least number of words. Others such as Traum [12] see effectiveness as achieving the best outcome. But it is not clear what a good dialogue is for a conversational agent in an interaction that is not task based.

The Companions Project is aiming to develop a new kind of dialogue system, one that goes beyond task-oriented dialogue systems and is capable of intelligence, persistence and personalisation through a history of interaction [15]. In this paper we are simulating a Companions Project prototype dialogue system called PhotoPal [8]. PhotoPal discusses a selection of photographs with a person with the aim of soliciting information about the photographs and engaging the person in a natural conversation, much as two people might have a conversation about photos. We have chosen a Wizard of Oz approach to collect sample dialogues and interactions with the companion prototype systems in order to inform the design and development of companionable technologies. Also, we can investigate and assess novel evaluation metrics on what is a complex interaction. Evaluating companionable systems cannot rely on known dialogue metrics such as task completion and user satisfaction. The goal of the dialogue is to build and maintain a relationship or to affect the persons' mood and hence we need a collection of realistic human-machine interactions to determine the impact of utterances over the course of an interaction. In this paper, however, we are only interested in the difference in dialogues if the participants believe they are talking to a machine or to a human. After running many Wizard of Oz experiments simulating companionable speech systems, the authors believe that whether or not the participant knows that a human is faking the dialogue management does not significantly affect the dialogue spoken by the participant. We believe that Wizard of Oz experiments can be run under full disclosure without affecting the dialogues collected. Obviously, there will be a difference in the participant's belief but this does not necessarily translate into observable differences in a participant's dialogue.

To investigate this hypothesis we ran six Wizard of Oz experiments simulating the Companions Project's PhotoPal prototype. Three of the participants would take part under full disclosure and the other three participants would not be told that all processing would be done by a human behind the interface. The collected dialogues were analysed by a "blind" reviewer, a dialogue expert, to determine if the participants could be assigned to either group correctly and to suggest what features of the dialogue were different. The participants were interviewed after the experiments and asked specifically about what difference it would have made to their behaviour if they had been in the other group.

2. EXPERIMENTATION

Six participants from two groups talked to PhotoPal. Participants from both groups were given a similar description of the experiment and signed the same consent form. After reading through an introductory information sheet about PhotoPal and the experiment they were about to take part in, three of the participants were made aware that they would be talking to a human. All the participants were told that the experiments were being run in order to collect and analyse their conversation with PhotoPal. The three participants aware of the human wizard's involvement were told that the human would be playing the role of PhotoPal and approximately simulating the level of the Companion Project's prototypes. The other three participants were told that they would be talking to a companion prototype but were not told about the role of the human wizard. Each participant was asked to talk in whatever way they felt was appropriate and not to feel as if they needed to speak in a "computer like" fashion. Each participant supplied four or five personal photos of their own choosing. Subjects included pets, friends, holiday snaps, weddings and family as well as more esoteric subjects such as white-water kayaking and music band members.

The wizard attempted to roughly replicate the capabilities of current Companion Project prototypes except for current constraints on speech recognition and the limited domain knowledge. Often the Companion prototypes fail completely due to poor speech recognition or use of language outside of the dialogue management domain. The person playing the wizard role was the same throughout the experiments and aimed to remain consistent in his behaviour throughout the six experiments. The wizard was aware of each participant's belief (that they were talking to a human or machine) and, in hindsight, this could maybe have been avoided so as not to introduce any possible variability in the wizard's behaviour and hence participant dialogue due to that knowledge.

3. WIZARD OF OZ ARCHITECTURE

Our Wizard of Oz architecture has been in ongoing development in order to support the evaluation and design of companionable dialogue systems [1]. This consists of two web-browser based interfaces. The participant and the wizard are in different locations and communicate through the interfaces over the Internet. The wizard replaces many broad functions of the PhotoPal prototype: real-time speech recognition, natural language understanding, dialogue management and natural language generation. The text-to-speech technology was kept to hide the human wizard's involvement. The participant sees the interface as shown in Figure 1.



Figure 1: The user's interface

This interface shows an avatar, a list of previous system utterances and the photo viewing area. The avatar is an embedded CrazyTalk web-player (<http://www.reallusion.com/crazytalk/>) which lip-syncs along with each system utterance. The photo viewer is a custom built Adobe Flex (<http://www.adobe.com/products/flex/>) application which shows the participant's photos. The wizard sees the interface as shown in Figure 2. The wizard's interface shows the photo viewing area, a text box for typing text to be spoken by the avatar (on the participant's interface) and a streaming video from the participant's web-camera (not shown in Figure 2). The wizard's speech input box keeps a record of previous utterances and ranks them in terms of popularity of use. When the wizard starts typing the speech box shows a selection of matching previous utterances. This feature of the wizard's interface has made it much easier for the wizard to reply in a reasonable amount of time and also makes it simple for the wizard to replicate sentences often in order to make them sound more "computer like." The more the wizard's interface is used the more useful the library of previous utterances becomes.

Participant	Actual Group	Reviewer guess	Utterances	Average words	Average unique words
One	Strict	Non-strict	59	7.5	4.2
Two	Strict	Strict	50	4.5	2.4
Three	Strict	Non-strict	46	26.8	9.6
Four	Non-strict	Non-strict	41	37.7	12
Five	Non-strict	Strict	52	9.8	4.7
Six	Non-strict	Strict	58	7.6	4

Table 1: Basic dialogue metrics

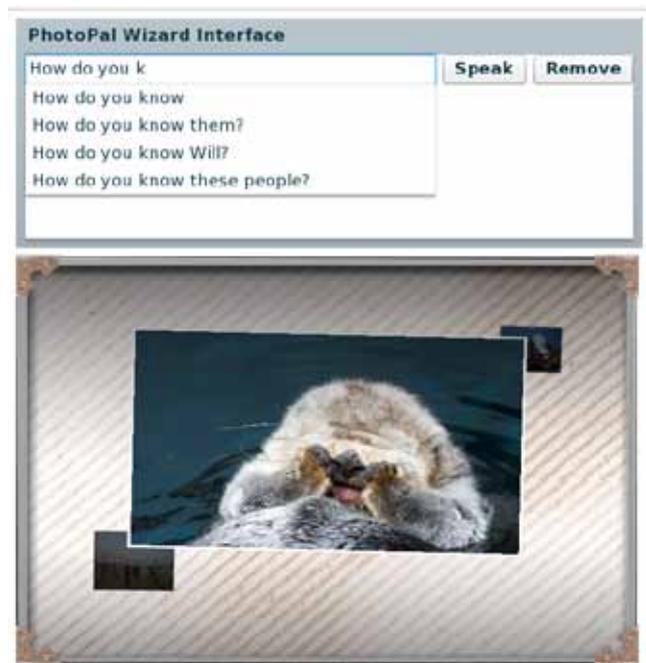


Figure 2: The wizard's interface

The two interfaces are linked by a Red5 server (<http://osflash.org/red5>) which is used to synchronise the state of the photo viewer and the system utterances and also streamed audio and video one way – from the participant to the wizard. Lag time was less than a second and so did not noticeably add to the system utterance delay. This meant that the wizard could accurately choose both how long to wait to reply as if to simulate processing time and also to interrupt the user to see how they reacted to these annoyances.

4. ANALYSIS

On average the dialogues contained 51 user utterances, had a vocabulary size (unique words) of 296 words and an average utterance length of 15.7 words. Some basic metrics of the dialogues are shown in Table 1. The reviewer was shown the six dialogues and asked to assign each dialogue

to either group along with some reasoning for the choice. He was not given audio or video of the participant's session so the analysis was based only on the transcribed dialogue.

In general, the reviewer (a computer scientist with significant experience in building spoken language dialogue systems [14], [5]) looked at the "style" of the interaction. Dialogues with machines are often shorter (in utterance length and overall length) than comparable conversations with humans, and invoke significantly less use of dysfluencies (such as "er" and "um") and politeness (apologising, for example). Utterances tend to be more direct, with little explanation or follow-up. These are observed behaviours when people interact with artificial agents, caused in part by the common sense belief that it is not worth adding additional information, as of course the computer cannot understand. There are many counter-intuitive examples in real dialogue corpora; for example in the OASIS call steering corpus collected by British Telecom [4], even though the artificial agent clearly identifies itself as a computer at the outset of the interaction, callers (real users of British Telecom's directory assistance program) would still occasionally engage in social conversation. However, for the most part, it is readily accepted that interactions with machines are shorter, more focused with the minimum of social conventions.

Below is the reviewer's assessment of the six transcribed dialogues in turn, the group he guessed the participant was in, his reasoning, the actual group and some discussion of the dialogue and participant interview.

4.1 Participant One

- inexact corrections e.g. "That's almost correct."
- follow up utterances e.g. "But not quite."
- lots of politeness;
 - back-channels e.g.
 - System: "Thank you."
 - User: "That's okay."
 - System: "Where was it?"
 - User: "Where? I ..."
- longer turns/ many separate utterances in one turn e.g. "Where? I can't remember exactly. On [name]'s mother's farm. In Buxton."
- asked questions e.g. "Did you understand that? Am I talking too much?" and "Are you sure that's okay?"
- implicit confirmation e.g.
 - User: "[Name A]."
 - System: "[Name A]?"
 - User: "Yes. The groom."
 - System: "In the previous photo?"
 - User: "And the photo before that."
- engaged in the dialogue e.g.
 - System: "Bath?"

- User: "In the bath? Maybe, although he has his clothes on so I think it's unlikely."
- testing behaviour e.g. "He's my mother." and
 - User: "That's right, I'm playing the viola."
 - System: "Thank you."
 - User: "And acting the fool."
- humour e.g.
 - User: "She's my girlfriend."
 - System: "Sorry."
 - User: "That's quite alright. I'm quite glad of it actually."

The reviewer guessed that participant One was in the full disclosure group but actually, participant One was unaware he was talking to a human.

4.2 Participant Two

- a small amount of politeness e.g. "You're welcome."
- short, goal directed utterances;
- no initiative.

The reviewer guessed participant Two believed that he was talking to a machine and this was correct.

4.1.3 Participant Three

Participant Three was the only non-native English speaker although her level of English was very good. She was also the only female participant.

- a significant number of overloaded answers. i.e. utterances with more information than asked for e.g.
 - System: "Are there any people in the photo?"
 - User: "Erm. I'm afraid not it's just the monument."
- very long utterances;
- lots of dysfluency;
- self-talk e.g. "...how do you say that..." This is common with non-native speakers but, possibly not when speaking to a machine;
- politeness;
- makes explicit use of personal pronoun i.e. "you";
- feedback e.g.
 - "That's fun." and,
 - "You're really clever aren't you? Yes it is. Although it took you a little bit of time to take this you know in the processing but yeah it is."
- asked questions e.g. "Do you want me to choose?"
- surrendered initiative explicitly e.g. "Okay you choose."
- used back-channels e.g.
 - System: "Okay."
 - User: "Okay."

Being a non-native English speaker might explain some of participant Three's dialogue characteristics. The reviewer guessed participant Three was in the full disclosure group but participant Three believed that she was speaking to a fully autonomous prototype.

4.4 Participant Four

- used politeness;
- asked clarification questions e.g.
 - “Select my name?” and,
 - “Do you want me to select one of them?”
- made explicit use of personal pronoun i.e. “you”
- very long utterances;
- lots of dysfluency;
- significant number of overloaded answers;
- self talk/narration e.g. “I can’t find the picture of the sheep.”
- took initiative e.g. “Do you want me to talk about it?”

The reviewer thought that participant Four was in the full disclosure group which was correct. i.e. Participant Four knew he was talking to a human.

4.5 Participant Five

- a small amount of politeness e.g. “You’re welcome.”
- apologised;
- gave some limited additional information e.g.
 - System: “Please tell me more.”
 - User: “Okay we, my girlfriend and I got [name of pet] whenever we moved into our flat back in er June two thousand and eight, sorry two thousand and seven er he’s about four years old. And he’s very playful.”
- mostly answered question directly.

The reviewer thought that participant Five believed that he was talking to a machine but he was in fact aware that he was talking to the human wizard.

4.6 Participant Six

- used absolutely no initiative;
- always answered the question;
- no politeness;
- very short, goal directed utterances;
- very little additional information e.g.
 - System: “I see. Please go on.”
 - User: “We were out for his birthday and we were all dressed up as seventies cops. Just for a laugh.”

The reviewer thought that participant Six was not aware that he was speaking to the human wizard but actually, he was aware.

The reviewer correctly assigned only two of the six participants to the correct group. Participants Five and Six essentially role-played as if they were talking to a computer and this meant that the reviewer thought that they were unaware they were talking to a human. This was primarily because of their relative lack of politeness and direct answering. Participant Two also spoke similarly but did believe that he was talking to a machine so this was as expected. Participants Two, Five and Six each spoke similar

enough dialogue that they were each categorised as believing they were speaking to a machine.

The dialogue from participants One, Three and Four all showed aspects of human-human conversation such as dysfluency (the use of, “erm” for example), relative politeness, asking of questions and more. Participant Four was told that he was talking to a human and this obviously showed through in his dialogue but participants One and Three believed they were talking to a machine. Participants One and Three did not speak to the computer in a “computer-like” way. They “chatted” to the system in a reasonably informal manner. There is no doubt that the Wizard of Oz experiment facilitated this conversation as the existing Companion prototypes have quickly discouraged users from talking naturally or at length due to their poor speech recognition, limited dialogue domain and repetitive speech. This is to be expected of speech systems when compared to humans of course. Participant Three said during her interview that she would have definitely talked differently if she had known that she was speaking to a human behind the interface. This is interesting as she was classified by the reviewer as one of the participants who knew she was talking to a human. Participant One complained about being interrupted by the system and said that the poor quality avatar (in his opinion) made it easier to believe that he was talking to a computer. Participant One was the only person to purposefully attempt to break the system with extraneous utterances, contradictions, humour and inexact statements yet he was certainly engaged with the conversation. Perhaps this has something to do with the presence of the avatar (discussed later). These three participants (One, Three and Four) were categorised as knowing they were talking to a human by the external reviewer despite the fact that only one of them knew they were actually talking to a human.

This analysis indicates that there is little or no difference in the dialogue produced between the two conditions. It is though, probably true that there are other factors that play a larger role in determining a participant’s behaviour than whether or not they know that the technology is being faked by a human in Wizard of Oz experiments. These factors may just be masking the smaller effect looked for in these experiments but still maybe one should address these other factors first. One such factor could be the quality of the simulated interaction.

Participants Five and Six seemed from their dialogue to be talking to a computer but they both knew that they were talking to a human. None of the candidates were given explicit advice on how to talk to the system or wizard other than being asked to talk “naturally.” When interviewed, all but one of the participants said that they felt the pause after they spoke and before the avatar started speaking was unnatural and meant the system had to lead the

conversation. The wizard tried to simulate delays due to processing by waiting a number of seconds before responding to a user utterance. The wizard waited longer if the user utterance was longer and waited less time if the user's utterance was short. The participants complained that the avatar was very simplistic and used no paralinguistic devices such as leaning in or opening her mouth prior to interrupting. Participant Two said that talking to an avatar was better than having no avatar but that the avatar should be capable of subtlety. Participant Five said that the experience of talking to the avatar (when knowing that it was controlled by a human) was similar to using instant messaging clients. Participant One said that the avatar would have to be better quality than both what the system said and how the text-to-speech output sounded, so as not to detract from the experience. However, he went on to say that his perceived low-quality of the avatar made the system appear more "dumb" and helped him to believe that he was talking to a machine (he was unaware of the wizard's involvement until after the experiment). The avatar was certainly in uncanny valley territory with its semi-photo-realistic looks but lack of meaningful movements other than lip-syncing. Also the text-to-speech system used lacked any variability in prosody. These hindrances or drawbacks normally associated with computerised speech systems dragged down the level of communication achievable by the human wizard, thus making the speech system more low-fidelity. A low quality delivery of system utterances may be the reason why participants Five and Six treated the human as a machine despite knowing they were talking to the human wizard. Certainly it is likely that lower quality simulations may lead to lower quality or unnatural behaviour from users [10]. Another interesting idea that may significantly affect a user's behaviour is that the presence of the avatar [6] affects the behaviour of users more so than what they are told they are talking to. Presence can be looked at in two parts: place illusion and plausibility illusion [11]. Place illusion is the feeling of being there and comes from research into virtual environments. For the purposes of simulating speech systems, it is more or less given that place illusion will hold because the simulation (using a wizard) is so similar to the real prototype if the wizard is hidden. It is plausibility illusion that is most interesting. Plausibility illusion is, "the illusion that what is apparently happening is really happening (even though you know for sure that it is not)."[11] Given that the external reviewer failed to categorise four of the sample dialogues one could say that the participants all felt that using the PhotoPal Wizard of Oz system was plausible and that their behaviour reflected this. This is even true with participants Five and Six who knew that they were talking to a human but still acted as if they were talking to a machine (according to the reviewer's analysis). Presence is a qualia and it is difficult to

say that the presence of the avatar or system as a whole affected the participants more than what they were told they would be talking to. It may be worth investigating in the future though.

Of course it is obvious that much variation in dialogue will come from the participants themselves. People talk differently to each other and it is probably safe to assume that they will talk differently to computers. Although people interact with current interfaces in similar ways this can be largely attributed to the narrow-band interaction technologies such as the keyboard and the mouse that constrains user choice. A broadening of variation in interaction styles can be seen with more broad-band input technologies such as multi-touch screens and speech systems. Speech can be a very broad input modality and we should expect large variability in the way a user approaches such a novel system.

5. HUMAN-WIZARD DIALOGUE

Dialogue differs depending on the qualities of the interlocutors and raises the question: should one try to build or assess human-computer speech systems by analysing human-human dialogue [3]? Human-human dialogue is simple to find or produce reference examples of. However, "language directed to children is different from language directed at grown-ups" [3] and this is true for other categories of interlocutor such as those speaking a foreign language and presumably, computers. Human-computer dialogue would be ideal but it is difficult (impossible) to record samples of humans talking to the speech system that one is attempting to build. The problem of obtaining such human-computer dialogue is that machines can't hold reasonable conversations with humans primarily because of the restrictions in breadth and depth of reasoning. Human-human conversations are not suitable because that is a human talking to a human so maybe it is possible to investigate the space between human-human and human-computer dialogue. This human-wizard dialogue could be suitable for analysis and/ or learning methods if one can better understand the factors that affect the qualities of the dialogue as talked about above.

Participants One, Two and Three believed they were talking to a computer yet participants One and Three conversed similarly to human-human dialogue. They were affected by the human wizard behind the interface in such a way that their behaviour differed from that of participant Two by being more like expected human-human dialogue instead of human-computer dialogue. Participants Four, Five and Six knew they were talking to a human yet the dialogue from participants Five and Six was more like that expected of human-computer dialogue than the expected human-human dialogue of participant Four. Participants Five and Six were

affected by the experimental set-up in some way so as to change their behaviour. The participants of these experiments experienced a humanisation of the computer or a computerisation of the human wizard. This type of experiment and the human-wizard dialogue produced may be suitable for use in analysis or modelling of companionable speech systems. The quality of simulation, presence of the system and the mindset of individual users will affect the dialogue produced and these features should be explored in the future.

6. ACKNOWLEDGMENTS

This work was funded by the Companions project (www.companionsproject.org) sponsored by the European Commission as part of the Information Society Technologies (IST) programme under EC grant number IST-FP6-034434.

7. REFERENCES

- [1] Bradley, J., Mival, O. and Benyon, D. (2008) A Novel Architecture for Designing by Wizard of Oz. In: Proceedings of CREATE08.
- [2] Bradley, J., Mival, O., and Benyon, D. (2009) Wizard of Oz experiments for Companions. In Proceedings of the 2009 British Computer Society Conference on Human-Computer interaction.
- [3] Dahlbäck, N. & Jönsson, A. (1992) An Empirically Based Computationally Tractable Dialogue Model. Proceedings of the Fourteenth Annual Meeting of The Cognitive Science Society.
- [4] Durston, P.J., Farrell, M., Attwater, D., Allen, J., Kuo, H.J., Afify, M., Fosler-lussier, E. and Lee, C. 2001. OASIS Natural Language Call Steering Trial. In Proceedings of Eurospeech 2001.
- [5] Hardy, H., Biermann, A., Bryce Inouye, R., McKenzie, A., Strzalkowski, T., Ursu, C., Webb, N. and Wu, M. The AMITIES System: Data-Driven Techniques for Automated Dialogue. In *Speech Communication* 48 (3-4), pages 354-373. Elsevier. 2006.
- [6] Heerink, M., Kröse, B., Weilinga, B. and Evers, V. (2009) Measuring the influence of social abilities on acceptance of an interface robot and a screen agent by elderly users. In Proceedings of the 2009 British Computer Society Conference on Human-Computer interaction.
- [7] Kelley, J. F. (1984) An Iterative Design Methodology for User-Friendly Natural Language Office Information Applications. *Transactions on Office Information Systems*, 2(1).
- [8] Mival, O., O'Keefe, B., Bradley, J., Roa-Seiler, N. and Benyon, D. (2008) PhotoPal: Companionship, Sharing and the Digital Echo. In proceedings of the Collocated Social Practices Surrounding Photos Workshop at CHI 2008.
- [9] Molin, L. (2004) Wizard-of-Oz Prototyping for Cooperative Interaction Design of Graphical User Interfaces. In proceedings of Nordichi 2004.
- [10] Pettersson, J. S. and Siponen, J. (2002) Ozlab: A simple demonstration tool for prototyping interactivity. In Proceedings of Nordichi 2002.
- [11] Slater, M. (2009) Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philos. Trans. R. Soc. Lond.* 364, 3549–3557 doi: 10.1098/rstb.2009.0138.
- [12] Traum, D., Robinson, S. and Stephan, J. 2004. Evaluation of multi-party virtual reality dialogue interaction. In Proceedings of Fourth International Conference on Language Resources and Evaluation (LREC2004), Lisbon, Portugal.
- [13] Webb, N., Benyon, D., Bradley, J., Hansen, P. and Mival, O. (2010) Wizard of Oz Experiments for a Companion Dialogue System: Eliciting Companionable Conversation. In Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC), Valletta, Malta.
- [14] Webb, N., Benyon, D., Hansen, P. and Mival, O. Evaluating Human-Machine Conversation for Appropriateness. In proceedings of the 7th International Conference on Language Resources and Evaluation (LREC2010), Valletta, Malta. 2010.
- [15] Wilks, Y. (2006) Artificial Companions as a new kind of interface to the future Internet. Oxford Internet Institute, Research Report 13, October 2006.