Data-driven model of virtual patient
for doctor social training

Magalie Ochs[1], Evelyne Lombardo[1], Soisik Verbog[5], Marie-Christine Moll[5], José Hureaux[5], Daniel Francon[6], Jean-Marie Pergandi[3], Daniel Mestre[3], Gregoire De Montcheuil[2], Catherine Pelachaud[4], Brice Donval[4], Jorane Saubesty[2], Chloé Clavel[4] and Philippe Blache[2]

Aix Marseille Université CNRS ENSAM, Université de Toulon [1]LSIS UMR 7296, [2]LPL UMR 7309, [3]ISM UMR7287 ; [4]CNRS LTCI Télécom ParisTech ; [5]Centre Hospitalier Universitaire d'Angers, Aquarel Santé ; [6]Institut Paoli Calmette

Abstract. The way doctors deliver bad news has a significant impact on the therapeutic process. In this paper, we present an ongoing project that aim at developing an embodied conversational agent simulating a patient to train doctors to break bad news. The embodied conversational agent is incorporated in an immersive virtual reality environment integrating several sensors to detect and recognize in real time the verbal and non-verbal behavior of the doctors interacting with the virtual patient. The virtual patient behavior as well as the virtual environment are constructed based on a methodology mixing a corpus-based approach, real medical data and expertise and empirical and theoretical studies on human-machine interaction.

1   Introduction

Doctors should be trained not only to medical or surgical acts but also to develop competences in communication for their interaction with patients. For instance, they often face the announcement of undesirable events to patients, as for example damage associated to the care. A *damage associated to the care* is the consequence of an unexpected event that can be due to complication connected to the pathology of the patient, unforeseeable medical complication, dysfunction or medical error. The damage can have physical, psychological, or even social and material repercussions. Such undesirable event is frequent: an undesirable event with damage arises every five days in a unit of 30 beds (Michel *et al.*, 2011).

The way *doctors deliver bad news related to damage associated to care* has a significant impact on the therapeutic process: disease evolution, adherence with treatment recommendations, litigation possibilities (Andrade *et al.*, 2010)[1]. However, both experienced clinicians and medical trainees consider this task as difficult, daunting, and stressful. Nowadays, training health care professional to break bad news, recom-

---

[1] Note that this study has been conducted in the context of breast cancer. We suppose that the results can be extended to other contexts of breaking bad news.

mended by the French Haute Autorité de la Santé (HAS), is organized as workshops during which doctors disclose bad news to actors playing the role of patient. The training by the simulation (Granry and Moll, 2012), *i.e.* put the doctor in a simulated situation of consultation with actors playing patients' role, allows to develop the level of understanding of the professionals of this potentially conflicting and painful situation for them, for the patient and his close relationships. However, this training solution requires a huge amount of human and financial and its time consuming (each 30 mn. session requires an hour of preparation).

The objective of the project is to develop an *immersive* platform that enables doctors to train to break bad news with a virtual patient. For this purpose, we aim at developing an embodied conversational agent (ECA) simulating a patient. Such a platform would play a decisive role for institutions involved in training (hospitals, universities): the needs concern potentially thousands of doctors/students. Organizing such training at this scale is not realistic with human actors. A virtual solution would be then an adequate answer. The methodology used in the project is based on a multidisciplinary approach gathering computer scientists, linguists, psychologists and medical doctors. Moreover, we adopt a *data-driven methodology* to model the verbal and nonverbal behavior of the virtual agent as well as the virtual environment. One goal in this project with this multidisciplinary and data-driven approach is to try to simulate as believable as possible the environment of breaking bad news and the virtual patient behavior. The paper is organized as follows: a state of art in this domain is presented in Section 2, the overall architecture of the project is introduced in Section 3, the global methodology that has been adopted to develop such a training platform is presented in Section 4. We discuss the perspectives of the project in Section 5.

2   Existing virtual patients for doctors' training

Several ECAs embodied the role of virtual patients have already been proposed for use in clinical assessments, interviewing and diagnosis training (Lock et al., 2006; Kenny et al., 2008; Andrade et al., 2011). Indeed, previous research has shown that doctors demonstrate non-verbal behaviors and respond empathetically to a virtual patient (Deladisma et al., 2006). In this domain, the research has mainly focused on the anatomical and physiological models of the virtual patient to simulate the effects of medical interventions or on models to simulate particular disorder (e.g. Kenny et al., 2008; Lock et al., 2006 or the eViP European project[2]. In our project, we focus on a virtual patient to train doctors to deliver bad news. A first study (Andrade et al., 2010) has analyzed the benefits of using a virtual patient to train doctors to break the diagnosis of breast cancer. The results show significant improvements of the self-efficacy of the medical trainees. The major limit of the proposed system, highlighted by the participants, is the lack of non-verbal behaviors of the patients simulated in the limited environment Second Life[3]. Our objective in this project is to simulate the non-

---

[2] http://www.virtualpatients.eu
[3] Linden Labs, San Francisco, CA

verbal expression of the virtual patient to improve the believability of the virtual character and the immersive experience of the doctor.

Most of the embodied conversational agents used for health applications have been integrated in 3D virtual environment on PC. Virtual reality in health domain is particularly used for virtual reality exposure therapy (VRET) for the treatment for anxiety and specific phobias (e.g. Parsons and Rizzo, 2008). In our project, in order to offer an immersive experience to the doctor, we have integrated the virtual patient in a virtual reality environment. In the next section, we present the overall architecture of the training platform.

3   Architecture of the training platform

The overall architecture of the training platform is illustrated Figure 1. The speech of the doctor is recognized by an Automatic Speech Recognition (ASR) *system* (Nocera et al., 2002). The audio is preprocessed to eliminate non-speech segments, and cut the speech into shorter segments call IPU segments (Inter-Pausal Units) separated by small silences (200 ms). Each IPU is streamed to the *ASR system* to be transcribed and then sent to the *Scenario Controller*. Moreover, in order to avoid waiting for the result of the ASR to generate a virtual patient's reaction, voice activity information is sent to the supervisor (e.g. *doctor_is_talking*) in order to trigger listening behavior of the virtual patient. When doctor's speech is not well recognized, the scenario controller triggers specific verbal or non-verbal behavior (e.g. "I don't understand" or only a shrug of the shoulders). We are currently improving the ASR by learning specific vocabulary related to this medical context. To enable a multimodal interaction, we are currently integrating various other *sensors* to detect the non-verbal behavior of the doctor (gesture, gaze, posture, etc.) using a Kinect. This information will be used by the *Scenario controller* to coordinate the virtual patient's behavior (*e.g.* to mimic a head shake). Moreover, the *Scenario controller* contains different elements describing the pre-defined pedagogical scenario of breaking bad news: the bad news to announce (digestive perforation during an endoscopy[4]), the attitude of the virtual patient (aggressive versus accommodating[5]), the pedagogical objectives (i.e. the information that the doctor should transmit to the virtual patient). The *Scenario controller* contains also a set of rules to determine the appropriate reaction of the virtual patient (verbal and non-verbal) given the elements of the scenario (e.g. predefined attitude of the virtual patient) and the historical of the dialog. For instance, the virtual patient may ask a specific question if the doctor didn't inform the patient on a specific point or may express only aligned behavior if it is defined with an accommodating attitude. The *Scenario controller* communicates with the dialog system to determine the verbal reaction of the virtual patient. The set of rules characterizing the

---

[4] The scenario has been carefully chosen with the medical partners of the project for several reasons (e.g. the panel of resulting damages, the difficulty of the announcement, its standard characteristics of announce).

[5] These attitudes correspond to those played by the human actor endowing the role of the patient during doctor's trainings.

verbal and non-verbal behavior of the virtual patient is defined based on analyzed of corpus and with expert in breaking bad news training (Section 4). The non-verbal behavior is selected from the *behavior library* specifically designed for the scenario (Section 4.2) and sent to the non-verbal behavior animation system VIB. *VIB* (Virtual Interactive Behavior) is a generic platform for creating Embodied Conversational Agents (ECA) (Pelachaud *et al.*, 2009). VIB computes the animation parameters (Facial Animation Parameters – FAP – and Behavioral Animation Parameters – BAP) to animate the face and body of the virtual patient. Moreover, the *Non-verbal behavior system Greta* contains a text-to-speech system (Aylett and Pidcock, 2007) to generate the speech synchronizes with the non-verbal behavior (including the lips animation).

The virtual patient is finally animated through different players on different platforms: PC, Virtual reality environment and 3D glasses. The virtual reality environment is constituted of a 3m deep, 3m wide, and 4m high cubic space with three vertical screens and a horizontal screen (floor). Using a cluster of graphics machine, the system is able to deliver stereoscopic, wide-field, real-time rendering of 3D environments, including spatial sound, and thus trying to achieve optimal sensorial immersion of the user. The virtual agent based on the VIB platform has been integrated in these different environments through the Unity player. In the next section, we describe the methodology used to design the virtual environment as well as the virtual patient behavior.
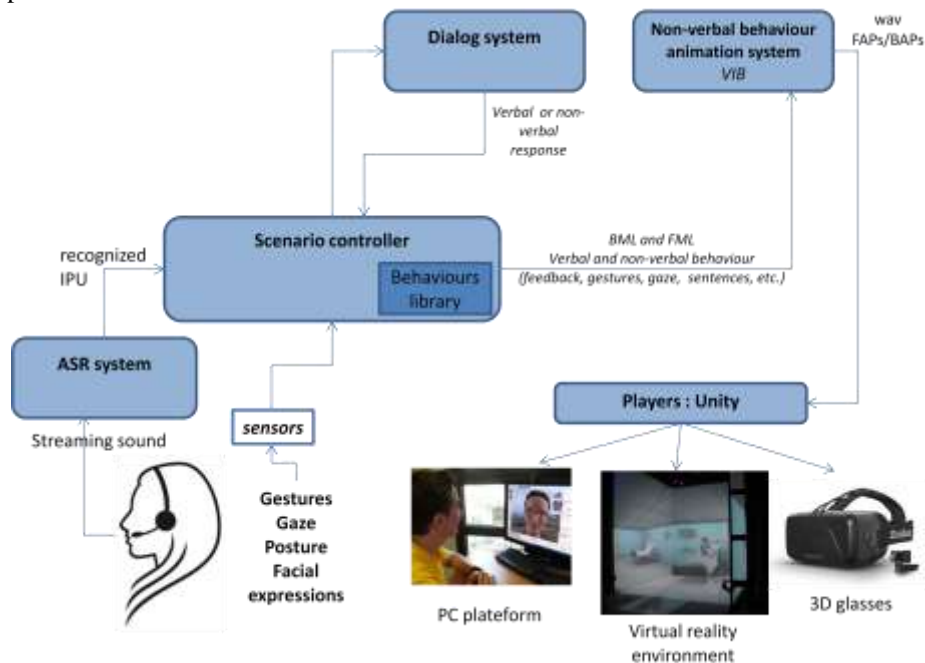

Figure 1: Overall architecture of the training platform

4 Methodology

The methodology adopted in this project to develop the training platform aims at mixing a top-down approach and a bottom-up one. From a *top-down perspective*, the objective is to exploit the theoretical and empirical studies on human-machine and human-human interaction to guide some choices concerning the development of functionalities in the trainings platform (e.g. exploiting presence factors to design the doctor-patient interaction, Section 4.1). From a *bottom-up perspective*, we aim at exploiting real medical data and expertise to model the patient behavior and its environment (Section 4.2 and 4.3).

4.1   Starting from presence factors to design virtual training platform

In order to provide an immersive experience to doctors to train to break bad news with a virtual patient, we have analyzed the different elements that will enable us to create such an experience. As highlighted in (Slater et al., 2001), the psychological immersion is independent of the device (for example, a book, projecting us in a virtual world can provoke a psychological immersion, without technological and physical immersion). This type of immersion is called *sense of presence* which gives the user the impressiom to lose the sense of time and space. In the literature, there are 7 identified factors that affect this type of sense of presence. In the following, we propose to present each of these factors and to explain how we aim at handling this factor in the training platform: (1) the *ease of interaction*: interaction correlates with the sense of presence felt in the virtual environment (Billinhurst, & Weghorst, (1995)). In the training platform, the doctor will interact in a multimodal way (e.g. voice, gesture) with the virtual patient; (2) *the user control*: the sense of presence increases with the sense of control (Witmer and Singer, 1998). During the training, the doctor will feel independent in his actions (e.g. he will be able to draw as in real environment) ; (3) the *realism of the image*: the more realistic virtual environment is, the more the sense of presence is strong (Witmer *et al.*, 1998). For this purpose, we construct the virtual environment based on the characteristics of the real environment (Section 4.3); (4) *Duration of the exhibition*: prolonged exposure beyond 15 minutes with the virtual environment does not give the best result for the sense of presence with HMD (Head Mounted Display) (Stanney (2000)) and there is even a negative correlation between the prolonged exposure in the virtual environment and the sense of presence (Witmer and Singer, 1998). Consequently, we aim at designing training session of 15 minutes; (5) *Social presence and social presence factors*: the social presence of other individuals (real or avatars), and the ability to interact with these individuals increases the sense of presence (Heeter, 1992). The virtual patient will be endowed with social competencies (social signals, non-verbal alignment, etc); (6) the *Quality of the virtual environment*: quality, realism, the ability of the environment to be fluid, to create interaction are key factors in the sense of presence of the user (Hendrix *et al.*, 1996). In the training session, the fluidity of the virtual environment implies the absence of bugs. Two other factors are more particularly related to the individual perception, and contextual and psychological factors that should be taken into account during the

evaluation of presence (Mestre, 2015). Indeed, these 7 factors of presence will also be evaluated through questionnaires (Witmer and Singer, 1998; Bouchard *et al.*,2014) addressed to doctors after the training sessions.

4.2 Corpus-Based approach to define virtual patient's behavior

In order to model the behavior of the virtual patient (verbal and non-verbal), we propose a corpus-based approach to identify precisely the reactions of the patient (when and what) to replicate it accordingly on the virtual one. For ethical reasons, it is not possible to videotape real breaking bad news situations. Instead, simulations are organized with actors playing the role of the patient. A corpus of such interactions has been collected in different medical institutions (the *Institut Paoli Calmette*[6] and the *hospital of Angers*). Simulated patients are actors trained to play the most frequently observed patients reactions (e.g. accommodating or aggressive). The actor follows a *pre-determined scenario*. The total volume of videos is 5 hours 43 minutes and 8 seconds for 23 videos of patient-doctor interaction with different scenario (e.g. patient aggressive or accommodating, announce of cancer, announce of digestive perforation, etc.).

These simulated interactions of the collected corpus are manually transcribed and annotated (with Elan software[7]) at *several levels*: at the *discourse level* and at the *non-verbal level*. At the *discourse level*, the objective is to annotate the different *dialog phases* (e.g. introduction, announce, future implications) identified by the French Haute Autorité de la Santé (HAS)[8], the turn taking, and the used of medical vocabulary. At the *non-verbal level*, the gestures (absence or presence), the posture, the head movements, the smiles, the eyebrow movements, and the gaze direction (interlocutor, away and self) are annotated. The coding scheme has been defined based on a preliminary analysis of the corpus (Saubesty and Tellier, 2015). Both the doctor and the patient verbal and non-verbal behavior are annotated and transcribed. By this way, we can analyze precisely the coordination of their behavior and identify when the virtual agent should trigger which behavior during the interaction. The annotated corpus is exploited in different ways. The corpus has been used to construct the non-verbal behavior library of the virtual patient.

Indeed, the VIB architecture includes a common library of gestures and facial expressions (e.g. facial expressions of emotions, head nods, etc.) but also more specific non-verbal behaviors. To simulate a virtual patient, new gestures have been created to allow the ECA to play the scenario of a patient to whom doctor announces s/he has an intestinal perforation. Most of the created gestures are used to indicate pain. To identify and describe these gestures, we analyzed videos of the corpus described above. By analyzing these practical cases of videos featuring a doctor and patient, we can observe a number of recurring features and gestures of the patient in such situation. We illustrate some of these gestures below (Figures 2) and the corresponding gestures

---

[6] http://www.institutpaolicalmettes.fr/
[7] http://tla.mpi.nl/tools/tla-tools/elan/
[8] http://www.has-sante.fr/

created for the patient. Finally, we have created 16 specific gestures to simulate the virtual patient that have been integrated in the *Behavior Library* of the *Scenario Controller* (Figure 1) containing already generic non-verbal signals such as facial expressions or backchannels.

The annotated corpus is also currently exploiting to analyze the multimodal coordination of the doctors and patient behavior. Indeed, we are developing temporal sequences mining algorithms to identify the temporal relations between inter-individual modalities, for instance to identify the potential reactions of the patient when the doctor is using a medical term or the coordination of head movements of the patient depending on the patient's attitude.

Figure 2: The patient locates a pain in the stomach, with both hands

Note that an explicit computational representation of emotions is not integrated in the virtual patient. In the virtual patient model and in the Scenario Controller, we manipulate directly social signals (e.g. head movements, gaze direction, gestures) annotated in the corpus.

4.3 From real medical environment to virtual one

The breaking bad news is performed by the doctor in the recovery room. The recovery room welcomes patients after endoscopic or surgical interventions realized under general anesthesia after operating room. This specific area allows increasing safety of patients and is used to detect quickly surgery or medical complications. In order to simulate this medical environment, we have designed a "realistic" virtual environment, from the doctor's point of view. To address that point, we start from images of real recovery room (Fig. 3, left) and develop equivalent virtual environments (Fig. 3, right).

Figure 3: Recovery room real (left) and virtual (right).

5   Perspectives

The presented platform to train doctors to break bad news is currently under development. Several research axes are explored to create a natural multimodal interaction. We aim at developing a dialog model to enable the virtual patient to interrupt the doctor to ask specific questions not handled by the doctor but corresponding to pedagogical objective. We have developed a corpus illustrating the different behaviors of the doctor and the patient and containing multi-level annotations of these behaviors. We will rely on these annotations in order to automatically extract temporal multimodal sequences and learn model of the non-verbal behavior for both the detection of the doctor's behavior and the generation of the virtual patient's behavior.

To evaluate the platform, two types of evaluation will be considered: one, intrinsic to the system, aiming at estimating the quality of the conversational virtual patient, in terms of communication skills, the other proposing a measurement of the trainee performance. As for the first evaluation, different measures will be done, according to the component to be evaluated. Speech recognition will be measure with classical techniques (recall/precision), against the general corpus. The virtual patient will be evaluated thanks to user perceptive tests concerning the verbal and non-verbal behavior of the agent: the relevance of its behaviors, the coherence of the attitude, the adaptation to the input information, etc. The trainee evaluation is an entire research subject. An evaluation grid will be defined, starting from the existing one currently used in the hospitals: the « Affective Competency Score » (Quest, Ander et al. 2006). The ACS will be scored by the trainees to measure their self-efficacy before and after a session with the virtual patient. Professional observers will also rate the ACS to evaluate the trainees' performances. Moreover, a presence test questionnaire (such as the Witmer questionnaire (Witmer et al., 1998)) will be used to assess the interactive experience of the trainees on different platform more or less immersive (PC, CAVE and HMD).

6 References

1. Andrade AD, Bagri A, Zaw K, Roos BA, Ruiz JG. Avatar-mediated training in the delivery of bad news in a virtual world. J Palliat Med 2010;13:1415-9.
2. Aylett, Matthew P., and Christopher J. Pidcock. "The CereVoice characterful speech synthesiser SDK." *IVA*. 2007.
3. Billinhurst, M., & Weghorst, S. (1995). The use of sketch maps to measure cognitive maps virtual of environments. In Proceeding of *Virtual Reality Annual International Symposium* (VRAIS '95), pp. 40-47.
4. Bouchard S., Robillard G., St-Jacques J., Dumoulin S., Patry M.J & Renaud P., (2014) Reliability and Validity of a Single -Item Measure of Presence in VR.
5. Cruz-Neira, C., Sandin, D. J., & DeFanti, T. A. (1993). Surround-screen projection-based virtual reality: the design and implementation of the CAVE. In *proceedings of the 20th annual conference on Computer Graphics and interactive techniques*, 135-142.
6. Heeter, C. (1992). Being there: The subjective experience of presence. Presence: *Teleoperators and Virtual Environments*, *1*(2), 262-271.
7. Hendrix, C. & Barfield, W. 1996. Presence within virtual environments as a function of visual display parameters. *Presence: Teleoperators and Virtual Environments*, *5*(3).
8. P. Kenny, T. D. Parsons, J. Gratch, and A. A. Rizzo. 2008. Evaluation of Justina: A Virtual Patient with PTSD. In Proceedings of the 8th international conference on Intelligent Virtual Agents (IVA '08), Helmut Prendinger, James Lester, and Mitsuru Ishizuka (Eds.). Springer-Verlag, Berlin, Heidelberg, 394-408.
9. Michel P, Lathelize M, Quenon JL., Bru-Sonnet R, Domecq S, Kret M., Comparaison des deux Enquêtes Nationales sur les Événements Indésirables graves associés aux Soins menées en 2004 et 2009. Rapport final à la DREES (Ministère de la Santé et des Sports) – Mars 2011, Bordeaux.
10. Granry, J-C and MC Moll. HAS 2012 Rapport de mission, état de l'art en matière de pratiques de simulation dans le domaine de la santé.
11. Nocera, P., Linares, G., Massonié, D., and Lefort, L. (2002) "Phoneme lattice based a* search algorithm for speech recognition" in International Conference on Text, Speech and Dialogue, pages 301–308, Brno, Springer
12. C. Pelachaud, Studies on Gesture Expressivity for a Virtual Agent, Speech Communication, special issue in honor of Björn Granstrom and Rolf Carlson, 51 (2009) 630-639
13. T.E. Quest TE, DS Ander, JJ Ratcliff: The validity and reliability of the affective competency score to evaluate death disclosure using standardized patients. J Palliat Med 2006;9:361– 370.
14. Saubesty, J. & Tellier, M. (2015) Multimodal analysis of hand gesture back-channel feedback. Gesture and Speech in Interaction, Sept. 2015 Nantes, France.
15. Slater, M., Linakis, V., Usoh, M., Kooper, R., & Street, G. (2001). Immersion, presence, and performance in virtual environments: An experiment with Tri-Dimensional Chess. ACM Virtual Reality Software and Technology (VRST), 163–172
16. Stanney, K.M. (2000). *Unpublished research data*. University of Central Florida.
17. Witmer, B., & Singer, M. (1998). Measuring presence in virtual environments: A presence questionnaire. Presence, 7(3), 225–240.