

A model to generate adaptive multimodal job interviews with a virtual recruiter

Zoraida Callejas *, Brian Ravenet **, Magalie Ochs ***, Catherine Pelachaud ***

* Dpt. Languages and Computer Systems, University of Granada, CITIC-UGR, Granada, Spain

** Institut Mines-Télécom, Télécom ParisTech, CNRS LTCI, Paris, France

*** CNRS LTCI, Télécom ParisTech, Paris, France

zoraida@ugr.es, {ravenet,ochs,pelachaud}@telecom-paristech.fr

Abstract

This paper presents an adaptive model of multimodal social behavior for embodied conversational agents. The context of this research is the training of youngsters for job interviews in a serious game where the agent plays the role of a virtual recruiter. With the proposed model the agent is able to adapt its social behavior according to the anxiety level of the trainee and a predefined difficulty level of the game. This information is used to select the objective of the system (to challenge or comfort the user), which is achieved by selecting the complexity of the next question posed and the agent's verbal and non-verbal behavior. We have carried out a perceptual study that shows that the multimodal behavior of an agent implementing our model successfully conveys the expected social attitudes.

Keywords: Virtual agents, Affective interaction, Dialog management, Conversational agents, Multimodal behavior

1. Introduction

Employment interview training has a strong relationship with candidate performance (Macan, 2009), as it influences the ability of applicants to explain themselves with better verbal and non-verbal skills, to answer difficult questions and provide more organized responses (Huffcutt et al., 2011), which provokes more favorable evaluations.

The research presented in this paper is part of the European TARDIS¹ project (Anderson et al., 2013), which objective is to develop a serious game to train youngsters for job interviews. The youngsters would be able to practice playing a game with a high variety of interviews of different complexity where a conversational agent acts as the recruiter.

Undergraduates experience sustained anxiety levels immediately before and during mock interviews with peers (Young et al., 2004). Applicants experiencing anxiety during the interview may receive lower scores despite that their job performance could have been successful (Macan, 2009). Thus, anxiety is an important factor to be considered for interview training, especially with young populations, and embodied artificial recruiters have shown to provoke a sense of presence that has the power to induce anxiety in virtual interviews (Kwon et al., 2013).

In this article, we propose a computational model for a virtual recruiter that is able to generate a wide variety of job interviews. At each turn, the model decides whether to try to challenge or comfort the user according to the detected user's anxiety and the difficulty level selected for the interview. Then, it selects the best system dialog act (DA) and social attitude, and renders the multimodal output accordingly. We have developed a virtual recruiter endowed with the proposed computational model using the Greta embodied conversational agent platform (Niewiadomski et al., 2009) and conducted a perceptual study that shows that the virtual recruiter successfully conveys the expected social attitudes through its multimodal behaviors.

2. Related work

Some virtual recruiters have already been developed. However, in the existing approaches the interview has predefined structure in which there is a collection of questions from which the agent selects the one that is going to be posed to the user in the next turn.

This way, the MACH agent (Hoque et al., 2013), developed to provide social skill training for job interviews, selects the questions from a list of 15 frequently employed in human-human interviews. In this work, the focus with respect to the agent behavior is in providing feedback while it is listening, mainly through head nods and arm movements.

The virtual reality interview presented in (Brundage et al., 2006) supports two types of interview: challenging and supportive, in which the recruiters behavior is tuned mainly through the use of eye contact and the interruptions to the users turns. The questions asked during the interview are the same in both conditions and are chosen from a restricted list of open questions.

Similarly, (Kwon et al., 2013) present an immersive environment for virtual job interviews. Their research is also focused on the anxiety experienced by students during their first job interviews. However, the verbal behavior of the virtual agent is limited to a list of 12 general-purpose questions.

Our objective is to be able to generate a wider variety of different interviews that adapt to the users anxiety. In order to do so, we do not only consider a textual form per question, but we adjust the reactions of the agent with respect to the type and difficulty of the next question to be posed, its wording, and the agent's non-verbal behavior. By balancing these dimensions we can select adaptive interaction strategies and social attitudes, considering combinations that are seldom present in the literature.

3. Our model

We present a computational model of virtual recruiters that are responsive to the anxiety experienced by interviewees

¹<http://researcher.tardis-project.eu/the-project/presentation>

modulated by choosing a level of difficulty for the job interview.

The model follows the SAIBA architecture (Kopp et al., 2006), an international common multimodal behavior generation framework. Instead of the Intent Planner, we have introduced a Dialog Manager that selects the appropriate system response (defined in terms of dialog acts) and the virtual recruiter’s social attitude to express given the user’s anxiety level. The virtual recruiter’s social attitude is stored in the Agent Mind, which is queried by the Natural Language Generator and the Behavior Planner. The Natural Language Generator selects a phrase that reflects the attitude selected. The phrases correspond to one of the possible wordings for the dialog act selected. The phrase is included in a FML file (Heylen et al., 2008) containing the associated dialog act. This file is used by the Behavior Planner to instantiate the appropriate non-verbal behaviors depending on the attitude and dialog act (communicative intention) of the agent. Finally, the Behavior Realizer and the Text-To-Speech (TTS) engine display the animation of the agent. Figure 1 shows a summary of the steps involved and two examples with different difficulty levels for the game.

The objective of the dialog manager changes according to the different combinations of two inputs: the users anxiety level and the difficulty level of the game. We suppose that the anxiety recognizer provides an anxiety level ranging in $[0, 1]$, and consider three intervals: low (below 0.25), medium (between 0.25 and 0.75) and high (higher than 0.75). The detection of anxiety can be done using a combination of audiovisual and/or physiological cues (Baur et al., 2013). However in the study presented here, the anxiety level is an input provided at each turn; it does not rely on a particular anxiety recognizer. Additionally, we consider six possible difficulty levels from 1 to 6, where 1 is the lowest difficult one and 6 the highest. The difficulty level is selected at the beginning of the game and does not vary during the interview.

At each turn, the dialog manager computes the objective of the system, which may be to comfort or challenge the user. As shown in Table 1, with higher difficulty levels the system is more prone to challenge the user, whereas for lower levels of difficulty it tries to calm the user down. However, in order to explore a wider space of dialog strategies, it is possible to consider different objectives according to a certain probability distribution. Additionally, the dialog strategy depends on the tendency of the anxiety level of the user, i.e. whether during the whole interaction the user tends to be relaxed or tense. This tendency can be computed as the slope of the linear regression of all the anxiety values up to the current moment.

The system objective is implemented by selecting the complexity of the next dialog act (the type of question it will ask) and the social attitude it displays. The complexity of the dialog act depends on two factors: the focus of negative facts and the openness of the question. This way, a question is considered more complex to respond if it is focused on negative facts (e.g. asking about a weakness of the interviewee) and if the response requires a long elaboration instead of a concise one. We consider these factors in the two phases of human anxiety processing described in (Beck and

Difficulty level	Anxiety level in previous turn		
	Low	Medium	High
1 or 2	Comfort		
3 or 4	If tendency = decreasing Comfort 90%- Challenge 10% else Comfort 10%- Challenge 90%	If tendency = increasing Comfort 10%- Challenge 90% else Comfort 90%- Challenge 10%	Comfort 50%- Challenge 50%
5 or 6	If tendency = neutral Comfort else Challenge		Challenge

Table 1: Strategy for selecting the dialog objective

Clark, 1997): the perception of a threat, and the perception of the availability and effectiveness of coping resources.

Once the dialog manager has selected the DA and the system attitude in that turn (hostile, neutral or friendly), the natural language generator chooses a phrase that matches both, and the behavior planner chooses a non-verbal behavior that corresponds with the system attitude and with the selected DA.

The system keeps asking questions until the user stays in a medium or low anxiety level for a certain number of turns. This number increases for the interactions with high difficulty.

The natural language generator follows the guidelines of the Personage project (Mairesse and Walker, 2011). Concretely, we have taken into account the number of times that the agent talks about itself (friendly phrases contain more self-references), the variety of vocabulary (hostile phrases are better structured with more synonyms), the preference for nouns vs. verbs (in friendly phrases there is a preference for action), the formality and length of the expressions (hostile phrases are longer and more formal and friendly phrases give the impression of being more spontaneous), and the preference for negative vs. positive contents (in hostile phrases there are more negations and negative contents are predominant). This way, it is possible to render the same dialogue act with different attitudes in the wording. For example the phrases “We will answer you in about a week” and “You will receive an answer not earlier than a week from now” both correspond to the same dialogue act with friendly and hostile wordings respectively.

We have created a database that contains at least a friendly, hostile and neutral phrase per dialogue act (a minimum of 228 phrases). When the dialog manager has selected a dialog act and an attitude for the current turn, the natural language generator module queries the database to find a phrase. If several phrases are available for the selected dialog act and attitude, it chooses one randomly.

Our behavior planner works with a Bayesian Network in order to choose the nonverbal behavior corresponding to the dialog act and the social attitude selected by the dialog manager. This Bayesian Network was built from a corpus of agents non-verbal behavior. The corpus was collected through a user-perspective experiment where participants had to configure the non-verbal behaviors of an agent displaying different attitudes (Ravenet et al., 2013). The Bayesian Network contains the probabilities of displaying

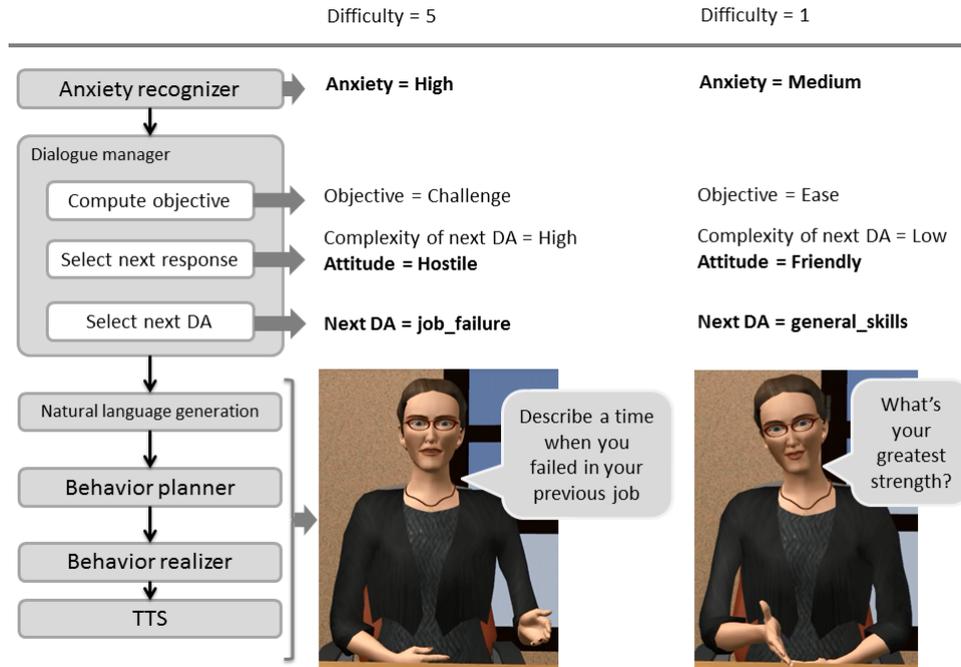


Figure 1: Steps in the computational model with two examples corresponding to two turns in different interviews

certain behaviors to communicate a dialog act type with a specific social attitude. The possible variations are the type of facial expression (positive, negative or neutral), the activation of gestures (arm movements, head movements, both or no movements), the amplitude of arm movements if activated (small, normal or wide), the strength of arm movements if activated (weak, normal or strong), the head orientation (downward, upward, tilted aside or straight) and the presence of gaze avoidance. These non-verbal parameters have been selected based on the studies in Human and Social Sciences about the perception of social attitude. It also makes sure that the inferred probability of the target attitude (using Bayesian inference on the model) is higher than the other alternative attitudes. By doing so, we ensure that the generated behavior corresponds to the desired attitude and we also keep the variability of the probabilistic model, which is interesting when modeling human-like behavior.

4. Evaluation of the model

In order to evaluate the capacity of our model to compute the appropriate verbal and non-verbal behavior for the specific attitudes, we have performed an initial user-perceptive study with 110 participants. They have rated 4 video-clips corresponding to interviews with a virtual recruiter displaying a neutral, friendly or hostile attitude in 4 conditions: verbal only (friendly or hostile verbal behavior, neutral non-verbal behavior), non-verbal only (friendly or hostile non-verbal behavior, neutral verbal behavior), multimodal (friendly or hostile both with verbal and non-verbal behavior) and control (neutral verbal and non-verbal behavior). For each video clip, we asked the participants to indicate their perception of the virtual recruiter by indicating their agreement with the following sentences. A 5 points Likert

scale (from “Totally disagree” to “Totally agree”) was set for each question.

1. The virtual recruiter behavior is believable.
2. The virtual recruiter gives the impression to want to hire the interviewee.
3. The virtual recruiter gives the impression to want to fail the interviewee.
4. The virtual recruiter tries to make the interviewee at ease.
5. The virtual recruiter wants to destabilize the interviewee.
6. The virtual recruiter expresses an hostile attitude.
7. The virtual recruiter expresses a friendly attitude.
8. The virtual recruiter expresses a dominant attitude.

We have conducted ANOVA and post-hoc HSD-Tukey tests to compare the opinions of the participants under the different conditions. In the verbal only condition, no significant difference appeared between the attitude ratings. In the non-verbal only condition, the friendly attitude was correctly perceived whereas the hostile attitude was partially perceived (the agent was identified as less friendly but not more hostile). However, in the multimodal condition the agent’s behavior was perceived as intended in all cases, showing significant differences between the friendly and hostile behaviors and the baseline.

Figure 2 shows the results for each question when the agent was rendering neutral, friendly and hostile behaviors. The lines go from the minimum (bottom) to the maximum (top) rating for each question, the boxes start in the first quartile (bottom) and finish in the third quartile (top), and the

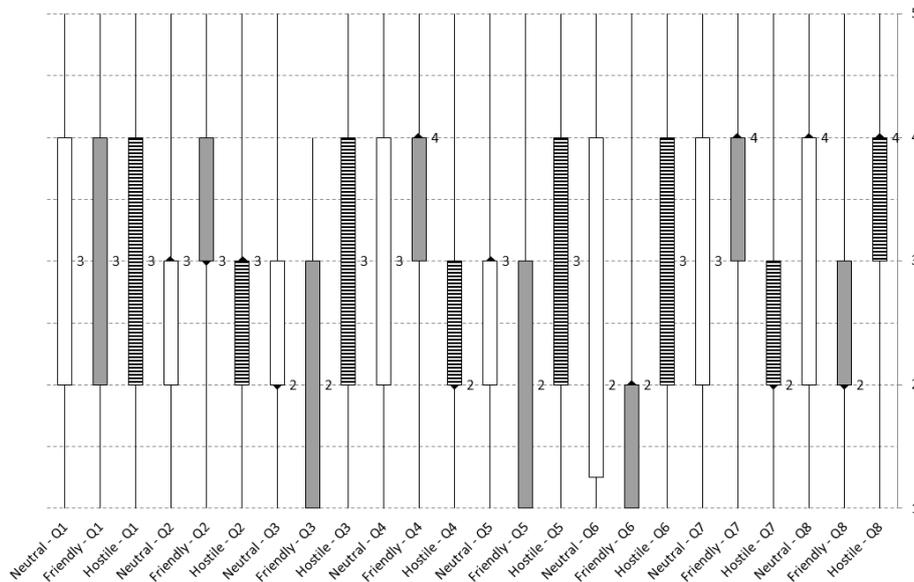


Figure 2: Results (min, max, q1, q3 and median) for the perceptive study across all conditions

points and values correspond to the median ratings. As can be observed, the virtual recruiter is perceived in average believable, but there are no significant differences in believability depending of the attitude (all neutral, friendly and hostile are considered believable). As expected, the virtual recruiter is perceived as dominant. However, the expression of friendliness significantly decreases the perceived dominant attitude. Also the agent gave the impression to try to destabilize and fail the user with hostile attitudes and was perceived as friendly and trying to ease the interviewee with the friendly behavior.

5. Conclusions

We have presented a model to generate adaptive multimodal job interviews with a virtual recruiter. Our model generates a wide a variety of interviews in which the objective of the virtual recruiter (either to comfort or challenge the user) is adapted at each dialog turn according to the variations in the perceived user anxiety. The objective selected is achieved by choosing questions of different complexity in combination with multiple verbal and non-verbal cues that successfully convey friendly, neutral and hostile attitudes.

We have conducted a perceptive study in which 110 individuals rated video clips corresponding to simulated job interviews with a virtual recruiter endowed with our model. The results validate our proposed model by showing that the agents multimodal behavior successfully conveyed the expected social attitudes.

For future work we plan to complete the evaluation by studying whether the objectives of our system at each turn (to challenge or comfort the user) are successfully accomplished by means of the behaviors selected, and are adequate to the difficulty level of the game.

6. Acknowledgments

This research has received partial funding from the European Union Information Society and Media Seventh Frame-

work Programme FP7-ICT-2011-7 under grant agreement 288578 (TARDIS) and 287723 (REVERIE). We are very grateful to Cereproc (www.cereproc.com) for letting us use their voice synthesizer.

Zoraida Callejas was supported by the Spanish Ministry of Education, Culture and Sport under the programme “Programa Nacional de Movilidad de Recursos Humanos del Plan Nacional I-D+i 2008-201”, José Castillejo mobility grant no. CAS12/00227.

7. References

- Keith Anderson, Elisabeth Andre, T. Baur, Sara Bernardini, M. Chollet, E. Chryssafidou, I. Damian, C. Ennis, A. Egges, P. Gebhard, H. Jones, M. Ochs, C. Pelachaud, Kaka Porayska-Pomsta, P. Rizzo, and Nicolas Sabouret, 2013. *The TARDIS Framework: Intelligent Virtual Agents for Social Coaching in Job Interviews*, pages 476–491. Number 8253 in Lecture Notes in Computer Science. Springer International Publishing.
- Tobias Baur, Ionut Damian, Patrick Gebhard, Kaska Porayska-Pomsta, and Elisabeth Andre. 2013. A job interview simulation: Social cue-based interaction with a virtual character. In *2013 International Conference on Social Computing (SocialCom)*, pages 220–227.
- Aaron T. Beck and David A. Clark. 1997. An information processing model of anxiety: automatic and strategic processes. *Behaviour research and therapy*, 35:49–58.
- Shelley B Brundage, Ken Graap, Kathleen F Gibbons, Mirtha Ferrer, and Jeremy Brooks. 2006. Frequency of stuttering during challenging and supportive virtual reality job interviews. *Journal of fluency disorders*, 31:325–339.
- Dirk Heylen, Stefan Kopp, Stacy C. Marsella, Catherine Pelachaud, and Hannes Vilhjmsson. 2008. The next step towards a function markup language. In *Proceedings of the 8th International Conference on Intelligent Virtual Agents*, page 270280. Springer-Verlag.

- Mohammed Ehsan Hoque, Matthieu Courgeon, J Martin, Bilge Mutlu, and Rosalind W Picard. 2013. Mach: My automated conversation coach. In *International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2013)*.
- Allen I. Huffcutt, Chad H. Van Iddekinge, and Philip L. Roth. 2011. Understanding applicant behavior in employment interviews: A theoretical model of interviewee performance. *Human Resource Management Review*, 21:353–367.
- Stefan Kopp, Brigitte Krenn, Stacy Marsella, Andrew N. Marshall, Catherine Pelachaud, Hannes Pirker, Kristinn R. Thorisson, and Hannes Vilhjalmsson. 2006. Towards a common framework for multimodal generation: The behavior markup language. In *International Conference on Intelligent Virtual Agents*, page 2123.
- Joung Huem Kwon, John Powell, and Alan Chalmers. 2013. How level of realism influences anxiety in virtual reality environments for a job interview. *International Journal of Human-Computer Studies*, 71:978–987.
- Therese Macan. 2009. The employment interview: A review of current studies and directions for future research. *Human Resource Management Review*, 19:203–218.
- Francois Mairesse and Marilyn A. Walker. 2011. Controlling user perceptions of linguistic style: Trainable generation of personality traits. *Computational Linguistics*, 37:455488.
- Radoslaw Niewiadomski, Elisabetta Bevacqua, Maurizio Mancini, and Catherine Pelachaud. 2009. Greta: an interactive expressive eca system. In *Proceedings of the 8th international conference on Autonomous Agents and Multiagents Systems*, pages 1399–1400.
- Brian Ravenet, Magalie Ochs, and Catherine Pelachaud. 2013. *From a User-created Corpus of Virtual Agents Non-verbal Behavior to a Computational Model of Interpersonal Attitudes*, pages 263–274. Number 8108 in Lecture Notes in Computer Science. Springer Berlin Heidelberg.
- Melissa J. Young, Ralph R. Behnke, and Yvonne M. Mann. 2004. Anxiety patterns in employment interviews. *Communication Reports*, 17:49–57.