RoomReader: A Multimodal Corpus of Online Multiparty Conversational Interactions

Justine Reverdy, Sam O’Connor Russell, Louise Duquenne, Diego Garai-Alde, 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.

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.

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.

Table 1: Summary of number of words, utterances, and floor time [minutes] per gender and role in the 30 recordings. The n in brackets is the number of sessions, and the number of recordings per gender and role.

<table>
<thead>
<tr>
<th>Auditory</th>
<th>Male</th>
<th>1500</th>
<th>1000</th>
<th>2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral</td>
<td>2250</td>
<td>1750</td>
<td>1250</td>
<td></td>
</tr>
<tr>
<td>Visual</td>
<td>2500</td>
<td>2000</td>
<td>1500</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>4750</td>
<td>3750</td>
<td>3250</td>
<td></td>
</tr>
</tbody>
</table>

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

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