



HAL
open science

The Acorformed Corpus: Investigating Multimodality in Human-Human and Human-Virtual Patient Interactions

Magalie Ochs, Philippe Blache, G. Moreau De Montcheuil, J M Pergandi, R. Bertrand, J Saubesty, D Francon, D Mestre

► To cite this version:

Magalie Ochs, Philippe Blache, G. Moreau De Montcheuil, J M Pergandi, R. Bertrand, et al.. The Acorformed Corpus: Investigating Multimodality in Human-Human and Human-Virtual Patient Interactions. CLARIN-2018, Oct 2018, Pisa, Italy. hal-01907582

HAL Id: hal-01907582

<https://hal.science/hal-01907582v1>

Submitted on 29 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The Acorformed Corpus: Investigating Multimodality in Human-Human and Human-Virtual Patient Interactions

M. Ochs¹, P. Blache², G. Montcheuil^{2,3,4}, J.M. Pergandi³,
R. Bertrand², J. Saubesty², D. Francon⁵, and D. Mestre³

Aix Marseille Université, Université de Toulon, CNRS,
¹LIS UMR 7020, ²LPL UMR 7309, ³ISM UMR 7287 ; ⁴Boréal Innovation,
⁵Institut Paoli-Calmettes (IPC), Marseille, France

Abstract

The paper aims at presenting the Acorformed corpus composed of human-human and human-machine interactions in French in the specific context of training doctors to break bad news to patients. In the context of human-human interaction, an audiovisual corpus of interactions between doctors and actors playing the role of patients during real training sessions in French medical institutions have been collected and annotated. This corpus has been exploited to develop a platform to train doctors to break bad news with a virtual patient. The platform has been exploited to collect a corpus of human-virtual patient interactions annotated semi-automatically and collected in different virtual reality environments with different degree of immersion (PC, virtual reality headset and virtual reality room).

1 Introduction

For several years, there has been a growing interest in Embodied Conversational Agents (ECAs) to be used as a new type of human-machine interface. ECAs are autonomous entities, able to communicate verbally and nonverbally (Cassell, 2000). Indeed, several researches have shown that embodied conversational agents are perceived as social entities leading users to show behaviors that would be expected in human-human interactions (Krämer, 2008).

Moreover, recent research has shown that virtual agents could help human beings *improve their social skills* (Anderson et al., 2013; Finkelstein et al., 2013). For instance in (Anderson et al., 2013), an ECA endowed the role of a virtual recruiter is used to train young adults to job interview. In our project, we aim at developing a virtual patient to train doctors to break bad news. Many works have shown that doctors should be trained not only to perform medical or surgical acts but also to develop skills in communication with patients (Baile et al., 2000; Monden et al., 2016; Rosenbaum et al., 2004). Indeed, the way doctors deliver bad news has a significant impact on the therapeutic process: disease evolution, adherence with treatment recommendations, litigation possibilities (Andrade et al., 2010). However, both experienced clinicians and medical students consider this task as difficult, daunting, and stressful. Training health care professional to break bad news is now recommended by several national agencies (e.g. the French National Authority for Health, HAS)¹.

A key element to exploit embodied conversational agents for social training with users is their *believability* in terms of socio-emotional responses and global multimodal behavior. Several research works have shown that non-adapted behavior may significantly deteriorate the interaction and the learning (Beale and Creed, 2009). One methodology to construct believable virtual agent is to develop model based on the analysis of corpus of human-human interaction in the social training context (as for instance in (Chollet et al., 2017)). In our project, in order to create a virtual patient with believable multimodal reactions when the doctors break bad news, we have collected, annotated, and analyzed two multimodal corpora of interaction in French in this context. Both human-human and human-machine interaction are considered to investigate the effects of the virtual reality displays on the interaction. In this paper, we present the two corpus in the following sections.

¹The 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.

2 Multimodal Human-Human Corpus Analysis to Model Virtual Patient's Behavior

The modeling of the virtual patient is based on an audiovisual corpus of interactions between doctors and actors playing the role of patients (called “Standardized patients”) during real training sessions in French medical institutions (it is not possible, for ethical reasons, to record real breaking bad news situations). The use of “Standardized Patients” in medical training is a common practice. The actors are carefully trained (in our project, actors are also nurses) and follow pre-determined scenarios defined by experts to play the most frequently observed patients reactions. The recommendations of the experts, doctors specialized in breaking bad news situations, are global and related to the attitude of the patient ; the verbal and non-verbal behavior of the actor remains spontaneous. Note that the videos of the corpus have been selected by the experts as representative of real breaking bad news situations.

On average, a simulated consultation lasts 9 minutes. The collected corpus, in French, is composed of 13 videos of patient-doctor interaction (the doctor or the patient vary in the video), with different scenarios².

The initial corpus has been semi-manually annotated, leading to a total duration of 119 minutes. Different tools have been used in order to annotate the corpus. First, the corpus has been automatically segmented using SPPAS (Bigi, 2012) and manually transcribed using Praat (Boersma, 2002). The doctors' and patient's non-verbal behaviors have been manually annotated using ELAN (Sloetjes and Wittenburg, 2008). Different gestures of both doctors and patients have been annotated: head movements, posture changes, gaze direction, eyebrow expressions, hand gestures, and smiles. Three experts annotated one third of the corpus each. In order to validate the annotation, 5% of the corpus has been annotated by one more annotator. The inter-annotator agreement, using Cohen's Kappa, was satisfying ($k=0.63$). More details on the corpus are presented in (Porhet et al., 2017).

The annotated corpus has been analyzed for three different purposes:

- to build the *dialog model of the virtual patient*: the dialog model of the virtual patient is based on the notion of “*common ground*” (Garrod and Pickering, 2004; Stalnaker, 2002), *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 (e.g. the cause of the damage), specific to breaking bad news situations, have been defined based on the manual analysis of the transcribed corpus and in light of the pedagogical objective in terms of dialog. The dialog model is described in more detail in (Ochs et al., 2017) ;
- to design *non-verbal behaviors of the virtual patient*: the corpus has been used to enrich the non-verbal behavior library of the virtual patient with gestures specific to breaking bad news situations.
- to design *the feedback behavior of the virtual patient*: in order to identify the multimodal signals triggering patient's feedbacks, we have applied sequences mining algorithms to extract rules to model the multimodal feedback behavior of the virtual patient (for more details (Porhet et al., 2017)).

3 Multimodal Human-Virtual Patient Corpus Analysis to Investigate the Users' experience with different virtual reality displays

Based on the corpus analysis presented in the previous section, we have implemented a virtual reality training system inhabited by a virtual patient and developed to give the capabilities to doctors to simulate breaking bad news situation. The system is *semi-autonomous* since it includes both automatic and manual modules, making it possible to simulate a fully automatized human-machine interaction (for more details on the semi-autonomous system (Ochs et al., 2018)). Implemented on three different virtual environment displays (PC, virtual reality headset, and an immersive virtual reality room), the doctors can interact in natural language with a virtual patient that communicates through its verbal and non-verbal behavior (Figure 1).

²The corpus is on Ortolang part of the CLARIN infrastructure



Figure 1: Participants interacting with the virtual patient with different virtual environment displays (from left to right): virtual reality headset, virtual reality room, and PC.

In order to collect the interaction and create the corpus of human-machine interaction in the context of breaking bad news, we have implemented a specific methodology. First, the doctor is filmed using a camera. His gestures and 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. As for the virtual agent, its gesture and verbal expressions are recorded from the Unity Player. The visualization of the interaction, is done through a 3D video playback player we have developed (Figure 2). This player replays synchronously the animation and verbal expression of the virtual agent as well as the movements and video of the participant.

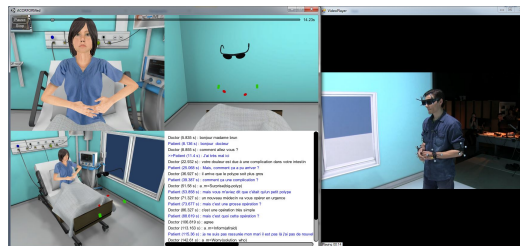


Figure 2: 3D video playback player

This environment facilitates the collection of corpora of doctor-virtual patient interaction in order to analyze the verbal and non-verbal behavior in different immersive environments.

Using the semi-autonomous system, we have collected 108 interactions in French of participants with the virtual patient. In total, 36 persons have participated to the experimentation. Ten of them are real doctors that already have an experience in breaking bad news to real patients. Each participant has interacted with the systems 3 times with three different devices: PC, virtual reality headset, and virtual reality room. The task of the participants was to announce a digestive perforation after a gastroenterologic endoscopy in immediate post operative period³.

The collected corpus is composed of 108 videos (36 per device). The total duration of the corpus is 5h34 (among which two hours with real doctors). In average, an interaction lasts 3mn16 (an example of interaction is presented on the <http://crvm.ism.univ-amu.fr/en/acorformed.html>). Note that thanks to the tools described in the previous section, some of the non-verbal participant behavior has been automatically annotated.

In order to evaluate the global experience of the users, we asked the participants to fill different questionnaires on their subjective experience to measure their feeling of presence (with the *Igroup Presence Questionnaire*, IPQ (Schubert, 2003)), feeling of co-presence (Bailenson et al., 2005), and perception of the believability of the virtual patient (questions extracted from (Gerhard et al., 2001))⁴. These subjective

³The 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 announcement).

⁴The analyze of the subjective experience of the participants is out of scope of this paper and is described in other articles

evaluations enabled us to *tag* the video of the corpus with the results of these tests and then to correlate objective measures (e.g. verbal and non-verbal behavior of the participants) to subjective measures (e.g. feeling of presence and perception of the virtual patient's believability).

Acknowledgements

This work has been funded by the French National Research Agency project ACORFORMED (ANR-14-CE24-0034-02) and supported by grants ANR-16-CONV-0002 (ILCB), ANR-11-LABX-0036 (BLRI) and ANR-11-IDEX-0001-02 (A*MIDEX).

References

- K. Anderson, E. André, T. Baur, S. Bernardini, M. Chollet, E. Chryssafidou, I. Damian, C. Ennis, A. Egges, P. Gebhard, et al. 2013. The tardis framework: intelligent virtual agents for social coaching in job interviews. In *Advances in computer entertainment*, pages 476–491. Springer.
- A. D. Andrade, A. Bagri, K. Zaw, B. A. Roos, and Ruiz J. G. 2010. Avatar-mediated training in the delivery of bad news in a virtual world. *Journal of palliative medicine*, 13(12):1415–1419.
- W. Baile, R. Buckman, R. Lenzi, G. Gloger, E. Beale, and A. Kudelka. 2000. Spikes—a six-step protocol for delivering bad news: application to the patient with cancer. *Oncologist*, 5(4):302–311.
- J. N. Bailenson, C. Swinth, K. Hoyt, S. Persky, A. Dimov, and J. Blascovich. 2005. The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments. *Presence: Teleoperators and Virtual Environments*, 14(4):379–393.
- R. Beale and C. Creed. 2009. Affective interaction: How emotional agents affect users. *International journal of human-computer studies*, 67(9):755–776.
- B. Bigi. 2012. Sppas: a tool for the phonetic segmentations of speech. In *The eighth international conference on Language Resources and Evaluation*.
- P. Boersma. 2002. Praat, a system for doing phonetics by computer. *Glott international*, 13(341-345).
- J. Cassell. 2000. More than just another pretty face: Embodied conversational interface agents. *Communications of the ACM*, 43:70–78.
- M. Chollet, M. Ochs, and C. Pelachaud. 2017. A methodology for the automatic extraction and generation of non-verbal signals sequences conveying interpersonal attitudes. *IEEE Transactions on Affective Computing*.
- S. Finkelstein, S. Yarzebinski, C. Vaughn, A. Ogan, and J. Cassell. 2013. The effects of culturally congruent educational technologies on student achievement. In *International Conference on Artificial Intelligence in Education*, pages 493–502. Springer.
- S. Garrod and M. Pickering. 2004. Why is conversation so easy? *Trends in cognitive sciences*, 8(1):8–11.
- M. Gerhard, D. J Moore, and D. Hobbs. 2001. Continuous presence in collaborative virtual environments: Towards a hybrid avatar-agent model for user representation. In *International Workshop on Intelligent Virtual Agents*, pages 137–155. Springer.
- N. Krämer. 2008. Social effects of virtual assistants. a review of empirical results with regard to communication. In *Proceedings of the international conference on Intelligent Virtual Agents (IVA)*, pages 507–508, Berlin, Heidelberg. Springer-Verlag.
- K. Monden, L. Gentry, and T. Cox. 2016. Delivering bad news to patients. *Proceedings (Baylor University Medical Center)*, 29(1).
- M. Ochs, G. Montcheuil, J-M Pergandi, J. Saubesty, B. Donval, C. Pelachaud, D. Mestre, and P. Blache. 2017. An architecture of virtual patient simulation platform to train doctor to break bad news. In *International Conference on Computer Animation and Social Agents (CASA)*.

currently under review.

- M. Ochs, P. Blache, G. Montcheuil, J.-M. Pergandi, J. Saubesty, D. Francon, and D. Mestre. 2018. A semi-autonomous system for creating a human-machine interaction corpus in virtual reality: Application to the acorformed system for training doctors to break bad news. In *Proceedings of LREC*.
- C. Porhet, M. Ochs, J. Saubesty, G. Montcheuil, and R. Bertrand. 2017. Mining a multimodal corpus of doctor's training for virtual patient's feedbacks. In *Proceedings of 19th ACM International Conference on Multimodal Interaction (ICMI), Glasgow, UK*.
- M. Rosenbaum, K. Ferguson, and J. Lobas. 2004. Teaching medical students and residents skills for delivering bad news: A review of strategies. *Acad Med*, 79.
- T. Schubert. 2003. The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness. *Zeitschrift für Medienpsychologie*, 15(69-71).
- H. Sloetjes and P. Wittenburg. 2008. Annotation by category: Elan and iso dcr. In *6th International Conference on Language Resources and Evaluation*.
- R. Stalnaker. 2002. Common ground. *Linguistics and Philosophy*, 25(5):701–721.