

Engaging Content Engaging People

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-

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**

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

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])		Total			
Female	Male	Female	Male	Other	
(percent)		(percent)	(percent)	(percent)	(percent)

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.

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

Student

Table 2: Set of metrics accompanying the multimedia files

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.

Tutor's behaviour



						Grid	Text	Subtitles	Lexicon	Comments	Recogn	izers M	etadata Co	ontrols	
							T002_C	harlie_Chec	k						
						> N	r		Annotatic	in		Begin Time	End Time	D	uration
	91		_	_			48 They k	ond of have th	ree answers	one is the car. It	Aake you	0 03 09 92	0 00 03 16 2	00 00	0.06.3
A CALCULAR							49				_	0 03 16 23	0 00:03:27.9	20 00.0	10:11.6
				199 - L 200			50 OK, m	saybe you're b	lased.			0 03 27 92	0 00:03:30.1	0 00.0	0.02.1
				1000		-	51 E2 Alricht	and before is set a	tionid manufa	e marie	-	30.03.30.11	00.03.34.9	00.0	0.04.8
AND OR A		RoomReader		LIV AND			62 Mingrit	, so let s just i	word monste	THOME.	-	30 03 37 16	00.03.37.1	00.0	10 02 0
	1.0						54 Yeah					30 03 41 05	0 00 03 41 7	0 00 0	0.001
and a second	and a second						55					0 03 41 76	0 00 03 46 0	0 00 0	0 04 2
						1	56 Yeah.	that's true act	ually. Louess	the yeah the thi	ng about	0 03 46.03	0 00:04:01.6	0 00 0	30:15
						1	57					0 04 01 68	0 00 04 03 7	00 00	0 02
				10	1	1	58 So wa	it, what did w	say? So we	think you you \$	guys ar.	0 04 03 73	0 00:04:11.0	0 00.0	0 07
				100 - SA		1	59					00.04:11.01	0 00:04:13.9	0.00	0.02
1				The local division of			60 All agr	eed on that				0 04:13.90	0 00:04:15.0	0.00	0.01
				1 m	- (A) 4m	1	61					0 04 15 01	0 00:04:21.2	50 00.0	10:06:
				1.2			62 Yeah.					0 04 21 25	0 00.04.21.7	30 00.0	10:00
					1 1 1 1	-	63					0 04 21 70	0 00:04:24.4	00.00	10.02
		10 14					64 Yeah,	that's true and	s you said it p	oretty quickly as	well. So I	30 04 24 49	0 00:04 29.8	0.00	0.05
					and the second		00	ala.			-	20 04 29 82	00.04.41.1.		0.11
10 Max 10 Max	100		and a start of				67 On yes	an.			-	0.04.41.12	0 00:04:49.4		10.00
							68 It all de	enends on ho	w you classif	these things 1	nuese .	0.04 48 40	00.04.51.0	0 00 0	0 02
							69	openses on no	1 100 000000	and be training b, it	100.00	0.04:51.01	0 00:04:52.3	0 00.0	0.01
							70 So Mo	nster number	one anyway			0 04 52 31	0 00:04:55.3	00 00.0	0.03.0
		a a a a a a a a	a. a. e			1	71					0 04:55.33	0 00:04:56.1	0 00.0	00:00
		60:03:27.697		Selection: 00:00:0	0.000 - 00:00:00.000 0										
	14 14 14	Fd -id b bis bF b	M N M	bs s -			election M	tote 10	n Mode	10					
				100 10 11		1 100 00									
	The state of the s			to share in a											





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

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 4: Processing of original captured signals

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

Continuous Engagement Annotation Framework



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

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



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



Data Usage

Figure 7: Continuous engagement annotations of 5 participants

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



Trinity College Dublin The University of Dublin

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.

