RoomReader: A Multimodal Corpus of Online Multiparty Conversational Interactions

Justine Reverdy, Sam O'Connor Russell, Louise Duquenne, Diego Garaialde, Benjamin Cowan, Naomi Harte

Sigmedia Group, ADAPT Centre, School of Engineering, Trinity College Dublin
ADAPT Centre, School of Information and Communication Studies, University College Dublin
{reverdyj, russelsa, nharte}@tcd.ie

Problem

The multiple lockdowns that were enforced in many countries to slow the spread of the COVID-19 pandemic forced workers, teachers and students to rapidly adapt to an online environment to continue their professional, educational and social activities.
This situation highlighted the multiple issues accompanying video-conferencing interactions (e.g., fatigue, time latency leading to missed or distorted turn-taking cues, or difficulties to perceive disengagement cues from interlocutors), and the lack of available datasets.

Conversational Engagement in Educational Context

The degree of involvement of students in a topic being discussed and their willingness to continue the interaction. It can be analysed along three dimensions: from visual cues, from linguistic and paralinguistic cues, and from elements of the dialogue structure relevant to group cohesion.

Scenario Design

Collaborative student-tutor scenario aiming to elicit spontaneous speech: a quiz of three to five questions inspired from the TV show Family Feud -- adaptation of the MULTISIMO task. The scenario has clear phases of question-answering and consensus building.

> *"I'm going to ask you a question that was also asked to 100 people in a survey and you guys are going to have to talk together and come up with the top three most popular answers the question. Name something people are often chased by in movies."*

Figure 1: Session S09 captured with OBS Studio.

Main Outcomes

- 8h44 of multimodal recordings of conversational group interactions
- Labelled dataset of 30 online tutorial-style sessions, 118 participants
- Full transcriptions with utterance, word and phoneme level boundaries
- Engagement Annotations + Associated Metrics

Ethical considerations

Ethical and privacy aspects were considered at all stages and integrated into the design process. The participant recruitment strategy, information provided to participants regarding the usage of their data, participant consent, data storage, and a licence agreement enabling the corpus to be shared with researchers worldwide, have been independently assessed by the School of Engineering Ethical Committee, TCD, and the TCD Data Protection Officer to ensure GDPR compliance.

Associated Metrics

Participants answered questionnaires before and after the sessions to obtain self-reported and externally observed metrics.

Table 2: Set of metrics accompanying the multimedia files

Measure	Respondent role
Self-reported engagement	Student
Perceived engagement	Tutor & External annotators
Group dynamic (Bespoke)	Tutor
Group dynamic (Validated)	Tutor
Personality Test	Student & Tutor
Tutor's behaviour	Student

Post-Processing and Transcriptions

The recordings were made via an online platform (Zoom) and a screen recorder (OBS Studio), which necessitated several post-processing step to synchronize audio and video files. ASR was applied and manually corrected, as well as reintroducing paralinguistic elements removed by the ASR.

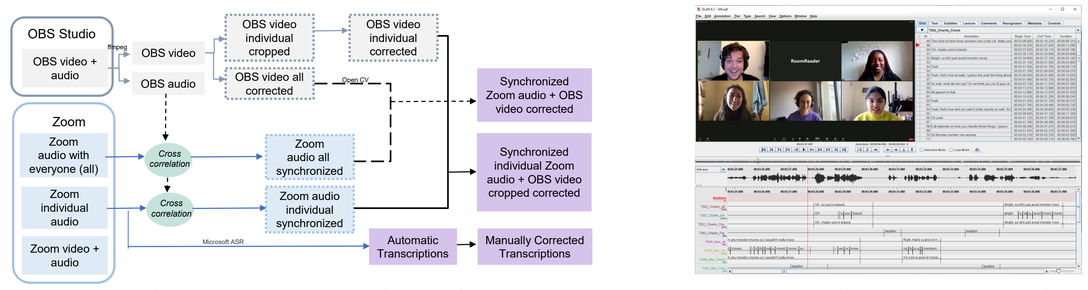

Figure 4: Processing of original captured signals

Figure 5: Example of a session in ELAN with participants transcription tiers

Descriptive Statistics

Table 1: Summary of number of words, utterances, and floor time (minutes) per gender and role in the 30 recordings. *Note: The 2 tutors did 15 sessions each, while the students participated only once.*

	Tutor (30[2])		Students (115)						Total
	Female		Male		Female		Male	Other	
		(percent)		*(percent)*		*(percent)*		*(percent)*	
n	15	*10.3%*	15	*10.3%*	64	*44.1%*	49	*33.8%*	2	*1.4%*	145
Word	25054	*32.7%*	17371	*22.7%*	17501	*22.9%*	16301	*21.3%*	305	*0.4%*	76532
Utterance	2231	*20.9%*	1754	*16.4%*	3553	*33.2%*	3089	*28.9%*	61	*0.6%*	10688
Floor time	134.27	*32.2%*	87.42	*21.0%*	101.42	*24.3%*	92.36	*22.1%*	1.7	*0.4%*	417.17

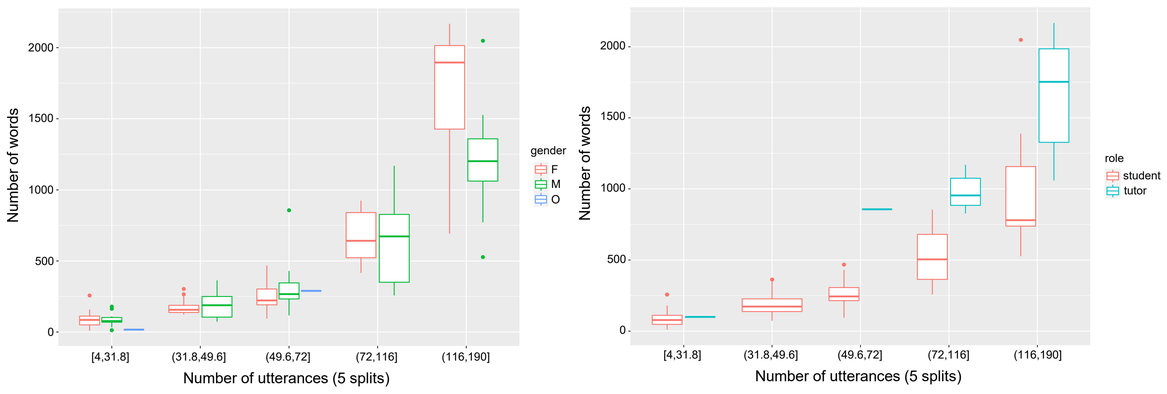

Figure 2: Word count per number of utterance per Speaker over each session | Per gender and role

Continuous Engagement Annotation Framework

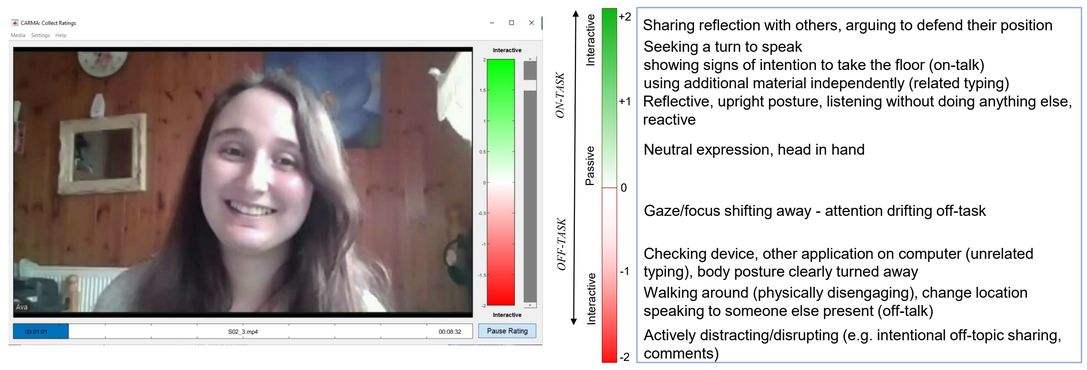

Figure 6: Continuous engagement annotation process using the online conversational engagement scale.

Expert annotators were recruited to individually annotate the participants along a scale adapted from Goldberg et al. (2019).

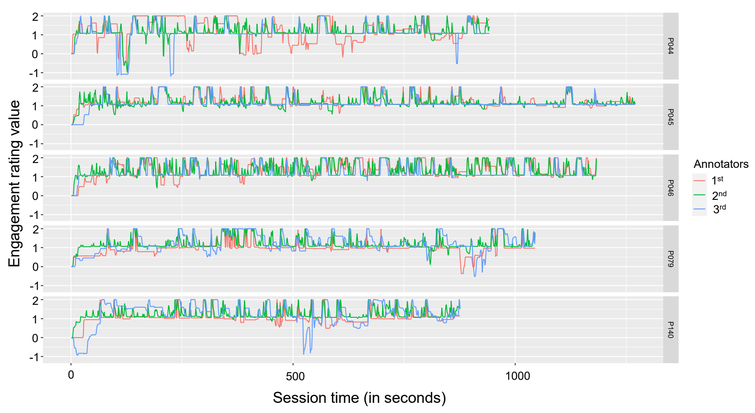

Figure 7: Continuous engagement annotations of 5 participants

Multiple Cognitive States Examples

A wide range of affective/cognitive states are present in the data, categorized below by visual engagement intensity, plus additional states.

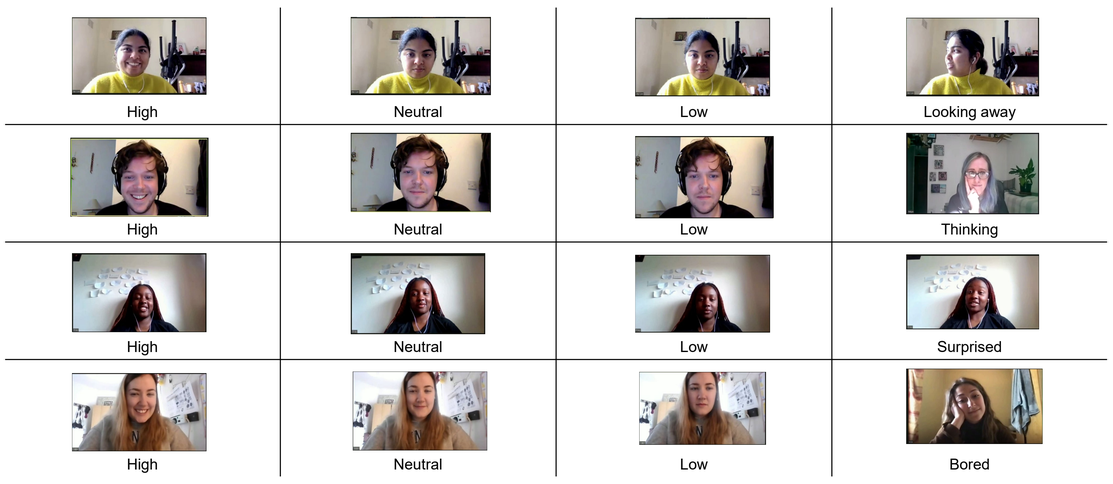

Figure 3: Examples of affective/cognitive states present in the RoomReader sessions.

Data Usage

- Automatic Engagement Detection
- Online Conversation Analysis
- Comparison Online vs Face-To-Face Multiparty Dialogue
- Assess engagement perception (self-reported and externally perceived)

This research was conducted with the financial support of Science Foundation Ireland under Grant Agreement No. 13/RC/2106_P2 at the ADAPT SFI Research Centre at TCD. ADAPT, the SFI Research Centre for AI-Driven Digital Content Technology, is funded by Science Foundation Ireland through the SFI Research Centres Programme.