Differences in Brain Activity During Turn Initiation in Human-Human and Human-Robot Conversation

Ekaterina Torubarova*, Caroline Arvidsson†, Julia Uddén‡, André Pereira*
*Department of Speech, Music and Hearing, KTH Royal Institute of Technology, Stockholm, Sweden
Email: ekator@kth.se, atap@kth.se
†Department of Linguistics, Stockholm University, Stockholm, Sweden
Email: caroline.arvidsson@ling.su.se
‡Department of Psychology, Stockholm University, Stockholm, Sweden
Email: julia.udden@psychology.su.se

Abstract—**In this paper, we investigated how the turn-taking mechanisms affect brain activity by comparing human-human and human-robot conversations. The current results within the initial exploratory analysis suggest higher involvement of an area previously associated with pragmatics during turn initiation in a conversation with a human in comparison to a robot, which might indicate deeper processing of the human agent's intention. We suggest that studying the turn-taking mechanism as a part of conversational dynamics can shed light on the differences in experiencing interaction with a human and with a robot.**

I. Introduction

Social robots are embodied agents used to communicate, assist, guide, and interact with a user. Unlike industrial robots, the aim of which is to substitute or facilitate human labor, a social robot usually engages in a shared task or a conversation with the user, to serve as a companion or an assistant. It is crucial to study how humans perceive robots. This field of research can both help to improve robotic design, and also to shed light on human communication and interaction mechanisms in general. While many studies use behavioral metrics or subjective reports to evaluate user experience during or after communicating with a robot, these methods rely on explicit parameters. However, not every aspect of interaction is experienced through explicit behavior. Inner states of the user such as conversational engagement may not be easily evaluated by, for example, external annotation of human-robot interaction.

Neuroimaging is a valuable approach for studying HRI because it allows the investigation of underlying neural processing in the absence of explicit behavior. This approach, unlike subjective rating or third-person annotation, can allow studying cognitive processes in terms of brain activity profoundly and objectively. Cross et al. [7] cover the potential of using neuroimaging in HRI research and how it can shed light on social cognition in general. The focus of many recent studies using neuroimaging for HRI has been focused on whether interaction with a robot can elicit involvement of social cognition brain mechanisms similar to interacting with a human [9, 21, 14, 10]. What has been less studied is the differences in the communicative process between a human and a robot interlocutor.

One of the key mechanisms involved in a dialogue is turn-taking. The turn-taking process is the mechanism that allows a conversation to unfold as a connected dialogue that requires estimation of the interlocutor's turn and planning of one's turn in a way to avoid long silences and overlaps [18]. The turn-taking process is crucial for establishing common ground and understanding the interlocutor's intention. While picking up the turn during a conversation, we rely on such cues as intonation [19], temporal and rhythmic alignment with the interlocutor [3, 20]. In brain imaging studies on human conversation, the turn-taking process has not been extensively studied before due to the limitations of studying speech production and concurring motion artifacts, and the neural correlates of turn-taking have not been clearly defined [4].

In human-robot interaction modeling turn-taking mechanisms that would allow a naturalistic conversation between the user and the robot is one of the crucial goals to make robots more human-like. For that reason, we consider turn-taking to be a valuable process to investigate inner user experience while talking to a robot, such as the level of conversational engagement. Hsiao et al. [12] found that level of engagement in a conversation can be successfully predicted from turn-taking (duration and count of turns and silence segments). Despite the presumable close connection between turn-taking and engagement, the effect of the timing of turn-taking can be ambiguous. For example, Cafaro et al. [5] showed that if interruption (i.e. untimely turn-taking) is used for cooperation between agents as opposed to disruptive strategy, the interrupting agent is perceived as more engaged. Similarly, in the study by Chao and Thomaz [6] adding interruptions during a collaborative task between a user and a robot increased task efficiency. On the other hand, an interruption can be severely disruptive in communication [11].

So far the brain activity differences in perceiving a human and a robot interlocutor in terms of conversational dynamics have not been extensively studied. Overall, how turn-taking mechanisms differ in communicating with a human and a robot and how it affects communication experience remains an open question. One of the aspects of this question is how can the agent's nature affect the preparation of one's response.

This paper aims to investigate the differences in brain activity during turn initiation in a free dialogue between two humans and a human with a robot.

II. Method

For this study, we took for analysis an openly available fMRI dataset of human-human and human-robot conversation provided by Rauchbauer et al. [16][1].

A. Dataset

The dataset consists of the recordings of 25 native French-speaking participants (21 participants included in the analysis in Rauchbauer et al. [15]). Each participant had several one-to-one conversations with an interlocutor, alternating between a human agent (the experimenter) and a robot agent (humanoid robotic head Furhat [1] whose utterances were pre-written based on a pilot human-human conversation, and the robot's responses were controlled using a Wizard-of-Oz procedure). The participants were presented with a 'cover story' for the experiment in which they were told that they were participating in an advertisement campaign, and they were supposed to discover the key message of the campaign together with the interlocutor. While being in the fMRI scanner, they were presented with images of anthropomorphized fruits, and they were instructed to freely discuss the images with the interlocutor who would be placed outside of the scanning room and connected via online video stream and bidirectional audio, to find out the key message of the campaign. Each conversation lasted one minute, after which the presented image and the agent changed. Each participant had four sessions of six one-minute conversations, three with a human agent and three with a robot agent; in total 24 minutes of conversation for each participant. During the conversations, their brain activity and audio of the conversation were recorded (other recorded parameters are out of the scope of the current analysis). For the details about data acquisition see Rauchbauer et al. [15].

The authors based their analysis on the differences related to the conversational agent. The fMRI events were divided based on the agent, taking the whole human-human (HHI) and human-robot (HRI) conversation as the contrast. The results showed that irrespective of the agent, a free conversation activated areas related to sensorimotor aspects of speech comprehension and production (such as posterior temporal cortex, inferior frontal gyrus), as well as visual processing (lateral and ventral occipital cortex). The human agent condition revealed higher activation in areas related to social cognition (such as the temporal cortex including TPJ and hypothalamus) in comparison to the robot condition. The opposite contrast revealed increased activation in visual areas, including the fusiform gyrus known for face processing.

While this analysis demonstrated higher involvement of social areas for HHI and, potentially, a higher effort for face processing in HRI, this analysis does not allow studying conversational dynamics, for example, turn-taking mechanisms. In the current analysis, we were interested in contrasting human and robot conditions during turn initiation.

B. Definition of Conversational Events

For the current analysis, the transcriptions of the participant's and the interlocutor's speech[2] were juxtaposed in a TextGrid format. The onsets and durations of the events were extracted from the transcriptions using a Python script. Based on the transcriptions, segments of production and comprehension with respect to the participant were defined for each conversation, where *production* was a time window of the participant's utterance, and *comprehension* was a time window of the participant's silence during the agent's utterance. After that, we defined a *turn initiation* class of events, which was defined as a 600 ms time window before the onset of the participant's utterance. Additional events were also defined and included in the model (thirteen event classes in total), but not reported in this paper.

C. Analysis

The preprocessing and the analysis of fMRI data were carried out using SPM12[3]. For the preprocessing, rigid body transformation (realignment) was performed using 6 parameters (translations and rotations). The head movements in x, y, and z were checked independently. One participant was excluded from the following analyses due to excessive head movement. The functional images were then coregistered to one of the anatomical images (T1) and normalized to a standard Montreal Neurological Institute (MNI) space. The normalization was performed with affine regularization and included a resampling of the voxels to 2x2x2 mm with a 4th degree B-spline interpolation. White and grey matter segmentation and bias correction were conducted during the normalization step. Finally, functional images were spatially smoothed using a 3D isotropic 5 mm full-with-at-half-maximum Gaussian kernel. A temporal high-pass filter (cycle cut-off at 128 sec) was utilized to account for various low-frequency effects.

The hemodynamic response function (HRF) was modeled for the conditions and six motion parameters (from the rigid body transformation) using a general linear model (GLM). The regressors were convolved with a canonical HRF using a 2 mm within brain mask.

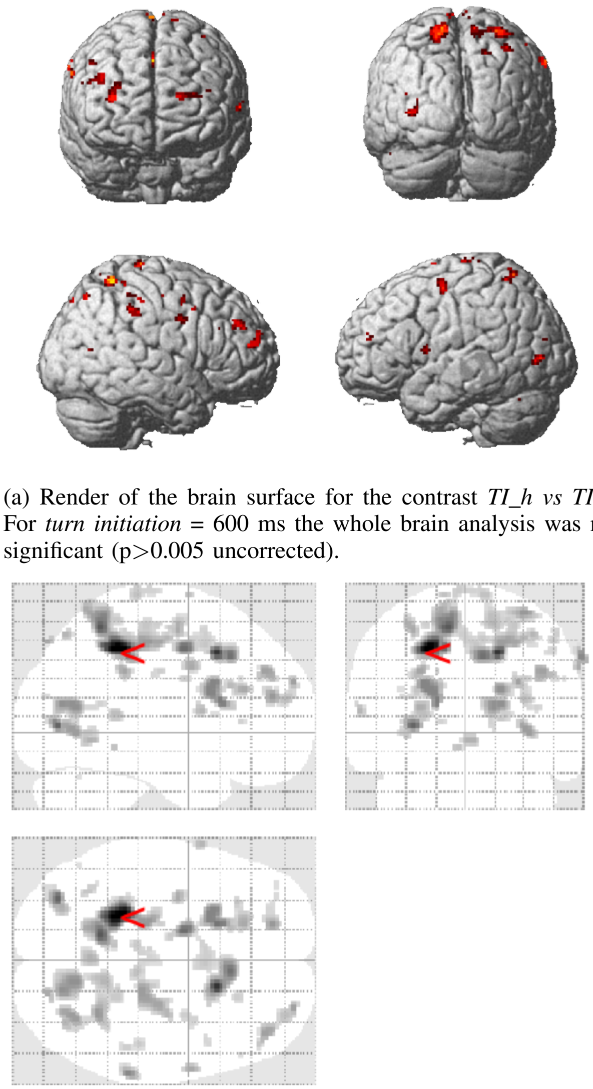
Two contrasts were used in this study: turn initiation in a conversation with a human vs. turn initiation in a conversation with a robot (*TI_h vs TI_r*) and the reverse contrast (*TI_r vs TI_h*).

For the second level analysis, the cluster-forming threshold of $p_{uncorrected}$ was set to .005 (no extent-level threshold, k = 0). Family wise error (FWE), as implemented in SPM12, was utilized as multiple comparison correction method (at the cluster and peak-level).

[1] Available at https://openneuro.org/datasets/ds001740/versions/2.1.0

[2] Available at https://www.ortolang.fr/market/corpora/convers/v2

[3] https://www.fil.ion.ucl.ac.uk/spm/software/spm12/



(a) Render of the brain surface for the contrast *TI_h vs TI_r*. For *turn initiation* = 600 ms the whole brain analysis was not significant (p>0.005 uncorrected).

(b) Significant activation at [-24, -40, 44] after small volume correction (p_{FWE} = 0.007)

Fig. 1: Results of the contrast between human and robot agent during turn initiation

D. Results

For the initial exploratory analysis, we started at the whole-brain level. No significant activation was found for the contrast *TI_r vs TI_h*. For *TI_h vs TI_r* contrast the exploratory analysis revealed a pattern of activity in the parietal area, albeit not significant (see Fig. 1a). We then performed small volume correction (search volume: 20 mm sphere at [-40, -48, 46]). The coordinate was based on a previous study [2] were this area showed activation in relation to a pragmatic task. A significant activation was found at the coordinates [-24, -40, 44] (T = 5.12, p_{FWE} = 0.007) (see Fig. 1b).

E. Discussion

In this study, we analyzed an open-source fMRI dataset of human-human and human-robot conversations extracting novel events from fMRI data. While previous analysis of the dataset used the whole conversation contrasting human and robot agents, we divided each conversation into shorter events. Given that online verbal conversation has not been extensively studied using fMRI due to noise and motion artifacts, our first aim was to test how well can short events be extracted from fMRI data of a free conversation with a human and a robot. The focus of the current paper was the turn initiation time window before the onset of the participant's speech.

Current preliminary results suggest higher involvement of an area previously associated with pragmatics for the human but not for the robot agent during turn initiation. This finding might indicate a higher level of the human agent's intention processing. The manual check of conversation recordings demonstrated that the human agent, unlike the robot agent, produced more 'lively' speech with naturally occurring prosodic and linguistic turn-ending cues, such as 'umm' at the end of the phrase, etc. The robot's utterances were prewritten and produced using the Wizard-of-Oz method, and manual errors by the robot's operator could have occurred while producing the robot's responses. In addition, the robot interlocutor was not expressing non-behavioral cues, as its non-human nature was emphasized by the study design. In addition, despite being based on a previously recorded human dialogue, the robot's utterances were not identical to the human agent's utterances, being shorter and more limited. While the Furhat robot can produce more engaging and versatile behavior, the emphasis on the robot's artificial nature in this dataset and the limitation of its expressions could have affected the differences in intention processing.

Our results require more detailed further exploration. As the first step, we aim to investigate different duration of turn initiation time windows. The duration of 600 ms was established based on the findings in psycholinguistics studies: picture naming tasks suggest 600 ms be a minimum window between seeing a picture and starting to articulate its name [13]. However, in a conversation based on a common topic of discussion that requires less automatic thinking and more complex sentence production, we can expect that turn initiation takes longer: Griffin and Bock [8] suggests 1500 ms latency for sentence production.

The current results showed that short events extracted from a free conversation can indeed be used for fMRI contrast. For further analysis, we are aiming to investigate other conversational events, for example, transitional gaps to see how the nature of the agent affected the change of turn between speakers. Also, a linguistic analysis of the content of the participant's utterances can shed light on whether and how the participants adjusted their speech depending on the agent, and how is it represented in terms of brain activity during production.

The current preliminary results suggest that the conversational dynamics can indeed differ for a human and a robot interlocutor during turn-taking. Given that, we can lead further analysis to investigate how it can be used for studying a user's inner state such as conversational engagement by performing conversation annotations. However, given the nature of the task, relatively short conversation duration, and lack of emotional expressions by the robot interlocutor, it is probable that the level of the participants' engagement would not greatly vary in this dataset. Due to that, we are going to collect new data of human-robot conversation a) using more dynamic and naturalistic conversational cues in the robot's speech, b) varying levels of engagement by alternating these cues to investigate to which extent the perceived nature of the agent affects the user's conversational processing and inner experience. While participant's engagement in human-robot interaction has been studied in terms of behavioral responses (such as gaze, head nods, verbal backchannels [17]), a brain imaging approach can shed light on the underlying neural processes without relying solely on external behavior annotations. This approach is important for improving robot's communication with atypical users such as autistic or ADHD users, whose behavioral cues differ from typical individuals and may not reliably predict their level of engagement. In this case, looking at brain activity during short conversational events such as turn initiation can shed light on how the agent's nature is processed by the user to eventually design the robot's behavior as more socially engaging.

III. Conclusion

We conducted an analysis of an openly available fMRI dataset of human-human and human-robot conversations, using a 600 ms turn initiation time window for contrasting human and robot conditions. Within the exploratory analysis the preliminary results suggest that during turn initiation, a conversation with the human interlocutor may stronger involve pragmatically relevant neural resources than a conversation with the robot. This can indicate deeper processing of the human agent's intention. These preliminary results require further research on the differences in the human brain activity in a free conversation with different agents.

References

[1] Samer Al Moubayed, Jonas Beskow, Gabriel Skantze, and Björn Granström. Furhat: a back-projected human-like robot head for multiparty human-machine interaction. In *Cognitive behavioural systems*, pages 114–130. Springer, 2012.

[2] Katarina Nanna Filippa Bendtz, Sarah Ericsson, Josephine Schneider, Julia Borg, Jana Bašnákova, and Julia Uddén. Individual differences in indirect speech act processing found outside the language network. *Neurobiology of Language*, pages 1–64, 2022.

[3] Štefan Beňuš, Agustín Gravano, and Julia Hirschberg. Pragmatic aspects of temporal accommodation in turn-taking. *Journal of Pragmatics*, 43(12):3001–3027, 2011.

[4] Sara Bögels and Stephen C Levinson. The brain behind the response: Insights into turn-taking in conversation from neuroimaging. *Research on Language and Social Interaction*, 50(1):71–89, 2017.

[5] Angelo Cafaro, Nadine Glas, and Catherine Pelachaud. The effects of interrupting behavior on interpersonal attitude and engagement in dyadic interactions. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*, pages 911–920, 2016.

[6] Crystal Chao and Andrea L Thomaz. Timing in multimodal turn-taking interactions: Control and analysis using timed petri nets. 2012.

[7] Emily S Cross, Ruud Hortensius, and Agnieszka Wykowska. From social brains to social robots: applying neurocognitive insights to human–robot interaction, 2019.

[8] Zenzi M Griffin and Kathryn Bock. What the eyes say about speaking. *Psychological science*, 11(4):274–279, 2000.

[9] Frank Hegel, Sören Krach, Tilo Kircher, Britta Wrede, and Gerhard Sagerer. Theory of mind (tom) on robots: A functional neuroimaging study. In *2008 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 335–342. IEEE, 2008.

[10] Ruud Hortensius and Emily S Cross. From automata to animate beings: the scope and limits of attributing socialness to artificial agents. *Annals of the new York Academy of Sciences*, 1426(1):93–110, 2018.

[11] Eric Horvitz, Johnson Apacible, and Muru Subramani. Balancing awareness and interruption: Investigation of notification deferral policies. In *International Conference on User Modeling*, pages 433–437. Springer, 2005.

[12] Joey Chiao-yin Hsiao, Wan-rong Jih, and Jane Yung-jen Hsu. Recognizing continuous social engagement level in dyadic conversation by using turn-taking and speech emotion patterns. In *Workshops at the Twenty-Sixth AAAI Conference on Artificial Intelligence*, 2012.

[13] Peter Indefrey and Willem JM Levelt. The spatial and temporal signatures of word production components. *Cognition*, 92(1-2):101–144, 2004.

[14] Ceylan Özdem, Eva Wiese, Agnieszka Wykowska, Hermann Müller, Marcel Brass, and Frank Van Overwalle. Believing androids–fmri activation in the right temporo-parietal junction is modulated by ascribing intentions to non-human agents. *Social Neuroscience*, 12(5):582–593, 2017.

[15] Birgit Rauchbauer, Bruno Nazarian, Morgane Bourhis, Magalie Ochs, Laurent Prévot, and Thierry Chaminade. Brain activity during reciprocal social interaction investigated using conversational robots as control condition. *Philosophical Transactions of the Royal Society B*, 374 (1771):20180033, 2019.

[16] Birgit Rauchbauer, Youssef Hmamouche, Brigitte Bigi, Laurent Prevot, Magalie Ochs, and Chaminade Thierry. Multimodal corpus of bidirectional conversation of

human-human and human-robot interaction during fmri scanning. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 661–668. European Language Resources Association, 2020.

[17] Charles Rich, Brett Ponsler, Aaron Holroyd, and Candace L Sidner. Recognizing engagement in human-robot interaction. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 375–382. IEEE, 2010.

[18] Harvey Sacks, Emanuel A Schegloff, and Gail Jefferson. A simplest systematics for the organization of turn taking for conversation. In *Studies in the organization of conversational interaction*, pages 7–55. Elsevier, 1978.

[19] Deborah Schaffer. The role of intonation as a cue to turn taking in conversation. *Journal of Phonetics*, 11(3): 243–257, 1983.

[20] Tanya Stivers, Nicholas J Enfield, Penelope Brown, Christina Englert, Makoto Hayashi, Trine Heinemann, Gertie Hoymann, Federico Rossano, Jan Peter De Ruiter, Kyung-Eun Yoon, et al. Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, 106(26):10587–10592, 2009.

[21] Eva Wiese, Giorgio Metta, and Agnieszka Wykowska. Robots as intentional agents: using neuroscientific methods to make robots appear more social. *Frontiers in psychology*, 8:1663, 2017.