ERiSA: Building Emotionally Realistic Social Game-Agents 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. We propose an integrated framework for social and emotional game-agents to enhance their believability and quality of interaction, in particular by allowing an agent to forge social relations and make appropriate use of social signals. The framework is modular including sensing, interpretation, behaviour generation, and game components. We propose a generic formulation of action selection rules based on observed social and emotional signals, the agent's personality, and the social relation between agent and player. The rules are formulated such that its variables can easily be obtained from real data. We illustrate and evaluate our framework using a simple social game called *The Smile Game*.

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

1 Introduction

Taking Alan Turing's famous question "Can machines think?" [16] to it's logical next step, researchers in computer science have started to challenge themselves to build not only computers that can think but also ones that have a virtual embodiment, can communicate and interact with humans, and "live" in a virtual world. Such embodied conversational agents are applied in a wide range of applications areas covering training and education [4], government and military [5], medical and health [3], films, and gaming [11].

In the area of gaming, Intelligent Virtual Agents (IVAs) research has led to major improvements in their believability. This is generally achieved by improving their visual appearance and more recently by making their behaviour more realistic. Yet there remain a number of improvements and contributions to be made including more natural expression of emotions [2], enhanced role of personality and social relations in the selection of the agent's behaviour [11], and better use of social signal processing [18] to drive the interaction.

In gaming applications, IVAs can be particularly interesting potential vehicles of affect, because players naturally engage with agents as part of a game. Over time, a pattern of interactions between player and agents may translate into a relationship if the game and its IVAs are designed to allow for this. It is this evolving relationship, expressed in terms of social signal that we propose to include in our IVA architecture.

In the real world, social interaction between people involve the encoding and decoding of social signals [18]. These signals are displayed through facial expressions, body gestures, and voice. We propose to model our game agents exactly like this: they will perceive and generate behaviour as we do, both when interacting with a user as well as when interacting with other IVAs.

While a number of works have moved towards building realistic IVAs in games (e.g. [11, 15]), so far there exists no fully integrated framework for a social and emotional game-agent, complete with the ability to develop simple social relations over time. Hence, in this paper we propose such a framework. In addition, we propose a formulation of action selection rules based on the IVA's personality, the social relation between player and agent, and the perceived (non-)verbal actions to enhance the agents' believability and the user's interaction in a game context. The rules are formulated so that they can be easily learned from data.

To illustrate the efficiency of our framework, we implemented a simple game called the Smile Game. In this game, the player can play, interact, and build a relationship with the agents (see Section 5).

2   Related Work

A number of IVA frameworks have been proposed to improve the social capabilities of agents. For example: The SSI (Social Signal Interpretation) framework [19] aims to integrate multi-modal affect recognition and interpretation. The framework consists of several components that enabled affect recognition and interpretation using tools that can recognise and interpret a user's social signals conveyed by their voice and face. The results of the recognition and interpretation can be processed as sensing information by the agents.

Another example of a fully integrated IVA framework is the SEMAINE framework [14] [10]. SEMAINE is an open-standards-based framework for building emotion oriented systems. It consists of a number of components that communicate with each other using XML based messages. This framework emphasises non-verbal behaviour, turn-taking, and back-channelling rather than understanding the topics of conversation verbally.

A similar framework is the ICT Virtual Human [7] framework. It contains a number of components that cover automatic speech recognition, emotion sensors, natural language processing, and behaviour generation. This framework is more complex than the SEMAINE Framework, incorporating a large number of modules developed both by ICT USC as well as by third parties. Similar to SEMAINE, it uses XML messages in an ActiveMQ framework as the means of communication between components.

A framework developed specifically for the gaming context is Koko [15]. Koko describes a conceptual framework to integrate affect recognition into games. The framework has seven components which communicate using arrays of vectors. With these components, the framework allows the game system to manage the agent's moods, user's emotional state, and game environment information. Unfortunately, this framework can only serve as an architectural connector between
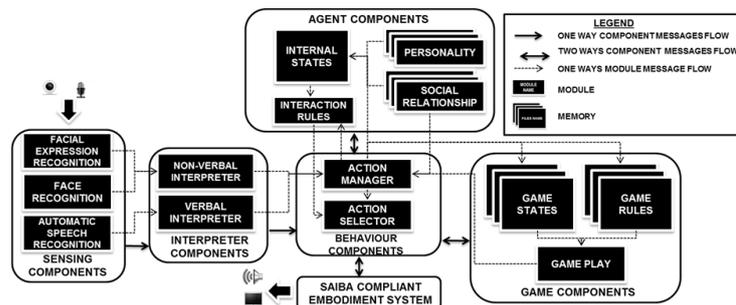
Fig. 1. Overview of the ERiSA Framework

a game framework and other affective modules/frameworks, hence by itself it is not sufficient for building complex games.

In addition to building a framework to improve the capability of the agents, a formal model of the agents' behaviour is essential to allow agents to reason about their own behaviour. Ochs et al. proposed such a dynamic model for Non-Player Characters (NPCs) in computer games [11], which studied the dynamics of emotions, personality, and social relationship between the NPCs. In this model, the agent's behaviour is influenced by their emotions which in turn are affected by their personality. For example: a negative emotion in i induced by j decreases the value of liking i has for j and vice versa. However, this model does not define how the player's relationship to the agents comes about in the first place.

In summary, while the existing frameworks focus on generic models of IVAs, none are capable of modelling evolving social relations between users and agents, and use these to generate behaviour rules.

3 The ERiSA Framework

We propose a framework for game-agents that extends existing frameworks such as SEMAINE or the ICT Virtual Human framework by including social relationship as a variable as well as the ability to learn social interaction rules from data. We adapt the framework for a game environment by adding components that manage the game play dynamically. The framework consists of modular components including sensing, interpretation, and behaviour planning (see Fig. 1).

The functions of the main components are as follows:

- The **Sensing Component** provides modules for Facial Expression Recognition (FACS and six basic emotions [1]), Automatic Speech Recognition, and Face Recognition. The Face Recognition module is used to recognise a player's face based on which an agent can build and interpret a social relationship with a player.
- The **Interpreter Component** receives data about the player's utterances, facial expressions, and identity. The player's utterances are interpreted in

the Verbal Interpreter module, while the user's emotions are interpreted by the Non-Verbal Interpreter module.

– The **Behaviour Component** proposes an agent behaviour based on the information of both verbal and non-verbal interpretation and the Interaction Rules. The proposed behaviour is then sent to a SAIBA compliant embodied system [9] by the Action Selector Module. Afterwards, the Action Manager module updates all the states and rules including Internal States, Game States, Interaction Rules, Game Rules, and Social Relationship.

– The **Agent Components** consist of Memories and a Social Relationship module. Memories include Internal States, Internal Rules, and Personality. The Interaction Rules are influenced by the agent's internal states such as personality, relationship , and emotion. The Social Relationship module manages the relationship status between two entities such as agent A-player I, player I - player II, or agent A-agent B.

– The **Game Components** consists of a Game Play module and related Memories including Game Rules, and Game States. The Game Play module regulates the game play from the Memories. The game play changes dynamically based on the agent's internal states.

An example in which the social relation is used in a game scenario is as follows: In an RPG game a player's mission is to convince guard Maximus to let the player pass the gate. The player has to do this by building a positive relationship, avoiding conflict. Suppose the player has met Maximus three times before, has a positive relationship with him, and Maximus has a high value of extraversion. The player greets the guard with a smile while saying "hello". The player's utterance and facial expression are captured and processed in the Sensing Components, and interpreted in the Interpreter Components. The Action Manager updates the values of the social relationship and the guard's internal states. In this case emotion is set to be more positive and the value of the social relationship is increased as well. Finally, Maximus reacts based on the information from the Interpreter Components and Interaction Rules.

4   Automatic Behaviour Modelling

Personality We based our model on the OCEAN personality model [13]. This model has been used in several works in the area of IVAs [11]. 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. For example: if the agent has personality with an extraversion value of 0.65 and conscientiousness of -0.65 then the agent tends to be talkative and shows a lack of self-discipline.

$$\mathcal{P} = \{O, C, E, A, N\} \in [-1, 1] \tag{1}$$

Social Relationship In the social relationship model, we highlight 2 variables of social relationship: *Likes* and *Knows*. The change in the degree of *Likes*

($Likes_t \in [-1, 1]$) at time t depends on the emotion E_t of the agent towards the player at that time (explained below), so a negative emotion at time t will lead to the decrease of *Likes*. The degree of *Knows* depends on how many times the agent and the player met, which in our game scenario is measured in terms of how often they played the game. The value of *Knows* is normalised to a real number between 0 to 1 ($Knows_{curr} \in [0, 1]$). The value of the Social Relationship at time t ($\mathcal{R}_t$) is calculated as:

$$\mathcal{R}_t = (Likes_t \times Knows) \tag{2}$$

where

$$Like_t = Likes_{t-1} + E_t \times a \tag{3}$$

Emotion Both personality and emotion play a role in the individual's behaviour [8, 12]. For example: an individual with high extraversion tends to express and perceive positive emotions, while an individual with high neuroticism trait tends to express and perceive negative emotions. In addition to personality, the social relationship also plays a role in the individual's emotional behaviour.

Subsequently, the agent's emotion at time t (E_t) is computed based on the average of events Evt_i at that time in addition to personality ($\mathcal{P}_E$ for extraversion and $\mathcal{P}_E$ for neuroticism) and social relationship at time t ($\mathcal{R}_t$). To model the dynamic properties of emotion perception in a simple yet convincing way, we implemented a decay function to our model. In the absence of any emotion evoking stimuli, the value of an affective dimension is assumed to decay over time [12]. The value of the decay rate r is set based on the agent's personality.

$$E_t = E_{t-1} \times e^{-rt} + V_t \tag{4}$$

where

$$V_t = \frac{\sum_{i=1}^{n} Evt_{(i,t)}}{n} \times (\mathcal{P}_E - \mathcal{P}_N + \mathcal{R}_t) \tag{5}$$

5 Experimental Methodology: The Smile Game

We implemented the proposed framework to create a simple game called *The Smile Game*. The main aim of this project is neither to detect nor synthesise laughter but to implement and evaluate the proposed framework and model that considers the relation between personality, social relationship, and emotion to enhance the agent's believability and quality of user interaction. Recent work in detecting and synthesising laughter can be found in the Ilhaire project [3] [17].

The Smile Game is inspired by the "Don't Smile" game that has two people facing each other. The goal of the game is simple: make your opponent laugh

---
[3] http://www.ilhaire.eu/

with jokes, and funny facial expression. The first person to smile uncontrollably loses. In *The Smile Game*, we replace one of the players with an ECA. We use two SEMAINE characters, Poppy and Spike [10] as the game agents. 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 ($\mathcal{P}_E = 0.2, \mathcal{P}_N = 0.5$). This game is considered to be a good case study since the game play is simple yet elicits rich non-verbal interactions between the agents and the player.

5.1   The Smile Game Recordings

In order to study what type of behaviour people playing the Smile Game would exhibit, we performed an experiment in which two people at a time play the game whilst being video recorded. This will allow us to model typical behaviour and replicate it in the agents. We recruited 10 participants, all of whom are students in our institute (5 male; 6 Asian, 4 Caucasian). We conducted five recording sessions, with a different number of games played per session.

During the game, there were three cameras recording both visual and audio information. The first and second camera each record a player from the front while the third camera records a profile perspective of both players. The images are used for facial expressions analysis and the audio recording for discourse analysis (i.e. presence of conversation, and who spoke first).

We annotated the recordings using the Facial Action Coding System (FACS) [6] for the players' facial expressions and created a transcript of the players' utterances[4]. Next we selected utterances and facial expression that were used for attacks to make the opponent laugh. These attack behaviours were then implemented in the agent.

There were a total of 14 distinct FACS combinations that succeeded in making the opponent smile/laugh. Unfortunately, not all FACS combinations can be generated by the agents. There were ten utterances that succeeded in making the opponent smile/laugh. Hence, we push all these utterances to the agents' attack vocabulary. In addition we added twenty additional funny sentences from various sources on the internet, such as "I feel like such a homo sapiens right now" or "Don't you have something better to do?".

5.2   The Smile Game Flow

An overview of the game flow is shown in Fig. (2). When the system starts, it searches for a face. When a face is detected, the agent tries if it can recognise it from the current library of faces. Based on this, the agent greets the player using the Welcome Dialogue Set. If the face is not recognised, the agent will ask for the player's name and will initialise a new entry in the face library as well as a relationship value $\mathcal{R}$. If the system recognises the face, it searches for the corresponding relationship value $\mathcal{R}$ and updates it. Afterwards, the agent starts

---

[4] The utterances in session 4 were in Indonesian, hence, we did a translation to English

a conversation using common topics such as the weather or news using the Chit-Chat Dialogue Set. After a fixed period of time, the agent will ask whether the player would like to play a game with the agent. If so, the game will commence and continues until someone smiles. The player can also asks the agent to play a game with them. When the game ends, the result of the game is saved by the agent, in addition to updating the value of $\mathcal{R}$. Both agent and player can ask to repeat the game, if the player does not want to repeat the game, the agent closes the interaction using its Goodbye Dialogue Set.
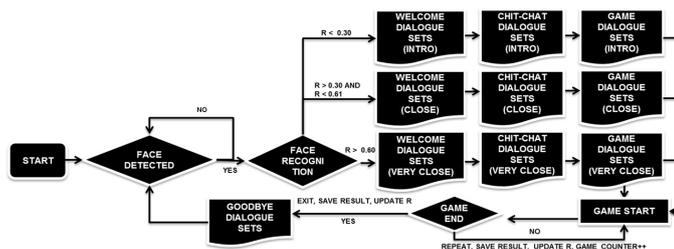

Fig. 2. Game flow. Action streams are selected based on the social relation value $\mathcal{R}$ which is either 'very close', 'close'. 'Intro' is used when the user is unknown.

In addition to the personality, relationship, and emotion model from subsection 4, we define a model to determine when the agent will smile. The agent's urge to smile S is derived from the model of emotion E from subsection 4. The difference is that while the model of emotion E considers all the possible player's verbal and non-verbal actions, the model of smile S only considers the possible player's verbal and non-verbal actions that could trigger the agent to smile. Currently, the same rules used by the agent to attack an opponent will incur an increase in S when observed by the agent.

To model the dynamic properties of the agent's smile in a simple yet convincing way, we again included a decay function to our model. This way the agent's urge to smile will diminish over time unless reactivated. Similar to the emotion model, the value of the decay rate r on this model is also set based on the agent's personality.

$$S_t = S_{t-1} \times e^{-rt} + V_t \qquad (6)$$

5.3   Evaluation

Before evaluating the system with real players, we created a simulation of the game. We experimented with two SEMAINE characters: Poppy and Spike [10]. The relationship value R between the simulated player and the agent is set to 0.65 and decay rate r to 0.02 for both agents. We performed random attacks to both characters for a certain time. Both characters were receiving the same sequence of attacks.
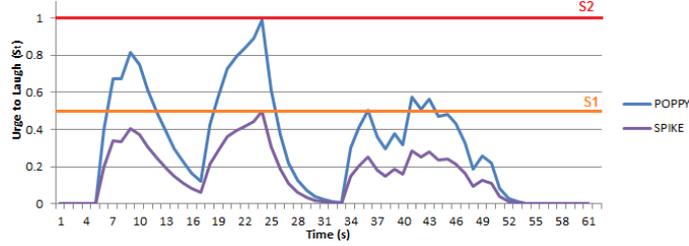
Fig. 3. Smile Simulation. The blue and purple lines show the character's urge to smile for Poppy and Spike, respectively. The orange line indicates the threshold to display a small smile and the purple line the threshold for smiling uncontrollably.

Figure 3 shows the result of the simulation of the both characters urge to laugh (S_t) as a result of three strikes. The first strike was a series of medium intensity attacks from the user from $t = 13$ to 17 s, the second one was a series of high intensity attacks from the user from $t = 25$ to 31 s, and the last strike was a series of low intensity attacks from the user from $t = 41$ to 57 s.

The S1 line indicates the threshold of generating a small smile by the agent, while the S2 line indicates the threshold for an uncontrollable smile, which means the agent loses the game. In this simulation, we used the same value of both thresholds for both agents. As we can see in the figure 3, from $t = 6$ to 11 s, Poppy who has a high extraversion value showed a small smile, and in the $t = 23$ s, Poppy was unable to hold her smile, and she lost, while Spike who has a high neuroticism only just showed a small smile at the second attack.

User evaluation In addition to the simulation, we performed a user evaluation of the game. We recruited 16 participants (9 male; all Asian; mean age = 24.440; age standard deviation = 3.759). All participants are students in our institute. Each participant played two sessions of the game, and in each session they played against both characters (Poppy and Spike). After the game, they were asked to fill in a questionnaire to rate their perceived relationship with the agent, the naturalness of the interaction during the conversation and the game, and the agent's response time. In addition, we also asked the participant to rate the agent's personality.

Figure 4 (a) and (b) shows the average scores for the social relationship and the naturalness of the interaction during the conversation and the game, while Fig. 4 (c) shows the score of the agents personality. A score of 1 is the lowest score, 3 indicates average score, and 5 is the highest score. Except for the rating of 'Social Relation', all scores were identical for session 1 and 2.

From the limited questionnaire results, we can tentatively draw two conclusions. Firstly, the users perceived the two characters to largely display the personality that we intended them to. Poppy was perceived to be twice as extrovert as neurotic, and Spike the other way around (see Fig. 4 (c)) Secondly, and more importantly, the participants were able to tell the changes in their
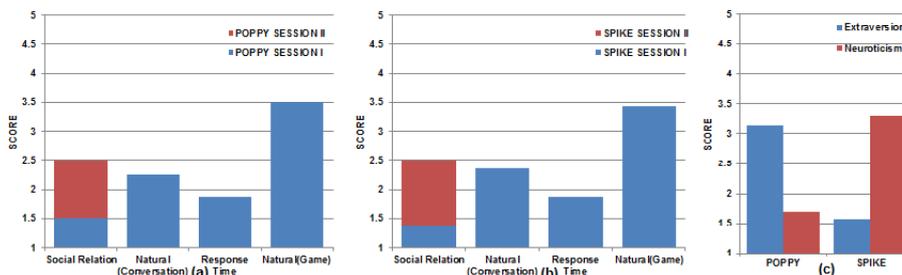
Fig. 4. User Evaluation results of the smile game. Sub-figures (a) and (b) show the perceived social relation, naturalness of behaviour, response time, and naturalness of the game for the game characters Poppy and Spike, respectively. Sub-figure (c) shows the results for perceived personality.

social relationship with both characters over time. The average user score for the social relationship between the agent and the player went from an average of about 1.5 at the first encounter, to 2.5 for the second encounter, when the agent was able to recognise the player and used the relevant dialogues for people it knows. However, the changes in the relationship are not very high because the participants only play with each character twice.

The rating of the naturalness in the interaction during the conversation is quite low for both of the characters due to the response time of the agents. Upon further analysis of the problem it turned out that this was caused by the low performance of the Speech Recognition module. In contrast, participants gave quite high scores for the naturalness of the interaction rating during the game.

6 Conclusion and Future Work

In this paper, we have presented an integrated framework for a social and emotional embodied game-agent. It has been evaluated in *The Smile Game*, a social game where the first person to smile loses. The results show that we successfully built an agent that is capable of creating and using the beginnings of a social relation with a player. However, we achieved quite low ratings for the naturalness in interaction during the conversation due to the low performance of the Speech Recognition module. We aim to explore better context-specific sensing components to enhance the interaction in our future work. Currently the agent's reactions are determined by simple hand-crafted interaction rules. In future work we aim for automatic interaction rule generation and adaptation by learning from (real-time) data obtained from actual interactions between the player and the game-agents.

References

1. T.R. Almaev and M.F. Valstar. Local gabor binary patterns from three orthogonal planes for automatic facial expression recognition. In *Affective Computing and*

Intelligent Interaction (ACII), pages 356–361, Sept 2013.

2. J. Bates. The role of emotion in believable agents. *Commun. ACM*, 37(7):122–125, July 1994.
3. T.W. Bickmore, L. Vardoulakis, B. Jack, and M. Paasche-Orlow. Automated promotion of technology acceptance by clinicians using relational agents. In *Intelligent Virtual Agents*, volume 8108, pages 68–78. Springer Berlin Heidelberg, 2013.
4. A. Bogdanovych, K. Ijaz, and S. Simoff. The city of uruk: Teaching ancient history in a virtual world. In *Intelligent Virtual Agents*, volume 7502 of *Lecture Notes in Computer Science*, pages 28–35. Springer Berlin Heidelberg, 2012.
5. J. Campbell, M. Core, R. Artstein, and et al. Developing inots to support interpersonal skills practice. In *Aerospace Conference*, pages 1–14, March 2011.
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. A. Hartholt, D. Traum, S. C. Marsella, A. Shapiro, G. Stratou, A. Leuski, L.P. Morency, and J. Gratch. All together now: Introducing the virtual human toolkit. In *International Conference on Intelligent Virtual Humans*, aug 2013.
8. E.R. Hilgard. *Introduction to psychology*. Harcourt, Brace, 1953.
9. S. Kopp, B. Krenn, S.C. Marsella, and et al. Towards a common framework for multimodal generation: The behavior markup language. In *Int'l Conf. Intelligent Virtual Agents*, pages 21–23, 2006.
10. 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.
11. M. Ochs, N. Sabouret, and V. Corruble. Simulation of the dynamics of nonplayer characters' emotions and social relations in games. *Computational Intelligence and AI in Games, IEEE Transactions on*, 1(4):281–297, Dec 2009.
12. R. W. Picard. *Affective Computing*. MIT Press, Cambridge, MA, USA, 1997.
13. G. Saucier and L. R. Goldberg. The language of personality: Lexical perspectives on the five-factor model. In J. S. Wiggins, editor, *The five-factor model of personality: Theoretical perspectives*, pages 21–50. New York: Guilford, 1996.
14. 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.
15. D.J. Sollenberger and M.P. Singh. Koko: An architecture for affect-aware games. *Autonomous Agents and Multi-Agent Systems*, 24(2):255–286, March 2012.
16. A. M. Turing. Computing machinery and intelligence. *Mind*, 59(236):pp. 433–460, 1950.
17. J. Urbain, R. Niewiadomski, M. Mancini, and et al. Multimodal analysis of laughter for an interactive system. In *Intelligent Technologies for Interactive Entertainment*, volume 124, pages 183–192, 2013.
18. A. Vinciarelli, M. Pantic, and H. Bourlard. Social signal processing: Survey of an emerging domain. *Image Vision Comput.*, 27(12):1743–1759, November 2009.
19. J. Wagner, F. Lingenfelser, T. Baur, I. Damian, F. Kistler, and E. André. The social signal interpretation (ssi) framework: multimodal signal processing and recognition in real-time. In *ACM Multimedia*, pages 831–834. ACM, 2013.