The WoNoWa Dataset: Investigating the Transactive Memory System in Small Group Interactions

Beatrice Biancardi
beatrice.biancardi@telecom-paris.fr
LTCI, Télécom Paris
Institut polytechnique de Paris
France

Lou Maisonnave-Couterou
loumaisonnave@gmail.com
Institut de Psychologie
Université de Paris
France

Pierrick Renault
pierrick.renault@outlook.com
Institut de Psychologie
Université de Paris
France

Brian Ravenet
brian.ravenet@limsi.fr
LIMSI-CNRS
Université Paris-Saclay
France

Maurizio Mancini
m.mancini@cs.ucc.ie
School of Computer Science & IT
University College Cork
Ireland

Giovanna Varni
giovanna.varni@telecom-paris.fr
LTCI, Télécom Paris
Institut polytechnique de Paris
France

ABSTRACT

We present WoNoWa, a novel multi-modal dataset of small group interactions in collaborative tasks. The dataset is explicitly designed to elicit and to study over time a Transactive Memory System (TMS), a group's emergent state characterizing the group's meta-knowledge about "who knows what". A rich set of automatic features and manual annotations, extracted from the collected audio-visual data, is available on request for research purposes. Features include individual descriptors (e.g., position, Quantity of Motion, speech activity) and group descriptors (e.g., F-formations). Additionally, participants' self-assessments are available. Preliminary results from exploratory analyses show that the WoNoWa design allowed groups to develop a TMS that increased across the tasks. These results encourage the use of the WoNoWa dataset for a better understanding of the relationship between behavioural patterns and TMS, that in turn could help to improve group performance.

CCS CONCEPTS

• **Information systems** → **Database design and models**; • **Applied computing** → Psychology.

KEYWORDS

Multi-modal Social Datasets; Multi-modal Group Behaviour Analysis; Social Signal Processing; Transactive Memory System

ACM Reference Format:

Beatrice Biancardi, Lou Maisonnave-Couterou, Pierrick Renault, Brian Ravenet, Maurizio Mancini, and Giovanna Varni. 2020. The WoNoWa Dataset: Investigating the Transactive Memory System in Small Group Interactions. In *2020 International Conference on Multimodal Interaction (ICMI '20), October 25–29, 2020, Virtual Event, Netherlands.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3382507.3418843

ICMI '20, October 25–29, 2020, Virtual Event, Netherlands

ACM ISBN 978-1-4503-7092-9/20/08…$15.00
https://doi.org/10.1145/3382507.3418843

1 INTRODUCTION

Emergent states are social processes resulting from affective, behavioral and cognitive interactions among group members [29, 39]. Until today, the study of group behavioural dynamics, and in particular, emergent states, in collaborative tasks has received little attention from scholars in Computer Science and Artificial Intelligence, as the study of these phenomena requires complex setups and expensive computational resources. One of these emergent states is the Transactive Memory System (TMS) [55], which describes the cooperative division of work that allows a group to learn, remember and communicate the group's knowledge. TMS allows group members to develop a representation of the distribution of knowledge among them ("who knows what"). Additionally, behavioural dynamics play a key role in a group. Group members adhere to specific social norms which govern, for example, their distance and body orientation, in order to coordinate and facilitate the interaction between them [26, 52]. While in Social and Psychological Sciences many researchers have studied the psychological processes characterising the development and dynamics of TMS in groups [34, 35], we are not aware of existing research focusing on what (and how) behavioural cues could predict the development of TMS. Researchers in Computer Science could benefit from this knowledge to automatically extract features and develop computational models of TMS to predict and enhance group performance.

In order to facilitate the study and understanding of this phenomenon, we present the **WoNoWa** (**Wh**o k**No**ws **Wha**t) dataset, a corpus of interactions between members of small groups performing several activities. The dataset is targeted on groups of 3 people to limit the complexity of the experimental setup and to reduce the number of interactions to be taken into account, ensuring, however, the good observation of group dynamics. The WoNoWa dataset includes multi-modal recordings, automated features and manual annotations of participants' non-verbal behaviours, as well as self-assessment measures of TMS and of leadership (which has been found to be related to TMS [3, 31]). It can be a valuable asset for researchers working on emergent states and group interaction.

With the WoNoWa dataset, we make the following contributions. First, to our knowledge, this is the first dataset whose protocol has been specifically designed to observe the emergence and development over time of a Transactive Memory System. To achieve this, we relied on psychological theories of TMS to reproduce the phases of TMS in our protocol (see Section 4). We also designed different tasks to vary the dynamics of the groups by proposing different types of collaboration.

Second, WoNoWa contains self-assessment measures of TMS given by the participants. This data can be exploited as ground truth in further analyses and in the development of computational models of TMS.

Third, even if the dataset was collected in a controlled setting, we tried to minimize the intrusiveness of the sensors, with the aim to have interactions that are as natural as possible and to enable the development of future real-time applications using commodity hardware only.

2 BACKGROUND

According to Moreland, groups are entities including at least three individuals sharing knowledge, activities and so on [44]. He outlines several reasons for which dyads and groups are different. Such differences concern the number and the structure of possible interactions, the strength of emotions and the phenomena developing during interactions. Group interactions, indeed, imply many one-to-one and one-to-many interactions, weaker emotions with respect to those in dyads, and the appearance of the so called emergent states: "constructs that characterize the properties of the group of a typically dynamic nature and that vary according to the context of the groups, inputs, processes and outcomes" [39]. Thus, emergent states "develop over time through the dynamic interactions of group members" [20].

One of these emergent states is the Transactive Memory System. In a group, at the individual level, people exploit their transactive memory, a sort of meta-memory determining one's memory skills. When this kind of meta-memory extends to the group level, that is, when group members are aware of each individual's transactive memory ("who knows what"), we talk about Transactive Memory System (TMS) [55]. TMS allows for obtaining a mental representation of the distribution of knowledge between the group members, that in turn allows each member to extend his/her knowledge beyond what he/she individually possesses. It is a property of the group, as a function of both individual memory systems and the communication among them.

Moreland [45] and Lewis [33] identify the three dimensions characterising TMS. These are: (i) Specialization of members' knowledge, that is, the extent to which network members are experts in areas that other members are not; (ii) Credibility: the belief about the reliability of other members' knowledge; (iii) Coordination: an effective, orchestrated knowledge processing.

Evidence show the benefits of TMS for group performance [28]: it improves coordination (people can predict others' behaviour) and work planning. In addition, task performance is enhanced, thanks to knowledge specialization: as knowledge overlaps are reduced, problem solving becomes more efficient and the group can reach its goals faster. Other evidence show a relationship between TMS and another emergent state, that is, leadership [3, 31].

Like individual memory, TMS includes the three phases of encoding, storage, and retrieval.

During the encoding phase, individuals identify other group members' expertise. They can learn "who knows what" by direct observation, explicit expert indication, by inferring roles or from a third party. In all these cases, the encoding phase requires interpersonal interactions.

During the storage phase, incoming information is allocated to members with matching expertise [36]. This phase requires acceptance and shared awareness of expertise.

During the retrieval phase, individuals obtain knowledge resources from relevant people through the shared mental map directory [36]. Transactive retrieval can only occur when there is awareness of knowledge distribution in the group.

Communication is an important factor for TMS development, as it positively impacts the accuracy of expertise recognition. For example, in [46], groups which communicated during a training could collectively recall more unique and specific information from the training than groups whose members were not allowed to communicate. In the existing studies about TMS, the measures used by researchers as indicators of TMS include recall, qualitative assessment and self-reports about members' expertise [34, 35, 46]. Since communication is crucial for TMS development, the analysis of group members' interactions could reveal important information about its level of TMS.

Non-verbal behaviour has been found to be an important indicator of the quality of the relationships between people. In particular, people arrangements in the physical space (F-formations, [26]) can reflect their relationships and are influenced by dynamic social contexts. People engaged in joint activities tend to arrange themselves into different spatial patterns according to the roles of the group members and their interactions [16]. Proxemics also shows that the interpersonal distance increases and decreases according to the degree of closeness among people [22]. At the same time, vocal turn-taking plays an important role during interactions. This complex phenomenon includes cues such as speech, silences, overlaps, and interruptions. Several studies show the association of turn-taking with social dimensions like competition and collaboration [23, 25]. In particular, interruptions are a relevant cue in face-to-face conversations: they can be considered as turn-taking violations [5], they can reflect interpersonal attitudes (e.g., dominance or cooperation) as well as involvement in the interaction [47].

Even though spatial and conversational patterns can reflect how members perceive each other, and besides their role has been investigated in the development of emergent states such as cohesion [24] and leadership [51], we are not aware of existing research on behavioural cues of TMS in a group.

With the WoNoWa dataset presented in this paper, we aim at filling this gap by providing a multi-modal dataset in which it is possible to observe the emergence and development over time of TMS and whose data would allow for getting new insights about the behavioural dynamics related to TMS.

3 RELATED WORK

Several datasets have been collected to study multi-modal behaviours in groups. In the following, we present a selection of these existing datasets, highlighting the different setups and dimensions of study.

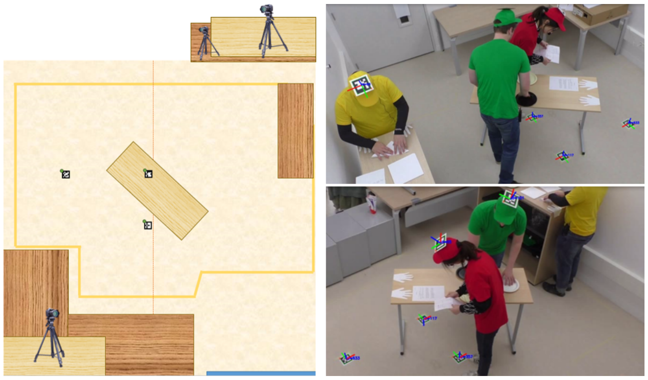

Figure 1: On the left: plan of the *Interaction Area*. On the top right: the view from the camera placed in the North-East corner of the area. On the bottom right: the view from the camera placed in the South-West corner of the area. On each camera view, we can see the Aruco Markers being detected (see Section 5.2.2).

The ELEA dataset (Emergent LEader Analysis) [50] is specifically designed to study emergent leadership. It contains audio-visual recordings of groups of 3 or 4 people sitting around a table and performing a survival task. Before and after the task, the participants filled out questionnaires to measure emergent leadership and related concepts. A simple non-invasive recording setup is used, including 2 webcams and a microphone.

The SALSA dataset (Synergetic sociAL Scene Analysis) [1] provides recordings of free-standing conversational groups participating in a poster presentation and a cocktail party in an indoor environment. This dataset is designed to address the challenges of the analysis of free-standing conversational groups (e.g., occlusions, bad lighting). Indeed, participants wear badges equipped with a microphone, an infrared beam and detector, a Bluetooth detector and an accelerometer.

The MatchNMingle dataset is specifically created to contribute to the efforts to overcome the challenges of the automatic analysis of social signals and interactions [12]. It includes conversations in sitting dyads and free-standing groups during speed dates and cocktail parties, collected in an indoor in-the-wild scenario. In addition to audio-visual recordings, it contains self-assessments of participants' personality, self-control and sociosexuality, as well as data from wearable sensors.

The review of these datasets is informative on how to study group behaviours and on the use of non-invasive techniques. However, none of these works was interested in the TMS. Throughout our review of the literature, we found many other works investigating group dynamics ([6, 13, 15, 30, 38, 48]) but again, the TMS was not their dimension of study.

WoNoWa is the first dataset directly addressing the study of the Transactive Memory System. The recording setup, without invasive sensors, provides naturalistic interactions between groups' members. Participants freely moved in the recording room allowing for the analysis of spatial arrangements. While other free-standing datasets (like SALSA and MatchNMingle) were designed without a specific research purpose other than improving social signal processing techniques, the WoNoWa dataset design is based on a primarily research purpose (i.e., to study TMS). It includes self-assessments of not only TMS but also leadership, useful to investigate its relationship with TMS.

4 METHODOLOGY

4.1 Protocol

The data collection lasted approximately 1 hour per group. The protocol has been approved by the Ethical Committee of Paris-Saclay University. The experimental room, located in Télécom Paris, was divided in different areas:

- The *Interaction Area*: it was the largest area of the room, where participants could interact together and accomplish the tasks during the *Encoding* and *Retrieval* phases (see Sections 4.1.2 and 4.1.4). It included a table in the center and three cameras, two on the top, at the opposite corners, and one at one side of the area. Other tables delimited this area from the rest of the room. Its plan is shown in Figure 1;
- The *Questionnaires Area*: it included a big table, three chairs, and three computers. Participants sat on the three sides of the table to silently fill out questionnaires during the data collection. They were not recorded (neither video nor audio) while being in this area;
- The *Experimenter Area*: placed in a corner of the room, the experimenter stayed there during the participants' activities, without interfering with them. From this area the experimenter could monitor the well functioning of the recording systems and the smooth running of the sessions.

.

The recording protocol was designed in order to reproduce the 3 phases of TMS, i.e., encoding, storage and retrieval (see Section 2). Moreover, we designed tasks requiring different types of collaborations, with the purpose to produce an increase of TMS. Additionally, a welcoming and a debriefing phase at the beginning and at the end of the data collection were present.

4.1.1 Welcoming [Duration: about 15 min]. Once the three participants arrived in the room, the experimenter briefly presented the experiment and let them sign the consent form. Then, participants wore a t-shirt and a cap of the same colour (allowing for automatic detection of their position and movements, see Sections 5.2.1 and 5.2.2), as well as a wireless microphone. Finally, they were invited to fill out the questionnaire Q1 (see Section 5.4.1);

4.1.2 Encoding [Duration: about 10 minutes]. This phase was designed for allowing participants to introduce each other and to identify each other's specific expertise. The participants entered the *Interaction Area* where they found a list of three fields of expertise: logistical, mathematical and manual. They were asked to discuss together about how to assign each skill to each member of the group. At the end, they filled out the questionnaire Q2 (see Section 5.4.2);

4.1.3 Storage [Duration: about 10 minutes]. This phase allowed the participants to develop their own expertise on the chosen field. Each participant watched on a computer in the *Questionnaire Area* a 5-minute tutorial about the field of expertise chosen during the

Encoding phase. More specifically, the tutorials concerned: setting up the table by following specific rules (logistical), the Imperial measurement system (mathematical) and making a heart and a frog origami (manual). The participants could watch the videos as many times as they wanted and they could take notes.

4.1.4 Retrieval [Duration: about 10 minutes]. The participants came back to the *Interaction Area* where they received instructions about 3 Steps to accomplish, requiring the acquisition of specific expertise:

- *Step 1.* During this Step, the participants were asked to perform 3 tasks, each of them related to one of the fields of expertise. These were: setting up the table by following the rules described in the tutorial (logistical), computing conversions between the Imperial and the International System (mathematical) and making the two origamis described in the tutorial (manual). They were free to choose how to assign the tasks within the group, they could look at their notes taken during the *Storage* phase, but they were not allowed to share them;
- *Step 2.* The participants were asked to modify the setup of the table and to do new origamis, this time following a list of dimensions (given in the Imperial system). In particular, these dimensions referred to: specific distances between the cutlery (logistical) and the size of the paper used to create the origamis (manual). Since the participants were only provided with measuring tools in the International System (meters), the mathematical expertise was needed to convert the dimensions and accomplish the task. They were not allowed to use their notes;
- *Step 3.* The participants were given new simplified instructions of the tasks proposed during *Step 1*, meaning without the dimensions. For instance, setting up 1 place at the table instead of 3. They were free to assign them in any way they wanted, as long as each participant was not given the task he/she already realised during *Step 1.*

These Steps were designed in order to elicit and develop TMS by proposing different types of collaboration. In *Step 1*, participants were supposed to work independently, whereas in *Step 2* the mathematical expertise was needed and, finally, in *Step 3* all the participants would need each other's expertise. After each Step they were invited to fill out the questionnaire Q3 (see Section 5.4.3).

4.1.5 End and debriefing [Duration: about 5 minutes]. The participants were invited to remove the microphones and take off the t-shirts and caps. The experimenter briefly explained the goals of the experiment and answered the participants' questions.

4.2 Technical Setup

All the interactions occurring in the *Interaction Area* (i.e., during the *Encoding* and *Retrieval* phases) were recorded by using 3 full HD handy video cameras (1920 x 1080, progressive scan, 50 fps). Two of them were installed at the opposite top angles of the *Interaction Area* (see Figure 1), at a height of about 3 meters, looking downwards. In such a way, there was always at least one camera able to capture the heads of participants in the *Interaction Area*. However, each video camera could capture only a part of the *Interaction Area*, so data fusion was performed for automatic extraction of features, see Section 5.2.2. An additional frontal video camera was positioned at a lower angle to have an additional global view of the participants, useful for the manual annotations.

Each participant wore a wireless microphone headset recording at 44.1k Hz, allowing us to record each one of them in a separate WAV file. The synchronisation between the video and audio streams was made by hand clapping at the beginning and at the end of each Step.

To summarize, for each group we collected 4 recordings: one for the *Encoding* phase, one for *Step 1*, one for *Step 2* and one for *Step 3* of the *Retrieval* phase. Each recording consisted in a set of 6 separated files: the videos from each of the three cameras, and the audio of each participant.

4.3 Participants

The participants were mainly recruited by a mailing list and were students and staff of Télécom Paris. Since we were not interested in manipulating the level of acquaintance among group members, each group was composed of 3 randomly chosen participants. We collected data from 17 groups, 2 of them being discarded since the recordings were incomplete. Around 6 hours of recordings have been collected.

Out of the 45 participants, 71% were male, 49% were in the range 18-25 years old, 42% were in the range 26-35 years old and 9% in the range 36-45. 58% were native French speakers, 16% were Italian speakers, 9% were Arabic speakers and 18% had another mother tongue. All groups whose members did not have the same mother tongue had a sufficient language proficiency to interact in English. 67% of the groups were composed by people who all knew each other. In 13% of the groups, the members never met any other member before. In 40% of the groups, all the members were colleagues. In 67% of the groups, no members were part of the same sport team or association of any other member. We asked the participants for how long they knew each other in months. Given that a person who did not know another answered 0 month, the average acquaintance duration per group was 10 months.

5 COLLECTED DATA

From audio and video recordings, we automatically extracted audio and video features, and we manually annotated specific non-verbal behaviours. Moreover, the dataset includes self-assessments gathered through questionnaires. All these features and annotations are available on request for research purposes. Having the objective of investigating the prediction of group TMS level in a real-time fashion, we chose to rely for the automatic extraction on methods and technologies which could be used in a real-time system.

5.1 Audio Features

5.1.1 Audio data post-treatment. For each group, we collected 12 audio recordings, one for each Step and participant. The audio tracks were synchronised with the videos to ensure a consistent time segmentation among modalities. We applied a normalization and compression on each track using Audacity[1] to enhance audio

[1]Audacity® software is copyright © 1999-2019 Audacity Team. The name Audacity® is a registered trademark of Dominic Mazzoni.

quality and improve noise level consistency among the tracks. Additionally, we applied a noise reduction to some tracks (on around 10%) to reduce repetitive parasite noises (e.g. heavy breathing, mobile phone interference, etc.). For each audio track, a matching label track was then generated using the Sound finder Audacity's built-in module, and each track segment exceeding a defined detection threshold was automatically delimited and tagged with a unique label. Whenever necessary, the automatic detection was manually adjusted to (1) remove irrelevant residual noises (e.g., direct contact with the microphone, noise due to an object falling on the ground during the experiment, etc.); (2) remove participant sounds which are not speeches (e.g. sigh, laughter, self-talking, etc.); (3) improve segmentation accuracy. The final segmentation was exported in a file containing the participant ID, the Step ID, the segments relative time (start and end) in seconds, and the segment label. From that file, we created binary temporal tracks where 1 represents speech and 0 non-speech.

5.1.2 Audio non-verbal cues. We used the binary segmentation to compute the following features grounded on previous work on group's analysis [24, 51]:

- **Total Speaking Turns (per minute)** - TST_i/min: The number of speaking turns accumulated over the entire Step for the participant *i*, divided by the Step total duration;
- **Total Speaking Length (per minute)** - TSL_i/min: The accumulated speaking time for participant *i*, divided by the Step total duration;
- **Average Speaking Turn duration** - AST_i: The average speaking turn duration for participant *i*, with $AST_i = TSL_i / TST_i$;
- **Total of Attempted Interruption (per minute)** - TAI_i/min: The number of attempted interruptions accumulated over the entire Step for the participant *i*, divided by the Step total duration. A turn taking is counted as an attempted interruption if participant *i* starts speaking while participant *j* or *k* was already speaking, resulting in an overlap;
- **Total of Successful Interruption (per minute)** - TSI_i/min: The number of successful interruptions accumulated over the entire Step for the participant *i*, divided by the Step total duration. A turn taking is counted as a successful interruption only if both of these conditions are met:
  1. participant *i* starts speaking while participant *j* or *k* was already speaking, resulting in an overlap;
  2. participant *j* or *k* (who was already speaking when participant *i* started) stops speaking before participant *i* ends his/her turn;
- **Successful Interruption Percentage** - SIP_i: The percentage of successful interruptions over attempted interruptions for the participant *i*, with $SIP_i = TSI_i / TAI_i$.

5.2 Video Features

5.2.1 Quantity of Motion. In complement of the manual annotation of groups behaviour (see section 5.3), we wanted to have an automatic estimation of each participant's non-verbal activity. This measure is the Quantity of Motion (QoM), an estimation of the amplitude of participant's body movements, including head, torso and arm movement. Its computation was based on computer vision algorithms for image color thresholding, as each participant was wearing colored t-shirts and caps, as illustrated in Figure 1. By applying thresholds on the image HSV color components, we obtained the participants' head and upper body silhouette area for each video frame. Then, we computed frame differencing between consecutive silhouette areas: if participants did not move, the resulting difference tended to be close to zero; conversely, if they moved, the difference was higher than zero. QoM is equal to the area resulting from frame differencing, normalized by the area of the participant's silhouette over two consecutive frames. Finally, to remove noise, we applied a Savitzky-Golay low-pass filter (order 1, frame size 75). Figure 2 reports an example of QoM extraction (1st plot from top) for participants of group 2 performing *Step 1*.

5.2.2 Head Position and Rotation. To extract head position and rotation of each participant, we implemented a solution in Python using OpenCV and its Aruco marker detection library [10]. The main limitation of Aruco marker detection approaches is that the markers need to be visible to be tracked, which is sometimes not possible because of the camera field of view or because of the occlusions. In order to overcome this limitation, we took advantage of the 2 cameras of the experimental setup by performing Aruco detection on the 2 video streams and by merging the extracted data. The process worked as follows:

First, calibration was performed on each camera to compensate any distortion. Since Aruco detection works by providing the coordinates of the markers in the camera space, we needed to convert their coordinates into the room space. Three reference markers were positioned on the floor in a way that they were visible by both cameras. These three markers positions in the *Interaction Area* were previously determined and constant throughout the whole data collection, giving us a referential for estimating the positions and rotations of the other markers in the space of the room, instead of the space of the camera. Since each participant was wearing a cap, we placed a unique marker to identify him/her on the top of his/her cap. The top-down view of the cameras allowed them to see the markers most of the time. In case of missing frames, we applied linear interpolation and average smoothing to fill the gaps.

With the South-West corner of the *Interaction Area* as the origin of our referential, our solution eventually extracted for each frame of the videos and for each participant:

1. $HP = (HP_x, HP_y, HP_z)$, the 3D head position in meters;
2. $HR = (HR_x, HR_y, HR_z)$, the 3D head rotation in radians.

Finally, we computed head velocity HV as the magnitude of the 1st derivative of the head position:

$$HV = \sqrt{(\frac{dHP_x}{dt})^2 + (\frac{dHP_y}{dt})^2 + (\frac{dHP_z}{dt})^2} \quad (1)$$

Figure 2 reports an example of head velocity extraction (2nd plot from top) for participants of group 2 performing *Step 1*.

5.2.3 Sample Entropy. While the QoM and HV could be used to estimate the overall non-verbal activity of a participant, we also aimed to measure the degree of regularity of a participant non-verbal activity. We expect such features to be informative about the degree of coordination (as a component of TMS, see Section 2) of the group. To measure it, we computed the Sample Entropy (SampEn) of

	Own task	Other's task	Not focused
Alone	*a* focuses on the task he/she is assigned to	*a* performs *b*'s task without *b*	*a*'s attention is not focused on one of the current activities
Interaction	*a* performs his/her task with the help of *b*	*a* directly helps *b* in his/her task	*a* and *b* discuss or work on topics unrelated to the current task

Table 1: Chronemics features manually annotated from the videos. Participants *a* and *b* are used as examples.

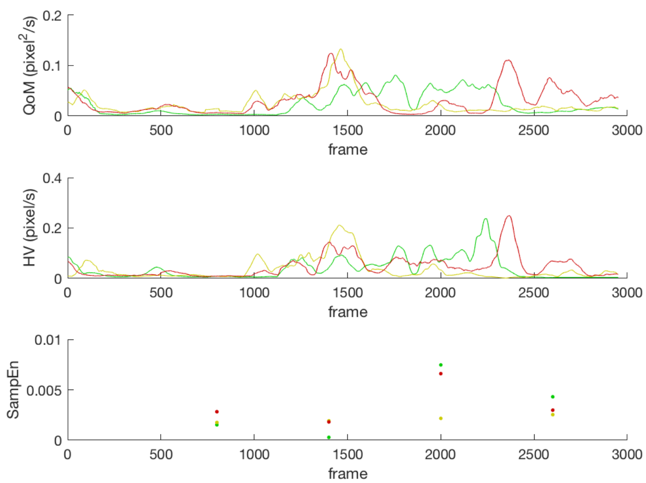

Figure 2: An example of movement features extraction on participants of group 2 performing *Step 1*. From top to bottom: Quantity of Motion (QoM) of participants' torso and head silhouette; Head Velocity (HV); Sample Entropy (SampEn) computed on HV.

each participant's head velocity (HV). SampEn is a particular type of entropy that takes into account the "recent" movement history [19, 49]. We computed SampEn by running the Matlab function implemented in [40] (Embedding Dimension $m = 3$, Tolerance $r = 0.2$) on the head velocity timeseries HV, segmented in time windows of 15 seconds with 3 seconds overlap. Figure 2 reports an example of head movement SampEn extraction (3rd plot from top) for participants of group 2 performing *Step 1*.

5.3 Manual Annotations of Group Behaviours

The manual annotations of the video-recordings concerned non-verbal behavioural cues of the following 3 categories described by Burgoon et al.[11], as we were interested in how they could be indicators of the group's TMS: chronemics, proxemics and kinesics.

5.3.1 Chronemics. Chronemics is the study of the role of time in communication [43]. With regard to our dataset, we were interested in determining how the participants organized their time across the Steps, as it could be informative about the degree of coordination and collaboration among the group members. For each participant, we annotated his/her type of *Focus* (working on his/her task, on the other's task or not focused on any task) and the presence of *Interaction* (working alone or interacting with another participant), resulting in the 6 features listed in Table 1.

5.3.2 Proxemics. Proxemics is defined as "the interrelated observations and theories of humans use of space" [22]. The experimental design used in our experiment promoted spatial arrangements between the group members. As introduced in Section 2, we are interested in the role of F-formations, that are spatial arrangements made by the members of a dyad or a group oriented towards the same object of attention [27]. Inspired from the work of Tong et al. [54] and Zhang and Hung [56], we annotated the following types of F-formations occurring in groups of 3 people: "L", "triangle", "semicircular" or "side-by-side".

5.3.3 Kinesics. Kinesics refers to all types of articular arrangements, positions of body parts in space, or motor performance, including posture, facial expressions and gestures [11]. These non-verbal cues are important in the analysis of learning [21], impression formation [7] and feedback expression [9].

Given the importance of the use and rate of arms' gesturing in expertise expression [7, 53], we are currently manually annotating the following types of gestures (we combined the taxonomies proposed by McNeill [41] and Bonaiuto [8]):

- *Adaptors*: touching behaviors and movements targeted toward the self, objects, or others (e.g. playing with an object, touching hair, interlacing fingers);
- *Ideationals*: non-repetitive complex gestures related to the semantic content of the speech, concrete or abstract (e.g. simulate a rectangular movement to describe a table or open the arms wide-open to talk about its large size);
- *Beats*: simple, repetitive, rhythmic movements that bear no obvious relation to the semantic content of the accompanying speech (e.g. hand swinging).

The manual annotations were performed by 2 annotators with experience in behaviour analysis, by following a consolidated approach already used for example in [14, 53]. Consequently, the first annotator rated all the data while the second one annotated 20% of randomly chosen data. Inter-rater agreement was computed in order to assess the reliability of the annotations of the first rater. Cohen's k indicated substantial/perfect agreement [32]: $k = 0.851$ ($p < .001$) for Chronemics features, $k = 0.684$ ($p < .001$) for F-formations and $k = 0.758$ ($p < .001$) for Gestures. For each of the cues described above, we computed: the total number of occurrences; the time percentage (i.e., obtained by dividing the total time of occurrence by the Step total duration); the average time (i.e., obtained by dividing the total time of occurrence by the total number of occurrences).

5.4 Participants' Self-Assessments

The dataset includes the participants' answers to well-established questionnaires administered during the data collection (5-point Likert scales were used for all questionnaires):

5.4.1 Questionnaire Q1. Completed by participants after signing the consent form, it contained demographics information (age, gender, native language), information about the degree of acquaintance

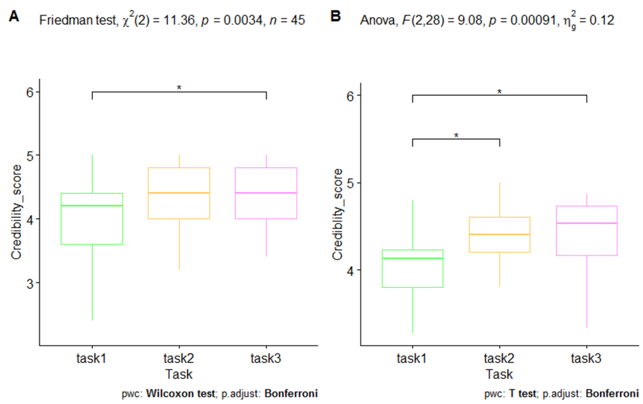

Figure 3: Results of the analyses for Credibility scores at individual level (A) and at group level (B). * stands for $p < 0.05$.

among the group's members, and a self-assessment of their own leadership level. The degree of acquaintance was measured through the following questions (each participant is asked to answer for each other member): *How long have you known participant i?*; *Are you a member of the same sport team or association as participant i?*; *Are you and participant i colleagues?*.

The self-assessment of leadership was measured through items from the "Multifactor leadership questionnaire-short form 6S" by Bass and Avolio [4].

5.4.2 Questionnaire Q2. Completed by participants after the *Encoding* phase, this questionnaire evaluated the impressions of each participant about each of his/her group partners. In particular, impressions about warmth and competence (i.e., the two fundamental dimensions of social cognition [17]) were measured, by using the adjectives identified by Aragones and colleagues (4 for warmth, 4 for competence) [2];

5.4.3 Questionnaire Q3. Completed by participants after each Step in the *Retrieval* phase (see Section 4.1.4), this questionnaire was composed of 2 parts. TMS perception of each member was assessed through Lewis' items [33] measuring the three dimensions of TMS (i.e., Specialization, Credibility, Coordination). For French participants, the French translation of Lewis' questionnaire, validated by Michinov [42] was used. The perceived level of leadership that participants attribute to each of their group partners was measured through 6 items directly inspired by the work of Gerpott et al. [18], and already used and validated by [37].

6 EXPLORATORY ANALYSES

The purpose of this paper is to present the design and characteristics of the WoNoWa dataset. In line with this, we report here the results of exploratory analyses of participants' self-assessments, investigating whether the design of the protocol effectively elicited TMS in the groups and whether there were any differences between the Steps of the *Retrieval* phase. In addition to TMS scores, we also investigated whether leadership perception (see Section 5.4.3) was influenced by the different type of collaborations proposed in Steps 1, 2 and 3 (see Section 4.1.4) and by the different expertise (logistical, mathematical, manual).

6.1 TMS Scores

We computed Cronbach α for each Step (*Step 1*, *Step 2* and *Step 3*) of the *Retrieval* phase and each subscale of TMS questionnaire (i.e., Specialization, Credibility, Coordination). We discarded 2 items from the Coordination subscale since they were not rightly interpreted by participants (their scores were negatively correlated with the others belonging to the same subscale). Since all the α computed on the remaining items indicated acceptable to very good levels of reliability (range between 0.63 and 0.88), for each *Step* we merged items of the same subscale. We analysed TMS scores in two ways: by considering scores of each individual separately and by grouping scores for each group. In all the results presented in this paragraph, no effects of group acquaintance (see Section 5.4.1) were found.

Individual Scores Since data did not meet the assumptions for ANOVA, we ran non-parametric tests on participants' scores for each TMS subscale. Credibility scores were significantly different at different Steps using Friedman test, $X^2(2) = 11.36, p = 0.003$, *Kendall W*= 0.13. No effect of participants' expertise (i.e., the field of expertise they chose during the *Encoding* phase) was found. Pairwise Wilcoxon signed rank test revealed significant differences in Credibility scores between *Step 1* and *Step 3* ($M_1 = 4.03 \pm 0.45, M_3 = 4.38 \pm 0.32; W = 186, p.adj = 0.01$). No differences were found between the Steps for Specialization and Coordination scores.

Group scores. In addition, we computed one score for each group by considering the mean of members' scores. Concerning Specialization scores, since data did not meet the assumptions for ANOVA, we ran a Friedman test with Step as within-subject variable and Specialization scores as dependent variable. Results did not show any significant differences of Specialization scores between the three Steps ($M_1 = 4.45 \pm 0.47, M_2 = 4.52 \pm 0.5, M_3 = 4.47 \pm 0.43$).

Concerning Credibility scores, since the assumptions of normality and sphericity of variances were met, we ran a one-way repeated measures ANOVA, with Step as the within-subject factor and Credibility scores as dependent variable. The Credibility scores were significantly different between the Steps, $F(2, 28) = 9.078, p < 0.0001, \eta^2 = 0.12$, see Figure 3. Post-hoc analyses with a Bonferroni adjustment revealed that the mean Credibility scores were significantly different between *Step 1* and *Step 2* ($M_1 = 4.03 \pm 0.39, M_2 = 4.31 \pm 0.45; t(14) = -2.97, p.adj = 0.01$), between *Step 1* and *Step 3* ($M_1 = 4.03 \pm 0.39, M_3 = 4.38 \pm 0.45; t(14) = -3.42, p.adj = 0.03$) and not between *Step 2* and *Step 3* ($t(14) = -1.17, p.adj = 0.77$).

Concerning Coordination scores, since the assumptions of normality and sphericity of variances were met, we run a one-way repeated measures ANOVA, with Step as the within-subject factor and Coordination scores as dependent variable. No significant differences were found.

6.2 Leadership Scores

We computed Cronbach α for each Step. Since all of them indicated very good levels of reliability (range between 0.86 and 0.89), for each Step we merged the scores of the 6 items about leadership perception in questionnaire Q3 (see Section 5.4.3).

For each participant, we computed a Leadership score as the sum of the merged scores given by the other two members of the group. Thus, the range of the Leadership scores goes from 2 to 10.

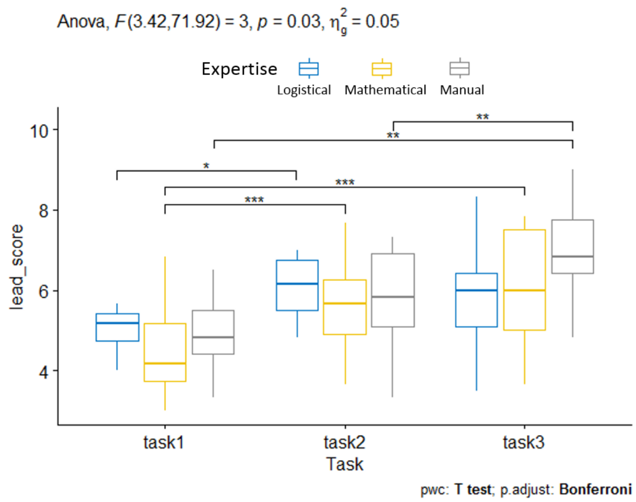

Figure 4: Results for Leadership scores across Expertise after each Step of the *Retrivial* phase.* stands for $p \leq 0.05$, ** stands for $p \leq 0.01$, * stands for $p \leq 0.001$.**

The data met the assumptions of normality, homogeneity of variances and homogeneity of co-variances (all $p - values > 0.1$). We thus ran a two-way 3x3 mixed ANOVA (with Greenhouse-Geisser sphericity corrections), with Step as the within-subject factor, Expertise as the between-subject factor and Leadership score as dependent variable.

Results showed a significant main effect of Step, $F(1.71, 71.92) = 32.27, p < 0.000, \eta^2 = 0.22$, and a two-way interaction between Expertise and Step, $F(3.42, 71.92) = 3, p < 0.03, \eta^2 = 0.05$, on Leadership score.

The main effect of Step on Leadership score has been found to be significant for each of the three Expertise ($F_{Logistical}(2, 28) = 5.19, p.adj = 0.036; F_{Mathematical}(2, 28) = 18.8, p.adj < 0.0000; F_{Manual}(2, 28) = 13.8, p.adj < 0.001$). We first studied the main effect of Step by not considering the participants' Expertise. Pairwise comparisons with Bonferroni correction ($t_{T1-T2}(44) = -5.5, p - adj < 0.0000; t_{T1-T3}(44) = -6.41, p.adj < 0.0000; t_{T2-T3}(44) = -3.01, p.adj = 0.013$) show that the mean Leadership score increased at each Step: $M_{T1} = 4.9 \pm 1, M_{T2} = 5.82 \pm 1.1, M_{T3} = 6.34 \pm 1.15$.

Considering the significant two-way interaction between Expertise and Step, pairwise comparisons showed that Leadership score significantly increased between *Step 1* and *Step 2* for Logistical and Mathematical, between *Step 1* and *Step 3* for Logistical and Manual, and between *Step 2* and *Step 3* for Manual. These results are reported in Figure 4.

6.3 Discussion

The results from exploratory analyses show an influence of the design of WoNoWa on TMS development. Participants' scores about Specialization subscale did not change between the Steps. This result is not surprising if we think that, due to the protocol design, the participants explicitly chose their field of expertise during the first phase of the data collection. The high scores for this subscale during the rest of the Steps (≥ 4.45) confirm that participants clearly distinguished each other's role.

Since we designed this data collection to explore in the future the potential relations between TMS scores and the non-verbal behaviours of group members, the increase of Credibility scores across the Steps of the *Retrieval* phase is a promising result.

We were also expecting an increase in Coordination scores across the Steps, however, from participants' self-assessments, we found no differences between them. One possible explanation could be that the participants completely misunderstood this dimension as we identified a low reliability among the items of this particular subscale (see Section 6.1). Moreover, keeping only 3 items of this subscale might have not given a proper representation of this dimension. An analysis of the other features of our dataset might be more informative about this particular dimension.

Concerning Leadership scores, results showed several differences across the Steps for the different expertise. This suggests that the design of the protocol manipulated the perception of leadership as it generally increased for all participants and all expertise across the Steps.

Our exploratory analyses might not have revealed an increase on all dimensions of the TMS, however, it did show a significant increase on Credibility, Leadership and high values on Specialization and Coordination. Leadership being linked to TMS [3, 31], this is encouraging for the future analysis of our dataset.

7 CONCLUSION AND FUTURE WORK

In this paper we presented WoNoWa, a novel multi-modal dataset of group interactions specifically designed for the study of Transactive Memory System (TMS). WoNoWa offers a rich set of extracted audio and video features, as well as manual annotations of non-verbal behaviours and participants' self-assessments, such as their perception of TMS, gathered trough questionnaires. All these data are available under request for research purposes.

The design of the protocol used for the data collection is strongly based on psychological theories of TMS. Exploratory analyses of participants' self-assessments showed that we successfully manipulated the level of Credibility, that is a component of TMS, and perceived leadership by proposing different types of collaborations.

In the short term we plan to exploit the extracted features and annotations to analyse the relationship between non-verbal behaviour and TMS. In addition, we will analyse the other questionnaires administered during the data collection to check for any effect of participants' characteristics, such as leadership disposition and warmth and competence levels, on TMS and group non-verbal dynamics. We will also include groups' performances in our analyses.

We believe that the WoNoWa dataset represents a precious resource for developing computational models of TMS to predict and enhance group performance. It could be exploited by different communities of researchers on group interactions and would encourage for a real-time analysis of this phenomena.

ACKNOWLEDGMENTS

Authors would thank the Fisher Leadership Initiative (FLI) for sending their MLQ-6S questionnaire, and the Multimedia group of LTCI Télécom Paris for the logistical support.

REFERENCES

[1] Xavier Alameda-Pineda, Jacopo Staiano, Ramanathan Subramanian, Ligia Batrinca, Elisa Ricci, Bruno Lepri, Oswald Lanz, and Nicu Sebe. 2015. Salsa: A novel dataset for multimodal group behavior analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38, 8 (2015), 1707–1720.

[2] Juan I Aragonés, Lucía Poggio, Verónica Sevillano, Raquel Pérez-López, and María-Luisa Sánchez-Bernardos. 2015. Measuring warmth and competence at inter-group, interpersonal and individual levels/Medición de la cordialidad y la competencia en los niveles intergrupal, interindividual e individual. *Revista de Psicología Social* 30, 3 (2015), 407–438.

[3] Daniel G Bachrach and Ryan Mullins. 2019. A dual-process contingency model of leadership, transactive memory systems and team performance. *Journal of Business Research* 96 (2019), 297–308.

[4] Bernard M Bass and Bruce J Avolio. 1992. Multifactor leadership questionnaire-short form 6S. *Binghamton, NY: Center for Leadership Studies* (1992).

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

[6] Cigdem Beyan, Nicolò Carissimi, Francesca Capozzi, Sebastiano Vascon, Matteo Bustreo, Antonio Pierro, Cristina Becchio, and Vittorio Murino. 2016. Detecting emergent leader in a meeting environment using nonverbal visual features only. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*. 317–324.

[7] Beatrice Biancardi, Angelo Cafaro, and Catherine Pelachaud. 2017. Analyzing first impressions of warmth and competence from observable nonverbal cues in expert-novice interactions. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*. 341–349.

[8] Marino Bonaiuto, Augusto Gnisci, and Fridanna Maricchiolo. 2002. Proposta e verifica empirica di una tassonomia dei gesti delle mani nell'interazione di piccolo gruppo. *Giornale italiano di psicologia* 29, 4 (2002), 777–808.

[9] Frederick Douglas Boyd Sr. 2000. *Non-verbal behaviors of effective teachers of at-risk African-American male middle school students*. Ph.D. Dissertation. Virginia Tech.

[10] G. Bradski. 2000. The OpenCV Library. *Dr. Dobb's Journal of Software Tools* (2000).

[11] Judee K Burgoon, Laura K Guerrero, and Valerie Manusov. 2016. *Nonverbal communication*. Routledge.

[12] Laura Cabrera-Quiros, Andrew Demetriou, Ekin Gedik, Leander van der Meij, and Hayley Hung. 2018. The MatchNMingle dataset: a novel multi-sensor resource for the analysis of social interactions and group dynamics in-the-wild during free-standing conversations and speed dates. *IEEE Transactions on Affective Computing* (2018).

[13] Lei Chen, R Travis Rose, Ying Qiao, Irene Kimbara, Fey Parrill, Haleema Welji, Tony Xu Han, Jilin Tu, Zhongqiang Huang, Mary Harper, et al. 2005. VACE multimodal meeting corpus. In *International Workshop on Machine Learning for Multimodal Interaction*. Springer, 40–51.

[14] Mathieu Chollet, Magalie Ochs, and Catherine Pelachaud. 2014. From non-verbal signals sequence mining to bayesian networks for interpersonal attitudes expression. In *International conference on intelligent virtual agents*. Springer, 120–133.

[15] Juan Abdon Miranda Correa, Mojtaba Khomami Abadi, Niculae Sebe, and Ioannis Patras. 2018. Amigos: A dataset for affect, personality and mood research on individuals and groups. *IEEE Transactions on Affective Computing* (2018).

[16] Y Den. 2018. F-formation and social context: how spatial orientation of participants' bodies is organized in the vast field. In *Proceedings of LREC 2018 Workshop: Language and Body in Real Life (LB-IRL2018) and Multimodal Corpora (MMC2018) Joint Workshop*. 35–39.

[17] Susan T Fiske, Amy JC Cuddy, and Peter Glick. 2007. Universal dimensions of social cognition: Warmth and competence. *Trends in cognitive sciences* 11, 2 (2007), 77–83.

[18] Fabiola H Gerpott, Nale Lehmann-Willenbrock, Sven C Voelpel, and Mark van Vugt. 2019. It's Not Just What is Said, but When it's Said: A Temporal Account of Verbal Behaviors and Emergent Leadership in Self-Managed Teams. *Academy of Management Journal* 62, 3 (2019), 717–738.

[19] RB Govindan, JD Wilson, H Eswaran, CL Lowery, and H Preißl. 2007. Revisiting sample entropy analysis. *Physica A: Statistical Mechanics and its Applications* 376 (2007), 158–164.

[20] Rebecca Grossman, Sarit B Friedman, and Suman Kalra. 2017. Teamwork processes and emergent states. *The Wiley Blackwell handbook of the psychology of team working and collaborative processes* (2017), 243–269.

[21] Marianne Gullberg. 2008. Gestures and second language acquisition. In *Handbook of cognitive linguistics and second language acquisition*. Routledge, 276–305.

[22] Edward Twitchell Hall. 1966. *The hidden dimension*. Vol. 609. Garden City, NY: Doubleday.

[23] Mattias Heldner and Jens Edlund. 2010. Pauses, gaps and overlaps in conversations. *Journal of Phonetics* 38, 4 (2010), 555–568.

[24] Hayley Hung and Daniel Gatica-Perez. 2010. Estimating cohesion in small groups using audio-visual nonverbal behavior. *IEEE Transactions on Multimedia* 12, 6 (2010), 563–575.

[25] Dinesh Babu Jayagopi, Sileye Ba, Jean-Marc Odobez, and Daniel Gatica-Perez. 2008. Predicting two facets of social verticality in meetings from five-minute time slices and nonverbal cues. In *Proceedings of the 10th international conference on Multimodal interfaces*. ACM, 45–52.

[26] Adam Kendon. 1990. *Conducting interaction: Patterns of behavior in focused encounters*. CUP Archive.

[27] Adam Kendon. 2010. Spacing and orientation in co-present interaction. In *Development of multimodal interfaces: Active listening and synchrony*. Springer, 1–15.

[28] Steve WJ Kozlowski and Daniel R Ilgen. 2006. Enhancing the effectiveness of work groups and teams. *Psychological science in the public interest* 7, 3 (2006), 77–124.

[29] S. W. J. Kozlowski. 2015. Advancing research on team process dynamics: Theoretical, methodological, and measurement considerations. *Organizational Psychology Review* 5, 4 (2015), 270–299.

[30] Wessel Kraaij, Thomas Hain, Mike Lincoln, and Wilfried Post. 2005. The AMI meeting corpus. (2005).

[31] Kibum Kwon and Daeyeon Cho. 2016. How transactive memory systems relate to organizational innovation: the mediating role of developmental leadership. *Journal of Knowledge Management* (2016).

[32] J Richard Landis and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *biometrics* (1977), 159–174.

[33] Kyle Lewis. 2003. Measuring transactive memory systems in the field: scale development and validation. *Journal of applied psychology* 88, 4 (2003), 587–604.

[34] Kyle Lewis and Benjamin Herndon. 2011. Transactive memory systems: Current issues and future research directions. *Organization Science* 22, 5 (2011), 1254–1265.

[35] Diane Wei Liang, Richard Moreland, and Linda Argote. 1995. Group versus individual training and group performance: The mediating role of transactive memory. *Personality and social psychology bulletin* 21, 4 (1995), 384–393.

[36] Jenny Liao, Nerina L Jimmieson, Anne T O'Brien, and Simon LD Restubog. 2012. Developing transactive memory systems: Theoretical contributions from a social identity perspective. *Group & organization management* 37, 2 (2012), 204–240.

[37] 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 (2020), 124185–124203.

[38] Nadia Mana, Bruno Lepri, Paul Chippendale, Alessandro Cappelletti, Fabio Pianesi, Piergiorgio Svaizer, and Massimo Zancanaro. 2007. Multimodal corpus of multi-party meetings for automatic social behavior analysis and personality traits detection. In *Proceedings of the 2007 workshop on Tagging, mining and retrieval of human related activity information*. 9–14.

[39] Michelle A Marks, John E Mathieu, and Stephen J Zaccaro. 2001. A temporally based framework and taxonomy of team processes. *Academy of management review* 26, 3 (2001), 356–376.

[40] Víctor Martínez-Cagigal. [n.d.]. *Sample Entropy*. https://uk.mathworks.com/matlabcentral/fileexchange/69381-sample-entropy

[41] David McNeill. 1992. *Hand and mind: What gestures reveal about thought*. University of Chicago press.

[42] Estelle Michinov. 2007. Validation de l'échelle de mémoire transactive en langue française et adaptation au contexte académique. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology* 57, 1 (2007), 59–68.

[43] Nina-Jo Moore, Mark Hickson, and Don W Stacks. 2014. *Nonverbal communication: Studies and applications*. Oxford University Press.

[44] Richard L Moreland. 2010. Are dyads really groups? *Small Group Research* 41, 2 (2010), 251–267.

[45] Richard L Moreland and Larissa Myaskovsky. 2000. Exploring the performance benefits of group training: Transactive memory or improved communication? *Organizational behavior and human decision processes* 82, 1 (2000), 117–133.

[46] Richard L Moreland and L Thompson. 2006. Transactive memory: Learning who knows what in work groups and organizations. In *Small groups: Key readings*. Psychology Press New York, NY, 327–346.

[47] Kumiko Murata. 1994. Intrusive or co-operative? A cross-cultural study of interruption. *Journal of Pragmatics* 21, 4 (1994), 385–400.

[48] Radoslaw Niewiadomski, Maurizio Mancini, Tobias Baur, Giovanna Varni, Harry Griffin, and Min SH Aung. 2013. MMLI: Multimodal multiperson corpus of laughter in interaction. In *International Workshop on Human Behavior Understanding*. Springer, 184–195.

[49] Joshua S Richman and J Randall Moorman. 2000. Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology* 278, 6 (2000), H2039–H2049.

[50] 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*.

[51] Dairazalia Sanchez-Cortes, Oya Aran, Marianne Schmid Mast, and Daniel Gatica-Perez. 2011. A nonverbal behavior approach to identify emergent leaders in small groups. *IEEE Transactions on Multimedia* 14, 3 (2011), 816–832.

[52] Albert E Scheflen and Norman Ashcraft. 1976. *Human territories: How we behave in space-time.* Prentice-Hall.
[53] Abishek Sriramulu, Jionghao Lin, and Sharon Oviatt. 2019. Dynamic Adaptive Gesturing Predicts Domain Expertise in Mathematics. In *2019 International Conference on Multimodal Interaction.* 105–113.
[54] Lili Tong, Audrey Serna, Simon Pageaud, Sébastien George, and Aurélien Tabard. 2016. It's not how you stand, it's how you move: F-formations and collaboration dynamics in a mobile learning game. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services.* 318–329.
[55] Daniel M Wegner. 1987. Transactive memory: A contemporary analysis of the group mind. In *Theories of group behavior*. Springer, 185–208.
[56] Lu Zhang and Hayley Hung. 2016. Beyond f-formations: Determining social involvement in free standing conversing groups from static images. In *Proceedings of the IEEE conference on computer vision and pattern recognition.* 1086–1095.