

# An architecture of virtual patient simulation platform to train doctor to break bad news

M. Ochs<sup>1</sup>, G. Montcheuil<sup>2</sup>, J.-M. Pergandi<sup>3</sup>, J. Saubesty<sup>2</sup>,  
B. Donval<sup>4</sup>, C. Pelachaud<sup>4</sup>, D. Mestre<sup>3</sup> and P. Blache<sup>2</sup>

Aix Marseille Université CNRS ENSAM, Université de Toulon, <sup>1</sup>LSIS UMR 7296,  
<sup>2</sup>LPL UMR 7309, <sup>3</sup>ISM UMR7287 ; <sup>4</sup> CNRS ISIR, Paris, France

## Abstract

The way doctors deliver bad news related to damage associated with care has a significant impact on the therapeutic process. In order to facilitate the doctor's training, a simulation system has been developed to train doctors to break bad news with a virtual patient. The virtual patient's dialog behavior, based on the notion of "common ground", has been specifically designed from the analysis of real training sessions in medical institutions and in light of pedagogical objectives. The doctors can interact in natural language with the virtual patient on three different platforms: PC, virtual reality headset, and virtual reality room. A tool has been developed to automatically record and then replay the interaction between the doctor and the virtual patient to debrief with an expert about this simulation.

**Keywords:** Embodied conversational agent, virtual reality, health domain, social simulation platform

## 1 Introduction

Doctors should be trained not only to perform 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 with care. A damage associated with care is the con-

sequence of an unexpected event that can be due to complication connected to the pathology of the patient, unforeseeable medical situation, dysfunction or medical error. The damage may 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 in France [1].

The way *doctors deliver bad news related to damage associated with care* has a significant impact on the therapeutic process: disease evolution, adherence with treatment recommendations, litigation possibilities [2]. However, both experienced clinicians and medical trainees consider this task as difficult, daunting, and stressful. Nowadays, training health care professional to break bad news, recommended by the French National Authority for Health (HAS)<sup>1</sup>, is organized as workshops during which doctors disclose bad news to actors playing the role of patients. Training by simulation [3], *i.e.* to put the doctor in a simulated situation of consultation with actors playing patients' role, allows developing the level of understanding of the professionals of this potentially conflicting and painful situation for them, for the patient and his/her close relationships. However, this training solution requires several persons: it is costly, and time consuming (each 30 mn. session requires

---

<sup>1</sup>The French National Authority for Health is an independent public scientific authority with an overall mission of contributing to the regulation of the healthcare system by improving health quality and efficiency.

an hour of preparation). We aim at developing a training system inhabited by an embodied conversational agent playing the role of a virtual patient to give the doctors capabilities to simulate breaking bad news situations.

In this paper, we present a multi-platform simulation system that has been designed to train doctors to break bad news with a virtual patient. The doctors can interact in natural language with a virtual patient that communicates through its verbal and non-verbal behavior. The dialog model of the virtual patient is based on the notion of “common ground” [4], *i.e.* a situation model represented through different variables that is updated depending on the information exchange between the interlocutors. The variables describing the situation model, specific to breaking bad news situations, have been defined based on the analysis of corpus of real training sessions in medical institutions and in light of the pedagogical objective in terms of dialog. The simulation training system can finally be run on three platforms: PC, virtual reality headset, and an immersive virtual reality room. An automatic recording system has been developed to replay the interaction between the doctor and the virtual patient to debrief with an expert about this simulation.

The paper is organized as follows. After a presentation of a state of art in this domain (Section 2), we present the architecture of the multi-platform simulation training system and its components (Section 3). In Section 4, we introduce the multimodal dialog model of the virtual patient based on the notion of “common ground” (Section 4.1), and based on the manual analysis of a transcribed corpus of real training sessions (Section 4.2). The non-verbal behavior model of the virtual patient is described in Section 4.4. The implementation of the model is detailed in Section 4.4. The different platforms to simulate an interaction of breaking bad news situation with a virtual patient are illustrated in Section 5.1 and the automatic recording for the simulation replay is introduced in Section 5.2. We conclude in Section 6.

## 2 Related works

Several virtual agents embodying the role of virtual patients have already been proposed for use in clinical assessments, interviewing and diagnosis training [5, 6, 2]. Indeed, previous research has shown that doctors demonstrate non-verbal behaviors and respond empathetically to a virtual patient [7]. 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. [6, 5] or the eViP European project<sup>2</sup>). In our project, we focus on a virtual patient to train doctors to deliver bad news.

A first study [2] 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 main 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. Our objective in this project is to simulate the non-verbal expressions 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 environments 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. [8]) but also for social patient perspective taking [9]. 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 virtual patient

The overall architecture of the training platform is illustrated in Figure 1 (at the end of the article). We present each module of the architecture in the following.

---

<sup>2</sup><http://virtualpatients.eu/>

The *speech of the doctor* is recognized by an *Automatic Speech Recognition (ASR)* system that we are currently integrating in the platform. We have chosen the commercial Dragon Natural speaking SDK solution developed by Nuance<sup>3</sup>. The audio is preprocessed to eliminate non-speech segments, and cut the speech into shorter segments called IPU segments (Inter-Pausal Units) separated by small silences (at least 200 ms for French). Each IPU is streamed to the ASR system to be transcribed and then sent to the *Supervisor*. Note that when the doctors speech is not well recognized (information sent by the ASR) the supervisor triggers specific verbal or non-verbal behavior (e.g. “I don’t understand” or only a shrug of the shoulders). Moreover, in order to avoid waiting for the result of the ASR to generate a virtual patients reaction, voice activity information is sent in real time to the supervisor (e.g. *doctor\_is\_talking*). This information enables the supervisor to trigger virtual patient’s listening behavior. The supervisor contains a listener model that defines the backchannel behavior of the patient. The listener model will also generate virtual patient’s reaction, if the silence of the doctor reaches a certain threshold of time, by sending a command to the dialog module. This model is currently under development. 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 Supervisor to coordinate the virtual patients’ behavior (e.g. to mimic a head shake). Note that, similarly to the process of voice activity, the doctor’s non-verbal behavior activity is sent in real-time to the supervisor (*doctor\_starts\_movement* and *doctor\_stops\_movement*). Consequently, the supervisor can trigger specific behavior without waiting for the computation of the movement interpretation. For instance, the virtual patient might look at the doctor when the doctor is moving.

The supervisor manages the synchronization of the recognized doctor’s utterance and his detected non-verbal behavior. The resulting multi-

modal utterance is then transferred to the dialog system. This module is described in more details in the next section (Section 4). Note that this module, based on a set of dialog rules and on the conversation log, determines the communicative intention of the virtual patient (what the agent intends to communicate). The communicative intention can be expressed through verbal, non-verbal or multimodal behavior. For instance, an agreement may be conveyed by a head nod, a verbal message or both. The virtual patient’s reaction is transferred to the supervisor. The supervisor selects the appropriate animation file from the behavior library depending on the output of the listener model and of the dialog system.

The animation is sent to the non-verbal behavior animation system VIB. VIB (Virtual Interactive Behavior) is a generic platform for creating Embodied Conversational Agents (ECA) [10]. VIB computes the animation parameters (Facial Animation Parameters FAP and Behavioral Animation Parameters BAP) to animate the face and the body of the virtual patient. The module receives XML file describing the communicative intention to perform (described in FML:Function Markup Language) or the non-verbal signals to express (described in BML: Behavior Markup Language). Moreover, the Non-verbal behavior system VIB contains a text-to-speech system [11] to generate the speech synchronizes with the non-verbal behavior (including the lips animation).

The virtual patient is finally animated on different platforms: PC, virtual reality headset, and virtual reality room. The virtual reality room 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 (Figure 2). The virtual agent based on the VIB platform has been integrated in these different environments through the Unity player. The communication between the modules is based on a Thrift protocol<sup>4</sup> to enable to connect modules developed in different lan-

<sup>3</sup><http://www.nuance.com/for-developers/dragon/index.htm>

<sup>4</sup><https://thrift.apache.org/>

Figure 2: Virtual reality environment



guage.

In the next section, we describe in more details the dialog system integrated in the architecture.

## 4 Corpus-based Virtual patient's multimodal dialog model

In order to model the virtual patient's behavior, we have analyzed audio-visual corpus of interactions between doctors and actors playing the role of patients during real training sessions in medical institutions. Indeed, 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 two different medical institutions. Simulated patients are actors trained to play the most frequently observed patients reactions. 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 scenarios (e.g. cancer diagnosis, digestive perforation's announcement, etc.).

The dialog model of the virtual patient aims at identifying automatically the dialog behavior of the virtual patient during the interaction with the doctor, that includes *verbal* (e.g. specific questions or remarks) and *non-verbal* (e.g. head nods, smiles) reactions to utterances of the doctor. In the following Section 4.1, 4.2 and 4.3, we focus on the verbal behavior of the virtual patient and present the non-verbal behavior model Section 4.4.

### 4.1 A dialog model based on the construction of a common ground

Concerning the verbal behavior, in order to identify the contents of the virtual patient's verbal reaction, we propose a dialog model based on the notion of *common ground* introduced by Garrod and Pickering [4]. Conversation is then viewed as a *joint activity* during which the interlocutors "work together to establish a joint understanding of what they are talking about" [4]. The joint activity is based on the alignment of their *situation models* containing information about space, time, causality, intentionality, etc. In other words, the interlocutors interact to construct a common representation of a situation, called an *implicit common ground*.

In our context, the common ground that the interlocutors (the doctor and the virtual patient) have to construct concerns about the situation of disclosure a damage associated with care. The French National Authority for Health (HAS) produces recommendations and best practice guidelines to facilitate the disclose of unfavorable information to patients [12]. Based on this guideline and on the analysis of the training corpus [13], five principal phases have been drawn from the data: "opening" (e.g. presentation, inquiring of the patient's state), "exposing the situation" (e.g. a reminder of the patient's care since he/she arrived in the hospital), "breaking the news" (e.g. clear exposition of known facts), "discussing the future" (e.g. what solution for the damage, who will perform it, where, ...) and "closing". For each phase, guideline describes the different information that the doctor should deliver to the patient concerning this breaking bad news situation. For instance, in the "breaking the news" phase, the doctor should, at least, inform the patient on the type of the problem (e.g. digestive perforation), when it occurred (e.g. during a surgical operation), the location (e.g. in the stomach), and the cause (e.g. the polyp wasn't positioned properly). In the "discussing the future" phase, the doctor should inform on the solution to cope with the situation (e.g. another surgical operation), who (e.g. the doctor that will operate), when (e.g. rapidly, tomorrow), why (e.g. risk of infection), what they should do before (e.g. radiology), what will be the consequences (e.g. stay some days at hospi-

tals), etc.

In order to construct the situation model, *i.e.* the common ground that the doctor and the patient should construct together, we have associated a variable to each information that the doctor should deliver to the patient. For instance, we have defined for the step “breaking the news”, 4 variables : `type_problem`, `when_problem`, `location_problem`, `cause_problem`. In total, we have defined 12 variables.

Finally, a situation model is described by this set of phases and associated variables. A common ground is constructed if all the variables are instantiated, *i.e.* if the doctor has provided all the information characterized by the variables. In the following, we call these set of phases and variables the common ground.

The dialog model is based on this common ground representation. The variables are used both to store the information provided by the doctor and to determine the reaction of the patient. Indeed, depending on the recognized verbal utterances of the doctors, the variables will be instantiated. For instance, if the doctor provides information on the location of the damage, the variable `location_problem` will be instantiated with the location. Moreover, the virtual patient will use the common ground, and in particular the non-instantiated variables, to determine his/her reactions. Indeed, the virtual patient will ask specific information to instantiate all the variables.

Note that the variables describing the situation correspond to *pedagogical objectives* of the breaking bad news situation in terms of dialog. Indeed, the variables correspond to the set of information that the doctors have to provide to the patient concerning the damage as specified by the French National Authority for Health (HAS). The dialog model based on this notion of common ground is then particularly suitable in a learning context since it has the advantage of integrating the learning objectives concerning the content of the conversation.

The described dialog model has been implemented for a specific scenario of breaking bad news situation described in the following.

## 4.2 The corpus-based dialog model

In order to test the dialog model, we have selected a specific scenario of breaking bad news situation. The situation is a digestive perforation that had occurred during an endoscopy. The scenario has been carefully chosen with the medical partners of the project for several reasons : the panel of resulting damages, the difficulty of the delivery, and the bad news standard characteristics.

To construct the dialog model for this specific scenario, we have manually analyzed transcribed corpus with this scenario with three objectives:

1. Validate the situation model: check that we can identify the different phases and variables of the situation model;
2. Identify the different values of the variables in this specific context of the digestive perforation;
3. Identify the appropriate verbal or non-verbal responses of the virtual patient.

For this purpose, we have analyzed 7 dialogs of a total duration of 108 mn (each dialog lasts from 8 mn to 27 mn).

Concerning the first objective, the dialogs of the analyzed corpus contain, indeed, the different identified phases and the doctors provide the different information corresponding to the defined variables. Consequently, the situation model that we have defined seems to enable us to represent this specific breaking bad news situation.

Concerning the second objective, we have manually analyzed the corpus in order to identify the different values of each variable. For instance, the type of problem (the variable `type_problem`) can be described as a digestive perforation, or a small hole. To identify the different values that the variables can take enables us to instantiate automatically the variables depending on the recognized speech of the doctor. For instance, if the ASR (Automatic Speech Recognition) recognizes the words “digestive perforation”, we can instantiate the variable `type_problem` with this value. All this val-

ues constitute the domain vocabulary in our specific context.

Concerning the third objective, we have manually identified the verbal reaction of the patient following an utterance of a doctor corresponding to the instantiation of a variable. We have observed several possible reactions as for instance questions to obtain more information on the delivered information or on non-instantiated variables. For instance, after the doctor informs the patient that the solution to cope with the damage (variable `solution_what`) is a surgical operation, the patient may worry about the operation and asks questions such as “which operation?” or “is it a big operation?” or ask about missing variables, e.g. “who will operate?” (variable `solution_who`) or “when” (`solution_when`).

For each variable, we have defined communicative intentions that the virtual patient could perform if the variable is not instantiated (e.g. `Ask(problem_type)`). To each communicative intention is associated one or several possible verbal or non-verbal reactions (e.g. “what exactly is this operation” or expression of surprise).

In order to align the doctor’s and patient’s vocabulary, the words used by the virtual patient depend on the words uttered by the doctor. For instance, if the doctor does not use the term “perforation” but only “small hole”, the patient will never use “perforation” but only “small hole”. The pre-defined responses of the virtual patient are then filled-in-the-blank sentences, the holes being filled to ensure that the virtual patient used the same vocabulary as the doctor did.

Finally, the situation model, and particularly the variables describing the model, enables us to make a link between the doctor’s recognized speech and the verbal reaction of the virtual patient.

### ***4.3 Implementation of the dialog model***

The dialog model presented below was implemented using OpenDial [14]. OpenDial is a java-based, domain-independent toolkit for developing spoken dialogue systems. The dialog is modeled in OpenDial through probabilistic rules. The rules are organized into three subsets corresponding the classical components of a dialog system : Natural Language Understanding

(NLU), Dialog Management (DM), and Natural Language Generation (NLG).

The Natural Language Understanding (NLU) rules translate the recognized speech provided by the ASR into doctor’s dialog acts. Moreover, the NLU rules determine the instantiation of the variables (that compose the common ground) depending on the recognized speech provided by the ASR. The rules enable us to also collect the words uttered by the doctor to constitute a shared vocabulary. The rules are designed based on the analysis of the corpus described in the previous section. For instance, the utterance “There is a small hole in your intestine” corresponds to two informative dialog acts `Inform(problem_type, small hole)` and `Inform(problem_where, intestine)`. Following this recognized sentence, we automatically add “small hole” to the shared vocabulary.

The Dialog Management (DM) rules determine the next patient’s communicative intention(s) based on the last doctor’s dialog acts and the current state of the common ground. The set of reactions following the instantiated variables and the questions about missing information are based on the corpus analysis (previous section). The probability of the different possible virtual patient’s reactions depends on the history of the dialog (if the virtual patient has already performed the dialog act) and the occurrence of such reactions in the corpus.

Finally, the Natural Language Generation (NLG) rules define for each communicative intention the different possible verbal or non-verbal expressions with associated probability. For example, the expression of an agreement could be either a non-verbal behavior (head nod), a verbal message (“yes”) or both. The shared vocabulary is also used by the generation to align the doctors and virtual patients vocabulary: the words used by the virtual patient depends on the words used by the doctor.

### ***4.4 From verbal to multimodal virtual patient’s behavior***

An important facet of our project is to simulate the multimodality of human-human conversation. Verbal and nonverbal behaviors are important means of communication. To this aim we have endowed the virtual character with the

capacity to speak and listen through his/her multimodal behaviors.

To ensure the character looks alive during the whole interaction, we have developed various idle motions whose amplitude and frequency for each body part can be adapted. Specific animations are related to the scenario (e.g. to show fatigue or pain, etc).

As shown in various studies [15], non-verbal communicative behaviours may serve different functions such as emphasizing some parts of what is said, transform the modality (i.e. (un)certainty), communicating an emotional state or a cognitive state (eg (un)certainty), etc. Nonverbal behaviours, specially gestures but also facial expressions and head movement, convey semantic content. In particular, the virtual character indicates where his/her pain is using deictic gestures. Deictic gestures can correspond to a precise location (pointing finger acting as an area) or to a region (the whole hand moves around to indicate the region). The character can illustrate its pain using iconic gestures that were extracted from the analysis of the corpus.

While listening, a great variety of multimodal behaviours, often called backchannels, are emitted to indicate the agents level of attention (eg by gazing at its interlocutor) and understanding (head nods or puzzled facial expression may provide information regarding level of understanding), to manage dialog turns (gaze direction and body orientation are important cues of turn-taking), to hold the turn while looking for words (e.g. scratching the back of the head). Backchannels are triggered by the scenario of the virtual character but also using an existing probabilistic model [16]. Currently, we are trying to improve the backchannel model by analyzing the feedback behavior of patients in the corpus using data-mining algorithms.

In a next step, we aim at integrating a real-time detection of the position of the doctor in order to control the eyes and head orientation of the virtual patient: for instance, to focus on the doctor and to follow him, or on the contrary to intentionally avoid the doctors gaze.

## 5 Virtual reality training platforms

The participant-virtual agent interaction quality may vary as a function of the visualization device (virtual reality room, virtual reality headset, or simple PC). Moreover, the available devices may depend on the medical institution. In order to offer a complete training tool for clinical personals, we have developed the simulation training system on different platforms: PC, virtual reality 3D glasses, and a virtual reality room. The environment has been designed to simulate a real recovery room where the breaking bad news is generally performed. The virtual patient is seated on the bed of the recovery room (Figure 3) and interacts in real-time with the doctor, using natural language and through non-verbal behavior (as described in the previous section). This virtual medical environment has been developed within Unity 3D. To deploy the environment on the different visualization devices, we used our home-made VR distribution platform (VRDistrib). In the next section, we describe in more details the different platforms (Section 5.1) and the tool to record and replay the simulated interaction (Section 5.2).

### 5.1 Multi-platorms of training

The immersive virtual reality room is a 4-sided CAVE. The projection system consists in 4 active-stereo DLP projectors feeding 4 screens (frontal, sides and floor) positioned in a “cubic” configuration (Figure 3). The CAVE involves a real-time optical tracking system, connected to a graphic PC cluster, enabling the updating of a stereoscopic visualization of the virtual world as a function of the participant’s viewpoint and actions (framerate is 60 Hz). This system affords optimal immersion of the participant. Besides this system, virtual worlds can be fed into an Oculus Rift (CV1) or into a single PC monitor, using the same rendering software and tracking system.

### 5.2 Automatic recording of the multimodal interaction

The goal of the platform is to provide a virtual environment to enable doctors to train break-

Figure 3: A user interacting with the virtual patient in the CAVE (left) and through the Oculus Rift (right)

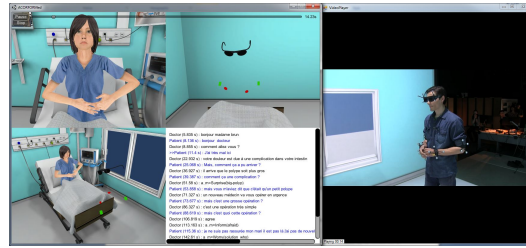


ing bad news situation. The system is not self-sufficient: we do not want to provide a mark to the doctor. The objective is for the doctor to debrief with an expert in breaking bad news situation. For this purpose, we aim at providing different tools to facilitate this debriefing between the trainee and the trainer :

- a tool to automatically replay the interaction ;
- a tool to compute some key elements of performance such as the use of complex medical terms, or the gaze direction of the doctor.

As a first step, we have developed a tool to replay the simulation between a doctor and the virtual patient. The doctor is filmed using a video camera. Moreover, his gestures and his head movements are digitally recorded from the tracking data: his head (stereo glasses), elbows and wrists are equipped with tracked targets. A high-end microphone synchronously records the participant’s verbal expression. Concerning the virtual agent, its gesture and its verbal expression are recorded in the Unity Player. To visualize the interaction, we have developed a 3D video playback player (Figure 4). This player reviews the animation and verbal expression of the virtual agent as well as the movements and video of the participant. The next step, based on the data recorded from the tracked targets and based on the recognized speech, is to automatically compute performance cues highlighted by the medical expert such as the quantity of gestures, the number of complex medical terms used, the head movements (nod, shake, etc.).

Figure 4: 3D video playback player



These different tools will moreover be used for research purposes to collect a corpus of doctor-virtual patient interaction in order to analyze the behavior of the doctor (verbal and non-verbal) interacting with a virtual patient in different immersive environments.

## 6 Conclusion

In conclusion, in this article, we have presented a training platform for doctor’s communicative skills. A virtual actor plays the role of a patient to train doctor to break bad news. Based on pedagogical conversational objectives specific to this medical situation, the dialog behavior of the virtual actor has been designed to trigger specific agent’s reactions in order to lead the doctor to provide the different information. Even if the dialog module has been implemented for a particular scenario, the proposed framework based on the notion of “common ground” could be applied to other breaking bad news situations.

The training platform can be run on different devices with different degree of immersion (PC, CAVE, and 3D headset). We are currently conducting experiments to evaluate the behavior and performance of naive participants and doctors depending on the used devices. We aim at analyzing both the subjective experience of the participants through questionnaires but also the different verbal and non-verbal behavior of the participants computing objective measures such as the quantity of gestures or the length of sentences.

## Acknowledgements

This work has been funded by the French National Research Agency project ACOR-



FORMED (ANR-14-CE24-0034-02).

## References

- [1] Quenon JL, Bru-Sonnet R, Domecq S, Kret M, Michel P, Lathelize M. Comparaison des deux enquêtes nationales sur les événements indésirables graves associés aux soins menées en 2004 et 2009. Technical report, DREES), 2011.
- [2] A.D Andrade, A. Bagri, K. Zaw, B.A Roos, and J.G Ruiz. Avatar-mediated training in the delivery of bad news in a virtual world. *Journal of palliative medicine*, 13(12):1415–1419, 2010.
- [3] J-C Granry and MC Moll. Rapport de mission, état de l'art en matière de pratiques de simulation dans le domaine de la santé. Technical report, HAS 2012, 2012.
- [4] Simon Garrod and Martin J Pickering. Why is conversation so easy? *Trends in cognitive sciences*, 8(1):8–11, 2004.
- [5] Benjamin Lok, Richard E Ferdig, Andrew Raij, Kyle Johnsen, Robert Dickerson, Jade Coutts, Amy Stevens, and D Scott Lind. Applying virtual reality in medical communication education: current findings and potential teaching and learning benefits of immersive virtual patients. *Virtual Reality*, 10(3-4):185–195, 2006.
- [6] Patrick Kenny, Thomas D Parsons, Jonathan Gratch, and Albert A Rizzo. Evaluation of justina: a virtual patient with ptsd. In *International Workshop on Intelligent Virtual Agents*, pages 394–408. Springer, 2008.
- [7] A.M Deladisma, M. Cohen, A. Stevens, P. Wagner, B. Lok, T. Bernard, C. Oxendine, L. Schumacher, K. Johnsen, R. Dickerson, et al. Do medical students respond empathetically to a virtual patient? *The American Journal of Surgery*, 193(6):756–760, 2007.
- [8] Thomas D Parsons and Albert A Rizzo. Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: A meta-analysis. *Journal of behavior therapy and experimental psychiatry*, 39(3):250–261, 2008.
- [9] Andrew Raij, Aaron Kotranza, D Scott Lind, and Benjamin Lok. Virtual experiences for social perspective-taking. In *Virtual Reality Conference, 2009. VR 2009. IEEE*, pages 99–102. IEEE, 2009.
- [10] Catherine Pelachaud. Studies on gesture expressivity for a virtual agent. *Speech Communication*, 51(7):630–639, 2009.
- [11] Matthew P Aylett and Christopher J Paddock. The cerevoice characterful speech synthesiser sdk. In *IVA*, pages 413–414, 2007.
- [12] C. Schnebelen, F. Pothier, and M. Furney. Annonce d'un dommage associé aux soins. Technical report, Haute Autorité de Santé., 2011.
- [13] J. Saubesty and M. Tellier. Multimodal analysis of hand gesture back-channel feedback. In *Gesture and Speech in Interaction, Nantes, France*.
- [14] P. Lison and C. Kennington. OpenDial: A Toolkit for Developing Spoken Dialogue Systems with Probabilistic Rules. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Demonstrations)*, pages 67–72, Berlin, Germany, 2016. Association for Computational Linguistics.
- [15] Laurent Prévot, Jan Gorisch, and Roxane Bertrand. A cup of coffee: A large collection of feedback utterances provided with communicative function annotations. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, may 2016.
- [16] 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.

Figure 1: Overall architecture of the virtual patient simulation platform

