

The TARDIS framework: intelligent virtual agents for social coaching in job interviews

K. Anderson, E. André, T. Baur, S. Bernardini, M. Chollet, E. Chryssafidou, I. Damian, C. Ennis, A. Egges, P. Gebhard, H. Jones, M. Ochs, C. Pelachaud, K. Porayska-Pomsta, P. Rizzo, and N. Sabouret

The TARDIS consortium

Abstract. The TARDIS project aims to build a scenario-based serious-game simulation platform for NEETs and job-inclusion associations that supports social training and coaching in the context of job interviews. This paper presents the general architecture of the TARDIS job interview simulator, and the serious game paradigm that we are developing.

1 Introduction

The number of NEETs¹ is increasing across Europe. According to Eurostat ², in march 2012, 5.5 million of 16 to 25 years old European youngsters were unemployed, amounting to 22.6% of the youngster global population. This unemployment percentage is 10 points above the entire world's population, highlighting European youth unemployment as a significant problem.

Current research reveals that NEETs often lack self-confidence and the essential social skills needed to seek and secure employment [BP02]. They find it difficult to present themselves in a best light to prospective employers, which may put them at further risk of marginalization. Social coaching workshops, organized by youth inclusion associations across Europe, constitute a common approach to helping people in acquiring and improving their social competencies, especially in the context of job interviews. However, it is an expensive and time-consuming approach that relies on the availability of trained practitioners as well as the willingness of the young people to engage in exploring their social strengths and weakness in front of their peers and practitioners.

The TARDIS project³, funded by the FP7, aims to build a scenario-based serious-game simulation platform that supports social training and coaching in the context of job interviews. It participates in the Digital Games for Empowerment and Inclusion initiative⁴. The platform is intended for use by young people, aged 18-25 and job-inclusion associations. Youngsters can explore, practice and improve their social skills in a diverse range of possible interview situations. They interact with virtual agents, which are designed to deliver realistic socio-emotional interaction and act as recruiters.

¹ NEET is a government acronym for young people not in employment, education or training.

² ec.europa.eu/eurostat

³ <http://www.tardis-project.eu>

⁴ <http://is.jrc.ec.europa.eu/pages/EAP/eInclusion/games.html>

The use of serious games for job interview simulations has two advantages: 1) repeatable experience can be modulated to suit the individual needs, without the risk of real-life failure; 2) technologies are often intrinsically motivating to the young [MSS04] and may be used to remove the many barriers that real-life situations may pose, in particular the stress associated with engaging in unfamiliar interactions with others. In this context, the originality of the TARDIS platform is two-fold. First, TARDIS is able to detect in real-time its users' emotions and social attitudes through voice and facial expression recognition, and to adapt the game progress and the virtual interlocutor's behaviour to the individual users. Second, it provides field practitioners with unique tools for 1) designing appropriate interview scenarios without reliance on researchers and 2) measuring individuals' emotion regulation and social skill acquisition (*via* the user modelling tools), thus enabling a more flexible and personalized coaching for young people at risk of social exclusion.

This paper presents the TARDIS architecture and game design. The next section briefly presents some related work in serious games and affective computing. Section 3 presents the TARDIS architecture. Section 4 presents the TARDIS game itself.

2 Related work

Addressing social exclusion of marginalized young people is a key issue that must be tackled on a wide range of different levels. The INCLUSO project [EDVDB10] outlined two main directions in supporting social integration: (1) to encourage and support personal development and (2) to encourage and support social participation. The TARDIS project focuses on the first challenge, *i.e.* on improving communication skills in one-on-one interactions. Based on the analysis that computer games as well as serious games are intrinsically motivating to users (e.g. see [MC05]), we have developed a virtual social coaching platform that follows a serious games paradigm, in which the youngster can train and evaluate his/her social interaction skills with a virtual agent.

The use of virtual agents in social coaching has increased rapidly in the last decade. Projects such as those by Tartaro and Cassell [TC08] or e-Circus [APD⁺09] provide evidence that virtual agents can help humans improve their social skills and, more generally, their emotional intelligence [Gol06]. In job interview situations, emotional intelligence of both the applicant and the recruiter plays a crucial role: in human to human job interviews, the personality of the applicant is inferred by the recruiter according to the mood, the emotions and the social attitudes expressed by the youngster [HH10]. Furthermore, visual and vocal perceptions of the interviewer have been shown to affect their judgments of the applicants during job interviews [DM99]. Youngsters' emotional intelligence, in particular their ability to self-regulate during the interview will affect the outcome of the interview.

In the TARDIS project, we address two challenges. First, in line with the objectives of the field of *Affective Computing* [Pic97], we aim to build virtual agents that react in a coherent manner (see also [MGR03,PDS⁺04,PSS09]). Based on the non-verbal inputs (smiles, emotion expressions, body movements) and their goals (making the applicant at ease or, on the contrary, trying to put him/her under pressure), the agent must select relevant verbal and non-verbal responses. Second, the model presented in this paper

seeks to take into account all the different dimensions of the socio-affective interaction, in the context of a job interview situation, including indepth real-time understanding of the individual youngsters emotional states while they are interacting with the system.

A number of serious games exist that relate to different aspects of TARDIS. First, the utilization of Microsoft Kinect has become popular as an affordable markerless motion capture system, and has been utilized successfully in many games with serious applications. Mostly, however, these games tend to be applied in the medical field, with applications for stroke rehabilitation patients [JYZ12] and people suffering with Parkinsons [SS12]. However, when it comes to reading more complex parameters, such as emotions, regular cameras are also frequently used.

Some examples of serious games that infer complex paradigms about the state of mind of the user from visual and audio inputs are *LifeIsGame* [Orv10] and *ASC Inclusion* [SMBC⁺13]. Both are ongoing projects within the EC framework and focus on helping children on the Autism spectrum to recognize and express emotions through facial expressions, tone of voice and body gestures. While these games access the facial expressions and behaviour of the user in a similar manner to TARDIS, virtual humans will interact based on low-level signal analysis as opposed to the virtual recruiter in TARDIS who will respond to high-level socio-emotional behaviours and mental states' representations.

Some serious games also focus on employment and the work place. A serious game with some similar objectives to TARDIS is interview training game *Mon Entretien D'embauche* [dR13]. Released by telecoms company SFR, the user is represented by an avatar and progresses through a number of events before the interview itself by selecting appropriate responses to interactions with other avatars. While the goal to relieve stress or anxiety is similar to that of TARDIS, the approach is quite different. While *Mon Entretien D'embauche* assesses progress based on the verbal responses of the player, TARDIS focuses on whether the player is completely engaged and displaying appropriate non-verbal social behaviours. Another workplace focused game is *iSpectrum* [Ins13], which aims to help people on the Autism Spectrum to prepare themselves for working in a new environment and increase their employability. It can also educate employers, increasing their knowledge about integrating someone with Autism within their company.

My Automated Conversation coach, *MACH* [HCM⁺13] utilizes a virtual human to coach the user's social communication skills. While one of the applications of this game is for interview situations, the focus here is on the facial expressions, speech and prosody. The intelligent agent here will mimic the users behaviour and respond verbally and non-verbally, as was done in *Semaine* [BDSHP12]. However, while TARDIS focuses on the underlying emotion or state-of-mind, MACH provides feedback on physical behaviours such as explicit facial expressions and tone and volume of voice.

3 The TARDIS architecture

3.1 General architecture

Social coaching involves three actors: the participant (the youngster), the interlocutor, which is replaced by an intelligent virtual agent in the TARDIS game, and the coach or

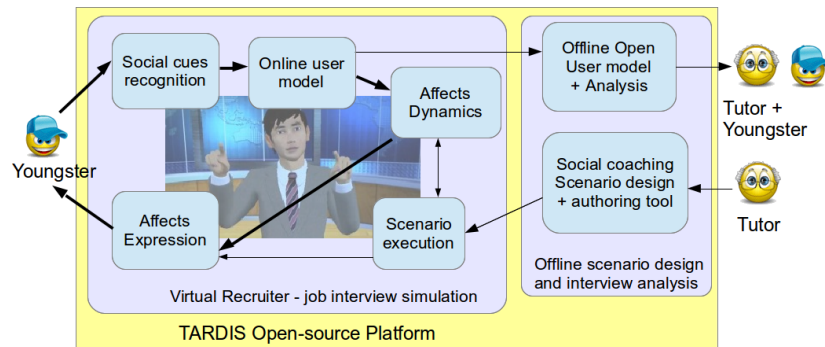


Fig. 1. TARDIS general architecture

practitioner from the social inclusion association. In this context, the TARDIS system architecture includes four main components:

- **The Scenario module** controls the discourse during the interview. It tells the virtual recruiter the expectation in terms of emotions and attitudes expressed by the youngster, depending on the stage of the interview (*Scenario Execution* in Fig. 1, Section 3.2).
- **The Social Signal modules** provide the affective model with information about the youngster’s emotions and social attitudes (we refer to them as *mental states*) that are detected by the system (*Social Cues Recognition and Online User Model* in Fig. 1, Section 3.3 and 3.4).
- **The Affective Module** is responsible for building a model of beliefs and intentions (for the virtual recruiter) about the mental states of the youngster (in terms of affects) and about the course of actions in the ongoing interview (*Affects Dynamics* in Fig. 1, Section 3.5).
- **The Animation module** is responsible for expressing the virtual recruiter’s affective state built by the affective model through verbal and non-verbal behaviour, both in terms of facial animation and body movements (*Affects Expressions* in Fig. 1, Section 3.6).

The interaction is recorded and post-game debriefing sessions can be organized between the practitioner and the youngster. Fig. 1 gives an overview of this architecture. The following subsections briefly present the different components of the platform. Section 4 presents the TARDIS game paradigm.

3.2 The Scenario Module

For the authoring of our interactive virtual recruiter’s behaviour, we rely on an authoring tool [GMK12] that allows to model and to execute different behavioural features at a very detailed and abstract level. A central authoring paradigm of this tool is the separation of dialog content and interaction structure, see Fig. 2.

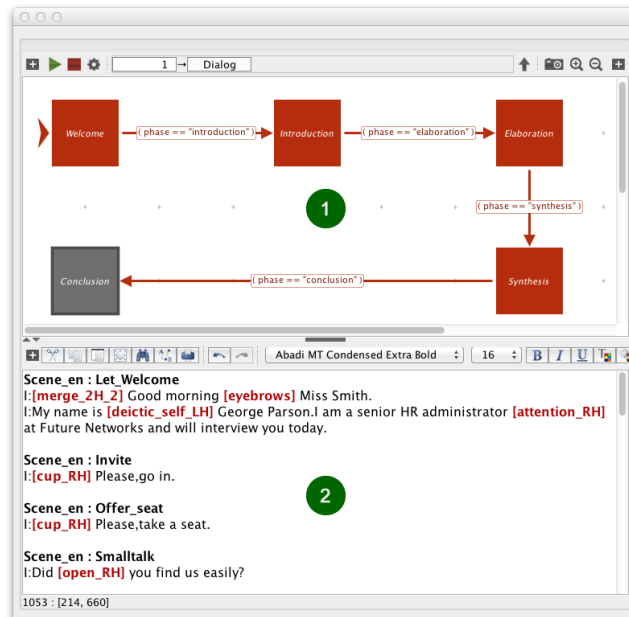


Fig. 2. Main five stages of the recruitment scenario modelled as HSFMs.

The multimodal dialog content is specified with a number of scenes that are organized in a *scenescrypt* (see Fig. 2, panel 2). The scene structure, which can be compared to those in TV or theatre playbooks, consists of utterances and stage directions for the actors. In our scenes, directions are animation commands for the non-verbal behaviour of the virtual recruiter (e.g. gestures, facial expressions, or postures). The (narrative) structure of our interactive recruitment simulation and the interactive behaviour of our virtual recruiter is controlled by parallel hierarchical finite state machines (HFSM) specifying the logic of the scenes played and the commands executed according to a user reactions and a given state of the interactive performance, see Fig. 2, panel 1.

In contrast with a linear theatre scene arrangement, an interactive presentation comes along with another degree of freedom, *i.e.* the reactions of the system to user's detected affects or to the virtual recruiter's mental state. Such reactions have to be considered in the scenario control. In order to realize this, we have enhanced a linear five stage recruitment story (see Fig. 2, panel 1.) with reactions to user input. During the five stages, depending on the attitude that the virtual recruiter wants to convey (see Section 3.5), the virtual recruiter may mimic specific behaviours of the user. For this task, we use two HFSMs (modelled within each recruitment stage) that allow the character to react in a similar way as the user's behaviours. The state machines react to the detected social cues and trigger an overlay behaviour. This behaviour is blended with any on-going animation that the virtual recruiter is performing at any time (Section 3.6). For example, if a user smiles while the virtual recruiter asks a question, the recruiter will also smile. Similar behavioural reactions are modelled in the case of head movements.

In order to detect and interpret the social cues expressed by the youngster during the interview, we have integrated social signal modules. In the next section, we first

present the module for the recognition of the youngster's social signals. In section 3.4, the module for the interpretation of the recognized social signals are introduced.

3.3 The Social Signal Recognition Module

The social signal recognition module is able to record and analyze social and psychological signals from users and to recognize predefined social cues, i.e. behaviours of the interviewee, conscious or unconscious, that have a specific meaning in a social context such as a job interview, in real-time. Examples of such signals are smiles, head nods and body movements. These social cues will allow us to determine the mental state of the user (Section 3.4) in real-time and will act as a basis for debriefing sessions between youngsters and practitioners after the interviews.

In TARDIS the social cues were selected based on two criteria: pertinence to the job interview context and automatic recognition viability from a technological point of view. In order to determine what social cues are pertinent to the desired context, we evaluated the data from multiple user studies involving real NEETs and practitioners. The technological viability was determined using a literature review as well as in-depth analysis of available state-of-the-art sensing hardware. This is an ongoing process and we are continuously looking to extend the list of social cues.

The system uses a combination of sensors and software algorithms which offer good results in terms of accuracy and low intrusion. High accuracy ensures that a youngster's social cues are correctly recognized and allows the virtual recruiter to react to them correctly. It is equally important that the approach has a low intrusion factor. For example, biological signal sensors are not feasible in this scenario because attaching various sensors to the skin of the users will most likely result in an increase in stress which might have a negative effect on the user's job interview performance, but may not be actually indicative of the user's actual abilities. Therefore, in the context studied, remote sensors are preferred.



Fig. 3. Examples of the gestures our system can recognise.

For recording and pre-processing human behaviour data, our system relies on the SSI framework⁵ which was developed as part of our previous work [WLB⁺13]. It provides a variety of tools for the real-time analysis and recognition of such data. Using SSI and the Microsoft Kinect sensor we are able to recognize various social cues [DBA13], such as *hand to face*, *looking away*, *postures* (arms crossed, arms open, hands behind

⁵ <http://openssi.net>

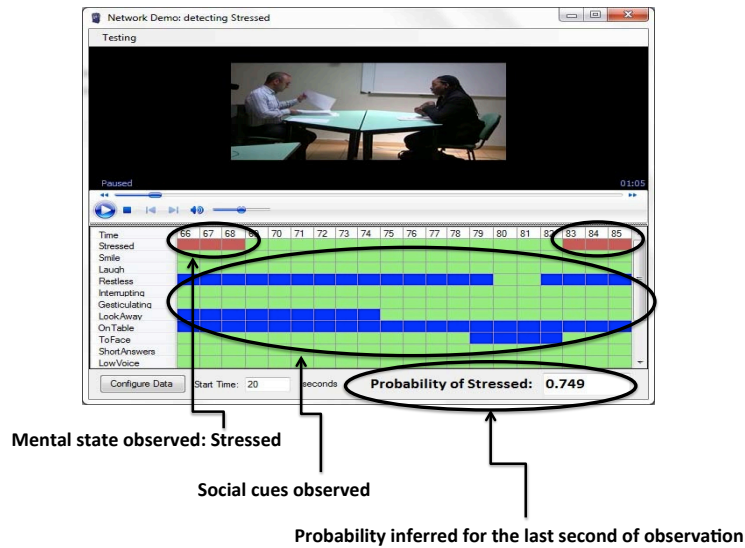


Fig. 4. Video playback facility, with the annotation data and Bayesian network output in the bottom-right corner of the screen.

head), *leaning back*, *leaning forward*, *gesticulation*, *voice activity*, *smiles* and *laughter*. In addition to these, our system is also able to compute the expressivity [CRKK06] of the user's movements (i.e. energy, overall activation, spatial extent and fluidity of the movement). Fig. 3 exemplifies some of the recognized social cues.

Given the recognized social signals, the next step consists in interpreting these signals to determine the socio-affective state of the youngster. In the next section, we present the process to interpret the youngster's social cues.

3.4 The Social Signal Interpretation Module

Inferring youngsters' mental states in real-time is important to determine the behaviours of the virtual recruiters. In order to gain an understanding of the social cues that occur in job interview situations and how these correspond to youngsters' mental state (in terms of affects and attitudes such as relaxed, stressed, enthusiastic...), we conducted a study at Mission Locale in Goussainville, France, with 10 real job seekers and 5 practitioners engaged in mock interviews. The data was video recorded and annotated manually.

Episodes as in Fig. 4 were annotated for fine-grained behaviours, e.g. looking away, lack of direct eye contact, smiling, etc. The annotations were mapped onto eight mental states, identified by the practitioners during post-hoc video walkthroughs and semi-structured interviews, as relevant to this context: *Stressed*, *Embarrassed*, *Ill-at-ease*, *Bored*, *Focused*, *Hesitant*, *Relieved*. The annotations have been fed through and used by a suite of four TARDIS bespoke C# applications: *Annotation Analyzer*, which calculates

the frequencies of the co-occurrences of groups of social cues with each mental state; *Net Populator*, which infers the probability tables of the Bayesian Networks (BNs), representing the mental states; *Net Demonstrator*, which allows to playback a video of an interaction along with the annotations used to train the networks and the real-time output of the network for a given mental state; *Net Tester* – a facility that feeds the annotations through the Bayesian Networks and produces statistics about the network’s performance in terms of correct and incorrect classifications.

The BNs and the Social Cue Recognition component (Section 3.3) are integrated, so that the inferences about mental states can occur in real-time. While the accuracy of the BNs themselves is relatively high when tested with video data, real-time recognition of social cues that would provide a sufficient and necessary evidence for the BNs remains a challenge for TARDIS and for the field of affective computing as a whole. Presently, we are testing several solutions which leverage and combine continuous stream of information from the Social Cue recognition module along with information about contextual information, in particular the type of question being asked by the recruiter and the expectations that relate to appropriate response on the part of the youngster.

Based on the inferred youngster’s mental states and based on the course of the interaction, the virtual recruiter should determine its affective behaviour. For this purpose, we have developed an Affective Module presented in the next section.

3.5 The Affective Module

The Affective Module for the virtual recruiter has two main computational functions:

- it will periodically compute the new affective states of the virtual recruiter, based on the perceptions, expectations from the scenario and its current affective states.
- it will select actions (*i.e.* different forms of utterances and specific branching in the scenario) according to its intentions following the youngster’s mental state.

Figure 5 illustrates the different elements of the Affective module as well as the links to the other modules.

The Affective model. This module is detailed in [JS13]. It provides a reactive model based on expectations from the youngster and perceptions (by Social Signal Interpretation) on the youngster. It allows a computation of the virtual recruiter emotions, moods and social attitudes. Emotion computation relies on the difference between expectations and perceptions. Emotions are modelled with OCC categories [OCC88]. Moods evolve on a middle term dynamic and are directly influenced by emotions following Mehrabian *Pleasure, Arousal, Dominance* framework [Meh96]. One of the originality of this model lies in the proposal for a computational model of the virtual recruiter social attitudes that rely on its personality and its current mood.

The Decision module. The goal of this module is to build a representation of the youngster’s beliefs (*a.k.a.* Theory of Mind or ToM [Les94]). The virtual recruiter considers the affective dimensions of the youngster’s answers (computed by the Social Signal Modules, see section 3.3 and 3.4) in a particular context (the question that has just been

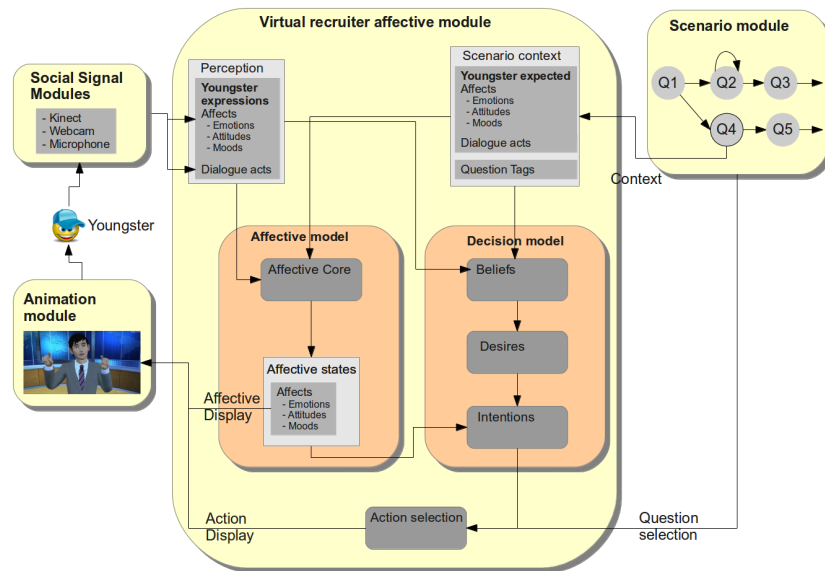


Fig. 5. The Architecture of the Affective Module

asked by the recruiter) to derive the positive or negative attitude the youngster has with respect to the considered topic. For instance, if the youngster reacts with too high a detachment to a question about the technical requirements for the job, then the agent might deduce that, on the topic *skill*, the user is not very confident. It will lower the value of $B_{Young}(skill)$. This model will, in turn, influence the next utterances of the virtual agents in function of the recruiter high-level intentions. The agent will select in the scenario subjects where the youngster is at ease (helpful/friendly) or not (confrontational/provocative).

3.6 The Animation Module

A person's affective state is mainly expressed through their non-verbal behaviour (e.g. facial expressions, body postures, audio cues, and gestures). In order to give the capability to the virtual recruiter to display its emotions and attitudes computed by the Affective Module (Section 3.5), we use the Greta system [BMNP07] to compute the animation parameters for the multimodal expressions of emotions [NHP11]. Moreover, the virtual recruiter will be animated using a combination of hand-animated gesture outputs (based on the animation parameters) and posed facial expressions, along with motion-captured animation techniques. This allows us to exploit the both methods: the essential flexibility relates to the character's gestures and movements when reacting to the input of the youngster, and the innate naturalness and subtleties of human motion that can be displayed through motion-captured data.

In order to ensure that the emotions conveyed by the virtual character are interpreted as intended, a number of studies on the perception of emotion through body motion have



Fig. 6. Stimuli from one perceptual study showing, from left to right synchronized facial and body motion, body motion alone and facial motion only for the basic emotion *fear*.

been carried out, with conversing virtual characters as the stimuli. We have found that some complex and subtle body motions are difficult to convey through body motion alone, and so have highlighted the importance of the appropriate combination of facial and body motions [EE12]. The literature suggests that the combination of facial and body motions does increase the perception of expressiveness of such emotions on video game characters (Figure 6). To this end, one major research challenge for the animation of the virtual character is the integration between the gestures, facial and body motions of the recruiter to ensure the maximum effectiveness of expressing emotions, while ensuring the character maintains a sense of naturalness about his/her motions. One way of making our virtual recruiter appear as natural as possible is through the inclusion of inherent human behaviours such as idle movements and interaction with objects and the environment around them.

The attitude of the virtual recruiter may be also expressed through non-verbal signals (such as smiles) but also through the dynamics of the behaviour. The global behaviour tendency (e.g. the body energy), the combination and sequencing of behaviours (e.g. order of display of smile and head aversion) as well as the mimicry of the the youngster’s behaviour may convey different attitudes [KH10]. In order to identify the non-verbal behaviours that the virtual recruiter may express to display different attitudes, we are analyzing the videos that were recorded at Mission Locale (Section 3.4). For this purpose, annotations have been added to consider the interaction context (e.g. turn-taking and conversation topic), the non-verbal behaviour of the recruiter (e.g. gestures, facial expressions, posture) and his expressed attitudes (friendliness of dominance using a bi-dimensional scale based on Argyle’s attitude dimensions [Arg88]). In order to give the capability to the virtual recruiter to display attitudes, we are extracting patterns and characteristics of behaviours associated with different levels of friendliness and dominance.

4 The TARDIS game

The TARDIS game is being developed with the aim to ensure that it meets both the educational and entertainment demands of the associated youngsters. The user will play him/herself applying for a fictional job, and is expected to behave as he/she would in real life. To this end, there is no “player” avatar and the input device used most frequently is the player him/herself. The movements and sounds made by the player will be the main method of input deciding how the game progresses.



Fig. 7. Our two different game settings: (top) flipchart/presentation scene with our male character and (bottom) our office desk environment with female recruiter.

4.1 Game setting and assets

The user will be seated in the real world, in front of a computer. Two different settings will be available in the game; a conference room and a typical office environment. In these settings, the virtual recruiter will be seated at a desk or standing in front of a flip-chart or use a presentation wall. The office/desk setting will be used as a prototypical interview arena, whereas the flip chart/conference room setting can be used for training purposes or feedback sessions, where information can be presented to the user. Small objects will also be present in the environment for the virtual character to interact with and present some variety for the youngster.

There will be two different character models in the TARDIS game, one male and one female. Different aspects of each character’s appearance can be modulated to give the impression of variety across interviews and scenarios; such as skin or hair colour,

glasses, clothing and age. This is important as the practitioners will have some control over the virtual recruiter and they will ensure that it is appropriate the individual scenarios. Examples of the TARDIS game assets can be seen in Figure 7.

4.2 Game modes

There will be a number of different game modes for TARDIS, to allow maximum interaction between the user and practitioner to permit the youngster to obtain the best results:

- Menu mode: Here, the user is presented with a menu and can select whether to modify settings or practice or play the game. There is also an option for the practitioner to enter edit mode.
- Settings/Options mode: Here, the user can select the level of difficulty at which they wish to play the game in, as well as set sound/visual options and check that their equipment is working adequately.
- Practice mode: Here, the user can practice a virtual interview, which does not count towards their progress. This mode will include a virtual interview as in the Play mode, but will allow for a “Pause” function, and will also include visual feedback to inform if their actions have been read by the system.
- Play mode: Here, they will be able to select which position they wish to interview for and take part in the virtual interview.
- Edit mode: This opens the editor for the practitioners, where they can modify interview scenarios to tailor individual needs of the user or job descriptions. They will also have access to change some dialog of the interaction, such as small-talk. This will result in a more varied experience for the youngster, but will not affect the expectations of the scenario.

4.3 Progress and Advancement

During the virtual interview, the user will receive “credits” for a range of properties concurrent with appropriate behaviour in job interviews. These credits will contribute to an overall score upon the conclusion of the interview. Completing a level will occur once a specified score has been achieved. They can then attempt a job interview at a more challenging level.

As the user progresses within the game, “levels” will not result in a different location in the physical world as in a platform game. Rather, the setting will remain the same, and the internal states of the virtual recruiter will reflect an advancement. For instance, the recruiter may become less patient, or more reactive to inappropriate behaviours. This can be reflected in the agent’s reaction to individual behaviours or movements of the player. It can also be built into the agent’s internal state. For instance, they could be aggressive or docile, friendly or aloof, and become less facilitative or forgiving of mistakes as the player progresses.

Since the objective of TARDIS is to achieve a sense of independence and inclusion for the youngsters who use it, we encourage players to assess their own progress, and will provide them with useful feedback to do so as the game progresses. To this end,

as well as a final score when the interview comes to an end, we will provide the player with a breakdown of how well they performed over a number of relevant factors. Physical parameters such as body or hand movement, head gaze or the evenness of the vocal intonation can be obtained from our SSI and reported to the user. So, the youngster can identify which aspects of their social interaction are satisfactory and they can acknowledge other areas where they may need to focus on improving. We also plan to translate the physical parameters calculated from the SSI in non-technical terms, such as gestures, posture, vocal, head gaze, or even more abstract personality traits like openness or professionalism. Then the user can be provided with further details about their performance along with suggestions on how to improve. For example, “you have scored xx% on comfort, maybe next time try to ensure that you keep your voice even and do not fidget too much to appear more at ease”. Any such feedback will be provided from a source outside of the virtual recruiter.

4.4 Post Interview Analysis

After each Game Session, the users receive feedback regarding their performance using our NOnVerbal behaviour Analysis tool (NovA) [BDL⁺13]. It enables the playback of the audiovisual recording of the interaction and visualizes recognized social cues, as well as continuous data streams in several modalities, such as graphs, time line tracks, heat maps, pie- and bar charts and others. Typically, different kinds of behaviours are coded on different parallel tracks so that their temporal relationships are clearly visible. Additionally, a series of statistics regarding the user’s performance is automatically computed which is meant to help users track their improvements over time.

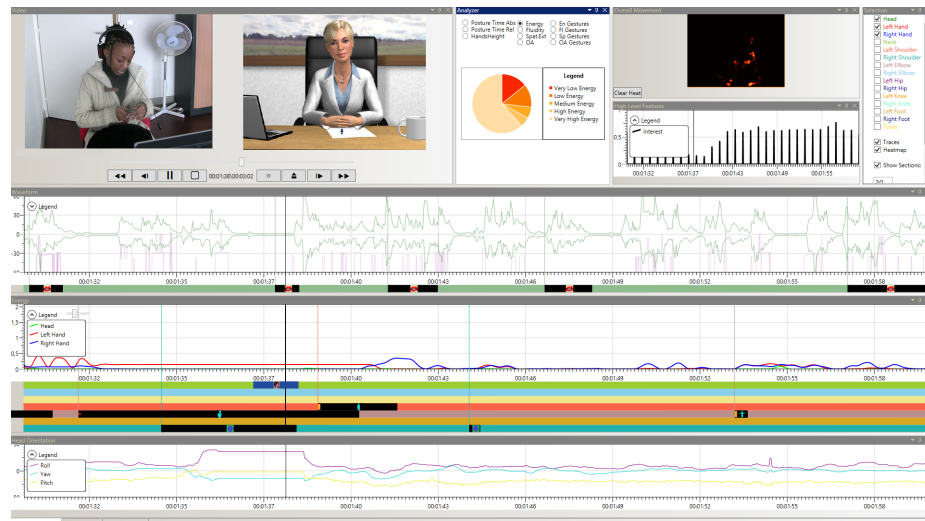


Fig. 8. NovA Analysis Tool

The analysis tool is designed to help youngsters reflect on their behaviour as well as to support the practitioners in identifying problems in the youngsters behaviour. The design of the tool has been informed during a workshop conducted in association with the practitioners at Mission Locale Goussainville. During the workshop, several key design points have been discussed. For instance, the practitioners pointed out that the ability of reviewing the performance of the user in parallel to the performance of the virtual agent is critical to interpreting the user's behaviours correctly. The discussions also yielded that it is equally important that the tool gives the practitioners information regarding when a certain behaviour of the user happened relative to the scenario's progress.

5 Conclusion

The originality of the TARDIS platform is threefold. First, we consider a participatory design for both the game and the simulator architecture. Workshops with practitioners and mock interviews with youngsters facing either human beings or virtual agents was used to define the expectations of the system. Second, the system is able to detect in real-time user's emotions and social attitudes through voice and facial expression recognition, and to adapt the progress of the game and the virtual interlocutor's behaviour to the individual users. Third, TARDIS gives practitioners new instruments to measure individuals' progress in being able to self-regulate emotionally and their social skill acquisition, thus facilitating reflection on their own practice and enabling a more flexible and personalized coaching for young people at risk of social exclusion.

We are currently evaluating the first prototype of the project and working on the next iteration of our opensource platform, implemented in C++ and Java using the Semaine⁶ framework with EmotionML⁷ messages.

Acknowledgement

This research was funded by the European Union Information Society and Media Seventh Framework Programme FP7-ICT-2011-7 under grant agreement 288578.

References

- [APD⁺09] Ruth Aylett, Ana Paiva, Joao Dias, Lynne Hall, and Sarah Woods. Affective agents for education against bullying. In *Affective Information Processing*, pages 75–90. Springer, 2009.
- [Arg88] Michael Argyle. *Bodily Communication*. University paperbacks. Methuen, 1988.
- [BDL⁺13] Tobias Baur, Ionut Damian, Florian Lingenfelser, Johannes Wagner, and Elisabeth André. Nova: Automated analysis of nonverbal signals in social interactions. In *Human Behavior Understanding (HBU) Workshop at 21st ACM International Conference on Multimedia*, 2013.

⁶ <http://www.semaine-project.eu>

⁷ <http://www.w3.org/TR/2010/WD-emotionml-20100729/>

- [BDSHP12] Elisabetta Bevacqua, Etienne De Sevin, Sylwia Julia Hyniewska, and Catherine Pelachaud. A listener model: introducing personality traits. *Journal on Multimodal User Interfaces*, 6(1-2):27–38, 2012.
- [BMNP07] Elisabetta Bevacqua, Maurizio Mancini, Radoslaw Niewiadomski, and Catherine Pelachaud. An expressive eca showing complex emotions. In *Proceedings of the AISB annual convention, Newcastle, UK*, pages 208–216, 2007.
- [BP02] John Bynner and Samantha Parsons. Social exclusion and the transition from school to work: The case of young people not in education, employment, or training (neet). *Journal of Vocational Behavior*, 60(2):289–309, 2002.
- [CRKK06] George Caridakis, Amaryllis Raouzaïou, Kostas Karapouzis, and Stefanos Kollias. Synthesizing gesture expressivity based on real sequences. *Workshop on multimodal corpora: from multimodal behaviour theories to usable models, LREC Conf. Genoa, Italy*, Mai 2006.
- [DBA13] Ionut Damian, Tobias Baur, and Elisabeth André. Investigating social cue-based interaction in digital learning games. In *Proceedings of the 8th International Conference on the Foundations of Digital Games. SASDG*, 2013.
- [DM99] Timothy DeGroot and Stephan J Motowidlo. Why visual and vocal interview cues can affect interviewers' judgments and predict job performance. *Journal of Applied Psychology*, 84(6):986–993, 1999.
- [dR13] Société Française de Radiotéléphonie. Mon entretien d'embauche game, June 2013.
- [EDVDB10] Jan Engelen, Jan Dekelver, and Wouter Van Den Bosch. Includo: Social software for the social inclusion of marginalised youth. *Social Media for Social Inclusion of Youth at Risk*, 1(1):11–20, 2010.
- [EE12] Cathy Ennis and Arjan Egges. Perception of complex emotional body language of a virtual character. In *Lecture Notes in Computer Graphics (Motion in Games'12)*, pages 112–121, 2012.
- [GMK12] Patrick Gebhard, Gregor Mehlmann, and Michael Kipp. Visual scenemaker - a tool for authoring interactive virtual characters. *Journal on Multimodal User Interfaces*, 6(1-2):3–11, 2012.
- [Gol06] Daniel Goleman. *Social intelligence: The new science of human relationships*. Bantam, 2006.
- [HCM⁺13] Mohammed Ehsan Hoque, Matthieu Courgeon, J Martin, Bilge Mutlu, and Rosalind W Picard. Mach: My automated conversation coach. In *Proc. ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2013)*, to appear, 2013.
- [HH10] Shlomo Hareli and Ursula Hess. What emotional reactions can tell us about the nature of others: An appraisal perspective on person perception. *Cognition & Emotion*, 24(1):128–140, 2010.
- [Ins13] Serious Games Insitute. ispectrum game, June 2013.
- [JS13] Hazaël Jones and Nicolas Sabouret. TARDIS - A simulation platform with an affective virtual recruiter for job interviews. In *IDGEI (Intelligent Digital Games for Empowerment and Inclusion)*, 2013.
- [JYZ12] Ying Jiao, Xiaosong Yang, and Jian J Zhang. A serious game prototype for post-stroke rehabilitation using kinect. *Journal of Game Amusement Society*, 4(1), 2012.
- [KH10] Mark L. Knapp and Judith A. Hall. *Nonverbal Communication in Human Interaction*. Cengage Learning, 2010.
- [Les94] Alan M. Leslie. ToMM, ToBy, and agency: Core architecture and domain specificity. In L A Hirschfeld and S A Gelman, editors, *Mapping the mind Domain specificity in cognition and culture*, chapter 5, pages 119–148. Cambridge University Press, 1994.

- [MC05] Micheline Manske and Cristina Conati. Modelling learning in an educational game. In *Proceeding of the 2005 conference on Artificial Intelligence in Education: Supporting Learning through Intelligent and Socially Informed Technology*, pages 411–418, 2005.
- [Meh96] Albert Mehrabian. Pleasure-arousal-dominance: A general framework for describing and measuring individual Differences in Temperament. *Current Psychology*, 14(4):261, 1996.
- [MGR03] Stacy Marsella, Jonathan Gratch, and Jeff Rickel. Expressive behaviors for virtual worlds. *Lifelike Characters Tools Affective Functions and Applications*, pages 317–360, 2003.
- [MSS04] Alice Mitchell and Carol Savill-Smith. The use of computer and video games for learning: A review of the literature. 2004.
- [NHP11] Radosław Niewiadomski, Sylwia Julia Hyniewska, and Catherine Pelachaud. Constraint-based model for synthesis of multimodal sequential expressions of emotions. *IEEE Transactions on Affective Computing*, 2(3):134–146, 2011.
- [OCC88] Andrew Ortony, Gerald L. Clore, and Allan Collins. *The Cognitive Structure of Emotions*. Cambridge University Press, July 1988.
- [Orv10] Verónica Orvalho. Lifeisgame: Learning of facial emotions using serious games. In *Proc Body Representation in Physical and Virtual Reality with Application to Rehabilitation*, 2010.
- [PDS⁺04] Ana Paiva, Joao Dias, Daniel Sobral, Ruth Aylett, Polly Sobreperetz, Sarah Woods, Carsten Zoll, and Lynne Hall. Caring for agents and agents that care: Building empathic relations with synthetic agents. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 1*, pages 194–201, Washington, DC, USA, 2004. IEEE Computer Society.
- [Pic97] Rosalind W. Picard. Affective Computing. *Emotion*, TR 221(321):97–97, 1997.
- [PSS09] Lena Pareto, Daniel L Schwartz, and Lars Svensson. Learning by guiding a teachable agent to play an educational game. in *Education Building Learning*, pages 1–3, 2009.
- [SMBC⁺13] Björn Schuller, Erik Marchi, Simon Baron-Cohen, Helen O’Reilly, Peter Robinson, Ian Davies, Ofer Golan, Shimrit Friedenson, Shahar Tal, Shai Newman, Noga Meir, Roi Shillo, Antonio Camurri, Stefano Piana, SSven Bölte, Daniel Lundqvist, Steve Berggren, Aurelie Baranger, and Nikki Sullings. Asc-inclusion: Interactive emotion games for social inclusion of children with autism spectrum conditions. In *Proc. 1st International Workshop on Intelligent Digital Games for Empowerment and Inclusion (IDGEI 2013)*, 2013.
- [SS12] Sandra Siegel and Jan Smeddinck. Adaptive difficulty with dynamic range of motion adjustments in exergames for parkinson’s disease patients. In *Entertainment Computing - ICEC 2012*, volume 7522 of *Lecture Notes in Computer Science*, pages 429–432. 2012.
- [TC08] Andrea Tartaro and Justine Cassell. Playing with virtual peers: bootstrapping contingent discourse in children with autism. In *Proceedings of the 8th international conference on International conference for the learning sciences-Volume 2*, pages 382–389. International Society of the Learning Sciences, 2008.
- [WLB⁺13] J. Wagner, F. Lingenfelter, T. Baur, I. Damian, F. Kistler, and E. André. The social signal interpretation (ssi) framework - multimodal signal processing and recognition in real-time. In *Proceedings of ACM MULTIMEDIA 2013*, Barcelona, 2013.