Play SMILE Game with ERiSA: A User Study on Game Companions

Andry Chowanda[1,2], Peter Blanchfield[1], Martin Flintham[1], Michel Valstar[1]

[1] School of Computer Science, The University of Nottingham, Nottingham, UK-GB
{psxac6,peter.blanchfield,martin.flintham,michel.valstar}@nottingham.ac.uk
[2] School of Computer Science, Bina Nusantara University, Jakarta, ID

Abstract. This paper describes the evaluation of our fully integrated virtual game companions framework (ERiSA) [4]. We conducted three user studies with different scenarios using two versions of The Smile Game[4] in semi-public and public spaces. In our study, we show that the game companions' personality was successfully perceived by the participants while interacting and playing with the game companions. Topic about the game itself was the most popular topic with total 598 occurrences in our studies. Moreover, facial expressions is the most performed type of attack in the game. Finally, from the large number of video data collected, we aim to automatically learn the interaction rules and additional attack movements.

Keywords: Social Relationship, Social Dialogue, Game-Agents, Interactions

1 Introduction

In the intelligent virtual humans research community, the focus of efforts appears to always be on improving the capabilities of the agents. This is resulting in increasingly complex systems. In particular social interactive agents, be they virtual humans or social robots, have the tendency of becoming increasingly, perhaps even excessively complex systems. This is mainly because these researchers aim at replicating human interactive capabilities, and thus in essence try to emulate all the complexities of the human social brain. To keep this daunting task feasible, this is often done in a reductionist manner, with separate modules responsible for the visual and audio sensing capabilities, dialogue management and mental state management, action planning, speech production and behaviour synthesis. One of an existing framework for virtual game companions is ERiSA (Emotionally Realistic Social Game-Agents) framework [4].

This paper describes the evaluation of our fully integrated virtual game companions framework(ERiSA) [4] by conducting a series of user studies using The Smile Game[4] in semi-public and public spaces. The Smile Game is inspired by the "Don't Smile" game that has two people facing each other and make each other laugh with jokes and funny facial expressions. The first person to smile uncontrollably loses. In the study we also collected initial data for human-agent

interactions. Total of 67 distinct participants (with total of 72 interactions for 3 hours 22 minutes and 31 seconds) interacted with two of SEMAINE virtual humans (Poppy and Spike) [15] [1] with different personalties in two versions of The Smile Game [4]. In this study we present work towards building fully automatic systems for integrated virtual game companions in semi-public and public areas to establish a natural space for social interaction.

Most of the participant (87.5 %) who interacted with both of the virtual game companions perceived Poppy as a nice and friendly "person", while Spike as a rigid, unpleasant "person" and hard to defeat in the game. Moreover, from the first and second study, we found out that topic about news was not a favourable topic. Eighty three percent of the participants turned down the conversation related to news by saying "no good news" and then the virtual game companions had to switch the topic immediately. On the other hand, topic related to the game itself was the most popular in our studies. Interestingly, in the third study most of the participants loved to talk about themselves, with total average duration of 208 seconds spent on this topic.

2   Related Work

It is not easy to evaluate a such complex system. There are several methods to evaluate the system depending on what aspects of the system we want to evaluate. Gratch and Marsella proposed a methodology to evaluate an emotion model for virtual humans by comparing the virtual human's behaviour to human's one. They also implemented the methodology to evaluate their proposed emotion Model [7].

Bickmore designed Tinker, a virtual museum guide with social intelligent [2]. The system was installed in an open space with enormous number of interactions for almost a year. The participants interacted with the virtual guide with multiple-choice touch screen. The researcher evaluated the effect of the virtual guide's social behaviour on the participants' engagement and learning in the museum. While Kopp evaluated their virtual museum's guide by deploying the system in a public museum, and recorded the interactions between the virtual guide and the visitors in log files. The log files then analysed for dialogues between the virtual guide and the visitors [10]. Similarly, Robinson [12] collected and analysed a corpus of favourite topics between a virtual guide and museum visitors.

Wizard of Oz (WoZ) method is also commonly used for evaluating a prototype system [8] as well as for initial data collection for human-computer interactions [5]. Vardoulakis collected conversation data to determine the favourite topics of older adults' with the agent as well as to evaluate their system using WoZ method [17]. Similarly, Sidner et al implements WoZ method for their social agents controlled remotely by an experimenter, to collect what topics are favourable to talk about with the agent [16].

Kelley [9] used WoZ to develop natural language computer system as a personal assistant agent to help business professionals manage their personal cal-

endars. Salber and Coutaz [13] also mentioned that the WoZ method has been largely used to build large corpora for natural language processing used in retrieval information systems (e.g. reservation and advisory systems).

The SEMAINE project team created a large number of audiovisual database of interactions with their Embodied Conversational Agent (ECA) using WoZ method, semi-automatic Sensitive Artificial Listener (SAL) and automatic SAL [11]. Concurrently, the interaction quality between the participants and the ECAs were also evaluated.

3   The User Study

We conducted three user studies of The Smile Game in both semi-public and public spaces. The first and second study were conducted as pilot studies where the system was installed at one of our school's semi-public spaces and any person (age of 18 or over) who was passing by, interacted with the system. Hence, it established a natural space for interaction with the virtual game companions, allowing for serendipitous encounters. Based on the evaluations of the pilot studies, we deployed an updated system and evaluated it during a public event (Mayfest 2015) held at The University of Nottingham. In this study, any person of any age who was passing by, played with the virtual game companions.

In the game, we use two SEMAINE characters, Poppy and Spike [11] as the game companions. Poppy has a high extraversion and low neuroticism personality ($\mathcal{P}_E = 0.5, \mathcal{P}_N = -0.1$, while Spike has a high neuroticism and low extraversion personality ($\mathcal{P}_E = 0.2, \mathcal{P}_N = 0.5$) and decay rate r to 0.02 for both game companions. We based our model on the OCEAN personality model [14]. The values of the traits are represented by a set of real numbers between -1 to 1, where 1 represents the strongest possible value in that particular trait (see [4] for details).

There were four phases of the interaction: Opening, Chit-Chat, Game, and Closing phase. The Opening Phase is a phase where the participant initiated their interaction with the virtual game companion. In this phase, the virtual game companion greeted the participant and asked the participant's name (if the virtual game companion hadn't met the participant before). Afterwards, the virtual game companion initiated a small chit-chat using "evaluating the day", weather, or news dialogue (Chit-Chat Phase). The next phase was the Game Phase where the participant and the virtual game companion played the game trying to make each other laugh by making weird facial expressions, funny body movements and jokes. Finally, the virtual game companion evaluated the game and said goodbye after they finished the game (Closing Phase).

During the interaction, a video and audio recording was made using a clearly visible camera and shotgun microphone. All the interaction videos were annotated using ELAN [3] and analysed for the evaluation. We annotated the interaction turn, participant's engagement, game strategy, movement type to evaluate the conversation and the game system. The conversation topics were also annotated to understand what do the participant usually talk about while interacting

with the virtual game companions. Figure 1 illustrates the interaction video annotated with ELAN. No 1 shows The Wizard/Debug screen, No 2 shows the participant screen, and No 3 shows the annotation block.
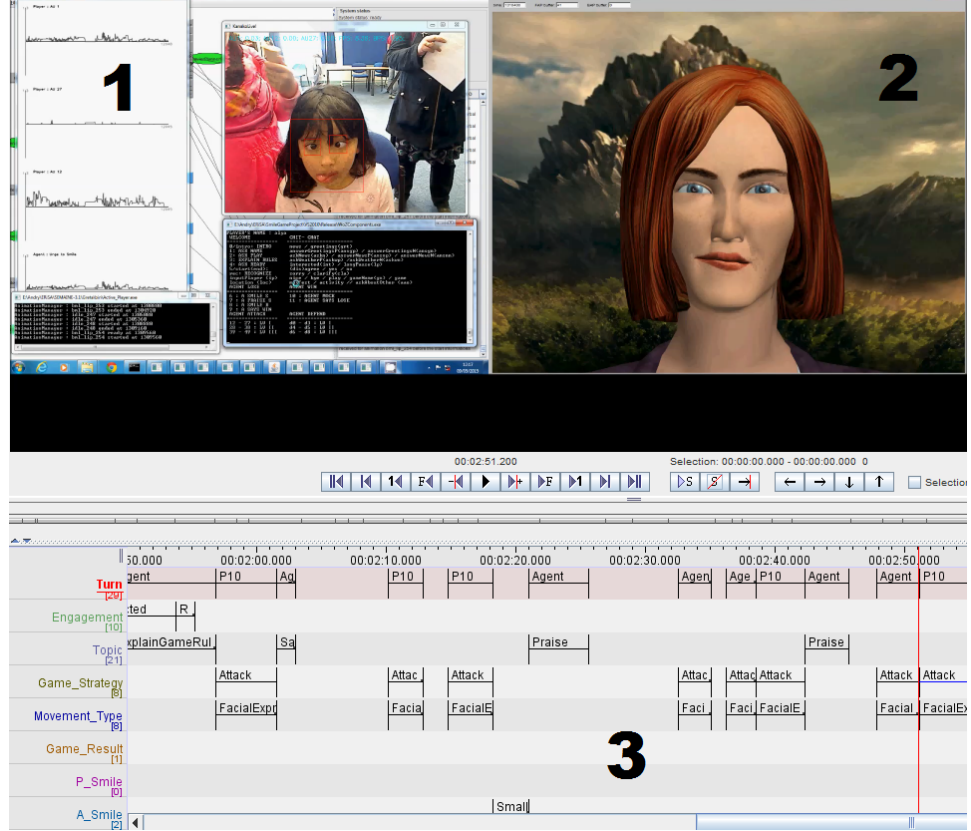

Fig. 1. The Smile Game User Study

3.1 Study 1: Wizard of Oz Pilot Study

Two different methods were applied for this study. In the first method, the participants (age of 18 or over) were recruited. In the second method, we run the game system at random times during working days for approximately 1.5 hour/day for 2 weeks, and let any person (aged 18 or over) who was passing by, interacted with the system. The participants were allowed to interact with the virtual game companions as long as they want and as many time as they like. The virtual game companions will forge a relationship link to the participant, depending on the quantity and quality of the interaction with the participant.

In this study, an operator called The Wizard controlled several system functions manually while a user interacted with the system, believing that they were interacting with a fully functional application system [8]. The Wizard replaced the Automatic Speech Recognition modules in Sensing Components, Verbal Interpreter modules in Interpreter Components, and all modules in Behaviour Components. While the other components remain automatic. In The Wizard screen (See Fig. 1 no. 1), The Wizard observes the participant's face through a web camera and hears the participant's voice through a microphone. In addition, The Wizard monitors participant's internal states through a graphical representation chart(as shown in Figure 1). The Wizard then interprets the participant's behaviour and makes a decision to reply to the participant using command line by choosing actions provided in the screen.

During any phases, The Wizard can either choose a specific action or just the category and let the system determines a specific action based on the participant's and virtual game companion's internal states. For example: In the greeting phase, The Wizard can either choose a specific greeting utterance or just type "greetings". The system will then choose a greeting utterance for The Wizard to greets the participant. Other example: In the game phase, The Wizard can specify what kind of facial expressions or jokes to perform or just type "attack" or "defend" to perform an attack or defend behaviour based on the virtual game companion's urge to smile. The closer the participant is, to make the virtual game companion laugh, the more aggressive the virtual game companion becomes.

In this study, there were 9 distinct participants (8 males) with 12 interactions with the virtual game companions (10 with Poppy, 2 with Spike). The total duration of the interactions was 30 minutes and 48 seconds (MIN = 36s; MAX = 333s; AVG = 153.42s; SD = 100.35). Six participants were invited to interact with the virtual game companions, while the others spontaneously played with the virtual game companions. In addition, there were 10 interactions where the participants interacted for the first time, with the virtual game companions.

Among the 5 participants who finished the game, there was only one participant who won the game playing with Poppy (winning rate 20%). Others couldn't help but laugh either because of the movements they did or the virtual game companion performed. Mostly, the virtual game companion won due to their funny facial expressions. There were 2 participants who laughed immediately after they performed funny facial expressions. This indicates, they were laughing because they thought they were doing some funny and stupid things. In four out of twelve game sessions, the participants also mimicked the virtual game companion's facial expressions.

3.2   Study 2: Fully Automatic System Pilot Study

The method of this study is similar to the first study (Wizard of Oz Pilot Study). The only difference in this study was that the system was running fully automatic. In this study, we run the game system at random times during working days for approximately 2 hour/day for two weeks.

In this study, there were 12 distinct participants (8 males) with 12 interactions with Poppy. The total duration of interactions was 19 minutes and 31 seconds (MIN = 3s; MAX = 351s; AVG = 97s; SD = 101.47). All participants spontaneously played with the virtual game companion without knowing what they should/can do with the virtual game companion, except 1 participant was invited to play with the virtual game companion. In addition, there were 10 interactions where the participants interacted for the first time, with the virtual game companions.

From all participants, only 2 participants successfully played with Poppy, and only 1 finished the game(winning rate 50 %). Ten out of twelve participants weren't sure what to do, and how to play the game while the virtual game companion failed to interpreted the user's speech due to the speech recognition technology.

3.3   Study 3: MayFest 2015

In this study, the system was ran approximately for 6 hours at the The University of Nottingham's Mayfest 2015 event. Due to the very noisy environment, it was not possible for the system to run fully automatic. Hence, we implemented a Semi-Automatic Wizard of Oz in this study. In contrast to the first study, in this study The Wizard only controlled Automatic Speech Recognition modules in Sensing Components and part of Interpreter modules in Interpreter Components, while the other components remained automatic.

In this study we did some improvements based on the evaluation of the previous studies. Firstly, we made the instructions for the game rules clearer as compare to the previous. Secondly, we put some fun elements in game (i.e. background image, music, and background effect). The tempo of the music is increasing dynamically based on the game level to create the tension. The background image (see Fig.2) also created a story that the virtual game companion is at the top of Ben Lawers Mountain, with a bird sound playing in the background. Several participants even asked the virtual game companion to walk and show them around what the mountain scenery looks like.

There were 46 distinct participants (25 males, 34 children) with 50 interactions (44 with Poppy, and 6 with Spike). The total duration of interactions was 2 hours 32 minutes and 12 seconds (MIN = 35s; MAX = 559s; AVG = 181.94s; SD = 90). All participants spontaneously played with the virtual game companions without knowing what they should/can do with the virtual game companions except 1 participant who had previously interacted with a virtual game companion. From all participants, there were 32 participants who finished their game with Poppy (winning rate 75 %). All participants who played with Spike, didn't want to continue playing with Spike because it was too hard to make Spike laugh and ultimately got bored.

Fig. 2. First (Left) and Second (Right) Version of Background

4   Result and Discussion

In all studies, participants who were interacting and playing with virtual game companions successfully perceived their personality and felt Poppy is more friendly, while Spike looked rigid, angry, and hard to defeat. Some participants said they never wanted to play with Spike again. This can be infered from the behavoiurs shown by the participants during the interactions. For example: Participant No. 41 from the third study said "He (Spike) seems unfriendly, I'll never want to play with him again." However, interestingly, most of the participants prefer Spike's voice because sometimes Poppy's words were hard to understand. Figure 4 shows top 5 topics that participants like to talked about, with the virtual game companions. Topic about the game itself was the most popular topic for conversation with the virtual game companions with total occurrences of 92 times (total AVG duration 178 sec), 49 times (total AVG duration 113 sec), and 457 times (total AVG duration 1330 sec) respectively for first, second, and third study.

In the first and second study we found out that topic about news was not a good topic to start the interaction. 83 % participants turned the conversation down with saying "no good news". Thus, in the third study we changed the topic into a context specific topic (i.e. about The Mayfest event) (e.g. Did you enjoy mayfest? or What is your favourite activity in Mayfest?). Participants responded positively about this topic with total AVG duration of 242 sec spent on this topic. In addition, in the third study, a number of people liked to tell the virtual game companion about themselves with 83 occurrences and total average duration of 208 sec spent on this topic.

Figure 4 illustrates top 5 movements preformed in all studies. In the first and second study top of the attack movements type when playing the game is facial expressions which occurred 34 times (total AVG duration 30 sec)and 26 times (total AVG duration 6 sec) respectively. Similarly, in the third study, facial expressions was the top choices to perform with total occurrence of 184 times (total AVG duration 139 sec). In addition, in the third study, a number of participants also performed the "knock-knock" jokes. Hand movements were also frequently seen in combination with other movements.

Table 1. Top 5 Topics in The Studies

Study #	Topics	Occurrences	Total AVG Duration (s)
STUDY I	The Game	92	178
	Greeting	56	69
	About News	21	54
	Goodbye	12	26
	Error	6	21
STUDY II	The Game	49	113
	Greeting	34	82
	About News	25	45
	Clarification	45	41
	Error	9	32
STUDY III	The Game	457	1330
	Greeting	122	248
	Mayfest	161	242
	About Player	83	208
	Goodbye	32	100

Table 2. Top 5 Movements in The Studies

Study #	Movement Type	Occurrences	Total AVG Duration (s)
STUDY I	Facial Expressions	34	30
	Jokes	26	29
	Body Movements + Facial Expressions	7	17
	Body Movements	5	11
	Head Movements	3	5
STUDY II	Facial Expressions	26	6
	Facial Expressions + Body Movements	1	4
	Jokes	9	3
	Head Movements	4	3
	Facial Expressions + Head Movements	2	2
STUDY III	Facial Expressions	184	139
	Facial Expressions + Hand Movements	63	105
	Jokes	76	92
	Head Movements	21	33
	Noise	16	19

All studies indicate that the virtual game companion is reasonably good in perceiving user's facial expressions especially for Action Unit 12 (Lip corner puller), a bit slow but reasonably accurate for Action Unit 27 (Mouth stretch), but sometimes give false positive for Action Unit 1 (Inner brow Raiser). Based on the debug screen (See Fig 1), the virtual game companion's urge to laugh (emotion perception) is reasonably good in reacting to the user's action from facial expressions. Furthermore, from total of 5 repeated interactions, only 3 participants were successfully recognized by the virtual game companion. In addition, there was one participant who was recognized as some other participant by the virtual game companion, despite that the participant has never interacted

with the virtual game companion before. As a result of the experimental setting of the MayFest and a relatively low turnout of the pilot studies, most of the participants were new to the system. Therefore, this setting did not allow us to evaluate the social relationship modules.

In the fully automatic system, the failure of the virtual game companion's perception of the verbal communication is a substantial obstacle to overcome. Seven out of twelve participants were not able to play with the virtual game companion because the virtual game companion failed to understand the user's utterances and repeatedly asked the participants to rephrase their word, or talked randomly.

5   Conclusion and Future Work

In order to achieve a fully automatic system for integrated virtual game companions, there are some problems need to be addressed. Firstly, it still needs to continuously improved in the game story and refine the game mechanics to make the game more fun and enjoyable. Secondly, a fully automatic system using speech recognition, is still tricky to implement in the open space area with noisy environment. The system turned down 58 % of interactions by failing to understand the participant's utterances in the fully automatic system.

A virtual game companion context based conversation will be the focus on our next development. Moreover, a trained model of noise in the background is also needed to filter the noise. In addition, from the large number of video data collected, we are now working to automatically learn the interaction rules and additional attack movements for the virtual game companions through the facial expressions performed by the participants. In order to achieve this, we annotated the recordings using the Facial Action Coding System (FACS) [6] for the participants' facial expressions and created a transcript of the participants' utterances. In addition the attack movements can be also rated and implemented as the virtual game companion's internal state of urge to smile using machine learning or statistical method.

Acknowledgements

References

1. E. Bevacqua, K. Prepin, R. Niewiadomski, E. de Sevin, and C. Pelachaud. GRETA : Towards an Interactive Conversational Virtual Companion. In *Artificial Companions in Society: perspectives on the Present and Future*, pages 1–17, 2010.

2. Timothy Bickmore, Laura Pfeifer, and Daniel Schulman. Relational agents improve engagement and learning in science museum visitors. In *Proceedings of the 10th International Conference on IVA*, IVA'11, pages 55–67, Berlin, Heidelberg, 2011. Springer-Verlag.
3. Hennie Brugman, Albert Russel, and Xd Nijmegen. Annotating multi-media/multi-modal resources with elan. In *LREC*, 2004.
4. Andry Chowanda, Peter Blanchfield, Martin Flintham, and Michel Valstar. Erisa: Building emotionally realistic social game-agents companions. In T. Bickmore, S. Marsella, and C. Sidner, editors, *Intelligent Virtual Agents*, volume 8637 of *Lecture Notes in Computer Science*, pages 134–143. Springer International Publishing, 2014.
5. Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg. Wizard of oz studies: Why and how. In *Proceedings of the 1st International Conference on Intelligent User Interfaces*, IUI '93, pages 193–200, New York, NY, USA, 1993. ACM.
6. P. Ekman, W.V. Friesen, and J.C. Hager. *Facial Action Coding System (FACS): Manual*. A Human Face, Salt Lake City (USA), 2002.
7. Jonathan Gratch and Stacy C. Marsella. Evaluating the modeling and use of emotion in virtual humans. In *International Conference on (AAMAS)*, New York, NY, August 2004.
8. Paul Green and Lisa Wei-Haas. *The Wizard of Oz: a tool for rapid development of user interfaces*. 1985.
9. J. F. Kelley. An iterative design methodology for user-friendly natural language office information applications. *ACM Trans. Inf. Syst.*, 2(1):26–41, January 1984.
10. Stefan Kopp, Lars Gesellensetter, Nicole C. Krmer, and Ipke Wachsmuth. A conversational agent as museum guide - design and evaluation of a real-world application. In *IVA*, volume 3661 of *Lecture Notes in Computer Science*, pages 329–343. Springer, 2005.
11. G. McKeown, M.F. Valstar, R. Cowie, M. Pantic, and M. Schröder. The semaine database: Annotated multimodal records of emotionally coloured conversations between a person and a limited agent. *T. Affective Computing*, 3(1):5–17, 2012.
12. Susan Robinson, David Traum, Midhun Ittycheriah, and Joe Henderer. What would you ask a Conversational Agent? Observations of Human-Agent Dialogues in a Museum Setting. In *International Conference on Language Resources and Evaluation (LREC)*, Marrakech, Morocco, May 2008.
13. Daniel Salber and Jolle Coutaz. Applying the wizard of oz technique to the study of multimodal systems. In *In Human Computer Interaction, 3rd International Conference EWHCI93*, pages 219–230. Springer Verlag, 1993.
14. G. Saucier and L. R. Goldberg. *The language of personality: Lexical perspectives on the five-factor model*, pages 21–50. New York: Guilford, 1996.
15. M. Schröder. The SEMAINE API: Towards a standards-based framework for building emotion-oriented systems. *Advances in Human-Computer Interaction*, 2010:Article ID 319406, 2010.
16. Candace Sidner, Charles Rich, Mohammad Shayganfar, Timothy Bickmore, Lazlo Ring, and Zessie Zhang. A robotic companion for social support of isolated older adults. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on HRI*, HRI'15 Extended Abstracts, pages 289–289, New York, NY, USA, 2015. ACM.
17. Laura Pfeifer Vardoulakis, Lazlo Ring, Barbara Barry, Candace L. Sidner, and Timothy Bickmore. Designing relational agents as long term social companions for older adults. In *Proceedings of the 12th International Conference on IVA*, IVA'12, pages 289–302, Berlin, Heidelberg, 2012. Springer-Verlag.