Convers 2: A Multimodal Corpus of Natural Human–Robot and Human–Human Conversations During fMRI

Authors: Camilla Di Pasquasio, Aix-Marseille University
Bruno Nazarian, Aix-Marseille University, CNRS
Thierry Chaminade, Aix-Marseille University, CNRS
Brigitte Bigi, Aix-Marseille University, CNRS
Abstract ID: 419
Event: Linguistics 2026 Paris
Subject: 15. Corpus linguistics
Presenter Name: Camilla Di Pasquasio

Abstract

Conversational corpora are essential for studying natural spoken interaction and developing models of human dialogue, providing valuable resources for investigating conversational structure and interactional dynamics. More recently, this approach has been extended to interactions with robotic agents. However, existing corpora have primarily focused on teleoperated robotic agents, whereas recent advances in large language models now make it possible to study conversations with autonomous embodied AI.

Convers 2 is a multimodal corpus of naturalistic conversations collected during real-time interactions between a human participant lying in an MRI scanner and a humanoid robotic head (Furhat). It extends the Convers 1 corpus, which documented interactions between human participants and a teleoperated humanoid robot.

While the robot's appearance and voice remain constant across interactions, its responses are generated either by a hidden human interlocutor or by a large language model, enabling the comparison of human–human and human–AI autonomous conversational behavior under controlled conditions. To minimize potential gender-related biases, the robotic interlocutor was designed with a gender-neutral appearance and voice.

A sample of twenty-five participants discussed six images across three fMRI sessions (four with a human interlocutor and four with the AI).

The corpus comprises 8 hours of multimodal raw data, including audio and video recordings, neurophysiological signals and eye-tracking data. Derived resources include speech transcriptions, conversational turn annotations, and behavioural questionnaires assessing individual differences in personality (Big Five) and attitudes toward robots (NARS).

Speech recordings from both participants and interlocutors were segmented into Inter-Pausal Units (IPUs), defined as speech segments separated by pauses of at least 200 ms. The corpus includes manually verified IPU annotations, automatic speech transcriptions generated with Whisper, and time-aligned TextGrid files containing conversational turn annotations. All linguistic annotations are synchronized with the corresponding multimodal recordings, including eye-tracking, physiological signals, and structural and functional MRI data.

The corpus enables investigations of conversational phenomena such as turn-taking, interactional timing, lexical alignment, and multimodal communication in both human–human and human–AI dialogue.

To enable autonomous interaction, a real-time perception–action loop compatible with the MRI environment was developed. The system integrates online audio acquisition, automatic speech transcription, LLM-based response generation, and speech synthesis through the robotic head.

Convers 2 provides a unique resource for investigating natural human–AI conversations and their neural, linguistic, and behavioral correlates, supporting future research in linguistics, cognitive neuroscience, and human–robot interaction.