Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 498–507
Marseille, 11–16 May 2020

Multimodal Analysis of Cohesion in Multi-party Interactions

Reshmashree B. Kantharaju[1], Caroline Langlet[2], Mukesh Barange[1], Chloé Clavel[2], Catherine Pelachaud[1]
[1] CNRS-ISIR, Sorbonne Université, [2] LTCI, Télécom Paris-Tech
Paris, France
{bangalore_ kantharaju, barange, catherine.pelachaud}@isir.upmc.fr
{caroline.langlet, chloe.clavel}@telecom-paris.fr

Abstract

Group cohesion is an emergent phenomenon that describes the tendency of the group members' shared commitment to group tasks and the interpersonal attraction among them. This paper presents a multimodal analysis of group cohesion using a corpus of multi-party interactions. We utilize 16 two-minute segments annotated with cohesion from the AMI corpus. We define three layers of modalities: *non-verbal social cues*, *dialogue acts* and *interruptions*. The initial analysis is performed at the individual level and later, we combine the different modalities to observe their impact on perceived level of cohesion. Results indicate that occurrence of *laughter* and *interruption* are higher in high cohesive segments. We also observe that, *dialogue acts* and *head nods* did not have an impact on the level of cohesion by itself. However, when combined there was an impact on the perceived level of cohesion. Overall, the analysis shows that multimodal cues are crucial for accurate analysis of group cohesion.

Keywords: Cohesion, Dialogue acts, Non-verbal behaviours, Interruptions

1. Introduction

Group conversation is a prominent form of human communication. Often, humans discuss, make decisions and exchange ideas in groups, through different settings (e. g., meeting, conference, council, party etc.). Literature in sociology and psychology have studied the various aspects of group dynamics i. e., the action, process and changes that occur within the group (Forsyth, 2018). While research questions concerning human behaviour in groups are manifold, in this research work we focus on group cohesion.

Cohesion describes the tendency of group members' shared bond or attraction that drives the members to stay together and to want to work together (Casey-Campbell and Martens, 2009). A cohesive group can be defined as a group that sticks together and is accompanied by feelings of solidarity, harmony and commitment (Mudrack, 1989). It is a group phenomenon that emerges over time in teams (Santoro et al., 2015). Several existing works in literature have associated group cohesion with group performance, team satisfaction and adherence (Beal et al., 2003).

Automatic estimation of cohesion can be useful for multimedia tagging and automatic analysis of meeting data. This information can be useful to measure the performance of teams. This article is a first step towards developing a computational model of cohesion estimation in multi-party human-human and human-agent interactions. In order to do this, we need to consider several factors on the higher level i. e., turn strategies, dialogue acts and on the lower level i. e., non-verbal behaviours. For the ease of reading, in our paper we will refer to the low level behaviours e. g., gaze, head nods and laughter as non-verbal social cues. This paper provides a preliminary analysis of how these low and high level multimodal behaviours are linked to the group cohesion in a corpus of human-human interactions. Our goal is to highlight the most relevant features of group cohesion.

In multi-party interactions, humans communicate and coordinate with each other via a number of verbal and nonverbal behaviours. They take *turns* and these *turns* mostly begin and end smoothly, with short lapses of time between them. However, this is not always the case since there are overlaps, interruptions and silences (Schegloff, 2000). Literature on cohesion estimation has shown a strong correlation between cohesion and interruption. In this paper, we define three layers of modalities: *non-verbal social cues*, *dialogue acts* and *interruptions*. Each layer is first analysed individually to assess their impact on the perceived level of cohesion. Then we observe how the certain behaviours from these three layers affect the perceived level of cohesion when combined.

In Section 2., we present group cohesion from a psychological perspective, and the communicative behaviours that can be associated to it from a dialogue perspective. In Section 2.4., we describe our three layer approach. Then, in Section 3., we present the data utilised for the analysis and the relevant annotations. Section 4., presents the results and discussion of the analysis of the three layers individually. And finally, Section 5., provides an analysis of the specific behaviours combined, and the relation between them and the level of cohesion.

2. Background and Related Work

This section presents cohesion from a theoretical perspective, dialogue perspective and the related work on automatic cohesion estimation.

2.1. Cohesion

Several definitions of cohesion have been presented in specific contexts such as sports team (Carron and Chelladurai, 1981) and group psychotherapy (Braaten, 1991). One of the earliest definitions of cohesion was proposed by Festinger et. al., "as the total field of forces that act on members to remain in the group" (Festinger et al., 1950). Several other researchers provided definitions that included "attractiveness to the group" (Back, 1951) or "commitment to the group" (Piper et al., 1983) or "commitment of members to group task" (Goodman et al., 1987). However, these definitions present cohesion as a uni-dimensional construct. Carron et. al., defined cohesion as "a dynamic process that is reflected in tendency of group to stick together and remain united in pursuit of its goals and objectives" (Carron, 1982) that presented it as a multi-dimensional construct. A multi-dimensional model was proposed: group-individual and task-social (Carron et al., 1985). The group-individual distinction recognizes that cohesion results from both a member's desire to remain part of the group as a unit (group integration, GI) and from a member's personal attraction toward being a group member (interpersonal attraction to the group, ATG). The task-social distinction reflects the perceived task and social aspects of the group. Social cohesion can be defined as the interpersonal attraction among members and task cohesion can be defined as the degree to which group members work together to achieve common goals and objectives. In total, four dimensions i. e., ATG-task, ATG-social, GI-task and GI-social were recognised. Braaten proposed a five-factor model for group cohesion in group psychotherapy: attraction and bonding, support and caring, listening and empathy, self-disclosure and feedback, process performance and goal attainment (Braaten, 1991). Another model was proposed by Carless and De Paola (Carless and De Paola, 2000) which is a three-factor model with task cohesion, social cohesion and attraction to group. An observation of the existing models and definitions helps identify two constructs of cohesion i. e., attraction to the group or interpersonal attraction (analogous with social cohesion) and commitment to the task (analogous with task cohesion).

2.2. Cohesion from Dialogue Perspective

For analysing cohesion from a dialogue perspective, we need to look at the behaviours that express the interpersonal attraction of the locutors in the group. However, in linguistic studies, the concept of cohesion is not related to group cohesion, but to the cohesion of the discourse itself. In (Taboada, 2004), the author describes linguistic cohesion as an occurrence "when the interpretation of some element in the discourse depends on the interpretation of another one". A discourse is cohesive if it functions as a unity. Verbal cohesion is realized through relation between parts of the discourse such as relations of coreferentiality or similarity. Thus, from a linguistic point of view, cohesion describes how different parts of discourse are linked to each other and how they build a cohesive and meaningful unit. To understand how locutors interact in a cohesive way, we have to look at dialogue studies. Dialogue studies describe *dialogue* as a joint activity, a task performed in collaboration (Mills, 2014). Cohesion is not explicitly mentioned in these studies, but we hypothesize that some specific communicative behaviours might be related to group cohesion. These communicative behaviours are presented as follows.

Alignment Studies on alignment focus on how locutors adjust their communicative behaviour for either diminishing or enhancing the social and communicative differences. Alignment comprises of several communicative behaviours, both verbal and non-verbal. Regarding verbal alignment, most of the studies investigate "dialog as an imitation-like coordination" and how the alignment of linguistic production can affect the social connection between locutors. Several studies have shown that dialogue participants automatically align their behaviour at different levels i. e., lexical, syntactic and semantic. (Reitter et al., 2006) have shown that locutors reuse lexical as well as syntactic structures from previous utterances. As a natural feature of human-human dialogue (Pickering and Garrod, 2004), verbal alignment has been used in human-machine interaction for improving the communicative skills of the agent (Campano et al., 2015). As alignment is about coordination and social connection, our hypothesis is that it might be a verbal indicator of cohesion between the dialogue participants.

Interpersonal Synergy While alignment focuses on how dialogue participants coordinate by using local turn-by-turn repetition at linguistic level, interpersonal synergy focuses on how dialogue participants complete each other's utterances in order to build a coherent and meaningful conversation (Mills, 2014). Interpersonal synergy deals with "how interdependence between speaker behaviours in conversation relies on complementarity" (Fusaroli et al., 2014). In (Fusaroli et al., 2014), the authors consider that, in a conversation, the pre-existing and locally negotiated procedural scripts or routines make the interlocutors interdependent in their linguistic behaviour. Routines are patterns of behaviours organized at the interaction level and they rely on complementarity dynamics. Complementarity in dialog can occur as the "structured sequences of speech turns, such as adjacency pairs: questions are normally responded to with an answer, not with another question; offers and invitations are usually followed by acceptances or declinations" (Fusaroli et al., 2014). In this study, we aim to focus on verbal phenomena related to interpersonal synergy. As structured sequences of speech turns, like adjacency pairs, rely on complementarity between dialogue participants, we hypothesize that they might give an indication about the level of cohesion.

Act4Team Act4Team is a coding scheme for the annotation of problem-solving group conversations. It focuses on verbal content and relies on both group dynamics and dialogue organization. It aims to underline the problem solving dynamic in the conversation and distinguishes four broad facets of verbal statement in groups: problem-focused statements, procedural statements, socio-emotional statements and action-oriented statements. The Act4team coding scheme has been used for annotating verbal expressions of cohesion by (Nanninga et al., 2017). According to an annotation campaign of the verbal content using the scheme, the authors identify several Act4team categories that are characteristic for social and task cohesion.

Turn Taking and Interruption An effective multi-party interaction depends on the coordination of team members in conversation using turns (Bohus and Horvitz, 2010). In dialogue interaction, turn taking refers to the ability of participants to alternate speaking turns. In multi-party interaction, overlaps can occur where more than one participant intends to speak simultaneously (Heldner and Edlund, 2010). This overlapping can be a characteristic of cooperation (Tannen, 1994) as well as conflict (West and Zimmerman, 2015) in the group. This violation of basic turn-taking rule can result in an interruption, where one speaker disrupts the turn of another with a new utterance. Interruption appears to be more common in multi-party conversations than in dyadic conversations (Beattie, 1981). In a multi-party interaction, participants tend to take turn to speak often since the current speaker can yield the turn to more than one listener. In addition to this, interruption is not limited to the current speaker(as interrupter) and primary addressee (as interruptee) which is a trivial case in dyadic interaction. Other participants can also interrupt the current speaker by addressing one of the listeners (non-speaker) in the group (Pontecorvo et al., 2000; Bangerter et al., 2010) which we refer to as *interruption-other*.

Based on the content, the interruption can be distinguished as either cooperative or disruptive (Li, 2001). Cooperative interruption includes utterances of support, agreement, finishing current speaker's phrase, or asking for clarification. Disruptive interruption includes utterances of showing rejection, topic change, or disagreement. Cafaro *et* al., observed the effects of interruption in dyadic interaction and found that the types of interruption i. e., cooperative and disruptive have an impact on the user's perception of interpersonal attitudes (Cafaro et al., 2016). Several works in literature have studied the perception of interruptions with respect to gender and status of the participants in a group (Beattie, 1981), (Tannen, 1994). Results showed that interruptions are not necessarily a display of dominance in group interactions. Therefore we hypothesize that interruptions have an effect on perceived level of cohesion.

2.3. Automatic Analysis of Cohesion

Several studies in literature have employed various techniques to collect and analyse cohesion data indirectly i. e., not via self-reports. For example, sociometric badges were used to infer cohesion based on temporal proximity, interaction duration and frequency (Zhang et al., 2018). In (Hung and Gatica-Perez, 2010) cohesion estimation and annotation of the level of cohesion as perceived by external observers was presented. Results showed that the best performing feature was the total pause time between each individual's turns and that a strong correlation existed between cohesion level and turn-taking patterns. It also indicated that automatically extracted behavioural cues could be used in estimation of perceived level of cohesion in meetings. In (Fang and Achard, 2018), the relation between cohesion and personality of participants was studied. Results indicated a high correlation between agreeableness (a personality trait) and cohesion. Additionally, speaking turn and variation of speech energy, were shown to be related to cohesion. Wang et. al., categorized cohesiveness of a group into cohesive, divisive, or mixed interactions (Wang et al., 2012). A variety of linguistic phenomena e. g., language use constituents (LUC), discourse markers, disfluencies were utilised. They found that cohesive interactions comprised of agreement and alignment with minor disagreements and other forms of rejection. Inferring cohesion based on content analysis i. e., examining linguistic and paralinguistic mimicry and convergence, in group discussion was presented in (Nanninga et al., 2017). They found that paralinguistic mimicry was useful in estimating social cohesion which is more openly expressed by nonverbal vocal behaviour than task cohesion.

2.4. Our Approach

In this paper, we analyse the relation between verbal, non-verbal behaviours and group cohesion in multi-party interactions. In order to do this, we take a multi-layer approach where each layer corresponds to a behaviour type. We first study each layer separately to understand how particular behaviours are associated with the perception of high and low cohesion. Then, we perform a multi-layer analysis to measure how their combination impacts the perception of cohesion in multi-party interaction. As mentioned earlier, we consider three layers: *non-verbal social cues*, *dialogue acts* and *interruptions*. Since our goal is to provide a computational model of cohesion estimation, our analysis focuses on semi-automatically detectable behaviours that are annotated in multi-party interaction corpora.

Non-verbal Social Cues Non-verbal behavioural cues like gaze, facial expressions, gestures, and body postures etc., indicate the attitude of a given individual in any social situation (Richmond et al., 1991) and convey information about affect, mental state, personality, and other traits (Vinciarelli et al., 2009). While works in literature provide a detailed analysis of the features e. g., prosody, visual energy that measure cohesion, they do not look at social signal cues per se e. g., gaze, head movement. Therefore, for our preliminary study, we focus on gaze behaviour, head nods, facial action units and laughter. Since cohesion is associated with bonding, feedback and support, we hypothesize that behaviours corresponding to these are frequent in highly cohesive segments. We also look at the presence of action unit AU4 i. e., brow lowerer which is often associated with negative emotions e. g., anger, disgust (Ekman, 1997).

Dialogue Acts As explained in the theoretical background (Section 2.), two kinds of interpersonal process in dialogues can be related to group cohesion i. e., *alignment* and *interpersonal synergy*. However, these two processes embed very different behaviours i. e., shared vocabulary, lexical and syntactic repetitions for alignment, and routines and adjacency pairs for interpersonal synergy. As a first step, this study only focuses on the interpersonal synergy and considers dialogue acts as an essential part of interpersonal synergy. It relies on routines and structured sequences of speech turns, as adjacency pairs. Dialogue acts are necessary elements to build such structured sequences. Our analysis exploits a dataset of group interactions and their related dialogue act annotations, which is presented in Section 3.. We choose to rely on the dialogue act annotation of the AMI corpus as the annotation schema used is similar to DIT++ (Bunt, 2011). We think it is more relevant to rely on well-known dialogue categories than Act4Team which is not commonly used in dialogue studies.

Turn Taking and Interruption Turn taking and interruptions are important for effective group interaction. Interruptions are not always dyadic in nature in a group interaction. Literature presented in Section 2. illustrates the effects of turn taking and interruption on group interactions and provides an insight into human behaviours during interactions. However, there are only few studies in the context of group cohesion. Therefore, the objective of this study is to analyse the relation of turn taking and interruption with group cohesion in multi-party interactions. We hypothesize that occurrence of turns, overlaps and interruptions are higher in highly cohesive groups.

3. Dataset

In this section, we present the dataset and the annotations that we utilised for our analysis. The Augmented Multi-party Interaction (AMI) corpus (Carletta et al., 2005) consists of 100 hours of multimodal recordings of four participants in realistic and scenario-driven meetings. The corpus has been annotated for speech transcription, dialogue acts, head and hand gestures, focus of attention along with several other properties.

Existing annotations

Cohesion: A portion of AMI corpus was annotated for task and social cohesion values by Hung et. al., (Hung and Gatica-Perez, 2010). The meetings were divided into two minutes segments. The authors selected 100 segments from the 10 meetings where the teams are asked to design a remote control and 20 segments from two groups involved in real discussions. The data was annotated manually by 21 annotators using a 27-item questionnaire on a 7-point Likert scale. Each segment was annotated by three different annotators and a kappa agreement was calculated. In total, 61 segments with a kappa score above 0.3 was retained. This consisted of 50 segments with high cohesion rating and 11 segments with low cohesion rating.

Dialogue Acts: 15 categories of dialogue acts (DACT) are annotated in the AMI corpus[1]. In the corpus, the DACT are segmented according to the intention expressed in an utterance i. e., each time a new intention is expressed, a new segment is marked. Each of the 15 categories belongs to one of the four main classes.

- **Minor** comprises of *Backchannel*, *Stall* (filled paused) and *Fragment*.
- **Task** is about information exchange and actions that an individual or group might take. It comprises of categories *Inform*, used by a speaker to give information, *Suggest*, related to the actions of another individual or the group as a whole, and *Assess*, any comment that expresses an evaluation.
- **Elicit** is about requesting someone to give information or completing some action. It includes three categories *Elicit-Inform*, requests some information, *Elicit-Assessment*, elicits an assessment about what has been said or done, and *Elicit-Comment-Understanding*.
- **Other** is about DACT that expresses social acts or comments about things that have been said previously. It includes *Offer*, intention related to the speaker's own actions, *Comment-About-Understanding*, commenting on a previous DACT, *Be-Positive*, acts that are intended to make an individual or the group happier, *Be-Negative*, acts that express negative feelings towards an individual or the group.

Among the 61 segments annotated with cohesion, only 25 are annotated with dialogue act annotations. Specifically, these annotations are available for eight of the eleven low cohesion segments. Therefore for our work, we consider a total of 16 segments i. e., eight high cohesion (M= 2.995, SD= 0.3276) and eight low cohesion(M=5.994, SD= 0.1929) segments with W= 0.94, p = 0.62 and W= 0.92, p = 0.45 respectively.

Our annotations

Gaze We define four different gaze targets for a given participant i. e., the other three participants in the group and "others" class e. g., looking at the table, slides. The manual annotations are performed at frame level using the ELAN annotation tool. Further, we compute *MutualGaze* i. e., overlapping gaze between any two participants at a given point in time and *OverallGaze* i. e., total amount of time spent by each participant in a group looking at the other participants.

Head nods We annotate the vertical up-and-down movements of the head, rhythmically raised and lowered. The head nods are annotated for all the four participants. The manual annotations are performed at frame level using the ELAN annotation tool.

Facial AUs We automatically extract three different Action Units i. e., AU2, AU4, AU12 using OpenFace (Baltrušaitis et al., 2016). The tool offers two kinds of scores for the AU i. e., intensity and presence. We segment the video data based on continuous presence of a given action unit and calculate the duration and intensity of activated AU for each segment.

Laughter We extract laughter instances from the transcription files already available with the corpus.

Annotation	Low Cohesion	High Cohesion
Mutual Gaze	202	258
Outer Brow Raiser (AU2)	28	26
Brow Lowerer (AU4)	77	59
Lip Corner Puller (AU12)	52	113
Head Nods	100	106
Laughter	31	108

Table 1: Total number of instances annotated for 16 low and high cohesion segments

Table 1 shows the number of instances annotated for all the 16 segments. For each behavioural cue, we calculate the number of instances for each segment, the total duration, the mean duration and additionally, mean intensity for Action Units.

[1]For the description of the dialogue acts annotation in the AMI corpus, we rely on the annotation manual available at http://groups.inf.ed.ac.uk/ami/corpus/annotation.shtml

Interruption In order to annotate interruption, we follow the three layer schema described in (Cafaro et al., 2019). This schema was developed for dyadic interactions and we adapt it for multiparty interactions. **Communicative layer** defines the interlocutors' speaking activities which includes *none* (no one is talking), *speaker*, *both* (two speakers are talking), *multi* (more than two speakers are talking). **Transition layer** defines the transition events from silence to speech and vice versa for the same speaker or multiple speakers. *Pause within* is a (long) silence within a speaking turn of speaker without a speaker switch; *Pause between* is a speaker switch from current speaker to another participant (or vice-versa) with a silence in between; *Perfect* is a speaker change without silence or an overlap in-between; *Overlap within* is an overlap without speaker switch; *Overlap between* is an overlap with a speaker change. This layer also distinguishes between overlaps and backchannels using the available dialogue act information along with the start and end time of the speech. **Interruption layer** defines the type of interruption based on the interruption time. It includes *overlapped interruption* – interruption having an overlap with speaker change; and *paused interruption* – interruption having a speaker switch from current speaker to another participant (or vice-versa) with a silence in between where the speaker does not manage to complete the sentence; *interruption-other* – interruption due to interrupter addressing the non speaker e. g., speaker A is interrupted by B addressing C (Schegloff, 2000). Figure 1 shows an example of the schema described.

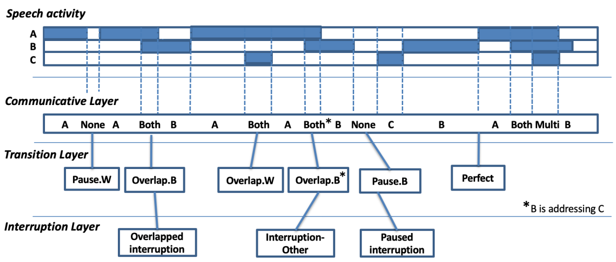


Figure 1: Example annotation at Communicative, Transition and Interruption layer, adapted from (Cafaro et al., 2019)

In order to annotate the data, we perform semi-automatic annotation at the communicative and transition layers based on the start time and end time of each utterance and dialogue act information. Then, we manually annotate the interruption layer with the help of multimodal information i. e., speech, verbal transcriptions, and the visual focus of attention i. e., direction of speaker's gaze.

4. Mono-layer Cohesion Analysis

4.1. Cohesion and Non-verbal Social Cues

In order to verify our hypothesis for this preliminary study, we perform an independent t-test on the data. Initially, we verify the assumption of normality of the data distribution using Shapiro-Wilk test. For the non-normal data we perform Mann-Whitney test.

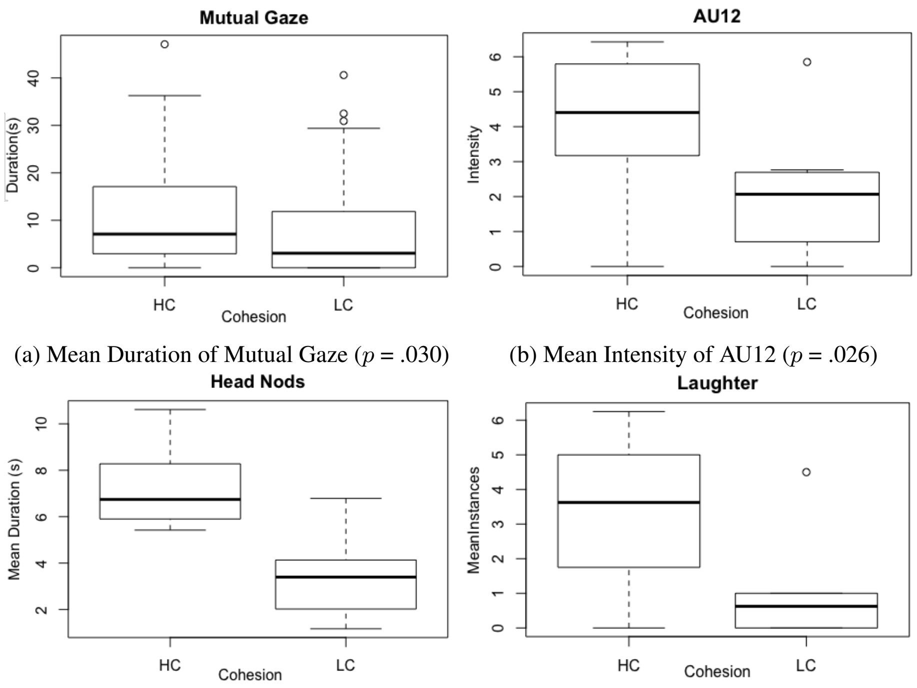
Gaze We did not find any significant difference in the gaze behaviour at the segment level between the low and high cohesive segments with p <0.1. Therefore, we observed the gaze behaviour at participant level. The duration of gaze for any given participant was significantly higher among participants, ($t(64)$ = -2.67, *df* = 60.75, p = .006), in the high cohesion segments (M = 76.64, SD = 27.83) than the participants in the low cohesion segments (M = 59.25, SD = 24.09). Similarly, participant pairs mutually gazed at each other longer in high cohesion segments than in low cohesion segments and this difference was statistically significant, (U = 857, p = .03, r = .31).

Facial Action Units From our data annotations we observe that AU12 i. e., Lip corner puller was activated more frequently in highly cohesive groups. The duration of activation was significantly higher ($t(16)$ = -2.57, *df* = 10.35, p= .026) in the high cohesive segments (M = 65.05, SD = 42.25) than low cohesive segments(M = 21.91, SD = 21.34). Further, the mean intensity of the activated AU12 was higher as well but the difference was not significant, ($t(16)$ = -2.04, *df* = 13.77, p = .060). There was no significant difference in the duration or intensity of activation of AU2 i. e., Outer brow raiser and AU4 i. e., Brow lowerer.

Head Nods Even though there wasn't a huge difference in the occurrence of head nods for both the groups, there was a significant difference in the duration of head nods, ($t(16)$ = -4.33, *df* = 13.99, p = .0006). In general, head nods in high cohesion segments lasted longer (M= 7.23, SD = 3.09) than low cohesion segments (M = 3.38, SD = 3.23).

Laughter Laughter was observed more frequently in high cohesion segments. The duration of laughter was not significantly different but the average occurrence of laughter per segment was lower ($t(16)$ = -2.59, *df* = 12.45, p = .022) in low cohesion segments (M = 0.96, SD = 2.22) than in high cohesion segments (M = 3.37, SD = 4.64).

Discussion As explained in Section 2.4., our aim was to recognize non-verbal social cues that are associated with low and high cohesion groups. In order to do this we looked at gaze behaviour, facial action units, head nods and laughter. Our initial assumptions were that behavioural cues associated with positive affect, involvement and support e. g., gaze at locutor, laughter, head nods, will be higher in cohesive groups. The main finding of our study is that instances of laughter are frequent in high cohesive groups. We observed that instances of shared laughter i. e., where two are more participants laugh simultaneously, was also higher. This is in line with several studies on laughter in groups which state that "laughter establishes a form of bond in social groups and makes people feel more comfortable" (Glenn, 2003). Additionally, we observe that AU12, that is associated with happiness and smile (Ekman et al., 1990), had a higher intensity value in these segments. Further, we observed that AU4, that is often associated with anger and contempt (Tian et al., 2001), occurred more frequently in low cohesion segments, however, the differences were not significant. This could be attributed to the fact that we observe the interaction for short duration of time (2min) and perhaps by considering more segments in the dataset



(a) Mean Duration of Mutual Gaze (p = .030)   (b) Mean Intensity of AU12 (p = .026)

(c) Mean Duration of Head Nods (p = .0006)   (d) Average instances of Laughter(p = .022)

Figure 2: Box plots of mean values of non-verbal social cues for low and high cohesion segments

this effect could be strengthened. The next assumption we looked at was head nods. The presence of head nods in conversation often creates a favorable environment (Hadar et al., 1984) and is commonly associated with attentive listening. In our data, there was almost no difference in the frequency of occurrence of head nods between the two groups. However, we did observe a significant difference in the average duration of the head nods. The final cue that we observe is the eye gaze of the participants. Overall, we assumed that in cohesive groups participants spend higher amount of time gazing at others and holding mutual gaze. Our results show that participants in high cohesive groups gazed at fellow members for longer duration than in low cohesive groups. This result supports studies that state that eye-gazing regulates understanding in multi-party scenarios and is important for managing the flow of interaction (Kendon, 1967). Further, low cohesive groups spent a shorter amount of time holding the gaze with other participants, which is in line with Exline et. al., (Exline, 1963), where they state that the duration of eye-contact decreased in non-collaborative conditions.

4.2. Cohesion and Dialogue Acts

The average number of DACT per segment in our dataset is 52. The highest number of DACT per segment belonged to *Task* (54 at most and 18 at least). *Other* had the lowest number of DACT per segment (6 at most and 0 at least). To understand how DACT can be linked to cohesion, we check whether the number of DACT for each specific category has an impact on the level of cohesion e. g., some categories might be positively correlated and some others negatively correlated.

Cohesion and the four Main Classes In each of the 16 segments annotated with a correlation score, there are several DACT that belong to the four main classes and 15 subcategories. We first measured the correlation between the number of DACT for each of the four main classes (*Task*, *Elicit*, *Minor* and *Other*) in each segment and the cohesion score of each segment. We considered the number of DACT for each category as independent variable and the cohesion score as dependent variable. To measure the correlation between the two variables, we used Pearson's correlation. We did not find a correlation between the cohesion score and the number of DACT for each main class. The p-values obtained was superior to .05, and hence the results were not significant and the correlation coefficient could not be interpreted.

Cohesion and the 15 Specific Categories A correlation could exist between a specific DACT category and cohesion. Since we did not find any correlation between cohesion and the number of DACT for any of the four main classes, we think that these classes might have been too broad to show any significant results. We used Pearson's correlation to measure the relation between the level of cohesion and the number of DACT for each of the 15 specific categories. For most of the categories, the results were not significant since the p-values were superior to .05. Only one category, *Be-Positive* showed a significant value with p = .030. The correlation coefficient was superior to 0.5, so we can argue that the correlation is high between the Be-Positive DACT and the level of cohesion. These results attest to the assumption made in (Nanninga et al., 2017) about the expression of feelings linked to the level of cohesion.

Linear Regression with Contrast between Main Classes In order to verify the results obtained with the Pearson's correlation, we computed a linear regression model with contrasts between the four different main classes[2]. This test shows the difference between the mean cohesion score obtained with one class in contrast with the mean cohesion score obtained with the three others. The results confirmed the correlation coefficient introduced above; when we contrast each of the four classes to the three others, none of them showed a significant impact on the cohesion score. The difference between the mean cohesion score obtained with one class compared to the mean cohesion score of the three others was never superior to 0.1 or inferior to -0.1.

DACT	Correl. Coef.	p-value
Inform	-0.485	.056
Suggest	0.373	.154
Assess	0.452	.078
Elicit-Inform	-0.194	.470
Elicit-Offer-or-Suggestion	0.373	.155
Elicit-Comment-Understanding	0.237	.377
Elicit-Assess	-0.097	.721
Offer	-0.388	.138
Comment-About-Understanding	-0.316	.232
Be-Positive	0.542	.030

Table 2: Correlation coefficients and p-values for the Pearson's test between cohesion score of each segment and the number of DACT of each specific category in each segment

Discussion Except for the *Be-positive* DACT, our analysis does not show any significant results regarding the correlation between the number of DACT of a specific class or category and the cohesion level. The results can be explained by the structure of the conversation. Since interactions are task-oriented – groups aiming to organize team work – the speaker changes occurred frequently (33 times in each segment on average). Each new speaker did not provide a DACT which could form an adjacency pair (Sacks et al., 1974) with the previous one at all times (we estimate that only half of them formed an adjacency pair). This type of conversation structure can create difficulty for an annotator that relies only on verbal behaviour such as dialogue acts for rating the cohesion level. In the next study, we aim to focus on interpersonal synergy that can be analysed through grounding mechanisms, as described in (Dillenbourg and Traum, 2006). Another hypothesis for explaining these results is that the DACT might be related to cohesion when we consider how they combine with other multimodal features. However, it was necessary to check whether verbal behaviours had an impact by themselves. We hypothesize that DACT might have an impact on cohesion when they are associated with other non-verbal behaviours (see Section 5.).

4.3. Cohesion and Interruption

Our aim was to analyse the relation between turn taking, interruption and cohesion in multiparty interactions. We utilise Pearson's correlation to observe the relation between cohesive segments and the independent variables and perform a one-way ANOVA to measure the differences between the two groups.

Turns The number of turns was positively correlated with cohesion score, Pearson's (r = 0.624, p = .01). A one-way ANOVA showed that there was a statistically significant effect of cohesion score on the number of turns during interaction ($f(1, 14)$ = 6.465, p = .023). High cohesive groups alter turns more frequently (M = 23.75, SD = 7.741) than low cohesive group (M = 15.125, SD = 5.667).

Overlaps The number of overlaps had a positive correlation with cohesion, Pearson's (r = 0.519, p = .039). The number of overlaps in high cohesive groups (M = 27.62, SD = 4.92) was significantly higher than in low cohesive groups (M = 16.5, SD = 11.46), with ($F(1, 14)$ = 5.327, p = .037).

Overlapped Interruption A positive correlation between number of overlapped interruptions and cohesion (r = 0.613, p = .008) was observed. A one-way ANOVA showed statistically significant difference in number of overlapped interruptions in low and high cohesion ($F(1, 14)$ = 9.847, p = .007) segments. High cohesive groups appeared to have more *interruptions* (M = 9.75, SD = 3.327) in comparison to low cohesive groups (M = 4.62, SD = 3.20). We did not find any correlation between cohesion and *paused interruptions*. The number of *paused interruptions* (M = 2.937, SD = 2.205) was significantly smaller ($t(16)$ = 3.5, *df* = 15, p = .003) than the number of *overlapped interruptions* (M = 7.187, SD = 4.118).

Interruption-other The occurrence of these interruptions had a positive correlation with cohesion, Pearson's (r = 0.674, p = .004). A one-way ANOVA indicated that ($F(1, 14)$ = 0.994, p = .007) participants used higher number of *interruption-other* in high cohesive groups (M = 2.125, SD = 1.124) than low cohesive groups (M = 0.50, SD = 0.756). However, the number of *interruption-other* (M = 1.31, SD = 1.30) during the interaction was significantly smaller ($t(16)$ = 7.388, p <.01) than the number of *overlapped interruptions* (M = 7.187, SD = 4.11).

Feature	Correl. Coef.	p-value
Turns	0.624	.010
Overlaps	0.519	.039
Overlapped interruption	0.613	.008
Paused interruption	0.258	.334
Interruption-other	0.674	.004

Table 3: Pearson's Correlation coefficients and p-values between cohesion and features related to turn taking and interruption

Discussion Our aim was to analyse the relationship between turn taking, interruption and cohesion summarised in Table 3. Our hypothesis that the number of turns are higher in high cohesive groups and lower in low cohesive groups during multi-party interaction is validated. This result supports the findings of (Hung and Gatica-Perez, 2010). Results show that participants exchange turns more frequently in high cohesive groups since all the members of the group

[2]Due to the high number of sub-categories (15 sub-categories of DACT), we only measure the contrasts between the four main classes

are actively participating in the interaction, thus increasing the number of turns. It also results in reducing the duration between two successive speaking turns compared to the duration in low cohesive group. The occurrence of overlaps during interaction is positively correlated with the group cohesion. Our hypothesis that the number of interruptions is high in cohesive groups is validated. Tannen (Tannen, 1994) describes that interruptions can be good indicators of cohesion in group e. g., when people are able to complete each other's sentences. Since the subset of the AMI corpus that we have utilized consists of task-oriented meetings, where participants collaborate and discuss with each other to achieve their common objective, the tendency to have interruptions is higher. Results regarding *interruption-other* are in-line with findings in psychology studies (Pontecorvo et al., 2000; Bangerter et al., 2010). That is, the interruption of a conversation by third parties are common during group interaction. Although, the number of *interruption-other* are relatively smaller in comparison to the number of *overlapped interruptions* in our data, they are still an important cue of cohesion in multiparty interactions.

5. Multi-layer Cohesion Analysis

In the previous section, our analysis considered the three layers (verbal, non-verbal social cues and interruption) separately and checked the impact of each on the level of cohesion. Non-verbal social cues like mutual gaze, laughter and AU12 were associated with cohesive segments. For dialogue acts, the results showed that the number of occurrences of specific categories did not have an impact on the level of cohesion except for *Be-Positive*, which appears to be positively correlated with cohesion. The number of turns, overlaps and interruption were positively correlated with cohesion. The analysis of the three layers shows that the perception of cohesion relies on several behaviours from different modalities. However, for multimodal analysis of the group cohesion, we need to analyse how these behaviours co-occur and how this co-occurrence affects the level of cohesion. Inspired by existing literature, we look at the relation between specific behaviours: (i) interruption – gaze and cohesion (ii) dialogue act – head nods and cohesion.

Interruptions and Gaze Eye gaze significantly helps in predicting the partner's turn taking activity (Jokinen et al., 2013). The result in section 4.1. shows that participant pairs mutually gazed at each other longer in high cohesion groups. We analysed the relation of mutual gaze between the interruptee and interrupter during interruption with cohesion. Mutual gaze instances occurring during interruption were positively correlated with group cohesion, Pearson's (r = 0.731, p = .001). A one-way ANOVA shows a statistically significant difference in number of mutual gaze instances (F(1,14)=15.868, p = .001) i. e., the number of mutual gaze is higher in high cohesive groups (M = 4.875, SD = 2.167) than low cohesive groups (M = 1.25, SD = 1.388). Participants during interruption gazed at each other more frequently in high cohesive groups in comparison to low cohesive groups.

Dialogue Acts and Head Nods During conversation, verbal and non-verbal signals are at play. In this section we present the analysis of the co-occurrence of head nods and DACT with relation to cohesion. In order to do this, we extracted the instances of head nods performed by listeners and the corresponding dialogue act types expressed by speakers for each specific DACT. We then computed a linear regression model with contrasts between the four main classes. The first model contrasts *task* to the three other DACT classes when occurring with head nods. The mean cohesion score obtained by these DACT when occurring with head nods was 4.955. The results showed that a listener's head nod occurring when the speaker is performing a DACT from the category *task*, is related to a lower cohesion score than head nods occurring with one of the other three classes (-0.200). In the same model, the residual contrast between *elicit* and *other*, when co-occurring with a head nods, showed that *elicit* produces a higher cohesion score than *task* (1.042). The second model contrasts *other* with *task*, *elicit* and *minor* when occurring with head nods. The results showed that the DACT *other* produce a cohesion level lower than the mean of the three other (-0.298). In the same model, the residual contrast between *task* and *elicit* shows that *task* produces a lower cohesion score for *task* than for *elicit*.

From the analysis of behaviours at a multimodal level i. e., interruption – gaze and dialogue act – head nods, we see that certain behaviours that did not have an impact by itself, have an impact on the perceived level of cohesion when they were combined. From this analysis we can conclude that multimodal behaviours can provide new insight into their relation with cohesion and enhance its estimation in multiparty interaction.

6. Conclusion and Future Work

In the present article, we provide an analysis of cohesion in multi-party interactions which focuses on three layers i. e., *non-verbal social cues*, *dialogue acts* and *interruption*. When considered separately interruptions and certain non-verbal social cues have an impact on level of cohesion. This paper also shows the importance of combining multiple modalities for effective cohesion analysis. The results from this work will contribute towards developing a computational model to simulate a cohesive group of virtual agents. Future work will include replicating the results with another multi-party corpus and development of an automatic cohesion estimation model.

Acknowledgments This project has received funding from the European Unionâs Horizon 2020 research and innovation program under grant agreement number 769553 and a grant overseen by the French National Research Agency (ANR17-MAOI) and by the European project H2020 ANIMATAS (ITN 7659552). This result only reflects the authors' views and the EU is not responsible for any use that may be made of the information it contains. We are grateful to Dr. Hayley Hung of TUDelft for sharing the cohesion annotation dataset with us. We would also like to thank Céline Caristan and Amandine Guillin for their thoughtful reviews on the statistical analysis.

7. Bibliographical References

Back, K. (1951). Influence through social communication. *The Journal of Abnormal and Social Psychology*, 46(1):9.

Baltrušaitis, T., Robinson, P., and Morency, L.-P. (2016). Openface: an open source facial behavior analysis toolkit. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–10. IEEE.

Bangerter, A., Chevalley, E., and Derouwaux, S. (2010). Managing third-party interruptions in conversations: Effects of duration and conversational role. *Journal of Language and Social Psychology*, 29(2):235–244.

Beal, D., Cohen, R., Burke, M., and McLendon, C. (2003). Cohesion and performance in groups: A meta-analytic clarification of construct relations. *Journal of Applied Psychology*, 88(6):989.

Beattie, G. W. (1981). Interruption in conversational interaction, and its relation to the sex and status of the interactants. *Linguistics*, 19(1-2):15–36.

Bohus, D. and Horvitz, E. (2010). Computational models for multiparty turn taking. Technical report, Microsoft Research Technical Report MSR-TR 2010-115.

Braaten, L. (1991). Group cohesion: A new multidimensional model. *Group*, 15(1):39–55.

Bunt, H. (2011). The semantics of dialogue acts. In *Proc of the Ninth International Conference on Computational Semantics*, pages 1–13. ACL.

Cafaro, A., Glas, N., and Pelachaud, C. (2016). The effects of interrupting behavior on interpersonal attitude and engagement in dyadic interactions. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*, pages 911–920. International Foundation for Autonomous Agents and Multiagent Systems.

Cafaro, A., Ravenet, B., and Pelachaud, C. (2019). Exploiting evolutionary algorithms to model nonverbal reactions to conversational interruptions in user-agent interactions. *IEEE Transactions on Affective Computing*.

Campano, S., Langlet, C., Glas, N., Clavel, C., and Pelachaud, C. (2015). An eca expressing appreciations. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 962–967. IEEE.

Carless, S. and De Paola, C. (2000). The measurement of cohesion in work teams. *Small group research*, 31(1):71–88.

Carletta, J., Ashby, S., Bourban, S., Flynn, M., Guillemot, M., Hain, T., Kadlec, J., Karaiskos, V., Kraaij, W., Kronenthal, M., et al. (2005). The ami meeting corpus: A pre-announcement. In *International workshop on machine learning for multimodal interaction*, pages 28–39. Springer.

Carron, A. and Chelladurai, P. (1981). The dynamics of group cohesion in sport. *Journal of Sport Psychology*, 3(2):123–139.

Carron, A., Widmeyer, N., and Brawley, L. (1985). The development of an instrument to assess cohesion in sport teams: The group environment questionnaire. *Journal of Sport Psychology*, 7(3):244–266.

Carron, A. (1982). Cohesiveness in sport groups: Interpretations and considerations. *Journal of Sport Psychology*, 4(2):123–138.

Casey-Campbell, M. and Martens, M. (2009). Sticking it all together: A critical assessment of the group cohesion–performance literature. *International Journal of Management Reviews*, 11(2):223–246.

Dillenbourg, P. and Traum, D. (2006). Sharing solutions: Persistence and grounding in multimodal collaborative problem solving. *The Journal of the Learning Sciences*, 15(1):121–151.

Ekman, P., Davidson, R. J., and Friesen, W. V. (1990). The duchenne smile: Emotional expression and brain physiology: Ii. *Journal of Personality and Social Psychology*, 58(2):342.

Ekman, R. (1997). *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, USA.

Exline, R. V. (1963). Explorations in the process of person perception: Visual interaction in relation to competition, sex, and need for affiliation. *Journal of Personality*.

Fang, S. and Achard, C. (2018). Estimation of cohesion with feature categorization on small scale groups. WACAI.

Festinger, L., Schachter, S., and Back, K. (1950). Social pressures in informal groups; a study of human factors in housing.

Forsyth, D. (2018). *Group dynamics*. Cengage Learning.

Fusaroli, R., Raczaszek-Leonardi, J., and Tylén, K. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology*, 32:147–157.

Glenn, P. (2003). *Laughter in interaction*, volume 18. Cambridge University Press.

Goodman, P., Ravlin, E., and Schminke, M. (1987). Understanding groups in organizations. *Research in Organizational Behavior*.

Hadar, U., Steiner, T. J., Grant, E. C., and Rose, F. C. (1984). The timing of shifts of head postures during conservation. *Human Movement Science*, 3(3):237–245.

Heldner, M. and Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, 38(4):555 – 568.

Hung, H. and Gatica-Perez, D. (2010). Estimating cohesion in small groups using audio-visual nonverbal behavior. *IEEE Transactions on Multimedia*, 12(6):563–575.

Jokinen, K., Furukawa, H., Nishida, M., and Yamamoto, S. (2013). Gaze and turn-taking behavior in casual conversational interactions. *ACM Transactions on Interactive Intelligent Systems*, 3(2):12:1–12:30, August.

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta psychologica*, 26:22–63.

Li, H. Z. (2001). Cooperative and intrusive interruptions in inter-and intracultural dyadic discourse. *Journal of Language and Social Psychology*, 20(3):259–284.

Mills, G. J. (2014). Dialogue in joint activity: Complementarity, convergence and conventionalization. *New Ideas in Psychology*, 32:158–173.

Mudrack, P. (1989). Defining group cohesiveness: A legacy of confusion? *Small Group Behavior*, 20(1):37–49.

Nanninga, M. C., Zhang, Y., Lehmann-Willenbrock, N., Szlávik, Z., and Hung, H. (2017). Estimating verbal expressions of task and social cohesion in meetings by quantifying paralinguistic mimicry. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, pages 206–215. ACM.

Pickering, M. J. and Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(2):169–190.

Piper, W., Marrache, M., Lacroix, R., Richardsen, A., and Jones, B. (1983). Cohesion as a basic bond in groups. *Human Relations*, 36(2):93–108.

Pontecorvo, C., Pirchio, S., and Sterponi, L. (2000). Are there just two people in a dyad? dyadic configurations in multiparty family conversations. *Schweizerische Zeitschrift für Bildungswissenschaften*, 22(3):535–558.

Reitter, D., Keller, F., and Moore, J. D. (2006). Computational modelling of structural priming in dialogue. In *Proc. of the Human Language Technology Conference of the NAACL, Companion volume: Short papers*, pages 121–124. ACL.

Richmond, V. P., McCroskey, J. C., and Payne, S. K. (1991). *Nonverbal Behavior in Interpersonal Relations*. Prentice Hall Englewood Cliffs, NJ.

Sacks, H., Schegloff, E., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(9-10):696–735, November.

Santoro, J., Dixon, A., Chang, C., and Kozlowski, S. (2015). Measuring and monitoring the dynamics of team cohesion: Methods, emerging tools, and advanced technologies. In *Team cohesion: Advances in Psychological Theory, Methods and Practice*, pages 115–145. Emerald Group Publishing Limited.

Schegloff, E. (2000). Overlapping talk and the organization of turn-taking for conversation. *Language in Society*, 29(1):1–63.

Taboada, M. T. (2004). *Building coherence and cohesion: Task-oriented dialogue in English and Spanish*, volume 129. John Benjamins Publishing.

Tannen, D. (1994). *Gender and Discourse*. Oxford University Press.

Tian, Y.-I., Kanade, T., and Cohn, J. F. (2001). Recognizing action units for facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2):97–115.

Vinciarelli, A., Pantic, M., and Bourlard, H. (2009). Social signal processing: Survey of an emerging domain. *Image and Vision Computing*, 27(12):1743–1759.

Wang, W., Precoda, K., Hadsell, R., Kira, Z., Richey, C., and Jiva, G. (2012). Detecting leadership and cohesion in spoken interactions. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5105–5108. IEEE.

West, C. and Zimmerman, D. H. (2015). Small insults: A study of interruptions in cross-sex conversations between unacquainted persons. In *American Sociological Association's Annual Meetings, Sep, 1978, San Francisco, CA, US;*.

Zhang, Y., Olenick, J., Chang, C., Kozlowski, S., and Hung, H. (2018). Teamsense: Assessing personal affect and group cohesion in small teams through dyadic interaction and behavior analysis with wearable sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(3):150.