

On Conversational Agents with Mental States

Tibor Bosse^{1,2(✉)} and Simon Provoost¹

¹ Department of Computer Science, VU University Amsterdam,
De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands
{t.bosse, s.j.provoost}@vu.nl

² Department of Training and Performance Innovations, TNO, Kampweg 5,
3769 DE Soesterberg, The Netherlands

Abstract. Embodied conversational agents (ECAs) have been put forward as a promising means for the training of social skills. The traditional approach to drive the behaviour of ECAs during human-agent dialogues is to use conversation trees. Although this approach is easy to use and very transparent, an important limitation of conversation trees is that the resulting behaviour of the ECAs is often perceived as predictable. To provide ECAs with more sophisticated behaviour, the current paper proposes an approach to endow them with mental states. The approach is illustrated by a motivational example in the domain of aggression de-escalation training.

Keywords: Virtual training · Aggression de-escalation · Cognitive modelling

1 Introduction

Embodied conversational agents (ECAs) are 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’ [5]. ECAs have been put forward as a promising means for the training of social skills [8]. Recent applications can be found in domains varying from negotiation [9] to sales conversations [3].

To effectively train users in developing such social skills, an important requirement for ECAs is *believability*, as believable agents permit their conversation partners to ‘suspend their disbelief’, which is an important condition for learning [2]. Although much progress has been made with respect to the physical appearance of ECAs, it still remains difficult to develop agents with believable behaviour. The traditional approach to drive the 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 quite 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 very 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., [3, 7]). Using such models, agents base their behaviour not only on their current observations (or input), but also on internal mental states, for example an emotional state that resulted from previous interactions. The abstract nature of cognitive models however, makes it difficult to unify them with conversation trees.

Elaborating upon similar approaches (like [3, 7]), the current paper makes a step towards building a bridge between the traditional conversation tree approach (transparent, but rigid) and cognitive models (dynamic, but abstract). The approach is illustrated by an example in the domain of simulation-based training for aggression de-escalation.

2 Aggression De-escalation Training

Aggressive behaviour against employees in the public sector, such as tram drivers, police officers, and ambulance personnel, is an ongoing concern worldwide. 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 skills¹. 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 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.

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. Here, an important factor is the distinction between *reactive* and *proactive* aggression made within psychological literature: reactive aggression is characterised as an emotional reaction to a negative event that frustrates a person's desires, whereas proactive aggression is the instrumental use of aggression to achieve a certain goal [6]. 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 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, e.g. by making it clear that aggressive behaviour is not acceptable [1, 4, 10]. By ensuring that the ECAs respond in an appropriate manner to the chosen responses, the system provides implicit feedback on the chosen communication style.

¹ More information on this project, called 'Simulation-based Training of Resilience in Emergencies and Stressful Situations', can be found at <http://stress.few.vu.nl>.

3 Conversational Agents with Mental States

The proposed training system is based on the InterACT software², developed by the company IC3D Media³. InterACT is a software platform that has been specifically designed for simulation-based training. 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. The atomic ECA behaviours consist of pre-generated fragments of speech, synchronised with facial expressions and possibly extended with gestures.

Each decision node is implemented as a multiple choice menu. Via such a menu, the user has the ability 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 that were explained above. The other two options are more ‘extreme’ interventions, which should be applied, respectively, in case the aggressor has calmed down or in case the aggression is about to escalate, for example when personal threats are being made [10]. Additionally, the choice of the user determines how the scenario continues (or whether it ends immediately) by triggering a corresponding branch in the tree. Because a correct or wrong user choice is always followed by, respectively, a positive or negative ECA response, this approach is potentially predictable and repetitive.

We therefore propose to endow the ECA with an internal *state of aggression* that is represented numerically. Additionally, each ECA has a personality, which specifies whether it is a reactive or a pro-active aggressor. Based on this, the dynamics of the ECA’s state of aggression are influenced by the observed communication style of the user in the following way: if a reactive aggressor is approached in a supportive manner, he calms down, but if he is approached in a directive manner, he becomes more aggressive. For the proactive aggressor, this works exactly the other way around.

This approach allows us to create a large variation in scenarios with relatively limited effort, because the ECA’s internal states keep track of the *history* of the conversation. To start with, threshold values determine which ECA verbal response matches which level of aggression. 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. Because we no longer require a new user choice leading to the new ECA response, we can actually create more different scenarios with half the work (see Fig. 1).

Lastly, under the proposed approach, the precise path that is taken through the conversation tree no longer solely depends on what the user does, but also on the

² <http://www.interact-training.nl/>.

³ <http://ic3dmedia.com/>.

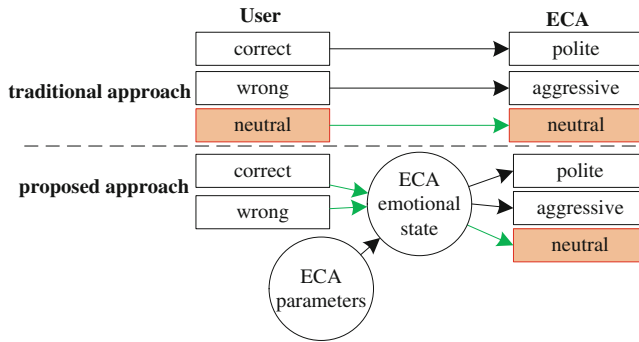


Fig. 1. The benefits of the proposed approach. The red verbal behaviours entail the extra work involved when adding a new ECA response, the green arrows the resulting new scenarios.

ECA's personality, i.e. the nature of the ECA's aggression and the parameter configuration that determines the rate at which its internal states change.

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. Anderson, L.N., Clarke, J.T.: De-escalating verbal aggression in primary care settings. *Nurse Pract.* **21**(10), 95–102 (1996)
2. Bates, J.: The role of emotions in believable agents. *Commun. ACM* **37**(7), 122–125 (1994)
3. Van den Bosch, K., Brandenburgh, A., Muller, T.J., Heuvelink, A.: Characters with personality! In: Nakano, Y., Neff, M., Paiva, A., Walker, M. (eds.) IVA 2012. LNCS, vol. 7502, pp. 426–439. Springer, Heidelberg (2012)
4. Bosse, T., Provoost, S.: Towards aggression de-escalation training with virtual agents: a computational model. In: Zaphiris, P., Ioannou, A. (eds.) LCT. LNCS, vol. 8524, pp. 375–387. Springer, Heidelberg (2014)
5. Cassell, J., Sullivan, J., Prevost, S., Churchill, E.: *Embodied Conversational Agents*. MIT Press, Cambridge (2000)
6. Dodge, K.A.: The structure and function of reactive and proactive aggression. In: Pepler, D., Rubin, H. (eds.) *The Development and Treatment of Childhood Aggression*, pp. 201–218. Erlbaum, Hillsdale (1990)
7. Gebhard, P., Kipp, M., Klesen, M., Rist, T.: Adding the emotional dimension to scripting character dialogues. In: Rist, T., Aylett, R.S., Ballin, D., Rickel, J. (eds.) IVA 2003. LNCS (LNAI), vol. 2792, pp. 48–56. Springer, Heidelberg (2003)
8. Kenny, P., Hartholt, A., Gratch, J., Swartout, W., Traum, D., Marsella, S., Piepol, D.: Building interactive virtual humans for training environments. In: *Proceedings of 2007 Interservice/Industry Training, Simulation and Education Conference*, Orlando (2007)

9. Kim, J., Hill, R.W., Durlach, P., Lane, H.C., Forbell, E., Core, C., Marsella, S., Pynadath, D., Hart, J.: BiLAT: a game-based environment for practicing negotiation in a cultural context. *Int. J. AI Educ.* **19**(3), 289–308 (2009)
10. Ministry of the Interior and Kingdom Relations: Handboek agressie en geweld - voorkomen, beperken, afhandelen. Technical report (in Dutch) (2008)