Integrating Conversation Trees and Cognitive Models within an ECA for Aggression De-escalation Training

Tibor Bosse[1,2] and Simon Provoost[1]

[1]VU University Amsterdam, Department of Computer Science
De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands
[2]TNO, Department of Training and Performance Innovations
Kampweg 5, 3769 DE Soesterberg, The Netherlands
t.bosse@vu.nl, s.j.provoost@vu.nl

Abstract. Traditionally, Embodied Conversational Agents communicate with humans using dialogue systems based on conversation trees. To enhance the flexibility and variability of dialogues, this paper proposes an approach to integrate conversation trees with cognitive models. The approach is illustrated by a case study in the domain of aggression de-escalation training, and a preliminary evaluation in the context of a practical application is presented.

Keywords: virtual training, aggression de-escalation, cognitive modelling.

1 Introduction

Embodied Conversational Agents (ECAs) can be defined as computer-generated characters 'that demonstrate many of the same properties as humans in face-to-face conversation, including the ability to produce and respond to verbal and nonverbal communication' [7]. ECAs have been put forward as a promising means for the training of social skills [11]. Indeed, in recent years, various systems have been designed involving ECAs that enable users to develop their social abilities (e.g., [12]).

An important requirement to effectively train users in developing their social skills, is *believability* of ECAs, as believable agents permit their conversation partners to 'suspend their disbelief', which is an important condition for learning [3]. In [8], believability is defined by three dimensions, namely aesthetic, functional, and social qualities of agents, which can be related, respectively, to the agent's physical appearance, behaviour, and interaction style. With respect to physical appearance and interaction style, much progress has been made in recent years: graphics are becoming increasingly realistic, and the mechanisms to interact with ECAs are changing from purely text-based systems to sophisticated multi-modal interfaces [17].

With respect to the behaviour of ECAs, some steps forward have been made as well, regarding verbal as well as and non-verbal aspects. For non-verbal behaviour, recent work addresses systems to generate realistic facial expressions, head movements, and body gestures [13]. Instead, the focus of the current work is on verbal behaviour, i.e., on dialogues. The traditional approach to drive the verbal behaviour of ECAs during a human-agent dialogue is to use *conversation trees*, i.e. tree structures representing all

possible developments of the dialogue, where users can decide between different branches using multiple choice. Although this approach can be successful due to its transparency, an important limitation of conversation trees is that they are quite rigid. Consequently, the resulting behaviour of the ECAs is often perceived as stereotypical and predictable. This can be overcome by constructing large conversation trees with many branches, but this approach is highly labour-intensive and difficult to re-use.

As an alternative, several authors have proposed the use of cognitive models to endow ECAs with more sophisticated behaviour (e.g., [10,14,16]). Using such models, agents base their behaviour not only on their current observations (or input), but also on internal states, such as their emotions and personality. As a consequence, this approach potentially results in more varied and human-like behaviour from the perspective of the human conversation partner. Elaborating upon these ideas, the current research attempts to further bridge the gap between traditional approaches based on conversation trees (which are transparent, but rigid), and more novel approaches based on cognitive models (which are flexible, but abstract). It does so by presenting an approach that not only enables flexible dialogues, but that can also easily be integrated with existing systems in the gaming industry based on conversation trees. The approach is illustrated by a specific case study in the domain of simulation-based training for aggression de-escalation.

The remainder of this paper is structured as follows. In Section 2, the context in which this research was conducted is described, namely a project on aggression de-escalation training. Next, Section 3 introduces the underlying dialogue system that is used within this project, and Section 4 presents an approach to integrate the system with a cognitive model. Section 5 describes a practical application that has been used to test the resulting behaviour of the system. Section 6 concludes the paper with a discussion.

2 Aggression De-escalation Training

In domains such as law enforcement and public transport, aggressive behaviour against employees is an ongoing concern. According to a recent study in the Netherlands, around 60% of the employees in the public sector have been confronted with such behaviour in the last 12 months [1]. Being confronted with (verbal) aggression has been closely associated with psychological distress, which in turn has a negative impact on work performance. Responses to aggression range from emotions like anger and humiliation through intent to leave the profession. In case of severe incidents, employees may even develop post-traumatic stress syndrome [4].

To deal with aggression, a variety of techniques are available that may prevent escalation [2,5]. These include verbal and non-verbal communication skills, conflict resolution strategies, and emotion regulation techniques. The current paper is part of a project that explores to what extent simulation-based training using ECAs can be an effective method for employees to develop these types of social skills[1]. In the envisioned training environment, a trainee will be placed in a virtual scenario involving verbal aggression, with the goal of handling it as adequately as possible. The scenarios

[1] More information on this project can be found at http://stress.few.vu.nl.

emphasise dyadic (one-on-one) interactions. For instance, the trainee plays the role of a tram driver, and is confronted with a virtual passenger who starts intimidating him in an attempt to get a free ride. The trainee observes the behaviour of the ECA, and has to respond to it by selecting the most appropriate responses from a multiple choice menu[2]. Additionally, the trainee is 'monitored' during the task by an 'intelligent tutor', i.e. a piece of software that observes his behaviour and provides personalised support.

The main learning goal of the training system is to help trainees develop their *emotional intelligence*: they should be able to recognise the emotional state of the (virtual) conversation partner, and choose the right communication style. In this respect, an important factor is the distinction between *reactive* and *proactive* aggression that is made within psychological literature: reactive aggression is characterised as an emotional reaction to a negative event that frustrates a person's desires (e.g. a passenger becomes angry because the tram is late), whereas proactive aggression is the instrumental use of aggression to achieve a certain goal (e.g. a passenger intimidates the driver in an attempt to ride for free) [15]. Hence, one of the key differences between these two types is the presence or absence of anger. To decide whether they are dealing with a reactive or a proactive aggressor, trainees should pay attention to specific cues that point to the presence or absence of emotion in the (virtual) conversation partner, such as a trembling voice or frequent arm gestures.

Based on the type of aggressive behaviour that is observed, the trainee should select the most appropriate communication style. More specifically, when dealing with a reactive aggressor, empathic, *supportive* behaviour is required to de-escalate a situation, for example by ignoring the conflict-seeking behaviour, by actively listening to what the aggressor has to say, and by showing understanding for the situation. Instead, when dealing with a proactive aggressor, a more dominant, *directive* type of intervention is assumed to be most effective. In this case, one should make it clear that aggressive behaviour is not acceptable, and that such behaviour will have consequences [2,5]. By ensuring that the ECAs respond in an appropriate manner to the chosen responses (e.g. a reactive aggressor calms down when approached in a supportive manner, but becomes even more angry when approached in a directive manner), the system will provide implicit feedback on the chosen communication style.

3 Dialogue System

The proposed training system is based on the InterACT software, developed by the company IC3D Media[3]. InterACT is a software platform that has been specifically designed for simulation-based training. It uses state-of-the-art game technology that builds upon recent advances in the entertainment gaming industry. Unlike most existing software, it focuses on smaller situations, with high realism and detailed interactions with virtual characters. True-to-life animations and photo-realistic characters are used to immerse the player in the game. An example screenshot of a training scenario for the

[2] Although our research as a whole explores a variety of modalities (such as speech, facial expressions and gestures), the current paper has an emphasis on text-based interaction.

[3] See http://www.interact-training.nl/ and http://ic3dmedia.com/.

public transport domain is shown in Figure 1. In this example, the user plays the role of the tram driver that has the task of calming down an aggressive virtual passenger.

To enable users to engage in a conversation with an ECA, a dialogue system based on conversation trees is used. The system assumes that a dialogue consists of a sequence of spoken sentences that follow a turn-taking protocol. That is, first the ECA says something (e.g. "I forgot my public transport card. You probably don't mind if I ride for free?"). After that, the user can respond, followed by a response from the ECA, and so on. In InterACT, these dialogues are represented by conversation trees, where vertices are either atomic ECA behaviours or decision nodes (enabling the user to determine a response), and the edges are transitions between nodes.



Fig. 1. Example screenshot of the InterACT environment.

The atomic ECA behaviours consist of pre-generated fragments of speech, synchronised with facial expressions and possibly extended with gestures. Scenario developers can generate their own fragments using the motion capture software FaceShift[4], using a Microsoft Kinect camera. As the recorded fragments are independent from a particular avatar, they can be projected on arbitrary characters.

Each decision node is implemented as a multiple choice menu that allows the user to choose between multiple sentences. In the current version, for every decision node, four options are used, which can be classified, respectively, as *letting go*, *supportive*, *directive*, and *call for support*. Here, the supportive and directive option relate to the communication styles explained earlier. The other two options are more 'extreme' interventions, which according to a national protocol for aggression management should be applied, respectively, in case the aggressor has calmed down or in case the aggression is about to escalate into physical violence [5]. Figure 1 illustrates how these four options can be instantiated in terms of concrete sentences (option A-D). Additionally, the choice of the user determines how the scenario continues (or whether it ends immediately) by triggering a corresponding branch in the tree.

[4] See http://www.faceshift.com/.

Although this approach works well, there is a risk that the behaviour of the ECAs becomes predictable in the long term. For example, in the situation shown in Figure 1, choosing option C (the 'directive' option) will always result in the ECA becoming irritated, no matter how often the scenario is played, or what has happened before. This problem can be overcome by endowing the agent with *internal states* [6] that are either set beforehand (e.g. whether the agent is a reactive or a pro-active aggressor) or are the result of earlier interactions (e.g. a state of anger that gradually increases during the scenario). Our approach to realise this will be explained next.

4 Integration with a Cognitive Model

To endow the ECAs with internal states, an existing cognitive model of aggression is used [5,18]. Although the details of the model are left out of the current paper, a high-level overview of the knowledge on which the model is based is shown in Table 1. This table describes how the agent's mental state changes depending on the type of de-escalation approach that it observes.

Table 1. Impact of various de-escalation approaches on agent's mental state.

observed approach	reactive aggression	proactive aggression
letting go	*remains constant*	*remains constant*
supportive	*decreases*	*increases*
directive	*increases*	*decreases*
call for support	*remains constant*	*remains constant*

Our approach to connect this model of aggression to the dialogue system is depicted in Figure 2. The integrated system consists of three elements (dialogue system, human user, and cognitive model) that interact based on the following information flow:

- *from dialogue system to user*: The dialogue system continuously keeps track of which node in the conversation tree is active. As mentioned in Section 3, nodes are either atomic ECA behaviours or decision nodes (implemented as multiple choice menus). In a typical conversation tree, each ECA behaviour is followed by a decision node. This means that whenever the dialogue system shows information to the user, an ECA behaviour fragment is presented (i.e. the virtual character says something, accompanied with facial expressions and gestures) right before the multiple choice menu is displayed.
- *from user to cognitive model (1)*: Next, the user has to select an option from the multiple choice menu. The options correspond to the different types of observations used in the cognitive model (i.e., [letting_go, supportive, directive, call_for_support]).
- *from user to cognitive model (2)*: In addition, the user's emotional state is provided as input to the cognitive model as well. One of the simplest ways of achieving this is to ask the user to provide a subjective indication of how much emotion (s)he experiences during every interaction. A more advanced solution (which will be used in other stages of the current project) is to determine the user's emotional state based on various sensor measurements like heart rate, electrodermal activity, and facial expressions.
- *from cognitive model to dialogue system*: Based on the observed verbal and non-verbal behaviour of the user, the cognitive model determines the level of aggressiveness of the

verbal and non-verbal behaviour to be produced by the ECA. Next, these two variables (aggression intensity values of verbal and non-verbal behaviour) are transferred to the dialogue system, which uses them to decide how the conversation continues. Currently, this is done by defining for each point in the dialogue, a number of alternative sentences with varying levels of aggressiveness. For example, in case the user (playing the tram driver) has just chosen the sentence 'Your chip card is out of credit', and the ECA's verbal behaviour should have an aggressiveness level between 0 and 0.2, then it will respond with a statement like 'I understand that sir, but unfortunately I am in a hurry. Could you please let me hitch a ride?'. Instead, when it should have an aggressiveness level between 0.8 and 1, it will respond with 'Seriously? What do you want me to do, miss this ride? Come on, it's just one stop, man!'. In a similar manner, the variable for the ECA's non-verbal behaviour determines its amount of emotional expression (e.g. by using more arm gestures).

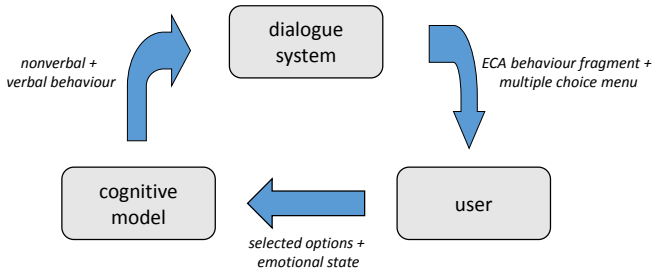


Fig. 2. Overview of the integrated system.

5 Practical Application

The integrated system as described in the previous section is currently being implemented in the InterACT environment. To already get an idea of the proposed mechanism's performance, its behaviour has been studied in the context of a practical application. To this end, a 'light' version of the system has been implemented in Matlab[5]. Basically, the application follows the same information flow as depicted in Figure 2, only the behaviour of the ECA is not visualised in a graphical environment; instead, its utterances and non-verbal behaviour are described in a textual format.

Additionally, a simple feedback module has been implemented. The goal of this module is to provide the user with feedback on his or her performance at the end of a training scenario. Essentially it checks whether the situation was successfully de-escalated or not, and in the latter case, it analyses what the cause of this unsuccessful de-escalation was. In this analysis, several types of mistakes are distinguished such as 1) the user failed to judge the type of aggression (i.e. reactive or proactive) correctly, 2) the user failed to apply the appropriate communication style (supportive or directive), and 3) the user failed to control his or her own emotional state. The decision tree that is used by the module is shown in Figure 3. Here, a green end state indicates successful de-escalation, whereas a red end state indicates unsuccessful de-escalation. Based on the specific end state a corresponding feedback message is generated,

[5] The application can be downloaded from http://www.cs.vu.nl/~tbosse/STRESS.

represented by the numbers in the figure. For example, in case a scenario is classified as category (6), the following feedback is presented:

> *6. The user applies the wrong approach towards a proactive aggressor.*
> "You correctly judged the nature of the aggression, but you used the wrong verbal approach. A proactive aggressor should always be approached in a directive manner. Acting supportively is likely to make the aggressor think he can walk all over you, and that his aggressive behaviour is going to get him what he wants."

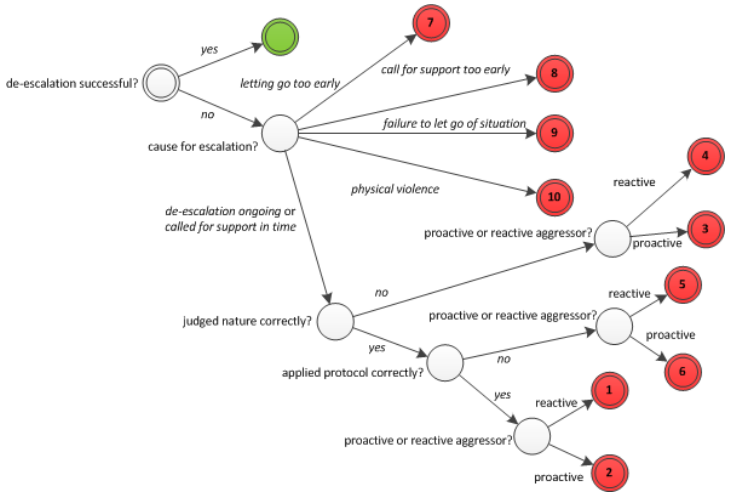


Fig. 3. Overview of the feedback system.

To test the Matlab application, a specific scenario has been worked out in the context of a man who is running late for the custody hearing for his daughter, and has no cash money to pay for a tram ticket. A group of users (students and researchers in Artificial Intelligence) have extensively played the scenario, by systematically varying the parameter settings of the cognitive model. A complete overview of the scenario, as well as an illustration of some of the resulting conversations, is provided in [18].

Based on this preliminary evaluation, we can conclude that the application was evaluated positively, in the sense that no instances of unrealistic dialogue flow were reported. This is a nice finding in itself, but it becomes more valuable in combination with the observation that the proposed approach allowed us to create a large variation in scenarios with relatively limited effort. This has to do with the fact that cognitive models enrich an ECA with internal states, and that these states basically keep track of the *history* of the conversation. To start with, we can now use threshold values to determine which verbal response to activate when the ECA has a certain internal emotional state. Moreover, by designing additional verbal statements that contain language of an increasingly aggressive nature but otherwise carry the same message, *every* user choice can now be followed by a wider variety of ECA responses. Lastly, adjusting the parameter settings that regulate the rate at which the ECA's internal state changes, makes it possible to endow the ECA with a virtually unlimited number of personality types. A more extensive explanation of these benefits is presented in [6].

6 Discussion

The current paper presented a system for simulation-based training of aggression de-escalation skills using Embodied Conversational Agents. By integrating a cognitive model of aggression within a dialogue system based on conversation trees, the system benefits from the advantages of both methods: on the one hand, the use of the dialogue system (based on pre-recorded conversation fragments) guarantees highly realistic animations. This is particularly important for a domain like aggression de-escalation, in which the ECAs ideally induce some kind of 'stress response' in their human conversation partners. On the other hand, the use of a cognitive model ensures that the ECAs are endowed with internal states, which enables them to take the history of the interaction into account when generating their behaviour. As a result, the resulting conversations provide more variation, and are therefore perceived as less predictable. Our first results based on the Matlab application pointed out that users indeed appreciated the conversations as interesting and not too predictable on the longer term.

This is by no means the first paper that attempts to enrich ECAs with more flexible behaviour. Without trying to provide an exhaustive overview, some related approaches are presented in [10,14,16]. The current paper does not attempt to compete with the above approaches in the sense that it claims to generate more variation in scenarios. Instead, one of the main assets of the proposed approach is its simplicity: it allows designers to generate variation in scenarios using a relatively light-weight and easy to use cognitive model. At the same time, it is compatible with state-of-the-art software in the gaming industry that uses traditional conversation trees, such as the InterACT environment. As a result, the approach takes the best of both worlds: on the one hand it can be connected to graphically realistic 3D environments, yet it offers more flexibility than most pre-scripted approaches that are typically used in industry. Another contribution is the fact that the approach has been implemented and tested in the context of a real world domain: aggression de-escalation.

Note that the approach is based on the deliberate decision to work with pre-generated conversation fragments. An interesting alternative is to generate utterances 'at runtime' using a combination of natural language generation techniques and text-to-speech software [9]. Such an approach has the advantage that it results in even less predictable ECA behaviour (from the user's perspective), but a drawback is that it is difficult to guarantee a natural development of the conversation, and that the resulting speech is typically perceived as less realistic by the user.

Regarding the interaction in the other direction (i.e., from user to ECA), our current system uses the easily controllable, but relatively rigid method of multiple choice menus. In ongoing research, we are exploring the possibilities to give the user more freedom by taking an intermediate approach: the user is still asked to choose between certain options; however these options are not completely pre-defined sentences, but 'classes of responses' corresponding to the different communication styles [letting_go, supportive, directive, call_for_support]. Using this approach, the user is free to choose his or her preferred wording, as long as it fits in a category. A sentiment analysis module will then relate the utterance to the right category, allowing for natural continuation of the dialogue. In addition to such extensions, future work will address a more extensive evaluation of the approach. This will not only be done with the presented Matlab application, but also with the complete ECA-based training system and with end users.

Acknowledgements. This research was supported by funding from the National Initiative Brain and Cognition, coordinated by the Netherlands Organisation for Scientific Research (NWO), under grant agreement No. 056-25-013. The authors would like to thank Karel van den Bosch for a number of fruitful discussions.

References

1. Abraham, M., Flight, S., and Roorda, W. (2011). Agressie en geweld tegen werknemers met een publieke taak (in Dutch). Amsterdam: DSP.
2. Anderson, L.N. and Clarke, J.T. (1996). De-escalating verbal aggression in primary care settings. Nurse Pract. 21(10):95, 98, 101-2.
3. Bates, J. (1994). The role of emotions in believable agents. Communications of the ACM, vol. 37, issue 7, pp. 122-125.
4. Bonner, G. and McLaughlin, S. (2007).The psychological impact of aggression on nursing staff. Br J Nurs. 16(13):810-4.
5. Bosse, T. and Provoost, S. (2014). Towards Aggression De-escalation Training with Virtual Agents: A Computational Model. In: Proc. of the 16th International Conference on Human-Computer Interaction, HCI'14. Springer Verlag, pp. 375-387.
6. Bosse, T. and Provoost, S. (2015). On Conversational Agents with Mental States. In: Proc. of the 15th Int. Conf. on Intelligent Virtual Agents, IVA'15. Springer Verlag, pp. 60-64.
7. Cassell, J., Sullivan, J., Prevost, S. and Churchill, E. (2000). Embodied Conversational Agents, MIT Press, Cambridge, MA.
8. De Angeli, A., Lynch, P., and Johnson, G. (2001). Personifying the e-market: A framework for social agents. In: M. Hirose (Ed.), Proc. of Interact 2001. IOS Press, pp. 198-205.
9. Deemter, K. van, Krenn, B., Piwek, P., Klesen, M., Schroeder, M., and Baumann, S. (2008). Fully generated scripted dialogue for embodied agents. Artificial Intelligence, 172/10, 1219-1244.
10. Gebhard, P., Kipp, M., Klesen, M., and Rist, T. (2003). Adding the Emotional Dimension to Scripting Character Dialogues In: Proc. of IVA'03, Springer, pp. 48-56.
11. Kenny, P., Hartholt, A., Gratch, J., Swartout, W., Traum, D., Marsella, S., and Piepol, D. (2007). Building Interactive Virtual Humans for Training Environments. In: Proc. of 2007 Interservice/Industry Training, Simulation & Education Conference, Orlando, FL.
12. Kim, J., Hill, R.W., Durlach, P., Lane, H.C., Forbell, E., Core, C., Marsella, S., Pynadath, D., and Hart, J. (2009) BiLAT: A game-based environment for practicing negotiation in a cultural context. International Journal of AI in Education, vol. 19, issue 3, pp. 289-308.
13. Lee, J. and Marsella, S. (2006). Nonverbal Behavior Generator for Embodied Conversational Agents. In: Proc. of IVA 2006, Springer LNCS, vol. 4133, pp. 243-255.
14. Mateas, M. and Stern, A. (2003). Façade: an experiment in building a fully-realized interactive drama. In: Game Developers Conference (GDC '03), San Jose, CA, USA.
15. Miller, J.D. and Lyna, D.R. (2006). Reactive and proactive aggression: Similarities and differences. Personality and Individual Differences, 41(8), 1469-1480.
16. Muller, T. J., Heuvelink, A., Bosch, K. van den, and Swartjes, I. (2012). Glengarry Glen Ross: Using BDI for Sales Game Dialogues. In: The Eighth Annual AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment.
17. Nijholt, A. and Heylen, D. (2002). Multimodal Communication in Inhabited Virtual Environments. International Journal of Speech Technology, vol. 5, issue 4, pp. 343-354.
18. Provoost, S. (2014). A Computational Model of Aggression De-escalation. M.Sc. Thesis, VU University Amsterdam. http://hdl.handle.net/1871/50480.