A multimodal corpus of Human-Human and Human-Robot conversations including synchronized behavioral and neurophysiological recordings

Thierry Chaminade

A multimodal corpus of Human-Human and Human-Robot conversations including synchronized behavioral and neurophysiological recordings

Thierry Chaminade
Aix Marseille Université, CNRS, INT, Marseille, France
thierry.chaminade@univ-amu.fr

Abstract

This paper presents a unique corpus intended to be shared with the community at the crossroad of linguistics and human-robot interactions (HRI). It is the result of an interdisciplinary collaborations within Cognitive Sciences combining social neurosciences, computational sciences, robotics and linguistics. It was recorded when brain activity of 25 participants was scanned using functional Magnetic Resonance Imaging (fMRI) while having unconstrained conversations with a fellow human or a conversational robotic head. Behaviors from the participant and the conversant were recorded synchronously (speech, eye-tracking, head and face movements). Manual and automatic analysis of these behaviors provide rich sets of data for the analysis of both behaviors and neurophysiological responses. Examples of results obtained with this corpus are provided.

1 Introduction

"Second-person social neuroscience" (Schilbach et al., 2013), which puts forward the importance of studying real-time social cognition in truly interactive scenarios, inspired a new experimental approach in which a natural conversation between two persons is recorded (Chaminade, 2017). It is now agreed that such interactive approaches are necessary to understand social cognition and its disorders. Efforts in the field of social neuroscience are therefore currently put in developing more ecological paradigms. In terms of methodology, the challenge is to investigate unconstrained behaviors. The classical scientific methodology requires controlling all experimental factors but one (or a few) to investigate its effect of the system. But this method has seen its limits to investigate social cognition. The current project is part of the endeavor to investigate natural, unconstrained behaviors as multimodal corpora.

Recorded behaviors were natural conversations participants were having with a confederate of the experimenter or a robotic head resembling the confederate while lying supine in a fMRI scanner. The robot was controlled by the confederate, unbeknown of the participant who believed the robot to be autonomous. Meanwhile the participant's brain was continuously scanned with functional magnetic resonance imaging (fMRI). This material should allow us to investigate the neural bases of language, of social interaction and the differences between human-human and human-robot interactions (HHI and HRI). But such unconstrained behavior poses one major difficulty: what do you use as explanatory variable? As many aspects of the behaviors as possible are recorded to be build time-series predictors for the analysis.

2 Methods

In order to investigate natural social interactions, it is essential that participants are unaware of the real purpose of the experiment. For this purpose, a cover story was developed there is a common, but loose, goal for the conversation, responding to a complex set of specifications: the conversations are truly bidirectional, and there is a legitimacy for talking with a robot. The experiment is described as a neuromarketing experiment, in which a company wants to know if discussing with a fellow or an artificial intelligence on the images of a forthcoming campaign is enough to guess the message of the campaign (Chaminade, 2017).

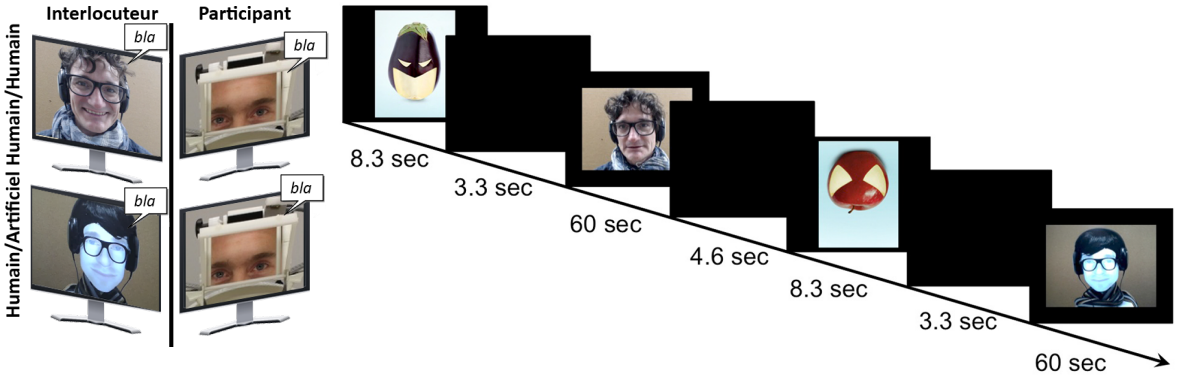

Figure 1: Experimental paradigm.

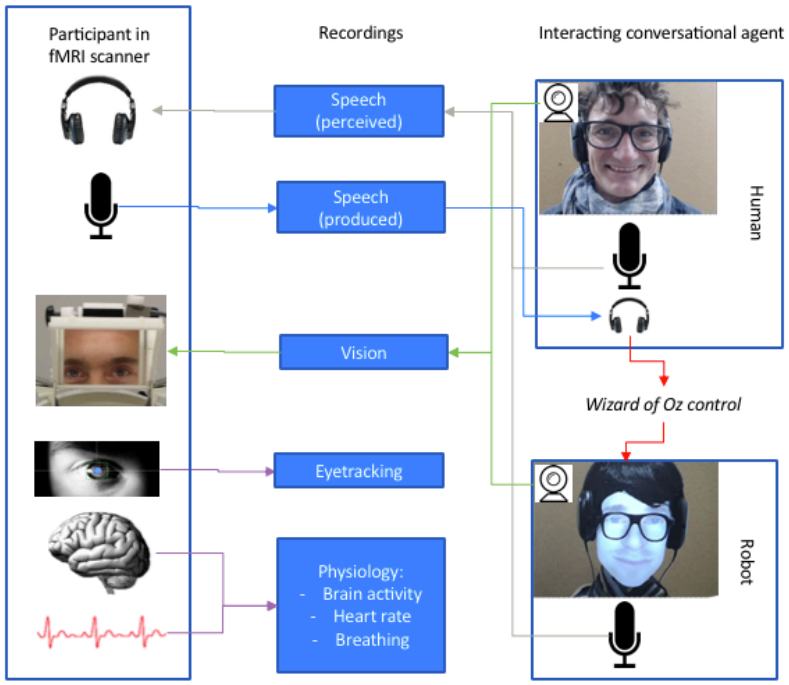

Figure 2: Summary of raw recordings.

Participants (n=25 in the full corpus) were welcomed to the MRI center and introduced to the cover story. It provided a fake rationale for the experiment and its set-up. Participants were explained that the study was sponsored by an advertising company to test the key message of a new campaign. The message was to be discovered through a conversation with another participant and a conversational robot, who also had information on the campaign and was able to speak autonomously (see Figure 1 for experimental design, taken from (Rauchbauer et al., 2019). Participants were informed about the study design, which presented images of anthropomorphized fruits and vegetables of the forthcoming advertisement campaign. The participants were told that they should talk naturally about each image with the other agent (alternating between the human and the robot), who would be outside the fMRI scanner room and connected via live video stream and bi-directional audio. They were presented with a fellow (a gender-matched confederate of the experimenter) and the robot, the retroprojected conversational robotic head with gender-matched face and voice synthetizer[1] (Al Moubayed et al., 2012). Participants were told that the robot had information on the advertisement campaign and could talk autonomously. Unknown to the participants,

[1] Furhat robotics: https://www.furhatrobotics.com

the robot was controlled by the human confederate in a Wizard of Oz configuration (Riek, 2012), and the robot's arguments were pre-written conversations based on a behavioral study (Chaminade, 2017). Participants were shown the conversational robotic head before being brought into the scanner room. The participant underwent four sessions of approximately 8 minutes of scanning comprising six experimental trials as follows: an image is presented, then the 1-minute discussion takes place alternatively with the human and the robot, totaling twelve trials of 1-minute conversations with the human and 12-minutes of 1-minute conversation for the robot for each participant (Figure 1). Noise-cancelling microphone and headphones allowed fluid conversation despite MRI noise. At the end of the study, the participants were debriefed in an open format. Participants could describe their impression of the interaction with both the human and the robot. Also, it was checked that participants still believed in the autonomous conversation of the robot. A total of 12 1-minutes trials per Agent (Human, Robot) and participant (n=25) are included in the corpus.

3 Data processing

3.1 fMRI data

fMRI data processing follows standard procedures and is described in previous work (Rauchbauer Birgit et al., 2019). In brief, for preprocessing, we apply to the acquired echo-planar images capturing the blood-oxygen level dependent (BOLD) signal slice-timing correction, realignment, unwarping to correct for local distortions of the magnetic field, and normalization to the standard Montreal neurological institute space. A number of nuisance covariates are calculated to control for movement artefacts, for potential artefacts from blood pulse and respiration, highly relevant in a paradigm involving speech, as well as global signal from grey matter, white matter and cerebrospinal fluid controlling for global fluctuations of the signal unrelated to the task [TAPAS toolbox (Kasper et al., 2017)].

fMRI data analysis first relies on the general linear model implemented in the toolbox Statistical Parametric Mapping (Penny et al., 2011). Each trial is modelled as a single regressor, and the images presented before each discussion are modelled as a single regressor. Single participants analyses are then imported into one model in the Conn toolbox (Whitfield-Gabrieli & Nieto-Castanon, 2012) that automates the extraction of BOLD time series in regions of interest. A continuous time-series of 385 points covering the ~8 minutes are extracted (repetition time: 1.205 seconds) for each session and each participant. We use a brain parcellation formed on the basis of functional and connectivity data, so that the regions of interest represent functionally homogeneous ensembles of voxels (Fan et al., 2016).

3.2 Behavioral data

Bidirectional conversation was performed via live unidirectional videoconference (the participant sees the conversant projected on screen) and bidirectional audio connection between participants inside the scanner and the human or a robot conversant outside the scanner room. We recorded the eye movements of the scanned participant (see Figure 2). All recordings were time-controlled by the clock of the MRI scanner.

Noise reduction was used to remove MR scanner's loud noise recorded on the participant's audio. Denoised participant and conversant data was segmented into Inter-Pausal Units (IPUs), defined as blocks of speech in between silences of minimum duration 200 ms. Files were uploaded into SPPAS[2] for manual transcription. Automatic Text normalization was performed using the SPPAS software tool (Bigi, 2015). Normalized transcribed files of the participants' and the interlocutors' (human and robot) conversations are available on the data repository Ortolang[3] (see example of conversations in Appendix A). From these normalized data, a number of linguistic variables can be calculated using tools from the Natural Language Processing.

Eye-tracking data was collected from the participant with an Eyelink 1000 Plus Long Range Mount with a temporal resolution of 1000 Hz[4]. The information recorded indicates fixations, saccades and blinks as well as the coordinates of gaze direction on screen in a standard image coordinate system (x, y). Combining eye- and face-tracking data indicates where the participant was looking throughout the conversations.
Videos recorded from the human confederate and the robot were recorded at 30Hz used for face-

[2] version 1.9.9, www.sppas.org/ (Bigi, 2015).
[3] https://hdl.handle.net/11403/convers.
[4] SR Research Ltd., Mississauga, Ontario, Canada, https://www.sr-research.com/products/eyelink-1000-plus/

tracking analyses. We used OpenFace (Baltrusaitis et al., 2018) to analyze each video separately. The output format of OpenFace is a .csv (comma separated value) containing 1800 observations (number of images per video). The .csv output file contains the 68 facial landmarks, 17 facial Action Units (AU), 3 features of gaze movements and 6 features of head pose rotations and translations. Detection of gaze shows where the confederate is looking.

3.3 Physiological data

Physiological data was recorded with the SIEMENS scanner's own system. A photoplethysmography unit was positioned on the left-hand index fingertip to record pulse oximetry and a breathing belt was positioned at the chest level. Data was acquired continuously at the frequency of 200Hz. Preprocessing using a specific toolbox (Kasper et al., 2017) output a csv matrix with three columns corresponding, to the time stamps of the observation, the cardiac signal and the respiration signal.

4 Examples of analyses

4.1 HRI vs HHI

The first simple contrast between HHI and HRI was published in the main publication for this corpus to date. Results confirmed that interacting with a human activates nodes of the social brain, in particular the temporoparietal junction, while interacting with a robot is associated with activation in the dorsolateral prefrontal cortex. Extended discussion can be found in Rauchbauer et al. (2019).

4.2 Speaking *vs* Listening

A simple use of the transcriptions allows investigating periods when the participant is talking *vs* listening, and therefore to map the speaking brain and the talking brain in action. With this contrast we identify the brain correlates associated with speech production, in the motor cortex bilaterally but also Broca's area and the cerebellum (Figure 3, top), while listening activates the temporal cortex bilaterally (bottom). While this data is unpublished, it shows that the same corpus can be used to ask different questions (related to social cognition or to language) in participants undergoing a natural interaction, through the use of the recorded behaviors and their processing.

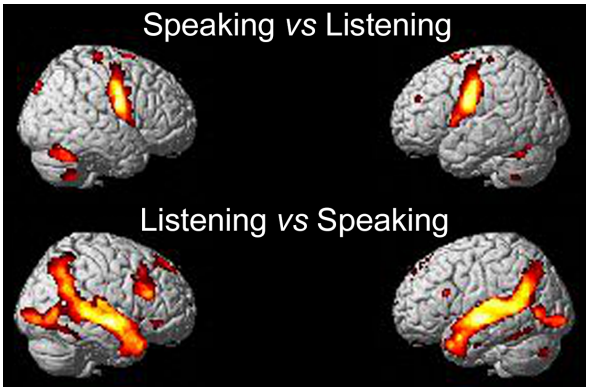

Figure 3: Analysis of the conversing brain.

4.3 Computational neuroscience analysis

We are currently developing a method to associate behavioral and physiological time-series (Hmamouche et al., 2020). In a nutshell, behavioral time-series are used to predict activity in brain regions of interest. The latest results, submitted to ICMI 2020, demonstrated that despite a difference in levels of activity, the same social features are used to predict the activity in areas of the social brain, while simpler features (such as speech activity) are sufficient to predict act activity in speech-related areas.

5 Conclusions

We acquired a corpus of natural conversational interactions with a human or a robot, designed to be shared. Combining brain neuroimaging and physiology, linguistic transcriptions and visual behaviors, it can be used to address multiple questions pertaining to social cognition, linguistics and human-robot interactions. We invite researchers with interest in this corpus to contact us in order to collaborate to make most of this unique dataset.

Acknowledgments

I thank principal investigators involved in this project, Laurent Prévot and Magalie Ochs, as well as all interns and post-doctoral scientists involved in recording and analyzing this dataset. Research supported by grants AAP-ID-17- 46-170301-11.1 by the Excellence Initiative of Aix-Marseille University (A*MIDEX) and the Institute for Language, Communication and the Brain (ILCB, ANR-16- CONV-0002).

References

Al Moubayed, S., Beskow, J., Skantze, G., & Granström, B. (2012). Furhat : A Back-Projected Human-Like Robot Head for Multiparty Human-Machine Interaction. In A. Esposito (Éd.), *Cognitive Behavioural Systems* (p. 114–130). Springer Berlin Heidelberg.

Baltrusaitis, T., Zadeh, A., Lim, Y. C., & Morency, L. (2018). OpenFace 2.0 : Facial Behavior Analysis Toolkit. *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*, 59–66. https://doi.org/10.1109/FG.2018.00019

Bigi, B. (2015). SPPAS - MULTI-LINGUAL APPROACHES TO THE AUTOMATIC ANNOTATION OF SPEECH. *The Phonetician*, *111-112*(ISSN:0741-6164), 54–69.

Chaminade, T. (2017). An experimental approach to study the physiology of natural social interactions. *Interaction Studies*, *18*(2), 254–275.

Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., Yang, Z., Chu, C., Xie, S., Laird, A. R., Fox, P. T., Eickhoff, S. B., Yu, C., & Jiang, T. (2016). The Human Brainnetome Atlas : A New Brain Atlas Based on Connectional Architecture. *Cereb Cortex*, *26*(8), 3508–3526. https://doi.org/10.1093/cercor/bhw157

Hmamouche, Y., Magalie, O., Prevot, L., & Thierry, C. (2020). Exploring the Dependencies between Behavioral and Neuro-physiological Time-series Extracted from Conversations between Humans and Artificial Agents. *9th International Conference on Pattern Recognition Applications and Methods*, 353–360.

Kasper, L., Bollmann, S., Diaconescu, A. O., Hutton, C., Heinzle, J., Iglesias, S., Hauser, T. U., Sebold, M., Manjaly, Z.-M., Pruessmann, K. P., & Stephan, K. E. (2017). The PhysIO Toolbox for Modeling Physiological Noise in fMRI Data. *Journal of Neuroscience Methods*, *276*, 56–72. https://doi.org/10.1016/j.jneumeth.2016.10.019

Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., & Nichols, T. E. (2011). *Statistical Parametric Mapping : The Analysis of Functional Brain Images*. Elsevier.

Rauchbauer Birgit, Nazarian Bruno, Bourhis Morgane, Ochs Magalie, Prévot Laurent, & Chaminade Thierry. (2019). Brain activity during reciprocal social interaction investigated using conversational robots as control condition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1771), 20180033. https://doi.org/10.1098/rstb.2018.0033

Riek, L. D. (2012). Wizard of Oz Studies in HRI: A Systematic Review and New Reporting Guidelines. *J. Hum.-Robot Interact.*, *1*(1), 119–136. https://doi.org/10.5898/JHRI.1.1.Riek

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, *36*(4), 393–414. https://doi.org/10.1017/s0140525x12000660

Whitfield-Gabrieli, S., & Nieto-Castanon, A. (2012). Conn : A Functional Connectivity Toolbox for Correlated and Anticorrelated Brain Networks. *Brain Connectivity*, *2*(3), 125–141. https://doi.org/10.1089/brain.2012.0073

A Appendices

Example of Human-Human conversation

Participant: clearly it's a pear
Participant: uh
Participant: a little dented
Participant: and uh
Participant: suddenly it seems that she is drunk
Participant: and
Conversant: yeah
Conversant: yeah yeah
Participant: and uh actually uh
Participant: at first I thought she was sad but in the end I have the impression that it's a little smirk
Conversant: ah yeah
Participant: yeah I don't see -end good sadness -end it's neutral uh
Participant: it's not uh so sad
Conversant: me she looked like uh, I don't know how to say maybe annoyed
Participant: how
Conversant: maybe disappointed for me

Example of Human-Robot conversation

Participant: then this is a strawberry that is also damaged
Participant: and and who looked
Participant: lost
Participant: and smashed
Participant: for me
Participant: uh
Conversant: like the other two
Participant: uh no
Participant: not too much the others they had more of an expression of pain or er
Participant: and uh
Participant: that's it
Conversant: maybe
Conversant: this strawberry is distorted
Participant: yes
Participant: on the sides
Conversant: the strawberry is also rotten