



HAL
open science

Investigating the dimensions of conversational agents' social competence using objective neurophysiological measurements

Thierry Chaminade, Birgit Rauchbauer, Bruno Nazarian, Morgane Bourhis, Magalie Ochs, Laurent Prévot

► **To cite this version:**

Thierry Chaminade, Birgit Rauchbauer, Bruno Nazarian, Morgane Bourhis, Magalie Ochs, et al.. Investigating the dimensions of conversational agents' social competence using objective neurophysiological measurements. 20 ACM International Conference on Multimodal Interaction, Oct 2018, Boulder, United States. pp.1-7, 10.1145/3281151.3281162 . hal-01973542

HAL Id: hal-01973542

<https://hal.science/hal-01973542v1>

Submitted on 22 Jan 2019

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.

Investigating the Dimensions of Conversational Agents' Social Competence Using Objective Neurophysiological Measurements

Thierry Chaminade

Birgit Rauchbauer

Bruno Nazarian

Morgane Bourhis

Institut de Neurosciences de la
Timone

CNRS – Aix-Marseille Université
Marseille, France

thierry.chaminade@univ-amu.fr

birgit.rauchbauer@univ-amu.fr

bruno.nazarian@univ-amu.fr

morgane.bourhis@univ-amu.fr

Magalie Ochs

Laboratoire Informatique et
Systèmes

CNRS – Aix-Marseille Université
Marseille, France

magalie.ochs@univ-amu.fr

Laurent Prévot

Laboratoire Parole et Language
CNRS – Aix-Marseille Université

Marseille, France

Laurent.prevot@univ-amu.fr

Abstract

Assessing the social competence of anthropomorphic artificial agents developed to produce engaging social interactions with humans has become of primary importance to effectively compare various appearances and/or behaviours. Here we attempt to objectify the social competence of artificial agents, across different dimensions, using human brain neurophysiology. Whole brain activity is recorded with functional Magnetic Resonance Imaging (fMRI) while participants, naïve to the real purpose of the experiment, discuss either with a human confederate or with an artificial agent, presently the robotic conversational head Furhat controlled with a Wizard of Oz procedure. This allows a direct comparison of local brain responses, not only at the cortical level but also in subcortical structures associated with motivational drive and impossible to investigate with non-invasive neurophysiology techniques such as surface recordings. The present data (n=21 participants) demonstrates the feasibility of this approach, and results confirm an increased activity in subcortical structures, in particular the amygdala involved in emotional processing and the hypothalamus, known to secrete, among others, the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

ICMI'18 Adjunct, October 16–20, 2018, Boulder, CO, USA

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-6002-9/18/10 \$15.00

<https://doi.org/10.1145/3281151.3281162>

neurohormone oxytocin involved in social bonding, as well as the temporoparietal junction bilaterally involved in the attribution of mental states, when interacting with a human compared to an artificial agent. The reverse contrast revealed more dorsal cortical areas. Altogether, these results support the use of fMRI to objectify the social competence of artificial agents along distinct dimensions.

Author Keywords

Humanoid Robot; Artificial Agents; Social Interaction; Social Competence; Neurophysiology.

ACM Classification Keywords

- Computer systems organization ~ Robotic autonomy.

Introduction

While artificial agents are increasingly present in contact with humans, the assessment of how they are accepted as social partners by their users is very poor. Conversational agents are used as web agents for e-commerce or tutors in e-learning applications, vocal assistants have entered our privacy with home assistants such as Amazon Echo, Google home or Apple HomePod, and humanoid robots have been proposed for cognitive therapies for children with autism spectrum disorder (e.g. [1]). Yet, there is no objective measure of their social competence, defined as their ability to trigger social motivation for the interaction [2]. As a matter of fact, "How social are artificial agents?" is a conundrum, as the adjective "social" should be restricted to interactive behaviours taking place between humans. The social acceptance of robots is usually addressed with questionnaires, for example the Negative Attitude towards Robot Scale (NARS [3]). Such an approach can be useful to compare various

artificial agents in terms of the subjective response they elicit, but is not sufficient to be interpreted in terms of their social competence, requiring a comparison with humans.

Here, we propose to objectify the social competence of artificial agents through the comparison of behavioural and physiological responses between conditions in which one naïve person has natural interactions alternatively with a fellow human or an artificial agent, allowing to investigate how different dimensions of social competence are impacted by the nature, human or artificial, of the agent we're interacting with. This allows us to distinguish not only behavioural, but also, and more importantly, neurophysiological features that are specific to interacting with a human to those that are preserved when the interacting agent is artificial (see also [4]). The experimental approach has been presented in details in a recent publication describing its foundations and implementation for the behavioral experiment [5], and a forthcoming publication will describe its adaptation to the MRI environment. We will focus here on the added value of recording neurophysiological responses of the brain using functional magnetic resonance imaging (fMRI).

In a nutshell, a naïve participant is recorded in the MRI scanner with an experimental setup using a videoconferencing to support conversation with another agent, a human or a robot, located outside of the scanner room. The first objective of the experiment being the exploration of natural social interactions, it is mandatory that participants are not aware of this objective. Here, we use a believable cover story to provide credible, but spurious, explanations to as many aspects of the experimental set-up as possible.

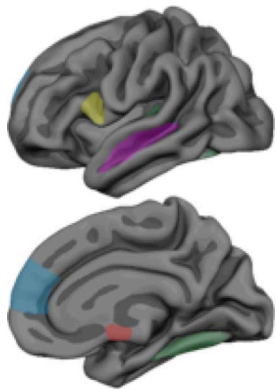


Figure 1: Inflated brain render with colored regions of interest described in the text.

Participants are made to believe they participate to a marketing experiment to validate an incoming advertising campaign. It is a similar approach that was used to provide evidence for the chameleon effect (i.e., alignment of postures and behaviour to that of an interaction partner [6]): participants believed they were participating to the development of a new psychological scale for which they were to work in pairs, uninformed of the actual goal of the experiment, investigating behavioral mimicry. The second objective is to have experimental control over some variables in order to be able to compare natural social interaction conditions to control conditions. As a conversation was chosen as support for the interaction, an artificial agent that is human like in appearance, able to have a discussion while being clearly identified as non human, was chosen as the agent for the control condition. The conversational robot Furhat from KTH was chosen for its ability to produce embodied conversation, in contrast to virtual agents that are by nature two-dimensional on-screen renders.

The main objective is to investigate neurophysiological correlates of natural social interactions. The underlying assumption is that neurophysiological responses can't be voluntarily controlled and can, by reverse inference, indicate cognitive processes at play in a given experimental condition. Indeed, the neural bases of a multitude of mechanisms involved in social cognition have been studied with functional neuroimaging, their neural correlates localized, and their response to artificial agents in controlled setups investigated. fMRI is the only neuroimaging technique that provides such a rich information, mainly because of its ability to locate spatially not only cortical but also subcortical

regions of increased activity. Here, some brain areas under investigation are presented (figure 1).

The Fusiform Face Area (FFA, green) plays a fundamental role in perceiving social information from others' faces [7]. Its response is increased when processing the actions of a humanoid robot [8]. The superior temporal sulcus (STS, purple) responds to the perception of social actions including gazes [9] and speech. We showed previously that the tendency to consider a motion depicted by a computer-animated character as biological correlated with activity in this area [10]. The ventral human premotor cortex (yellow) is central to speech production (posterior to or part of Broca's area), but also a key actor in social functions related to the mirror system, including action imitation and understanding. The importance of social motivation for normal social cognition was put forward as an explanation for social impairment in Autism Spectrum Disorder [2]. One candidate system supporting social motivation is the neurohormone oxytocin secreted by the hypothalamus (red). Accordingly, we reported that young adults with autism didn't exhibit differential activation of the hypothalamus (upper red area in medial render) when interacting with natural and artificial agent [11]. Finally, the Medial Frontal Cortex (MFC, blue) is associated with the attribution of mental states to humans, but not artificial agents [12]. Investigating brain responses in these areas allows us to identify how these various dimensions of social cognition are influenced by the natural or artificial nature of the agent we interact with, but also to investigate the influence changes in the artificial agents' appearance and behaviour on these dimensions of social interactions.

Methods

Participants

Twenty-one native French speakers, with no history of neurological or psychiatric disease, volunteered for this experiment, approved by the local ethics committee "Comité de protection des Personnes Sud méditerranée I". All participants gave their informed consent and received 40 euros.

Cover Story

In order to be able to observe a natural conversation, we developed a cover story. The experiment was presented as a study for an advertising campaign which sought to make the target audience talk about the message in order to understand it. They were hence told they would be presented with potential advertisement campaign images, that they would need to talk about these with two agents located outside the MRI: a fellow human and an conversational robotic head. Furthermore, the artificial agent was presented as autonomous and the "fellow participant" was in fact a confederate (TC for men and MB for women). After the experiment, the cover story was disclosed.

Conversation Robotic Agent

For this experiment we used the robotic head Furhat (Furhat robotics, <https://www.furhatrobotics.com/> [13]), in which we implemented prescript French conversation based on actual human conversational data in French [5] and adapted to the stimuli. The robot was controlled by a Wizard of Oz procedure implemented in Furhat web interface. This conversational script contained both answer specific of a given image, such as looks like a super hero", or "This is a red apple", and non specific feedback such as "yes", "no", "I think so". It also contained closed

questions, such as "Don't you think so?" and open ones such as "What do you think is the message?". In total we had around 50 possible sentences to implement according to the situation through a user interface showing tabs of the conversation. As for the confederate, Furhat face and voice was adapted according to the gender of the subject.

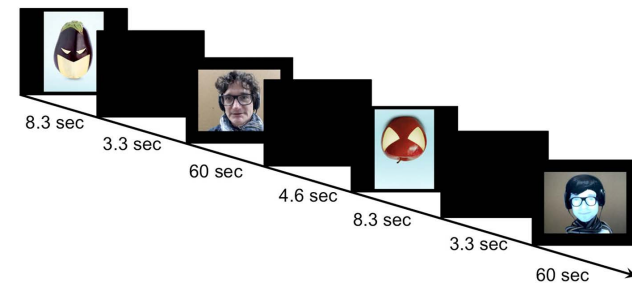


Figure 2: Timing of events in the experimental paradigm.

Experimental Paradigm

Two sets of each three stimuli were used in this experiment and presented to the participants. After the stimuli a black screen prepared the participants for their conversation. They then had one minute of conversation during which they talked either with the human or with the robot. The order of the stimuli was programmed in advance and the conversation partner alternated between consecutive trials (ie Human, Robot, Human, Robot...). The participants talked about each stimulus with each speaking partner two times, and had in total 24 minutes of conversation with the stimuli. Cameras were placed in front of the confederate and robot's face, allowing the subject to see his/her conversation partner live. Micros placed both in front of the subject and his/her conversation

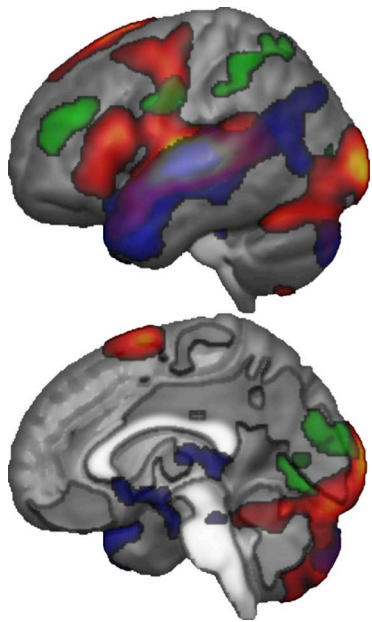


Figure 3: Lateral and medial renders of the activated clusters found in the left cerebral hemisphere for the main effect (Human & Robot, hot scale) and in the contrasts Human minus Robot (blue) & Robot minus Human (green).

partner, human or robot, allowed a bidirectional natural conversation. During each conversation the video and audio were recorded, as well as the direction of the participant's gaze, the blood pulse and respiration, and the fMRI data.

fMRI Acquisition

MRI data was collected with a 3T Siemens Prisma (Siemens Medical, Erlangen, Germany) using a 20-channel head coil. Blood oxygen level-dependent (BOLD) sensitive functional images were acquired using an EPI sequence in the 4 runs. Parameters were as follows: Echo time (TE) 30 ms, repetition time (TR) 1205 ms, flip angle 65°, 54 axial slices co-planar to the anterior / posterior commissure plane, FOV 210 x 210 mm, matrix size 84 x 84, voxel size 2.5 x 2.5 x 2.5 mm³, with multiband acquisition factor 3. After functional scanning, structural images were acquired (TE/TR 0.00228/2.4 ms, 320 sagittal slices, voxel size 0.8 x 0.8 x 0.8 mm³, field of view 204,8 x 256 x 256 mm³).

fMRI Analysis

Analysis was performed with SPM 12 running on Matlab, according to standard procedures. Preprocessing entailed correcting for delays in slice acquisition ("slice timing"), realignment and unwarping of the magnetic field distortion on the images, and smoothing with a 5-mm full-width half maximum kernel. First level, intra subject, statistical analysis was performed in each participant's native space. Normalization of beta estimates was based on DARTEL procedure, and second level analysis used a full ANOVA with Subjects, Sessions and Agents as factors. Localization of brain activation clusters was based on visual observation of the results rendered on the

average brain of our participants. Details of the analysis will be presented in a forthcoming neuroimaging publication. In the current publication, we focus on the added value of having this neuroimaging data to assess the social competence of the artificial agent.

Results And Discussion

Results

Testing for the main effect of conversation (including conditions from both conversations with the human and the robot) revealed activation clusters in the temporal lobe bilaterally (superior temporal gyrus (STG), superior temporal sulcus (STS)) extending dorsally into the ventral premotor cortex and into the depth of the dorsal part of central sulcus. Furthermore, we found activation in the occipital cortex, devoted to processing of visual information (including the extrastriate body area (EBA) and fusiform face area (FFA)). Also, we found activation in the left ventral premotor cortex (vPM) corresponding to Broca's area, involved in language perception and production (see Fig. 3), as well as in the medial premotor areas. The contrast comparing conversation when interacting with a fellow human to talking to the robot revealed cortical areas overlapping and extending beyond the temporal cortex found in the main effect of conversation, extending to the temporoparietal junction bilaterally, and subcortically in the hypothalamus, the amygdala and cerebellum bilaterally. The reverse contrast revealed activated clusters in the dorsal stream bilaterally, including the intraparietal sulcus and middle frontal gyrus, as well as in the depth of the central sulcus. A global view of the activations can be found in Figure 3.

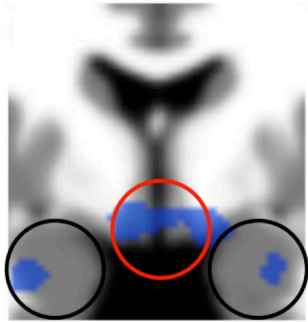


Figure 4: Section showing activation clusters in the amygdala (black circle) and hypothalamus (red circle) for the contrast Human minus

In the comparison between brain response to human and brain response to robot, we observed a large increase of the temporal cortex cluster bilaterally, in line with the general role of this cortical lobe in language processing, and in particular semantic. In addition, we found a number of subcortical responses, in particular in the hypothalamus and in the amygdala bilaterally.

In an oversimplification, the contrast between cortical and subcortical responses can be thought of as the contrast between conscious sensorimotor processing and autonomic motivational drives. While this simplification won't stand close scrutiny, it is helpful to provide an interpretation of our results in the light of our hypotheses. By experimental design, the human agent yields social motivation, and hence subcortical structures involved in motivational drive respond more to the human than robotic agent. This result is perfectly coherent with the hypothesis that the stance adopted with a natural and artificial agent is radically different [14], so that motivational aspects dominate in human interactions.

Effect of Robot: Cortical Processing

Interestingly, we observed increased activity in response to the robot compared to the human in cortical areas of the dorsal stream bilaterally. This can be explained *a posteriori*. Note that while linguistic analysis of the recorded discussions hasn't been carried out yet – but will be in the future – we expect a confirmation of the confederates feeling that participants discussed more and more lively with the human than the robot agent. While this wasn't expected and is in contradictions with the finding of a relative increase in cortical response during the robot interaction, it is fully coherent with the primary

question of the experiment, that is that humans adopt a specific stance when interacting with fellow humans that is radically different to the stance they adopt when interacting with an artificial agent, even though this agent reproduces a large number of the human's appearance and behavior.

Conclusion

In this paper, we report the first attempt to record the neurophysiology of natural conversation of a participant with a fellow human or an artificial agent with fMRI. Though preliminary, results suggest that despite similarities in basic sensorimotor processing, which were expected given that this is the reason why interactive artificial agents are designed anthropomorphic, different stances seem to be adopted as a function of the nature of the interacting agent. This experimental approach paves the way for an objective assessment of how humans perceive artificial agents, in particular distinguishing different dimensions of social cognition, and will be tested with other robotic embodiment and behaviors to confirm its use to assess different dimensions of social competence.

Acknowledgments

This research supported by grants ANR-16-CONV-0002 (ILCB) and ANR-11-LABX-0036 (BLRI) and has received funding from the Excellence Initiative of Aix-Marseille University - A*Midex, a French "Investissements d'Avenir" program (AAP-ID-17-46-170301-11.1).

References

1. Diehl, J. J., Schmitt, L. M., Villano, M., & Crowell, C. R. (2012). The clinical use of robots for individuals with Autism Spectrum Disorders: A critical review. *Research In Autism Spectrum Disorders*, 6(1), 249–262
2. Chevallier, C., Kohls, G., Troiani, V., Brodtkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in Cognitive Sciences*, 16(4), 231–239
3. Nomura, T., Suzuki, T., Kanda, T., & Kato, K. (2006). Measurement of negative attitudes toward robots. *Interaction Studies*, 7(3), 437–454
4. Wykowska, A., Chaminade, T., & Cheng, G. (2016). Embodied artificial agents for understanding human social cognition. *Phil. Trans. R. Soc. B*, 371, 20150375
5. Chaminade, T. (2017). An experimental approach to study the physiology of natural social interactions. *Interaction Studies*, 18(2), 254–275
6. Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of personality and social psychology*, 76(6), 893
7. Grill-Spector, K., & Malach, R. (2004). The human visual cortex. *Annu. Rev. Neurosci.*, 27, 649–677
8. Chaminade, T., Zecca, M., Blakemore, S. J., Takanishi, A., Frith, C. D., Micera, S., Dario, P., Rizzolatti, G., Gallese, V., & Umiltà, M. A. (2010). Brain response to a humanoid robot in areas implicated in the perception of human emotional gestures. *PLoS One*, 5(7), e11577
9. Allison T, Puce A, & McCarthy G (2000) Social perception from visual cues: role of the STS region. *TICS*, 4(7), 267
10. Chaminade, T., Hodgins, J., & Kawato, M. (2007). Anthropomorphism influences perception of computer-animated characters' actions. *Soc Cogn Affect Neurosci*, 2(3), 206–216
11. Chaminade, T., Fonseca, D., Rosset, D., Cheng, G., & Deruelle, C. (2015). Atypical modulation of hypothalamic activity by social context in ASD. *Research In Autism Spectrum Disorders*, 10, 41–50
12. Chaminade, T., Rosset, D., Fonseca, D. D., Nazarian, B., Lutcher, E., Cheng, G., & Deruelle, C. (2012). How do we think machines think? An fMRI study of alleged competition with an artificial intelligence. *Frontiers In Human Neuroscience*, 6
13. Al Moubayed, S., Beskow, J., Skantze, G., & Granström, B. (2012). Furhat: a back-projected human-like robot head for multiparty human-machine interaction. In *Cognitive behavioural systems* (pp. 114–130). Springer
14. Dennett, D. C. (1996), *The Intentional Stance* (6th printing), Cambridge (MA, USA): The MIT Press