The Brain-IHM Dataset: a New Resource for Studying the Brain Basis of Human-Human and Human-Machine Conversations

Magalie Ochs, Roxane Bertrand, Aurélie Goujon, Deirdre Bolger, Anne-Sophie Dubarry, Philippe Blache

▶ **To cite this version:**

Magalie Ochs, Roxane Bertrand, Aurélie Goujon, Deirdre Bolger, Anne-Sophie Dubarry, et al.. The Brain-IHM Dataset: a New Resource for Studying the Brain Basis of Human-Human and Human-Machine Conversations. Language Resources and Evaluation Conference (LREC), May 2020, Marseille, France. hal-02893898

HAL Id: hal-02893898
https://hal.science/hal-02893898v1

Submitted on 8 Jul 2020

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 Brain-IHM Dataset: a New Resource for Studying the Brain Basis of Human-Human and Human-Machine Conversations

Magalie Ochs[1], Roxane Bertrand[2], Aurélie Goujon[2], Deirdre Bolger[2], Anne-Sophie Dubarry[2], Philippe Blache[2]

[1]Aix Marseille Univ, Université de Toulon, CNRS, LIS, Marseille, France
[2]Aix Marseille Univ, LPL, Marseille, France
{first-name.last-name}@univ-amu.fr

Abstract

This paper presents an original dataset of controlled interactions, focusing on the study of feedback items. It consists on recordings of different conversations between a doctor and a patient, played by actors. In this corpus, the patient is mainly a listener and produces different feedbacks, some of them being (voluntary) incongruent. Moreover, these conversations have been re-synthesized in a virtual reality context, in which the patient is played by an artificial agent. The final corpus is made of different movies of human-human conversations plus the same conversations replayed in a human-machine context, resulting in the first human-human/human-machine parallel corpus. The corpus is then enriched with different multimodal annotations at the verbal and non-verbal levels. Moreover, and this is the first dataset of this type, we have designed an experiment during which different participants had to watch the movies and give an evaluation of the interaction. During this task, we recorded participant's brain signal. The Brain-IHM dataset is then conceived with a triple purpose: 1/ studying feedbacks by comparing congruent vs. incongruent feedbacks 2/ comparing human-human and human-machine production of feedbacks 3/ studying the brain basis of feedback perception.

Keywords: Feedbacks, multimodality, human-human/human-machine parallel corpus, brain basis of feedback, perception study, EEG, brain

1. Introduction

One of the main issues is the understanding of how interaction between two humans can be successful based on the behaviour of the different interlocutors during a conversation. In particular, among the parameters guaranteeing the quality and success of an interaction, the attitude of the participant when listening to the main speaker (nodding, facial expressions, etc.) is crucial: an absence of reaction or an inappropriate reaction leads to a loss of the speaker's engagement and therefore a failure of the interaction (Gratch et al., 2006; Bevacqua *et al.*, 2008). These attitudes of the interlocutor while listening are often conveyed by ***feedbacks***. As highlighted in (Allwood, 1992), *"The raison d'être of linguistic feedback mechanisms is the need to elicit and give information about the basic communicative functions, i.e. continued contact, perception, understanding and emotional/attitudinal reaction, in a sufficiently unobtrusive way to allow communication to serve as an instrument for pursuing various human activities"*. Feedbacks are very frequent, generally short and, multimodal (verbal and/or non-verbal). They ensure cohesion between interlocutors: they are a sign that the conversation is followed, understood, accepted and are therefore essential for natural interaction (Schegloff 1982; Alwood *et al.* 1992).

This question is crucial in the field of human-machine interaction based on Embodied Conversational Agent (ECA) and constitute a condition in our ability to develop environments in which the human participant perceives the behaviour of the artificial agent as believable and engaging. This question is also of deep importance when trying to determine the differences of perception of the other participant behaviours between human-human and human-machine situations. At the brain level, several works have been done to analyze the cerebral activities of users interacting or observing artificial agents (ECA or robots). For instance, in (Urgen *et al.*, 2018), users' EEG activities are analysed and compared depending on the level of realism of the robot, revealing a correlation between event-related brain potential and incongruent appearance and motions. In (Rauchbauer et al., 2019), the cerebral activities of users have been recorded in fMRI and compared depending on whether the participant interacts with a robot or a human.

The problem is that only few resources exist to compare human-human and human-machine interaction in general and the behavioral and neurophysiological users' activities in particular. One of the reasons explaining this lack of resources is the difficulty to create *comparable* (or even better *parallel*) corpora with human-human and human-machine interaction, with exactly the same context and the same dynamics in the interaction. In particular, the behaviour of the artificial agent should be comparable to a human, taking into consideration the virtual agent's expressive capacities (e.g. fluidity of the gesture, caricatured appearance).

In this paper, we first describe a method for creating a parallel corpus of human-human and human-machine conversation. In a second part, we describe how this method has been used for creating the Brain-IHM dataset. This resource is made of different conversations between a doctor and a patient. In these conversations, the doctor delivers bad news to the patient who listen and produces regularly feedbacks. The patient is played by a human actor and then re-played by an artificial agent. This corpus also comes with an original information: the brain activity of a participant watching the conversation.

The Brain-IHM dataset constitutes a unique resource for studying two types of questions:

- Feedback perception, including at the brain level
- Human-human vs. human machine interaction

The Brain-IHM dataset presented in this article aims more specifically at studying the feedback in general, the related cerebral activities during their perception, and to compare feedbacks produced by a human or a virtual agent. We propose in the last section of this paper a study done starting from the Brain-IHM dataset, trying to understand whether feedbacks are processed automatically. Today, most of the existing methods to assess artificial agent are based on subjective evaluations through questionnaires filled by the users after their interaction with the virtual agent. This work makes it possible to develop an objective measure of the virtual agent's believability, based on electroencephalography (EEG). The user's EEG activity related to the perception of virtual agent's feedback could constitute objective index of the perceived believability of the agent's behaviour.

2. Feedbacks

2.1 Types of feedback

In this project, we focus on two types of prototypical feedbacks (Schegloff, 1982): ***continuer*** (nods, yes, mhmh, ok, etc.) and ***assessment feedback*** (agreement, surprise, emotion, etc.). Continuer feedbacks generally appear within 200 ms after the speaker's statement and are performed more automatically without translating a semantic evaluation of the speech, unlike assessment feedbacks which generally appear later (beyond 200 ms after the speaker's statement). We provide for each of them a precise identification of their multimodal parameters (auditory and visual).

During a natural conversation, participants are expected to adopt typical responses. As for listeners, appropriate feedbacks are required for a successful interactional achievement. The appropriateness of feedback responses depends on different criteria such as their localization or their semantic value. For example, Bavelas *et al.* (2000) have shown that *continuers*, which help to show the construction of shared knowledge, are appropriate responses when they occur in the beginning of storytelling while *assessments*, which display an evaluative function concerning the events described, are rather provided in the end. The appropriateness of a feedback can also depend on its semantic value that can be identified with scalar attributes related to certainty/uncertainty, understanding/non-understanding for example (Prevot et al., 2016). In this paper, we consider the appropriateness of feedback according to this criterion of semantic value.

Note that we deliberately focus, in this first study, on these two types of feedbacks (continuers and assessments) since they have the advantages to be characterized (among other things) by two factors (semantic value and localization), that we can easily manipulate to create congruent and incongruent feedbacks.

Among the different feedbacks produced by the participants in the collected corpus, 3 types have been manipulated. Let's underline that the corpus has been collected in French, but the methodology as well as the type of modelling applied here can be applied to other languages.

Target feedbacks with corresponding functions and semantic axis:

- *tout à fait ("sure")* corresponding to a feedback of *confirmation* on the semantic axis expected / unexpected
- *oh non ("oh no")* reflecting disappointment on the axis expected/unexpected
- *ah bon ("really", "oh")* reflecting surprise on the semantic axis uncertainty/ certainty

Congruent feedback deal with items expressing the appropriate semantic value projected by the previous utterance while *incongruent feedback* deal with those expressing an inappropriate one.

Examples:

Doctor: Je suis votre médecin anesthésiste c'est moi qui vous ai endormi, vous vous en souvenez ? (*I'm your anesthesiologist, I put you to sleep, do you remember?)*

Patient:

- Congruent feedback: *Tout à fait*
- Incongruent feedback: *oh non*

Doctor: C'est un médicament qui permet de relâcher les muscles (It's a drug that relaxes muscles)

Patient:

- Congruent feedback: *Ah bon*
- Incongruent feedback: *Tout à fait*

Doctor: Ça a dû être un moment très pénible pour vous. (It must have been a very difficult time for you.)

Patient:

- Congruent feedback: *tout à fait*
- Incongruent feedback: *ah bon*

Doctor: Vous éprouvez toujours des difficultés à respirer ?(Still having trouble breathing?)

Patient:

- Congruent feedback: *tout à fait*
- Incongruent feedback: *ah bon*

Doctor: Avez-vous des questions ? (Do you have any questions?)

Patient:

- Congruent feedback: *tout à fait*
- Incongruent feedback: *ah bon*

The three studied feedback types are ***multimodal***, *i.e.* they are expressed through verbal and non-verbal signals. For instance, the feedback "*Tout à fait*" (sure) is associated with a head nod to strengthen the semantic value of confirmation in the face-to-face conversation. The feedback "*ah bon*" (really?) is produced with raised eyebrows to underline the semantic function of surprise. The feedback "*oh non*" is produced with frown eyebrows in order to reinforce the disappointment. Note that the incongruency considered in this study is the *semantic incongruency*, i.e. the use of a feedback with a specific function (for instance surprise) in a situation in which a feedback with another function is expected (for instance confirmation or disconfirmation). We do not consider incongruence in terms of contradiction between the verbal and non-verbal signal in the expression of a feedback (e.g. feedback with a verbal response "*Tout à fait*" (sure) and with a facial expression of surprise).

2.2 Virtual agent's feedbacks

The feedback described above have been reproduced on the Embodied Conversational Agent *Greta* of the platform VIB (Pelachaud, 2009). In this platform, the communicative signals, such as feedback, are described through an XML-based languages called FML (Function Markup Language) and BML (Behaviour Markup Language) of the SAIBA international framework (Vilhjálmsson et al., 2007). In order to simulate the identified feedbacks, we have enriched the library of the ECA with a set of files describing exactly the multimodal behaviour corresponding to the feedback.

Figure 1: *Two different feedbacks produced by the ECA*

To replicate the feedbacks on the ECA, linguists, experts on this phenomenon, have observed the feedbacks expressed by humans during the recordings of films (section 3) and have manipulated the prosody, gestures, head movements, and facial expressions of the ECA to obtain similar expressions. Figure 1 illustrates two examples of feedback. Finally, we have created 6 ECA's feedback corresponding to the needs. In addition to the 3 types of feedbacks studied, we have created 3 common feedbacks - "*D'accord* (I agree)*"*, "*Ok"*, "*Oui* (yes)" - to ensure a variability in the patient's behaviour and naturalness in the conversation. Note that the brain activity of the observer is not analyzed for these three feedbacks.

3. Creation of a *parallel* corpus (human-human and human-machine)

The Brain-IHM dataset has been designed to explore the question of feedback production in a controlled context by studying their perception by a third-party participant (hereafter the observer). Our first hypothesis is that the observer's perception of a feedback is comparable to that of the conversation's participants. In other words, when somebody hears (and sees) a conversation without attending to it, she/he will process the different elements of the dialogue in the same way as for the interlocutors. If this hypothesis is true, it broadens the possibilities of investigation. In particular, it becomes possible to study the brain's reactions of feedback perception.

Studying the brain signal of a conversation interlocutor in a natural environment is a very complex task, the signal being extremely noisy due to muscular activity, moves, etc. The idea of this Brain-IHM dataset has been then to record a dialogue between two actors reproducing a specific conversation, in which one of the interlocutors is the main speaker, the other producing regularly feedback.

3.1 Dialog scenarios and films recording

The theme chosen in this study is a medical context in which a doctor has to break bad news to a patient. Such a context is particularly suitable to the study of feedbacks since the patient is mainly in a listener position. We have first elaborated 3 scenarios, in collaboration with doctor partners of the project: intestinal perforation following endoscopy, regurgitation of gastric fluid during an operation, respiratory arrest during an operation. Each scenario led to the creation of a prototypical dialogue in which the doctor's turn taking are fully specified and followed by feedbacks produced by the listener. The dialogs have been validated by real doctors, used to be faced with such situations.

Figure 2: *Set-up for the recording of the doctor-actor and patient-actor dialog. A green background has been used to enable us to integrate the scene in a virtual environment.*

The following step was to train two actors, playing the role of each dialogue interlocutors (doctor and patient). The doctor had to follow strictly the general content of each turn, but with the possibility to adapt the way the content has been expressed. The patient had to produce exactly the feedback stipulated in the scenario.

In each dialog, the patient produced several feedbacks (at least one at each turn end), among which 18 are studied. In order to compare the perception of the different types of feedbacks, we asked the actors to play the same scenario (and then the same dialogue) twice: one in which all feedbacks are congruent, the second with 50% of

incongruent feedbacks (with semantic incongruency as described Section 2.1)

Example of incongruent feedback:

Doctor: *Do you have any questions?*
Patient: *Really?*

Note that in the incongruent condition, some feedbacks are produced canonically (i.e. with incongruency) in order to avoid an unnatural dialog. Moreover, in the perspective of varying the patient's production, we also integrated a lexicalized turn for the patient, in response to a question of the doctor. Two doctor-actors and two patient-actors played these scenarios, for about 3 minutes each.

Each scenario led to several films, varying the doctor/patient pair. Figure 2 illustrates the set-up for the recording. In addition, for each scenario, we established specific dialogs in which patients produced incongruent feedback (e. g. a surprise feedback in place of a confirmation feedback).

From this first collection of films, we selected the best ones. They were selected by external experts in human-human interaction, evaluating the most realistic interactions. Finally, two films have been chosen for each scenario (6 films in total): one in which the patient produces congruent feedbacks, a second in which he/she produces congruent and incongruent feedbacks.

3.2 Human-human and human-virtual agents videos

Each film has been edited, allowing the patient to be seen as illustrated in Figure 3.b. The goal is to make it possible for the observer to perceive as clearly as possible the feedbacks produced by the actors (virtual or real).

In order to create the parallel corpus with the virtual actor, we have annotated the video of human-human interaction described in the previous section using ELAN to identify precisely the timing (begin, end, duration) and the type of the feedbacks expressed by the actor-patient (Figure 3.a).

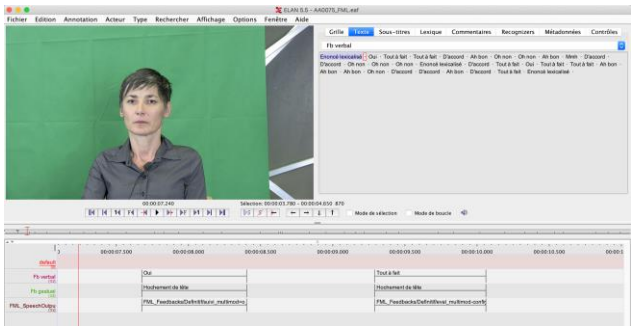

Figure 3.a: *Screenshot of the annotation of the patient's feedbacks in ELAN.*

The annotation consists in specifying the FML file in the library of multimodal feedbacks produced by the virtual agent (see section 2.2) corresponding to the feedbacks expressed by the patient in the film. Moreover, the annotation specifies the timing and the duration of the feedbacks (Figure 3.a). A tool has been developed to then generate automatically from the annotation file the animation of the virtual agent expressing the feedbacks at the exact timing. The tool takes as input the annotated files (Figure 3.a) and generates the film corresponding to the animation of the virtual patient with the appropriate feedbacks described in FML (Section 2.2) at the indicated moment and with the corresponding duration specified in the ELAN file. By this way, the feedback behaviour of the virtual agent is the same as the feedback behaviour of the actor-patient.

Moreover, the conversations were integrated into a virtual environment designed in Unity and validated by doctors to represent a recovery room where the breaking of bad news generally takes place in hospitals. In total, we obtain a corpus of 12 films, 4 per scenario. For each scenario, we have:

- two *human-human* films, one in which the actor-patient produces congruent feedbacks, another with congruent and incongruent feedbacks
- two *human/virtual agent* films, one in which the artificial patient produces congruent feedbacks (the same feedbacks produced by the actor-patient and at the same time) plus a second with congruent and incongruent feedbacks (similar to the actor-patient)

Figure 3.b: *Screenshots of the video of the Brain-IHM dataset (left, simulation of a actor-doctor /actor-patient interaction, right, re-synthesis of the same multimodal behaviour of the actor-patient on the virtual agent)*

It is interesting to note the triple originality of the Brain-IHM dataset: 1/ it is the first corpus with controlled production of feedback 2/ it contains parallel human-human and human-virtual agent conversations 3/ it contains both verbal, non-verbal and electro-physiological information

3.3 Verbal and non-verbal annotations

The fact that the production is almost fully controlled facilitates annotations. At the verbal level, in particular the transcription of the dialogue is aligned onto the signal, using SPASS (Bigi, 2015). This is done at the phoneme level, offering a precise segmentation in phonemes, syllables and tokens. The POS tagging is also provided, as well as a mid-level syntactic segmentation within the turns. The quality of the signal recording also offers the possibility to apply automatic prosodic analyzers.

The non-verbal level is only concerned with face attitudes: both actors were asked to be static, the listener only produced different facial expressions (nods, eyebrow raises, smiles, etc.). This information is directly annotated without using any tool, thanks to the specifications of the scenarios.

Beside these annotations, the main originality of the corpus is that it comes with two other types of information: one concerning the subjective evaluation of the virtual patient, the second with the recording of the brain signal of the participant watching the films. This two information are described in the next section.

4. Experimental Set-up and Design

The brain signal has been acquired thanks to a specific design. Each participant has to watch movies of the dialogue in 4 different conditions: human/human, human/virtual agent, with only canonical (congruent) or canonical and incongruent feedbacks. These conditions make it possible to do different comparisons: human vs. machine communication and congruent vs. incongruent productions. More precisely, the idea is first to confirm the hypothesis that no difference can be observed in the perception of feedbacks by human or virtual agent. This done, it becomes possible to explore the specific effects observed at the brain level.

Thirty-six participants (27 females and 9 males) have been recruited for the experiment; all participants signed a consent form at the start of the experiment and were remunerated for their participation. Subject recruited had between 18 and 40 years old.

Electroencephalography (EEG) data were recorded in a using the Biosemi Active2 64-electrode system, which amplifies the signal at the level of the participant's head via an AD-box, a DC amplifier. In this EEG system, the ground is replaced by two electrodes, the CMS (Common Mode Sense) and the DRL (Driven Right Leg), which ensure that the participant's average potential is as near as possible to the ADC voltage of the Biosemi AD-box. Signals were recorded at a rate of 2048Hz and the left mastoid was defined as the reference online. This sample rate was especially high because the same system was used to trig auditory and visual feedbacks. External electrodes were positioned on the face to record both horizontal and vertical eye movements to facilitate their offline processing

During EEG recording, the participants passively watched the 4 different movies while comfortably seated in front of a computer screen in a Faraday cage. Before beginning the presentation of the movies, a very short video with sound was presented to regulate the sound level of the videos. A free-field mode of sound presentation was used via loud speakers positioned to the left and right of the screen. Participants were instructed to passively watch the 4 videos; the task did not involve any response on their part. Each movie lasted approximately 5 minutes and, between each, there was a short break during which the participant responded to an online questionnaire based on the preceding movie.

This questionnaire was composed of 7 questions or affirmations on the participants perception of the patient (virtual or real depending on the video just watched): the *believability* of the patient ("*In your opinion, how believable is the patient compared to real patients?"),* the participant's *appreciation* of the patient (*"You like the patient"*), the *reactivity* of the patient (*"Did you find the patient responsive to what the doctor said?"*), the *naturalness* of the conversation ("*Have you found the natural conversation between the patient and the doctor?"*), the perceived *understanding* of the patient (*"Did you get the impression that the patient understood what the doctor said?"*), the perceived *performance of the doctor* in the task of delivering the bad news (*"Do you think the doctor well explained the problem and what was going to happen to the patient?"* and *"Do you think the doctor had difficulty telling the patient the bad news?"*). The participants respond to the questions or indicate her/his level of agreement to the affirmation through a 5-point Likert scale.

The analysis of the responses to the questions on the perceived believability of the agent and of the naturalness of the conversation enables us to validate the two different conditions, either with only congruent feedbacks (congruent condition) or with congruent and incongruent feedbacks (incongruent condition). In both conditions, for the human-human and the human-virtual agent conversation, the participants have significantly perceived the conversation with incongruent feedbacks as less natural with an agent less believable than the conversation with only congruent feedbacks. Figure 4 illustrates the average responses of the participants to the question on the *believability* of the patient ("*In your opinion, how believable is the patient compared to real patients?"*) on a Likert scale of 5 points. The results of the *T-test* show that the actor-patient expressing only congruent feedbacks (HHC) has been perceived significantly more believable than when she expressed incongruent feedbacks (HHI) ($p<0.05$). In the same way, virtual agent expressing only congruent feedbacks (HAC condition) has been perceived significantly more believable than when she expresses incongruent feedbacks (HAI condition) ($p<0.001$).

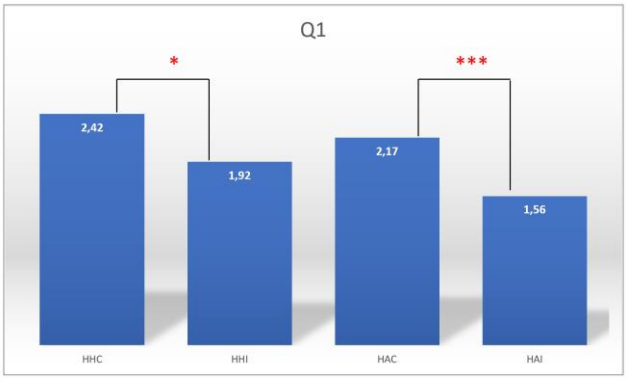

Figure 4: *Mean of the responses (between 0 to 5) of the question "In your opinion, how believable is the patient compared to real patients?" for each condition.*

Figure 5 illustrates the average response of the participants to the question on the *naturalness* of the conversation ("*Have you found the natural conversation between the patient and the doctor?"*) on a Likert scale of 5 points. The results of the *T-test* show a significant difference on the perception of the naturalness of the conversation depending on the feedbacks expressed: the conversations were perceived significantly more natural when the actors

(virtual or real) expressed congruent feedback than when they expressed incongruent one.

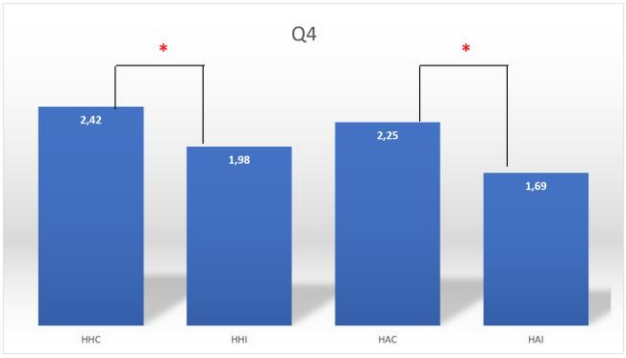

Figure 5: *Mean of the responses (between 0 to 5) of the question "In your opinion, how believable is the patient compared to real patients?" for each condition.*

Note that the results of the questionnaire are far from surprising but enable us to validate that the incongruent feedbacks have been perceived as such.

The presentation of the 4 types of videos (human-human congruent, human-human incongruent, human-virtual agent congruent and human-virtual agent incongruent) was counterbalanced across participants to avoid the possibility of confounds due to the order of presentation of the different scenarios. Crucially, to synchronize the onset of each feedback, auditory and visual, and the EEG signal, we integrated 2 small black squares into each video, one for auditory and one for visual feedbacks, which changed from black to white at the onset of each feedback. The change in luminosity of each colored square was detected by photodiodes positioned against the screen and the photodiode signals was recorded as two auxiliary EEG channels via the ERGO1 and ERGO2 inputs of the Biosemi AD-box (Figure 6). In addition, to ensure that the different feedback types could be distinguished offline, the duration of the colour change varied for visual and auditory feedbacks and for each of the 4 video thus types yielding a different photodiode signal for each video type (Human-Human, Human-Virtual Agent), feedback modality (auditory and visual) and feedback type (congruent and incongruent). Thus, each feedback corresponds to a step function whose t_0 is the onset of the feedback and the type of feedback can be determined by calculating the duration of the step function.

5. Application: the brain basis of feedback perception

Feedbacks can be considered at the lower level as discourse markers (more instructional than referential) or on the opposite as full linguistic objects. In the first case, feedbacks can be considered as automatic reactions to the discourse of the main speaker. Some studies have shown that such feedbacks can be predicted only from the time course (Penteado et al., 2019): the occurrence of a feedback seems to be mainly dependent on the realization of the previous one (see also (Ward and Tsukahara, 2000). However, some other works have also shown that feedbacks are dependent from different linguistic features such as prosody (in particular breaks) but also from morpho-syntactic information (some adverbs in the main speaker's discourse can increase the probability of a backchannel realization by the hearer (Bertrand et al., 2003; Bertrand et al., 2007). Such a contextual behaviour indicates that a certain type of linguistic processing has to be done by the hearer. Finally, some types of feedbacks (in particular assessment feedback) require a certain understanding of the main speaker's production, or in other words a form of semantic processing, which is the higher

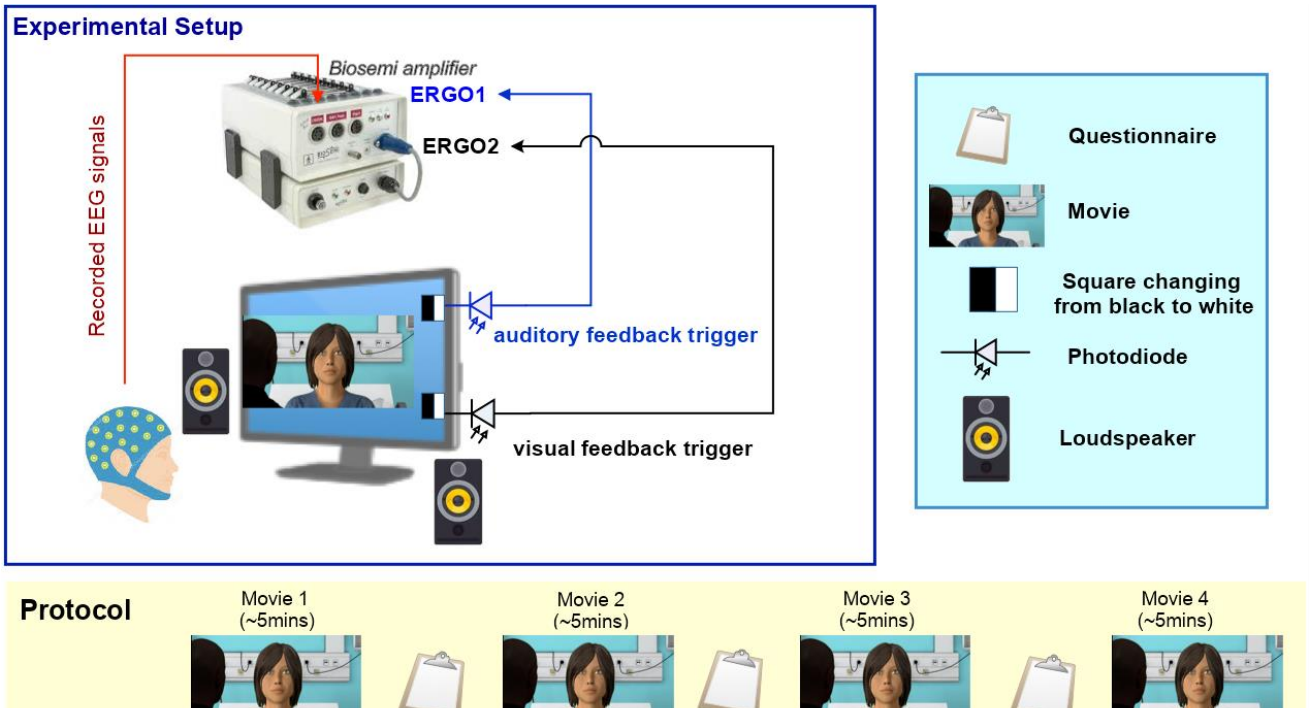

Figure 6: A graphical overview of the experimental setup and protocol.

linguistic level. Feedbacks can be then considered very differently, depending on their relationship with the previous context. In some cases, they seem to be at the lowest level, produced almost automatically where in some other cases, they seem to require deeper linguistic processing. In sum, the main question of this study is to uncover whether feedbacks are processed automatically or not.

Our first hypothesis in this study is that feedback perception relies both on automatic and deep mechanisms. We propose for doing that to analyze the brain signal in response to the production of feedbacks. Our hypothesis is that verbal and non-verbal feedbacks elicit specific Event-Related Potentials (ERP) depending on the way they are processed. Previous works have shown that a larger early posterior negativity (EPN) can be associated with the perception of facial expression, modulating the amplitude of the N170 and P100 components (Eimer, 2011; Herrmann et al., 2004; Sprengelmeyer et al., 2006; Krombholz et al., 2007; Righart et al., 2006). In the same vein, some works have uncovered similar effects in the perception of emotional words (Wang et al., 2014) and more precisely the sensibility of the N170 component to emotions and their intensity (Sprengelmeyer et al., 2006; Krombholz et al., 2007; Righart et al., 2006). These effects can be associated to a certain kind of automatic processing. The hypothesis in this experiment is that feedback perception is in a certain way made up of mechanisms including facial and emotional recognition. If so, it should be the case that feedback perception would elicit comparable early components such as N170 and P100. The second aspect of this experiment is to look for traces of deep mechanisms, involving in particular a certain level of semantic processing. The way we propose to approach this question is to study the effect of incongruity in feedback perception, The idea is that incongruent feedbacks, as described in the previous section, elicit similar effect as semantic incongruity. Such a phenomenon is associated with a strong negative wave elicited 400ms after stimulus onset, known as the N400 effect (Besson et al., 1992; Kutas et al., 1980; 2011), that can also be observed in the case of emotional incongruity (Leuthold et al., 2011; Bartholow et al., 2001). We also expect that the amplitude of the N400 could be correlated with the level of incongruity of the feedback (as observed with semantic incongruity)

Our second hypothesis is that there is no difference in the perception of feedbacks produced by a human or by a virtual agent, the signal signature of the brain activity should be the same. This hypothesis is important for different reasons. First, it is necessary to study more precisely the type of mechanisms that occur at the brain level when communicating with a human or with an avatar. For example, we know that facial expressions are processed in the same way, being them produced by an avatar or a human (Dyck et al., 2008). The Brain-IHM dataset will open new possibilities to compare human and machine communication. Moreover, at the methodological level, such type of resource and the way they are built also offer new possibilities to explore the role of some specific features in feedback production (e.g. the delay, prosody, verbal/non-verbal synchrony, etc.) that cannot be controlled easily or even produced by a human.

6. Conclusion

This paper presents a new and original dataset of controlled interactions, focusing on the study of feedbacks and more particularly two types of multimodal feedbacks: the *continuer* and the *assessment*. A *parallel* corpus of video of both human-human and human-virtual agent conversations have been created. A specific methodology has been developed to create such a corpus with a virtual agent's behaviour replicating the actor's behaviour but considering the expressive capacity of the virtual agent. In this article, we have particularly focus on the feedbacks in a French conversation between a doctor and a patient in the context of breaking bad news but the methodology could be replicated to analyse others multimodal behaviours in other language. This kind of corpus enables one to compare the perception of human-human versus human-machine interaction. Moreover, we have proposed a specific experimental setup and design to record the brain activities of the observers of the interaction.

In conclusion, the Brain-IHM dataset constitutes the first resource of this type, providing a controlled production of feedbacks, a parallel human-human/human-machine corpus, completely annotated at the verbal and non-verbal level and a set of EEG data making it possible to study the perceptive level.

7. Acknowledgements

This work has been funded by the CNRS (PEPS Brain-IHM) and supported by grants ANR-16-CONV-0002 (ILCB), ANR-11-LABX-0036 (BLRI) and ANR-11-IDEX-0001-02 (A*MIDEX).4

8. Bibliographical References

Allwood, J., Nivre, J., & Ahlsén, E. (1992). On the semantics and pragmatics of linguistic feedback. Journal of semantics, 9(1), 1-26.

Bailenson J.N., Swinth, K., Hoyt, C., Persky, S., Dimov, A., Blascovich, J. (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)

Bartholow B. D., Fabiani M., Gratton G., & Bettencourt B. A. (2001). A psychophysiological examination of cognitive processing of and affective responses to social expectancy violations. Psychological science, 12(3), 197-204.

Bertrand, Roxane; Espesser, Robert (2003). Prosodic cues of back-channel signals in French conversational speech. International Conference on Prosody and Pragmatics (NWCL)

Bertrand, Roxane; Ferré, Gaëlle; Blache, Philippe; Espesser, Robert; Rauzy, Stéphane (2007). Backchannels revisited from a multimodal perspective. Proceedings of Auditory-visual Speech Processing

Besson M., Kutas M., Petten CV. (1992) An Event-

Related Potential (ERP) Analysis of Semantic Congruity and Repetition Effects in Sentences, in Journal of Cognitive Neuroscience, 4(2):132-49.

Bevacqua E., M. Mancini, R. Niewiadomski, C. Pelachaud (2007), An expressive ECA showing complex emotions. In proceedings of AISB'07 ``Language, Speech and Gesture for Expressive Characters", 208-216

Bevacqua E., Mancini M., & Pelachaud, C. (2008) A listening agent exhibiting variable behaviour. In International Workshop on Intelligent Virtual Agents (pp. 262-269). SpringerBrigitte Bigi (2015). SPPAS - Multi-lingual Approaches to the Automatic Annotation of Speech. In "the Phonetician" - International Society of Phonetic Sciences,

Chollet M., Ochs M. and Pelachaud, C. (2014) Mining a multimodal corpus for non-verbal signals sequences conveying attitudes, Language Resources and Evaluation Conference (LREC)

Gardner, R. (2001). When listeners talk. Benjamins, Amsterdam.

Dyck M., Winbeck M., Leiberg S., Chen Y., Gur R. C., & Mathiak K. (2008). Recognition profile of emotions in natural and virtual faces. PloS one, 3(11).

Eimer M. (2011). The Face-Sensitive N170 Component of the Event-Related Brain Potential. In G.Rhodes, A. Calder, M. Johnson & J. V. Haxby (Eds.), Oxford Handbook of Face Perception, Oxford University Press.

Fox Tree J. E. (1999). Listening in on monologues and dialogues. Discourse Processes, 27, 35-53.

Gratch J., Okhmatovskaia, A., Lamothe, F., Marsella, S., Morales, M., van der Werf, R. J., & Morency, L. P. (2006). Virtual rapport. In International Workshop on Intelligent Virtual Agents (pp. 14-27). Springer.

Herrmann M. J., Ehlis A. C., Ellgring H., & Fallgatter A. J. (2004). Early stages (P100) of face perception in humans as measured with event-related potentials (ERPs). Journal of Neural Transmission, 112(8)

Kutas , M. , & Hillyard , S. A . (1980) Reading senseless sentences: brain potentials reflect semantic incongruity. Science, 207, 203 – 204 .

Kutas , M. , & Federmeier , K. D . (2011) Thirty years and counting: fi nding meaning in the N400 component of the event-related brain potential (ERP), Annual Review of Psychology , 62 , 621 – 647 .

Krombholz A., Schaefer F., & Boucsein W. (2007). Modification of N170 by different emotional expression of schematic faces. Biological Psychology, 76(3), 156-162.

Leuthold H., Filik R., Murphy K., & Mackenzie I. G. (2011). The on-line processing of socio-emotional information in prototypical scenarios: inferences from brain potentials. Social Cognitive and Affective Neuroscience, 7(4), 457-466

Morency, L. P., de Kok, I., & Gratch, J. (2010). A probabilistic multimodal approach for predicting listener backchannels. Autonomous Agents and Multi-Agent Systems, 20(1), 70-84.

Ochs M., de Montcheuil G., Pergandi J-M., Saubesty J., Pelachaud C., Mestre, D. and Blache P. (2017) An architecture of virtual patient simulation platform to train doctors to break bad news, Conference on Computer Animation and Social Agents (CASA)

Pelachaud, C. (2009). Studies on gesture expressivity for a virtual agent. Speech Communication, 51(7), 630-639.

Penteado, B. E., Ochs, M., Bertrand, R., & Blache, P. (2019, July). Evaluating Temporal Predictive Features for Virtual Patients Feedbacks. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*. ACM.

Poppe, R., Truong, K. P., Reidsma, D., & Heylen, D. (2010). Backchannel strategies for artificial listeners. In International Conference on Intelligent Virtual Agents (pp. 146-158 Springer

Porhet C., Ochs M., Saubesty J., de Montcheuil G., Bertrand R. (2017) Mining a Multimodal Corpus of Doctor's Training for Virtual Patient's Feedbacks, International Conference on Multimodal Interaction (ICMI).

Prévot L., Gorish J., Bertrand R. (2016) A CUP of CoFee: A large collection of feedback utterances provided with communicative function annotations. Language Resources and Evaluation Conference (LREC)

Rauchbauer, B., Nazarian, B., Bourhis, M., Ochs, M., Prévot, L., & Chaminade, T. (2019). Brain activity during reciprocal social interaction investigated using conversational robots as control condition. Philosophical Transactions of the Royal Society B, 374(1771), 20180033.

Righart R. & de Gelder B. (2006). Context influences early perceptual analysis of faces: An electrophysiological study. Cerebral Cortex,16,1249-1257

Schegloff, E.A., 1982. Discourse as an interactional achievement: some uses of ‘‘uhhuh" and other things that come between sentences. In: Tannen (Eds.), Analyzing Discourse: Text and Talk, Georgetown University Press, pp. 71--93.

Sprengelmeyer R. & Jentzsch I. (2006). Event related potentials and the perception of intensity in facial expressions. Neuropsychologia, 44(14), 2899-2906

Urgen, B. A., Kutas, M., & Saygin, A. P. (2018). Uncanny valley as a window into predictive processing in the social brain. Neuropsychologia, 114, 181-185.

Vilhjálmsson, H., Cantelmo, N., Cassell, J., Chafai, N. E., Kipp, M., Kopp, S., ... & Ruttkay, Z. (2007, September). The behavior markup language: Recent developments and challenges. In International Workshop on Intelligent Virtual Agents (pp. 99-111). Springer, Berlin, Heidelberg.

Ward, N. & Tsukahara, W. (2000). Prosodic features which cue back-channel feedback in english and japanese. Journal of Pragmatics, 32: 1177-1207.

Language Resource References

The Brain-IHM Dataset is available from the ORTOLANG DataWarehouse.