MOSAIC : a Corpus of Small-Group Interactions During a Collaborative Task

Amine Benamara[1], **Celine Clavel**[1], **Brian Ravenet**[1], **Nicolas Sabouret**[1], **Mathilde Sassier–Roublin**[2], **Julien Saunier**[2]
[1]Université Paris-Saclay, CNRS, LISN, [2]Normandie Université, INSA Rouen, LITIS
{benamara, clavel, ravenet, sabouret}@lisn.upsaclay.fr
julien.saunier, mathilde.sassier–roublin@insa-rouen.fr

Abstract

This paper presents MOSAIC (Multimodal Observations of Social Affect, Intimacy, and Cohesion), a multimodal interaction corpus of video and audio recordings of 17 groups of 4 participants (68 participants in total) playing a collaborative board game. The aim of this corpus collection is to support the design of socially interactive agents. We describe the experimental protocol of this corpus collection, the perceptive questionnaires completed by participants, the automatic annotation process of game specific elements and non-verbal behaviors, and the manual verbal annotations collected. We provide a preliminary characterization of this corpus with descriptive results for some of the perceptive scales used in this study. We illustrate the possibilities offered by this corpus on the question of interpersonal social relations and group dynamics.

Keywords: multimodal interaction, corpus, collaborative task

1. Introduction

Artificial agents are increasingly designed to collaborate with humans in small groups, for instance in education, training, or decision-making settings (Sebo et al., 2020). In such mixed teams, interaction quality depends not only on task performance but also on group cohesion, i.e., the extent to which members feel connected and committed to a shared goal (Mikalachki, 1969; Salas et al., 2015). Cohesion influences engagement, motivation, and collective performance (Beal et al., 2003), and therefore constitutes a central challenge for socially interactive agents operating in multiparty environments.

Recent work in social psychology indicates that cohesion partly emerges from situational intimacy i.e. short-term interpersonal processes involving positive affect, mutual understanding, and self-disclosure (Reis, 2017; Prager, 2000). For artificial agents to adapt their behavior in group interactions, computational models must thus capture how multimodal behaviors relate to perceived intimacy and cohesion over time.

However, this relationship cannot be directly studied using existing multiparty interaction corpora. Meeting corpora such as AMI focus on decision-making dynamics (Carletta et al., 2006), while datasets such as MULTISIMO target task-oriented collaboration (Koutsombogera and Vogel, 2018). Other resources, including GAME-ON, measure group cohesion but do not jointly capture interpersonal intimacy and multimodal behavioral signals (Maman et al., 2020). As a result, no current corpus aligns behavioral cues, perceived intimacy, and cohesion dynamics in short-term collaborative interactions. While previous corpora measure either group-level outcomes or behavioral signals, none enable the investigation of how interpersonal relational processes unfold over time and differentially contribute to group-level cohesion in short-term collaborative settings.

To address this gap, we introduce MOSAIC (Multimodal Observations of Social Affect, Intimacy, and Cohesion), a multimodal corpus of four-person collaborative interactions collected during a cooperative word-association game. The dataset contains recordings from 68 participants organized into 17 groups, resulting in 136 interaction sessions, adding up to more than 6 hours of multimodal data. Each session captures the emergence of short-term group relationships.

The corpus combines multimodal behavioral recordings with aligned social perception measures, enabling the study of how interpersonal behaviors relate to perceived intimacy and group cohesion. It includes audio-video recordings, transcriptions, gaze targets, facial behavior signals, interaction phases derived from game logs, and post-interaction questionnaires measuring cohesion and intimacy (self- and other-report), while also controlling for prior familiarity between participants.

The MOSAIC corpus supports research in social signal processing, multimodal interaction modeling, and the design of socially adaptive artificial agents in group settings.

This paper provides:

- A multimodal multiparty interaction corpus designed to study situational intimacy and group cohesion
- A role-based interaction phase annotation

Proceedings of Computational Affective Science (CAS) @ LREC 2026, pages 164–174
16 May 2026.

scheme automatically derived from game logs

- Multilevel social perception measures (self, interpersonal, and group) aligned with behavioral recordings
- Baseline analyses revealing differential temporal dynamics between intimacy and cohesion

Our preliminary analyses suggest that intimacy and cohesion follow partially distinct temporal dynamics, especially depending on prior familiarity between participants, highlighting the need for fine-grained multimodal corpora capturing both constructs simultaneously.

2. Related works

Understanding group interaction requires capturing interpersonal processes that are not directly observable but inferred from behavior. Psychological models describe intimacy as a relational process involving self-disclosure and partner responsiveness during interaction (Laurenceau et al., 1998; Reis, 2017; Prager, 2000). In computational and human–agent interaction research, this process has been operationalized through observable behaviors such as self-disclosure, positive affect, and mutual understanding (Potdevin, 2020). However, these approaches have primarily been investigated in dyadic settings, leaving open how such interpersonal mechanisms manifest in multiparty collaboration.

Psychological studies validate perceived intimacy without directly observing its multimodal behavioral realization, while HCI studies manipulate predefined behaviors assumed to express intimacy. Previous work has therefore not examined how intimacy is expressed through multimodal behaviors in spontaneous multiparty interaction.

In human–agent interaction, these mechanisms have been shown to influence interaction outcomes. Agents using self-disclosure or reciprocal behaviors increase trust, engagement, and social presence (Bickmore and Schulman, 2012; Lee and Choi, 2017; Potdevin, 2020). Similarly, robots expressing vulnerability improve human–human communication and cooperation within teams (Traeger et al., 2020). These findings highlight the functional importance of intimacy-related behaviors but do not provide observational data from natural multiparty interactions.

From a data perspective, social signal processing research relies on multimodal corpora to analyze group dynamics and train computational models (Bänziger et al., 2012; Soleymani et al., 2019; Kebe et al., 2024). Multiparty interactions are particularly challenging because participants simultaneously manage turn-taking, gaze, spatial positioning, and multiple conversational threads (Gillet et al., 2022). As a result, existing datasets have typically been designed around specific interaction contexts and target processes.

Meeting corpora such as AMI (Carletta et al., 2006) capture decision-making and role distribution in professional discussions using audio, video, and manual behavioral annotations. Other datasets focus on collaborative task solving: MULTISIMO (Koutsombogera and Vogel, 2018) records small groups performing a quiz task and provides multimodal annotations for affect, engagement, and interaction strategies, while linking them to individual participant traits. Some corpora investigate emergent group states, for instance leadership (Sanchez-Cortes et al., 2011) or cohesion. The GAME-ON dataset (Maman et al., 2020) specifically targets group cohesion in a cooperative escape-game scenario and combines multimodal behavioral recordings with group perception questionnaires.

Although these resources provide rich behavioral recordings and annotations, they typically capture either interaction behavior (e.g., coordination, engagement, roles) or collective perception (e.g., cohesion, leadership). They do not explicitly model interpersonal mechanisms underlying these group states, and in particular they lack aligned measures of perceived intimacy between participants during the interaction. Consequently, existing corpora do not allow studying how multimodal interaction patterns between individuals contribute to the temporal emergence of group cohesion in short-term collaborative settings.

3. Corpus Description and Data Acquisition

3.1. Data availability

Due to ethical constraints and participant consent limitations, video recordings cannot be publicly shared. Access to anonymized behavioral data is available upon request to the authors for research purposes, subject to a confidentiality agreement. This restriction is designed to protect participant anonymity while still providing access to the essential data needed for result reproducibility.

3.2. Participants

We recruited 68 participants, split into 17 groups of 4 (50 female, 18 male) aged between 18 and 19 (m=18.2 ; std=0.4). The participants were first and second year students from the University Bachelor of Technology. Participants received a cookie and 0.5 point bonus on their psychology course grade (small enough not to constitute a form of coercion

or pressure, while still serving as an incentive), with an alternative task offered for non-participants, ensuring equitable and ethical compensation. We collected a total of 136 videos (one video for each participant and for each session) with a total duration of 361.5 minutes (6h 1min 30s). Each game session lasted between 8min 43s and 14min 17s (m=10min38s, std=1min2s).

3.3. Data Collection and Recording Setup

The MOSAIC corpus was collected during structured collaborative interactions involving four participants engaged in a cooperative word-association task. The data collection protocol combined a controlled game environment with synchronized multimodal recording.

3.3.1. Collaborative task

Participants played two rounds of a computerized version of *Cross Clues*, a cooperative board game designed by *Blue Orange Games*. In this game, players have to make associations between words to score a maximum number of points during a restricted time.

The game board is composed of a 4x4 grid in which each column and each row is associated with a unique word. Each cell is thus associated with a pair of two words (w_c, w_r). At the beginning of the game, each player draws a card from the deck of 16 cards, each one designating the coordinates of a cell (e.g. "B3"), and keeps its content secret. Each player starts to think about a clue word linked to both words associated with the coordinates on their card. For instance, if column B is associated to the work "banana" and line 3 with the word "tissue", the player must find a single clue word that would suggest to the others these two words, so that they can guess the correct coordinate B3. At any time, any player can take the floor to suggest a clue word, taking the role of *suggester* during this turn. The 3 other players take the role of *guessers* during this turn, and have to guess the coordinates on the *suggester*'s card. The *suggester* can only announce their clue word to the other players and cannot provide any other information. The *guessers* discuss to agree on a proposition of coordinates, then select the cell corresponding to their choice on the touchscreen. The *suggester* can then validate or invalidate their guess. If the proposition is correct, the card is placed on the grid. If it is incorrect, the card is discarded without revealing it. A new turn then begins. The roles can either change, if a different player from the previous turn decides to suggest a clue word, or stay the same if the same player decides to suggest another clue word. The session ends when all the 16 coordinates cards have been played (placed on the grid or discarded), or when the time has elapsed. The task structure induces repeated cycles of clue production, collaborative reasoning, agreement negotiation, and feedback, creating rich multiparty interaction data.

The task was selected to elicit spontaneous yet structured multiparty interaction involving turn-taking, joint decision-making, and role alternation. The game's mechanics (time pressure, iterative turns, and role-based interactions) encourage participants to adapt their communication, offer support, and resolve ambiguities collaboratively. Success depends on effective communication, mutual understanding, and coordinated effort, as one participant (the *suggester*) provides clue words while the others (the *guessers*) collectively interpret and agree on the correct coordinates. This dynamic also requires active listening, perspective-taking, and consensus-building, as participants must align their interpretations and justify their reasoning to reach a shared decision. These interactions can also requires trust, cooperation, and emotional engagement between participants, reflected by reactions to incorrect proposition, celebration of success, and role adjustments. Additionally, the cooperative structure fosters shared emotional experiences, like tension, humor, or collective achievement, which strengthen interpersonal bonds.

3.3.2. Recording Environment

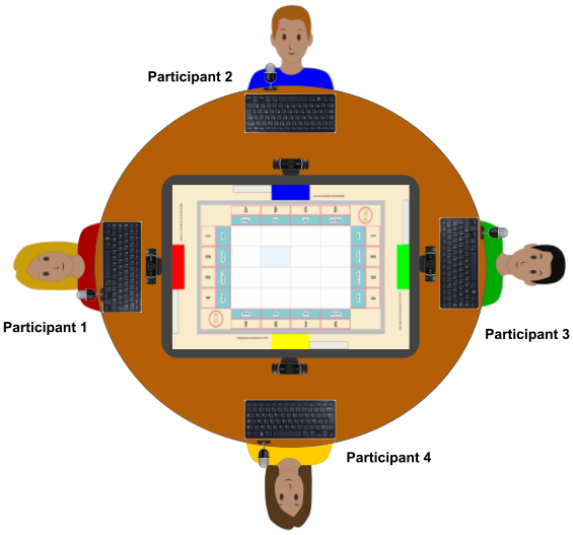

Figure 1: Illustration of the experimental setting.

An illustration of the setting is displayed in Figure 1. Four participants are seated around a round table, with a tactile screen in the center of the table. Each participant is assigned a color (red, green, blue or yellow), corresponding to a specific seat around the table. They are given a badge holder containing a card of the color assigned to them, which they must wear throughout the experiment.

A microphone (Lavalier II) is then attached to the necklace of the badge holder of each participant to capture their speech (AAC-LC 48 kHz - 162 kb/s). Four identical cameras (Logitech C922 Pro) are placed around the screen to capture each participant's image during the game from the opposite side (H.264 720p 60fps - 2500kb/s). Four keyboards are also placed near each participant. A video of the content of the screen is recorded using OBS Studio.

3.3.3. Experimental procedure

This study received an approval of the ethics committee of the Paris Saclay University (CER-Paris-Saclay-2024-100). Participants were recruited in groups of four and completed the experiment in a single session.

After providing informed consent, participants received instructions about the game rules and interface. A 5-minute familiarization round was conducted to ensure understanding of the mechanics.

Each group then played two consecutive game sessions of approximately 10 minutes each. After each session, participants completed a set of questionnaires measuring their perception of the interaction

3.4. Multimodal Data Streams

The corpus contains synchronized audio, video, game logs, gaze annotations, facial action unit activations, and speech transcriptions.

3.4.1. Game logs and annotation scheme

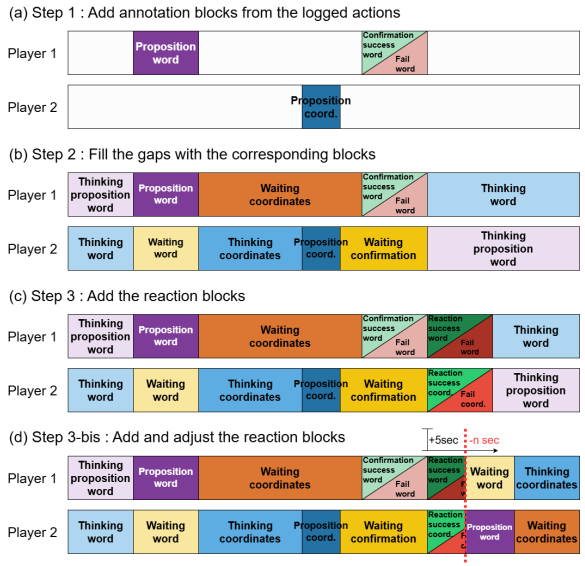

Figure 2: Annotation process of game phases from game logs.

We defined a role-based annotation scheme to automatically derive interaction phases from game logs. Game interface actions were logged with timestamps, including: clue word submission,coordinate proposition,confirmation of success or failure. Based on these logged events, we generated annotation blocks defined by a start time, end time, and categorical label. We distinguish actions made by the *suggester* and the ones made by the *guessers*. For the *suggester* role, five interaction phases are defined : thinking about a word, suggesting a word, waiting the coordinates proposition of the *guessers*, confirming the proposition (success or fail) and the reaction to the result (success or fail). For the *guessers*, six interaction phases are defined : thinking about a word, waiting for the *suggester* to give the word, thinking about possible coordinates, proposition of coordinates, waiting the confirmation from the *suggester* and finally the reaction to the result (success or fail). We also have one category associated to the end of the session and one to a participant canceling a proposition. The annotation scheme is thus composed of a total of 16 annotation categories, with 7 categories for the *suggester* role, 7 categories for the *guesser* role and 2 unrelated to the role (End session and Cancel proposition). For higher-level analyses, categories can be grouped into five broader process types (Collaboration, Observation, Thinking, Reaction, and Game-related actions). Their distribution across the corpus is presented in Section 4. The annotation process follows three steps: (1) Creation of base blocks from explicit logged actions (2) Temporal inference of intermediate phases (e.g., thinking, waiting) (3) Addition of reaction phases with a maximum duration of 5 seconds (empirically determined).

The process is illustrated on Figure 2, with a simplified example of two players' timelines. Using this annotation process, we annotated a total of 9168 blocks.

3.4.2. Performance scores

Using the logs of the game, we provide performance measures for each game session and for the combined session, from both group-level (number of successful and failed attempts) and individual-level, such as number of total propositions (fail or success) as *guesser* (coordinates) and *suggester* (clue-word).

3.4.3. Facial Action Unit Extraction

Facial behavior signals were automatically extracted using OpenFace (Baltrusaitis et al., 2018), a widely used toolkit for facial behavior analysis. OpenFace can, among others, detect and extract the intensity (value between 0.0 and 5.0) of 17 ac-

tion units (AU) on a human face image or video. We used this tool to work on 13 facial behaviors. Based on our expertise with OpenFace data, we provide, in addition to raw data, filtered and preprocessed data suitable for analysis. We excluded the AU 45 (blink), as we do not focus on blinks in our analysis, the AUs 25 and 26 (lips part and jaw drop), because it is often confused with speaking, and AU 14, because it has been shown that its detection is unreliable (Namba et al., 2021). Additionally, we used the AUs 6 and 12 together, to detect smiles, and the AUs 1 and 2, both separated, to distinguish inner brows raising from outer brows raising, and together, to detect brows raised from both sides.

OpenFace also provides a confidence value, that indicates how much confident (value between 0 and 1) the system is about the detection of the face. To minimize false detections and remove noise, we ignore all detections with a confidence lower than 0.8, .

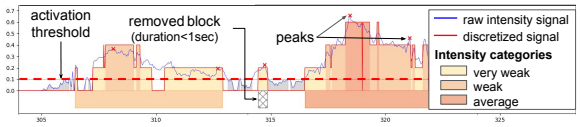

Figure 3: Illustration of the extraction process of Action Units (AUs). The raw signal from Openface (plain blue line) is first filtered using an activation threshold (dashed red line), to remove noise (grey areas). The peaks (red crosses) and their associated time boundaries (not represented to avoid cluttering the figure) are identified. The signal is then discretized (plain red line), and associated with intensity categories. Successive blocks are then merged in a single block with the maximum intensity of the merged blocks. Finally, blocks shorter than 1 second are removed (grey hatched blocks).

The processing pipeline consisted of: : (1) Normalization of raw AU intensity signals, (2) Peak detection using SciPy (minimum height = 0.05; minimum inter-peak distance = 80 frames), (3) Discretization of intensity values into five categories (Very weak [0.1-0.3[, Weak [0.3-0.5[, Average [0.5-0.7[, Strong [0.7-0.9[, Very strong[0.9-1]), (4) Conversion into temporal annotation blocks defined by sustained activation at the same intensity level, (5) Merging of successive blocks separated by gaps below 1 second, (6) Removal of blocks shorter than 1 second.

We illustrate this process on Figure 3, with a 20 seconds signal of a single AU, extracted from our corpus.

3.4.4. Gaze targets

Gaze vector estimation refers to the task of extracting gaze direction vectors from a video or an image of a human face. SpaTaTrack [1] (Spatial gaze Target Tracking) is a tool that uses the output from a camera to extract gaze direction vectors and identify the gaze targets of a human in a video. It allows to define the scene configuration by placing the different objects using a global coordinates system, and project the gaze vectors in this scene to identify the target of the gaze. A visual interface allows manual inspection and validation before extracting the targets in batch.

We used the SpaTaTrack tool configured with our experimental setup, and extracted for each participant and for each frame the gaze targets, which can take one of the following values :

- screen : when the target is either the table or the screen,
- red, blue, green or yellow : when the target is one of the participants,
- other : when the gaze vector does not intersect with any of the defined objects or zones.

We extracted a total of 35567 gaze target blocks, with 41.4% screen, 37.2% participant (red, blue, green or yellow) and 21.4% other.

3.4.5. Verbal transcriptions

Due to the close proximity of participants, individual audio tracks were not acoustically isolated, which compromised the performance of standard automatic transcription, specifically during crosstalk. To address this challenge, we used the HappyScribe service[2], which provided time-stamped annotations and diarization performed by professional linguists experienced in transcription tasks. The annotation process was conducted offline, allowing annotators to carefully review the audio. The service includes quality-control procedures to ensure the accuracy and reliability of transcriptions. While the task was generally straightforward, it became more challenging during segments where more than two participants spoke simultaneously. Each speech segment was divided into block units, defined as continuous utterances separated by pauses of at least 300 milliseconds. Each block is associated with a start time, end time, speaker identifier, and textual content. All transcripts are temporally aligned with the corresponding audio, video, and interaction phase annotations through shared timestamps.

3.4.6. Social perception measures

In addition to multimodal behavioral recordings, the corpus includes session level subjective measures

[1] https://github.com/mlamine21/SpaTaTrack

[2] More information is available at: https://www.happyscribe.com

collected after each game round. The following questionnaires were administrated :

Group Cohesion Questionnaire (GCQ) (Sassier-Roublin et al., 2026) : Measure of individual perception of group cohesion according to three dimensions : Social (13 items), Task (3 items) and Communication (3 item). Responses are provided on a 7-point Likert scale and reflect group-level perception.

Social Belonging Scale (Richer and Vallerand, 1996) : Measures perceived belongingness and acceptance within the group. Responses are provided on a 7-point Likert scale along 5 items for the Acceptation dimension, and 5 items for the Intimacy dimension. Scores reflect individual-level perception.

Virtual Intimacy Scale (VIS) (Potdevin, 2020). Measures perceived intimacy in social interaction. Two versions were administered: (1) a self-report (Self-VIS), in which participants rated their own intimacy, (2) and an other-report (Front-VIS), in which they evaluated the intimacy of the person facing them during the game. The use of these two complementary measures was intended to capture both their mutual influence and potential incongruence between self and partner perceptions. Responses are provided on a continuous 0–100 scale across three dimensions : honesty and genuineness (5 items), positivity (4 items) and mutual comprehension (6 items).

Pre-existing relationship : Participants rated how well they knew each other participant on a 5-point scale (1 = Not at all, 5 = Very well).

4. Results

4.1. Dataset characteristics

4.1.1. Annotation categories distribution

Interaction phase annotations were grouped into five higher-level process types:

Collaboration (53.2% of the blocks): This group concerns both the *suggester* ("Waiting coordinates") who can give clues using their non-verbal behavior to give indirect feedback to *guessers* during their thinking process, and the *guessers* ("Thinking coordinates"), who share in a more direct way their thinking process and their opinion to the other players, based on cues such as interruptions, overlapping speech, attention allocation, agreement, disagreement or doubt expression.

Observation (16.9%): moments when *guessers* observe the *suggester* writing a word ("Waiting word") and gather information about how the *suggester* feel about their word: "do they look confident? are they happy about their proposition?" or observe both the *suggester* and the other *guessers* reacting about their coordinates proposition ("Waiting confirmation").

Thinking (15.3%): as a mental processes before a word is suggested, indicating a reflexion process: "how can I link these words? Are there other possible coordinates that could confuse them?". Participants can either emphasize that they are about to suggest a word ("Thinking proposition word" or "Thinking word"), or that they probably won't suggest a word because they can't find an easy link between the words associated to their coordinates ("Thinking word").

Reactions (8.8%): emotional expression of the *suggester* right before they confirm coordinates ("Confirmation success/fail") and both *suggester* and *guessers* the moment after the answer is revealed ("Reaction success/fail coordinates/word").

Game-related (5.4%): mostly unrelated with group or interpersonal processes, refers to moments where the *suggester* is writing their word on the keyboard ("Proposition word"), when the *guessers* are clicking on the coordinates they chose ("Proposition coordinates"), the "End session" and the "Cancel proposition" categories.

Their distribution across the corpus (percentage of annotation blocks) and the categories they contain is illustrated in Figure 4.

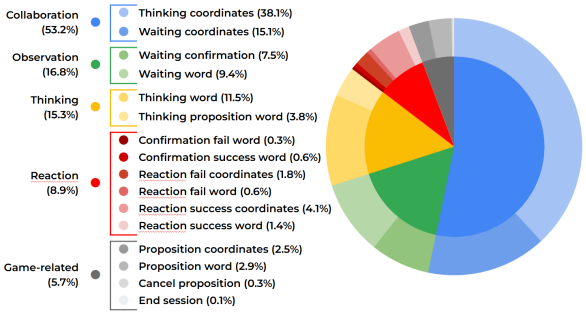

Figure 4: Distribution of interaction phase categories across the corpus.

4.1.2. Facial action unit distribution

A total of 27,648 facial activation blocks were extracted across the corpus. The facial behaviors we used can be found on the pie charts (Figure 5), with percentages indicating their respective proportions.

4.1.3. Gaze target distribution

A total of 35567 gaze target blocks were extracted across the corpus. Gaze was distributed toward the screen in 41.4% of annotated blocks, toward another participant (red, blue, green or yellow) in 37.2% of the blocks and categorized as other in

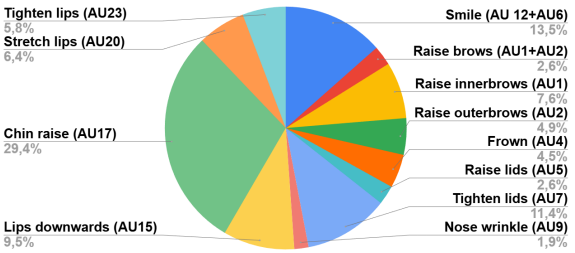

Figure 5: Distribution of extracted facial action units across the corpus.

21.4% of the cases. This distribution reflects the relative proportion of attention allocated to the shared task interface, co-participants, and non-defined targets.

4.1.4. Speech statistics

The corpus is composed of 11963 blocks, with a total speech duration of 3 hours, 33 minutes, 26.74 seconds (12806.74 seconds), and a word count of 44565. The average block duration is 1.07 second, with an average of 3.73 words per block.

4.1.5. Social perception measures

Cohesion: The overall cohesion score (Total dimension) indicates a high level of perceived cohesion across all participants (M = 5.97, SD = 0.84). Cohesion was particularly high along the Task dimension (M = 6.50, SD = 0.70) and the Communication dimension (M = 6.22, SD = 0.76) with lower variability compared to the Social dimension (M = 5.74, SD = 1.10).
Intimacy: Participants reported a high overall level (Total dimension) of intimacy for both the self-report (M = 78.79, SD = 13.04) and the front-report (M = 78.80, SD = 13.75). The Positivity dimension received the highest scores (Self : M = 83.93, SD = 13.02, Front : M = 84.43, S = 12.62), followed by the Mutual Comprehension dimension (Self : M = 79.02, SD = 15.95, Front : M = 78.50, SD = 17.27), while the Honesty dimension was rated comparatively low, yet still high (Self : M = 73.34, SD = 15.61, Front : M = 73.60, SD = 16.04).
Correlations: We computed correlation coefficients (Spearman when the data did not meet the assumption of normality, Pearson otherwise) to investigate how the components of intimacy and cohesion interact. Concerning the self-reported intimacy, the analysis revealed significant associations between all dimensions of both cohesion and intimacy. However, we can note substantial differences in the strength of these associations, with strong correlations with the social dimension of the cohesion (Honesty(H) : r = 0.520, Positivity(P) : r = 0.620, Comprehension(C) : r = 0.607; p $<$0.001 for all dimensions), followed by the communication dimension with moderate associations (H : r = 0.322, P : r = 0.414, C : r = 0.442; p $<$0.001 for all dimensions) and finally weak associations for the Task dimension (H : r = 0.235, P : r = 0.289, C : r = 0.273; p $<$0.001 for all dimensions). We observed a similar pattern with the front-reported intimacy, but with weaker associations for the Social dimension (H : r = 0.357, P : r = 0.474, C : r = 0.509; p $<$0.001 for all dimensions), the Communication dimension (H : r = 0.190 p $<$0.05, P : r = 0.314 p $<$0.001, C : r = 0.392 p $<$0.001) and even non significant for the Task dimension (H : r = 0.091 p $<$ 0.15, P : r = 0.208 p $<$0.01, C : r = 0.300 p $<$0.001).

4.2. Evolution of subjective measures across sessions

4.2.1. Statistical procedure

We analyzed the evolution of the scores across the 2 game sessions to investigate the dynamics created throughout the experiment, using one-sided Wilcoxon signed-rank tests when the data distributions did not meet the assumption of normality, paired t-test otherwise, with the 'less' alternate hypothesis (Session 1 $<$ Session 2), indicating a hypothesized increase after Session 2.

As the concepts studied (cohesion and intimacy) are dependent on existing links between the participants, we used the Links scale (see Section 3.4.6) to calculate a mean score (scale from 1 to 5) for each participant based on the strength of their existing relationship with the other three players. Participants were then divided into two groups: the Unfamiliar group, who reported not knowing the other participants well (mean score $<$ 3, n=21 , M = 2.16 SD = 0.54) and the Familiar group, who reported knowing them well (mean score $\geq$ 3, n=47, M = 3.77 SD = 0.58).

4.2.2. Evolution across sessions

Cohesion: Across all participants, significant increase was observed for total cohesion (p = 3.50e-02) score, reflecting a global increase in the perception of the cohesion after the second session. Social cohesion also exhibited a significant increase (p = 3.81e-02), but for the Task (p = 5.56e-02) and Communication (p = 5.53e-01) dimensions, the differences were not statistically significant, while still showing a tendency to increase for the Task dimension.
Intimacy: Perceived intimacy increased significantly across sessions for both self-report and front-report when considering all participants (p$<$0.001 for all dimensions).

4.2.3. Effect of Links

Cohesion: The participants in the Unfamiliar group registered a significant increase in the cohesion score, for the Social dimension ($p < 0.001$), the Task dimension (p = 4.51e-02) and Total cohesion ($p < 0.01$), while the Communication dimension did not show a significant increase (p = 2.85e-01).

For the Familiar group, the perceived cohesion did not show a significant increase for any of the dimensions across the sessions (Social : p = 5.50e-01, Task : p = 2.80e-01, Communication : p = 7.02e-01, Total : p = 5.48e-01).

Intimacy: For the Unfamiliar group, overall intimacy showed a significant increase (self and front : $p < 0.01$) as well as the Honesty dimension (self : $p < 0.01$, front : $p < 0.001$) and the Positivity dimension (self : $p < 0.01$, front : $p < 0.05$). No significant increase was observed for the self-reported Comprehension dimension (p=1.52e-01), whereas significant increase emerged in the front-report for the same dimension (p=3.98e-02).

For the familiar group, overall intimacy and all dimensions showed a significant increase ($p < 0.001$) for both the self and front report.

5. Discussion

Cohesion: The participants perceived high levels of cohesion along the Task and Communication dimensions. No significant evolution across sessions were observed for these dimensions. The social cohesion scores, while still above average, were relatively lower than the other dimensions, but showed a significant increase across sessions. According to the literature, cohesion can build around the shared objective of the group, later evolving and deepening through the social dimension, reinforcing bonds within the group, encouraging vulnerability exposure and risk taking to achieve better results (Forsyth and Burnette, 2010; Marmarosh and Sproul, 2021). These results could suggest an already established cohesion on the Task and Communication level, leading participants to reinforce and focus on social cohesion. This initial high level of task and communication cohesion, specifically in Familiar groups, could be due to participants having already interacted together in a collaborative context (e.g. group works in university setting) and possibly already sharing effective task and communication-level collaboration strategies.

The results also showed that Social and Task cohesion increased for Unfamiliar groups but not for Familiar groups. The Communication dimension seems to remain stable across the sessions, even if we can observe a slightly higher increase in the Unfamiliar group. This reveals an interesting impact of familiarity on cohesion, suggesting that building cohesion plays a more critical role among unfamiliar participants. As stated by (Ogrodniczuk et al., 2021), cohesion can move from initial uncertainty, starting from members getting to know one another and build trust, to deeper collaboration at a later stage, where cohesion gradually develops. This indicates that familiarity may impact the need for social bonding to enhance cohesion, as established relationships already reached a baseline level of cohesion, highlighting the importance of social skills in emerging groups. Moreover, these results indicate that the experimental setting allowed the observation of meaningful changes in perceived cohesion across sessions, providing a relevant framework for studying cohesion, specifically through the social dimension, and its dynamics in collaborative interactions. However, the high overall scores is a limitation of this dataset, as it makes it difficult to identify possible indicators associated to low cohesion scores.

Intimacy: The intimacy scores from both self and front report revealed high level of perceived intimacy. According to (Kozlowski and Chao, 2012), first impressions guide initial group interactions, often relying on visible traits such as race, age, and gender. They also suggest that groups with similar surface-level characteristics may experience stronger initial attraction among members. The high initial level of overall intimacy could then be partly explained by the shared characteristics (age, social class, education level, high semantic alignment and shared reference) of the participants, specifically for the Mutual Comprehension dimension scores. The collaborative game context, with low (for competitive participants who seek performance) to non existent stakes, probably helped in creating a positive and relaxed atmosphere, and could explain the Positivity dimension displaying the highest scores among the dimensions. The results should then be taken carefully by considering the specificities of our sample.

The perception of intimacy increased significantly across sessions, and familiarity did not seem to have an impact on intimacy evolution, which displayed a significant increase for all dimensions except for the self reported intimacy for the Mutual comprehension dimension in the unfamiliar group, which remained stable. This suggests that while participants did not perceived a change in their own attempt to enhance Mutual understanding, they perceived more easily this change in their interaction partner when social cohesion increased. Additionally, we can conclude that while cohesion seems to reach a peak in cohesion level, intimacy keeps increasing, potentially helping reach a higher peak of cohesion after longer period of collaborative interactions.

Correlations: The first conclusion suggested by

the results is that intimacy has few impact on the perception of cohesion along the Task dimension. For the Communication and Social dimensions of cohesion, results seem to indicate a strong association with both Positivity and Comprehension dimensions of intimacy. The results for the Honesty dimension appear more mitigated, and do not allow to draw conclusions on its link with the perception of cohesion.

6. Limitations

While the MOSAIC corpus provides valuable insights into multiparty interactions, several limitations must be acknowledged. First, the small sample size (68 participants) and demographically narrow population (18–19-year-old university students) restrict the generalizability of the findings to broader age groups or diverse cultural and social contexts. Additionally, the high and homogeneous scores for intimacy and cohesion across participants limit the dataset's ability to capture contrastive dynamics in terms of intimacy and cohesion. Finally, the automated extraction of gaze targets and facial expressions lacks human validation.

7. Perspectives and use cases

We introduced the MOSAIC corpus, an annotated video dataset aimed at exploring the links between cohesion, intimacy, and synchronized multimodal behaviors (action units, gaze targets and speech content) associated to these concepts. We described the segmentation used on the corpus based on the unfolding phases of the collaborative task, and suggested a grouping based on high-level processes, allowing the interaction to be studied from different angles. The collaboration interaction phases can be used to study group dynamics while focusing on decision-making or negotiation, and the observation and thinking phases can help gain insights on mental processes and their expression, and the reaction phases could be useful to explore emotional expressions in a collaborative context.

In the first analysis of questionnaires we provided, we focused on individual perceptions of cohesion and intimacy. This work could be extended to group-level analysis to explore for example disparity and homogeneity in the group's perception, or to examine how these constructs could be related to the group's performance in terms of success and communication quality. Additionally, we could further explore the interactions between the different measures by taking into account familiarity in the correlation analysis, and including the results for the social belonging scale.

This corpus could extends the contributions on the links between multimodal behaviors and high levels of intimacy (Shaver and Reis, 1988; Soleymani et al., 2019; Potdevin, 2020; Kang et al., 2012) and on the identification of markers of high levels of cohesion (Maman, 2022). The observation of an evolution of both cohesion and intimacy could additionally inform on the specific changes related to language and non-verbal behaviors that led to this evolution. Several additional annotations could also be introduced to extend this corpus, such as verbal features like interruptions, laughter, sentiments and intentions associated to utterances, or cross-modal features, like emotional states or communication quality.

Finally, this corpus can be used to inform the design of socially interactive agents displaying intimacy behaviors (verbal and non-verbal) during the same collaborative task. The annotation categories described in Section 3.4.1 and used in this work as an analysis framework also represent the different possible states of an agent interacting in this task. The contextual analysis of the participants behaviors during each of these states will help to identify and then reproduce behavior patterns associated with different intimacy levels on socially interactive agents. These generated behaviors will then be evaluated in interaction with humans to explore their relevance and realism, as well as the impact of different embodiments (humans vs robots or virtual agents) on the perception of cohesion and perceived intimacy.

8. Acknowledgments

This work was supported by a French government grant managed by the Agence Nationale de la Recherche as part of the France 2030 program, reference ANR-22-EXEN-0004 (PEPR eNSEMBLE / PC3 MATCHING).

9. References

Tadas Baltrusaitis, Amir Zadeh, Yao Chong Lim, and Louis Philippe Morency. 2018. OpenFace 2.0: Facial behavior analysis toolkit. In *Proceedings - 13th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2018*, pages 59–66.

Daniel J. Beal, Robin R. Cohen, Michael J. Burke, and Christy L. McLendon. 2003. Cohesion and Performance in Groups: A Meta-Analytic Clarification of Construct Relations. *Journal of Applied Psychology*, 88(6):989–1004.

Timothy Bickmore and Daniel Schulman. 2012. Empirical Validation of an Accommodation Theory-Based Model of User-Agent Relationship. In

David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Yukiko Nakano, Michael Neff, Ana Paiva, and Marilyn Walker, editors, *Intelligent Virtual Agents*, volume 7502, pages 390–403. Springer Berlin Heidelberg, Berlin, Heidelberg. Series Title: Lecture Notes in Computer Science.

Tanja Bänziger, Marcello Mortillaro, and Klaus R. Scherer. 2012. Introducing the Geneva Multimodal expression corpus for experimental research on emotion perception. *Emotion*, 12(5):1161–1179.

Jean Carletta, Simone Ashby, Sebastien Bourban, Mike Flynn, Mael Guillemot, Thomas Hain, Jaroslav Kadlec, Vasilis Karaiskos, Wessel Kraaij, Melissa Kronenthal, Guillaume Lathoud, Mike Lincoln, Agnes Lisowska, Iain McCowan, Wilfried Post, Dennis Reidsma, and Pierre Wellner. 2006. The AMI Meeting Corpus: A Pre-announcement. In Steve Renals and Samy Bengio, editors, *Machine Learning for Multimodal Interaction*, volume 3869, pages 28–39. Springer Berlin Heidelberg, Berlin, Heidelberg. Series Title: Lecture Notes in Computer Science.

Donelson R. Forsyth and Jeni Burnette. 2010. Group Processes. In *Advanced Social Psychology: The State of the Science*, pages 495–534. Oxford : Oxford University Press.

Sarah Gillet, Marynel Vázquez, Christopher Peters, Fangkai Yang, and Iolanda Leite. 2022. Multiparty Interaction Between Humans and Socially Interactive Agents. In *The Handbook on Socially Interactive Agents: 20 years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 2: Interactivity, Platforms, Application*, 1 edition, volume 48, pages 113–154. Association for Computing Machinery, New York, NY, USA.

Sin-Hwa Kang, Jonathan Gratch, Candy Sidner, Ron Artstein, Lixing Huang, and Louis-Philippe Morency. 2012. Towards building a virtual counselor: modeling nonverbal behavior during intimate self-disclosure. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*, pages 63–70.

Gaoussou Youssouf Kebe, Mehmet Deniz Birlikci, Auriane Boudin, Ryo Ishii, Jeffrey M. Girard, and Louis-Philippe Morency. 2024. GeSTICS: A Multimodal Corpus for Studying Gesture Synthesis in Two-party Interactions with Contextualized Speech. In *Proceedings of the ACM International Conference on Intelligent Virtual Agents*, pages 1–10, GLASGOW United Kingdom. ACM.

Maria Koutsombogera and Carl Vogel. 2018. Modeling collaborative multimodal behavior in group dialogues: The MULTISIMO corpus. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

Steve W. J. Kozlowski and Georgia T. Chao. 2012. The Dynamics of Emergence: Cognition and Cohesion in Work Teams. *Managerial and Decision Economics*, 33(5-6):335–354. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/mde.2552.

Jean-Philippe Laurenceau, Lisa Barrett, and Paula Pietromonaco. 1998. Intimacy as an interpersonal process: the importance of self-disclosure, partner disclosure, and perceived partner responsiveness in interpersonal exchanges. *Journal of Personality and Social Psychology*, 74:1238–51.

SeoYoung Lee and Junho Choi. 2017. Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity. *International Journal of Human-Computer Studies*, 103:95–105.

Lucien Maman. 2022. *Automated analysis of cohesion in small groups interactions*. Ph.D. thesis, Institut Polytechnique de Paris, Paris.

Lucien Maman, Eleonora Ceccaldi, Nale Lehmann-Willenbrock, Laurence Likforman-Sulem, Mohamed Chetouani, Gualtiero Volpe, and Giovanna Varni. 2020. GAME-ON: a multimodal dataset for cohesion and group analysis. *IEEE Access*, 8:124185–124203.

Cheri L. Marmarosh and Amy Sproul. 2021. Group cohesion: Empirical evidence from group psychotherapy for those studying other areas of group work. In *The psychology of groups: The intersection of social psychology and psychotherapy research*, pages 169–189. American Psychological Association, Washington, DC, US.

Alexander Mikalachki. 1969. *Group cohesion reconsidered : a study of blue collar work groups*. University of Western Ontario School of Business Administration.

Shushi Namba, Wataru Sato, Masaki Osumi, and Koh Shimokawa. 2021. Assessing automated facial action unit detection systems for analyzing cross-domain facial expression databases. *Sensors*, 21(12):4222.

John S. Ogrodniczuk, Joanna Cheek, and David Kealy. 2021. Group therapy development: Implications for nontherapy groups. In *The psychology*

of groups: The intersection of social psychology and psychotherapy research, pages 231–248. American Psychological Association, Washington, DC, US.

Delphine Potdevin. 2020. *Vers des agents conversationnels animés sociaux: Quelle influence de l'intimité virtuelle sur l'expérience utilisateur et la relation-client?* PhD Thesis, Université Paris-Saclay.

Karen J. Prager. 2000. Intimacy in personal relationships. In *Close relationships: A sourcebook*, pages 228–243. SAGE Publications, Inc.

Harry T Reis. 2017. The interpersonal process model of intimacy: Maintaining intimacy through self-disclosure and responsiveness. *In J. Fitzgerald (Ed.), Foundations for couples' therapy: Research for the real world*, pages 216—-225.

Sylvie F. Richer and Robert J. Vallerand. 1996. l'Echelle du sentiment d'appartenance sociale (ESAS). *Manuscrit inédit, Laboratoire de recherche sur le comportement social, Université du Québec à Montréal*.

Eduardo Salas, Rebecca Grossman, Ashley M. Hughes, and Chris W. Coultas. 2015. Measuring Team Cohesion: Observations from the Science. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 57(3):365–374.

Dairazalia Sanchez-Cortes, Oya Aran, and Daniel Gatica-Perez. 2011. An audio visual corpus for emergent leader analysis. In *Workshop on multimodal corpora for machine learning: taking stock and road mapping the future, ICMI-MLMI*.

Mathilde Sassier-Roublin, Amine Benamara, Céline Clavel, Julien Saunier, and Alexandre Pauchet. 2026. Measuring Group Cohesion: Development and Validation of the Group Cohesion Questionnaire (GCQ). Working paper or preprint.

Sarah Sebo, Brett Stoll, Brian Scassellati, and Malte F. Jung. 2020. Robots in groups and teams: A literature review. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2):1–36.

R. d Shaver and P. Reis. 1988. *Intimacy as an interpersonal process*. USA.

Mohammad Soleymani, Kalin Stefanov, Sin-Hwa Kang, Jan Ondras, and Jonathan Gratch. 2019. Multimodal Analysis and Estimation of Intimate Self-Disclosure. In *2019 International Conference on Multimodal Interaction*, pages 59–68, Suzhou China. ACM.

Margaret L. Traeger, Sarah Strohkorb Sebo, Malte Jung, Brian Scassellati, and Nicholas A. Christakis. 2020. Vulnerable robots positively shape human conversational dynamics in a human–robot team. *Proceedings of the National Academy of Sciences*, 117(12):6370–6375.