Virtual Agents for Professional Social Skills Training: An Overview of the State-of-the-Art

Kim Bosman[1], Tibor Bosse[2(✉)], and Daniel Formolo[1]

[1] Department of Computer Science, Vrije Universiteit, Amsterdam, The Netherlands
[2] Behavioural Science Institute, Radboud Universiteit, Nijmegen, The Netherlands
t.bosse@ru.nl

Abstract. Training of interpersonal communication skills is typically done using role play, by practising relevant scenarios with the help of professional actors. However, as a result of the rapid developments in human-computer interaction, there has been an increasing interest in the use of computers for training of social and communicative skills. This type of training offers opportunities to complement traditional training methods with a novel paradigm that is more scalable and cost-effective. The main idea of such applications is that of a simulated conversation between a human trainee and a virtual agent. By developing the system in such a way that the communicative behaviour of the human has a direct impact on the behaviour of the virtual agent, an interactive learning experience is created. In this article, we review the current state-of-the-art in virtual agents for social skills training. We provide an overview of existing applications, and discuss various properties of these applications.

Keywords: Review · Virtual agents · Social skills training · Serious games

1 Introduction

Having good interpersonal communication skills is an important ability for human beings to be able to function in daily life. According to [4], interpersonal communication skills is an umbrella term that covers a number of core competencies, including non-verbal communication, questioning, reinforcement, reflecting, explaining, self-disclosure, listening and humour. Unfortunately, the extent to which people possess these skills varies greatly per individual. The good news is that communication skills can be trained, at least to a certain extent. To enable professionals to practice and improve the social skills they require for their job, organizations invest a lot of time and money into training programs. Traditionally, such training programs make use of role play, through which participants can practice certain simulated scenarios either with professional actors or with teachers or classmates [10,14].

P. Cortez et al. (Eds.): INTETAIN 2018, LNICST 273, pp. 75–84, 2019.
https://doi.org/10.1007/978-3-030-16447-8_8

Although this type of training can be reasonably effective, it suffers from several drawbacks. First, organizing training sessions is very costly, both in terms of money and time. As a result, the frequency by which they are offered is low. Second, there are large differences in the successfulness of role-play-based training: for some students, the learning effect is large, whereas for others it is minimal. And third, training is never fully completed. As argued in [17], employees need frequent refreshing sessions, which often conflict with regular work schedules. In conclusion, existing approaches are hard to tailor to individual needs, and difficult to combine with work schedules.

As a complementary approach to role play, communication skills can be trained via serious games. According to [19], a Serious Game is "a mental contest, played with a computer in accordance with specific rules, that uses entertainment to further government or corporate training, education, health, public policy, and strategic communication objectives". Within the serious games domain, there has been increasing interest in the use of Intelligent Virtual Agents (IVAs) for training of social skills. IVAs can be defined as 'intelligent digital interactive characters that can communicate with humans and other agents using natural human modalities like facial expressions, speech, gestures and movement' [2].

The current article describes a number of representative applications in the area of IVAs for social skills training. The main purpose of this paper is not to describe new research results, nor to provide an exhaustive literature review. Instead, it is meant to provide a high level overview of existing approaches regarding IVAs for social skills training. A representative selection of recent applications is reviewed, and the systems are categorized according to a list of characteristics such as the application domain, the interaction modalities, and the extent to which the system has been evaluated.

2 IVAs for Social Skills Training

IVAs for social skills training are typically part of a larger system that can be called a Virtual Learning Environment. In [7], Virtual Learning Environments are defined as 'a multi-dimensional experience which is totally or partially computer generated and can be accepted by the participant as cognitively valid'. The key idea is that the user's senses are stimulated in such a way that the virtual environment is almost experienced as a real environment. Users or players are considered those who play the a serious game with the deliberate aim of improve the abilities for which the game is proposed.

Over the years, the graphics of virtual environments have become increasingly realistic, mainly due to the developments in the video games and military simulation industry [19]. Moreover, recent developments in Artificial Intelligence have paved the way for virtual environments for social skills training, which is the focus of the current paper.

Social skills are those abilities that people use to communicate with each other, both verbally and non-verbally, and it is important to be able to show the appropriate verbal and non-verbal behaviours. To be able to develop these skills in a simulation-based environment, it is important that the virtual environment closely resembles the real environment. Another important aspect that determines the user's experience is the extent to which the interaction between the user and the IVA is perceived as natural or *believable* [1]. That could be measured through surveys or biofeedback signals monitoring user's body reactions to the IVA's acts. If the virtual characters do not behave like people would normally behave in a particular situation, the credibility of the simulation decreases, which may in turn decrease the learning effect [3]. Other parameters that determine the quality of the experience include the graphics frame rate, the tracking capabilities, tracking latency (the time it takes before a head movement results in the correct change in the display), image quality, the amount of field a user can see, the behaviour of objects, and the range of sensory accommodations [12].

Furthermore, the success of a virtual learning environment depends to a considerable extent on the user's acceptance. As argued in [16], individually tailored e-learning environments will have a higher acceptance rate. Nevertheless, in many cases, e-learning has been found to be just as effective (e.g., [18]), or even more effective [15] than class-based learning.

In order to build IVAs for social skills training, typically a modular approach is taken, where developers first create separate modules for specific capabilities of the agent, and then integrate them into a coherent system. An overview of the various capabilities that an IVA might have, taken from [5], is shown in Fig. 1. This figure shows the possible capabilities of an IVA (represented by the rectangles) and their interactions at an abstract level.

The details of Fig. 1 are beyond the scope of this paper, but a rough summary is as follows. Typically, a human user (lower-left) interacts with a virtual agent of which the behaviour is displayed via a renderer (lower-right). The four rectangles on the left hand side of the figure represent processing of the user's *input*, of which the two modules on the left deal with non-verbal information and the modules on the right with verbal information. Similarly, the four rectangles on the right hand side of the figure are about generating the agent's *output*. Here, the two modules on the right deal with non-verbal information (e.g., displaying facial expressions on the agent's face) and the modules on the left with verbal information (i.e., determining what the agent says). The agent module, shown in the upper part of the picture, is an *internal* layer that connects the input to the output. For instance, a simple way to implement this would be to use a fixed question-answering mechanism that generates pre-defined responses for certain questions asked by the user [9]. However, more complex implementations make use of sophisticated dialogue managers that keep track of the progress of the conversation with the user.

As displayed in Fig. 1, on top of a system another layer is implemented to generate explicit *feedback* (e.g., using computational models of the task and the user's performance [6]). Providing feedback on the performance of the trainee is an important mechanism to facilitate learning. Within the context of IVA-based training, feedback may have the form of hints to inform the trainee that certain behaviour during the simulated scenario was appropriate or inappropriate. In

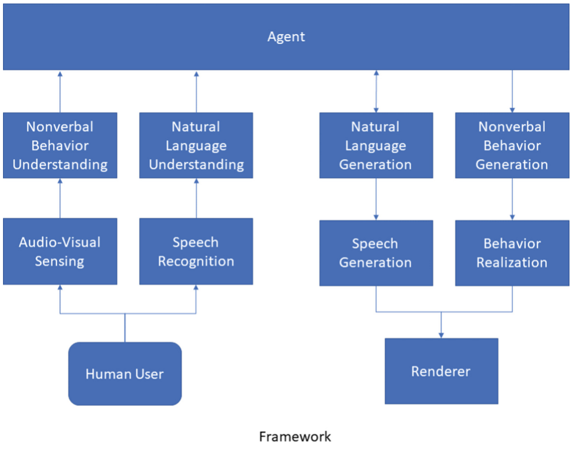

Fig. 1. ICT virtual agent architecture

addition, it is often claimed that after-session feedback is an effective method to enhance learning [13].

Finally, an important question to be addressed when developing IVA-based systems for social skills training is to what extent they are actually effective in enhancing a person's communication skills? When it comes to evaluation of training interventions, Kirkpatrick's framework is a useful instrument [8]. It distinguishes four evaluation levels, named (1) reaction ('do participants like the training?'), (2) learning ('do participants acquire the intended skills?'), (3) behavior ('do participants apply the learned behavior in practice?'), and (4) results ('does the training result in the targeted outcomes?').

The next sections discuss a number of recent projects involving IVAs for social skills training. First, in Sect. 3, the separate projects are briefly summarized. After that, in Sect. 4, they are compared according to a list of characteristics. The characteristics that are used for the comparison directly follow from the discussion above (and are related to the terms written in italics in the current section), namely: input, output, internal, feedback and evaluation. A sixth characteristic, *formalism*, has been added to provide information about the type of modelling framework or formal representation that has been used (e.g., AIML or Finite State Machines).

3 A Selection of Existing Applications

The search engines *Google Scholar*, *IEEE*, and *ACM library* were used to find relevant articles. Articles from before the year 2000 were not considered. Search terms like *Conversational Agent, Virtual Human, Virtual Agent* and *Avatar* were used in combination with terms like *Social Skills Training* and *Conversational Skills Training*. This resulted in slightly less than 1000 articles. However, a substantial number of search results turned out to be out of scope, so many of them were disregarded after reading the abstract. The main inclusion criteria were (1) the presence of a visually embodied agent, (2) the aim to train people's social or communicative skills, and (3) the presence of an implemented system. Hence, papers describing purely theoretical models or partially implemented systems were discarded. Also, papers describing commercial applications were excluded because they normally don't provide any details about the implementation of the system. Moreover, this literature study focuses on applications aimed to improve social skills required for professionals in work environments, e.g., in domains like healthcare, education, law enforcement and military. In contrast, it does not cover social skills training in psychotherapeutic context. Hence, also these articles were discarded. For this domain, an extensive review has been conducted by Provoost and colleagues [11].

Table 1. Summary of the twelve applications

Project	Input	Output	Internal
ASST *Decrease discomfort*	Speech	Speech Facial expressions	Scenario engine
Believable Suspect Agents *Police interrogations*	Free text Speech	Speech	Response model based on stance
BiLAT *Cross-cultural negotiations*	Specified menu choices (say, ask, give, do)	Speech Gestures	Rule-based responses
ColCoMa *Conflict management*	Free text	Text Facial expressions	AIML-processor
Communicate! *Communication skills*	Specified menu choices	Emotion Speech	Consultation graphs
deLearyous *Communication skills*	Free text	Pre-recorded audio Visual feedback	Scenario engine using sentiment
INOTS *Counseling*	Specified menu choices Heart rate, EDA	Speech	Conversation trees
MRES *Critical decision-making*	Speech	Speech, Gestures Facial expressions	Focus mechanism via Soar architecture
STRESS *Aggression de-escalation*	Specified menu choices Heart rate, EDA, EEG	Pre-recorded audio Facial expressions Gestures	Conversation trees
Virtual Patient *Medical interviews*	Free text Gestures	Facial expressions Text	AIML-processor
Virtual Recruiter *Job interviews*	Speech Multi-modal cues	Speech	Sequential behaviour planner based on stance
Virtual-Suspect *Police interrogations*	Free text	Text	Scenario engine

Once a relevant paper was found, the references used in this article were checked as well. This resulted in a selection of twelve papers. As the papers describe rather diverse approaches and application domains, we feel that this selection provides a fairly representative overview of the current state-of-the-art. However, we do not claim that this overview is exhaustive. The selected papers have been categorized and compared according to the criteria identified in Sect. 2; see Tables 1 and 2.

Table 2. Summary of the twelve applications (continued)

Project	Feedback	Evaluation	Formalism
ASST *Decrease discomfort*	Hints afterwards Scores	Level 2	MMDAAgent Snak Sound Toolkit
Believable Suspect Agents *Police interrogations*	Mood changes Thought bubbles Final reflection	Level 2	Interpersonal Circumplex NPCEditor
BiLAT *Cross-cultural negotiations*	Reflective tutor State changes	Level 2	PsychSim Intelligent tutoring system
ColCoMa *Conflict management*	Feedback from opponent Textual feedback Replay session	Level 1	AIML-Chatbot Facial animations C# and .NET
Communicate! *Communication skills*	Emotion changes Annotated textual feedback Scored goals	Level 1	Domain reasoner
deLearyous *Communication skills*	Position change	Unknown	NLP Interpersonal Circumplex Finite State Machine
INOTS *Counseling*	After-action review Homework review	Level 2	I-CARE framework LiSA-CARE framework
MRES *Critical decision-making*	Run-time adjustments	Unknown	Multigen-Paradigm's PSERT, DIRM
STRESS *Aggression de-escalation*	Run-time adjustments Hints afterwards	Level 2	InterACT Adaptive training
Virtual Patient *Medical interviews*	Annotated transcript Scores	Level 1	AIML-chatbot HTML, CSS, PHP, JS
Virtual Recruiter *Job interviews*	Mood changes Personality changes	Level 3	SSI Interpersonal Circumplex
Virtual-Suspect *Police interrogations*	After-action review Mood changes	Level 1	Not reported

4 Overview

This section provides an overview of the twelve applications that have been discussed. As mentioned in Sect. 2, the applications are compared according to six characteristics: input, output, internal, feedback, evaluation and formalism. The results are shown in Tables 1 and 2.

As becomes clear from the tables, there is a wide variety in the approaches used to train social skills using IVAs. However, one can also see that some methods are more commonly used than others. Below, these similarities and differences are discussed per characteristic. However, it is important to point out that the papers that were reviewed differed with respect to the level of detail in which the application was described. The comparisons made in the tables are solely based on the information that was available.

4.1 Input

There are roughly three approaches to allow the user to provide verbal input to the system, namely: free speech, free text input, and pre-determined multiple choice options. Obviously, each of these options has its pros and cons. Typically, interacting with an IVA using free speech is perceived as more natural than typing text, or selecting options from a multiple choice menu. In addition, the latter brings along a risk that users feel forced to select answers that they would never give in real life. On the other hand, free speech or text is clearly more difficult to process (on a semantic level) than pre-defined sentences, which increases the risk that the system generates inappropriate responses or 'backup' responses like 'I do not understand, please rephrase'.

Regarding non-verbal input, some systems (e.g. Virtual Recruiter) extract social cues from multi-modal input such as facial expressions or gestures. The purpose of this is to understand not only what is said, but also 'how' the user says something. In addition, INOTS and STRESS take physiological signals like heart rate, EDA or EEG signals into account.

4.2 Output

Also on the level of the output (i.e., the behaviour displayed by the IVA) a distinction can be made between verbal and non-verbal aspects. Regarding verbal aspects, all systems use either text or speech, with the latter being most popular. Speech is either generated based on pre-recorded audio files or is generated at run-time using text-to-speech engines. Non-verbal elements are not used by all systems, but in several cases they are used to enhance the believability of the agents. The non-verbal cues that are used are mostly facial expressions and (less frequently) gestures.

4.3 Internal

To determine how the virtual agent should respond, different methods are available. The trigger for activating these methods is the input from the player (see Sect. 4.1). However, depending on the method, the input is either directly mapped to output, or is first interpreted in terms of higher level intermediate constructs. More advanced systems first try to make an interpretation of the user's (verbal and/or non-verbal) input, for instance in terms of the atmosphere of the conversation (MRES or Virtual Patient), sentiment of the text (deLearyous), or the interpersonal stance that is taken (Believable Suspect Agents or Virtual Recruiter). Subsequently, this intermediate construct is then processed by some agent model. to determine on a high level how the agent should respond. An alternative method is to use markup languages for natural language generation as part of the internal module (ColCoMa or Virtual Patient).

Again, there is a large variety in the approaches that are taken, and there is no clear best approach. The advantage of simple input-output mappings clearly is that they are easy to handle by the developer. However, a drawback is that they may result in agent behaviour that is perceived as static and inflexible. This may be sufficient for some applications , but it is detrimental for others . In such cases, often more complex agent models are used, to allow for a wider variety of IVA behaviour. Another advantage of such complex approaches is that the history of the conversation can be taken into account (e.g., the IVA may still be in a bad mood because of something the user said some time ago), which is impossible with simple input-output mappings.

Finally, it is worth mentioning that almost all of the systems use a 'turn-taking' protocol, where the user is only allowed to provide input after the agent has finished speaking (and vice versa). An exception is ColCoMa, where the interaction between two human players is mediated by a chatbot.

4.4 Feedback

As mentioned earlier, providing feedback on the performance of the trainee is an important mechanism to facilitate learning. In the papers that were reviewed, various forms of feedback were encountered: feedback during the simulation, after the simulation, by a virtual coach, by human instructors, on paper, and via a replay of the simulation. Most of the applications offer either a textual summary (sometimes with notes) or an actual sit-down with a human coach to review the process. Besides these 'after-action reviews', it is also common to adjust the scenario while it is still running. The main idea behind these run-time adjustments is that the player receives immediate feedback on his or her choices during the interaction. For instance, if the learning goal of a system is to show empathy to frustrated customers, the IVA can be implemented in such a way that it calms down if the user takes an empathic stance, but otherwise becomes even more aggressive. This way, the behaviour of the IVA functions as an implicit reward or punishment, hence facilitating a kind of associative learning process.

4.5 Evaluation

The twelve applications are classified on the basis of Kirkpatrick's four levels of evaluation [8]. Most applications have been evaluated on the levels 1 or 2. Based on the information that was available, we categorized ColCoMa, Communicate!, Virtual Patient and Virtual-Suspect into Level 1; ASST, Believable Suspect Agents, BiLAT, INOTS and STRESS into Level 2; Virtual Recruiter reaches level 3. MRES and deLearyous did not provide enough information to be classified into one of the levels. The fact that only one of the projects went beyond level 2 can probably be explained by the difficulty to measure the real impact of training intervention, as well as the costs (in terms of time and money) involved in it. Also, as most of the applications were the result of academic endeavours from computer scientists, more extensive evaluation efforts were probably not high on their priority list. Nevertheless, to make IVAs for social skills training more widely adopted, it would be wise to spend more time on longitudinal studies with the aim to assess how effective these systems are in changing a person's behaviour.

4.6 Formalism

The column on formalisms has been included to provide an overview of the programming languages, modelling frameworks, and other tools that have been used to implement the IVA-based systems that were reviewed. As can be seen, the technology used varies from standard programming language (such as C#) and general AI tools (such as AIML) to more dedicated agent-based development frameworks (such as InterACT or PsychSim). Clearly, as each project uses its own approach, it is hard to draw any useful conclusion from this information. Perhaps the most important lesson that can be learned from this is that it is advisable to strive towards a more uniform standard for the development of IVAs (e.g., the Virtual Human Toolkit [5]).

5 Conclusion

This paper discussed twelve different applications which all share the aim to improve a user's social skills. The focus was on social skills training in the professional domain. Although there was a wide variety in the approaches taken to reach this goal, there are also some similarities between different applications. An overview of the differences and similarities can be found in Tables 1 and 2. However, it is important to note that not all papers provided the same amount of background information.

It is impossible to conclude that there is one single approach that works best in all situations. Rather, the choice for a certain paradigm or technology should depend on the purpose of the training application. As a general approach, when developing an IVA-based training system, it is useful to view the envisioned system in terms of the architecture displayed in Fig. 1. Then, for each module in the architecture, an entire spectrum of methods is available (e.g., for the 'input part' one can distinguish between free speech, free text, multiple choice, etc). The developer should select the method that is most suitable for the intended purpose, considering the relevant financial, temporal and other constraints. In addition, more effort should be spent on long-term studies that assess how effective IVA-based systems really are in changing a person's behaviour, and which factors contribute to that.

Acknowledgements. This research was supported by the Brazilian scholarship program Science without Borders - CNPq scholarship reference: 233883/2014-2.

References

1. Bates, J., et al.: The role of emotion in believable agents. Commun. ACM **37**(7), 122–125 (1994)
2. Beskow, J., Peters, C., Castellano, G., O'Sullivan, C., Leite, I., Kopp, S.: Intelligent Virtual Agents. LNCS. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67401-8

3. Bosch, K., Bosse, T., de Jong, S.: Trainen met gesimuleerde mensen: Effectief inzetten van gedragsmodellen voor militaire training. Militaire Spectator **9**(184), 358–373 (2015)
4. Hargie, O.: Handbook of Communication Skills. Routledge, Abingdon (1997)
5. Hartholt, A., et al.: All together now. In: Aylett, R., Krenn, B., Pelachaud, C., Shimodaira, H. (eds.) IVA 2013. LNCS (LNAI), vol. 8108, pp. 368–381. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40415-3_33
6. Heuvelink, A., Mioch, T.: FeGA: a feedback generating agent. In: Proceedings of the Seventh IEEE/WIC/ACM International Conference on Intelligent Agent Technology, IAT 2008, pp. 567–572 (2008)
7. Jense, G., Kuijper, F.: Virtual environments for advanced trainers and simulators. Z. pl.: Z. uitg. (1993)
8. Kirkpatrick, J.D., Kirkpatrick, W.K.: Kirkpatrick's Four Levels of Training Evaluation. Association for Talent Development, Alexandria (2016)
9. Leuski, A., Traum, D.: NPCEditor: creating virtual human dialogue using information retrieval techniques. AI Mag. **32**(2), 42–56 (2011)
10. Maguire, P., Pitceathly, C.: Key communication skills and how to acquire them. BMJ **325**(7366), 697–700 (2002)
11. Provoost, S., Lau, H.M., Ruwaard, J., Riper, H.: Embodied conversational agents in clinical psychology: a scoping review. J. Med. Internet Res. **19**(5), e51 (2017)
12. Slater, M.: Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. Phil. Trans. Roy. Soc. B: Biol. Sci. **364**(1535), 3549–3557 (2009)
13. Slovak, P., Thieme, A., Murphy, D., Tennent, P., Olivier, P., Fitzgerald, G.: On becoming a counsellor: challenges and opportunities to support interpersonal skills development. In: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work and Social Computing, CSCW 2015, pp. 1336–1347 (2015)
14. Stokoe, E.: Simulated interaction and communication skills training: the 'Conversation-Analytic Role-Play Method'. In: Antaki, C. (ed.) Applied Conversation Analysis, Palgrave Advances in Linguistics. Palgrave Macmillan, London (2011). https://doi.org/10.1057/9780230316874_7
15. Todorov, E., Shadmehr, R., Bizzi, E.: Augmented feedback presented in a virtual environment accelerates learning of a difficult motor task. J. Motor Behav. **29**(2), 147–158 (1997)
16. Van Raaij, E.M., Schepers, J.J.: The acceptance and use of a virtual learning environment in China. Comput. Educ. **50**(3), 838–852 (2008)
17. Wang, S., O'Brien-Pallas, L.L., Hayes, L.: A Review and Evaluation of Workplace Violence Prevention Programs in the Health Sector. Nursing Health Services Research Unit, Toronto (2008)
18. Wolf, T.: Assessing student learning in a virtual laboratory environment. IEEE Trans. Educ. **53**(2), 216–222 (2010)
19. Zyda, M.: From visual simulation to virtual reality to games. Computer **38**(9), 25–32 (2005)