Revealing the Secret to Successful Virtual Meetings: How Personality, Social Skills, and More Impact Conversational Involvement

What is the relation between lexical alignment and group involvement in online meetings?

Matthijs Rijm

Supervisor(s): Catholijn Jonker[1], Masha Tsfasman[1]

[1]EEMCS, Delft University of Technology, The Netherlands

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 25, 2023

Name of the student: Matthijs Rijm Final project course: CSE3000 Research Project
Thesis committee: Catholijn Jonker, Masha Tsfasman, Gosia Migut

Abstract

Involvement in an online setting has become increasingly more important in a rapidly digitizing world. Previous research has suggested that lexical alignment is an effective method for increasing interpersonal involvement. Multi-modal systems exist for analysing involvement, but the lexical alignment modality is not commonly used in these. This study examines the lexical alignment modality as a predictive measure of involvement in online group meetings. By analyzing participation indicators in virtual meetings, the aim is to identify moments of high involvement. Results indicate that lexical alignment does not correlate with moments of high involvement. The findings contribute to the scientific understanding of leveraging the verbal modality for analyzing and organizing online meetings.

KEYWORDS: engagement, involvement, lexical alignment, other-repetitions, group meetings, online meetings

1 Introduction

As the Covid-19 pandemic is transitioning to a less severe endemic form [1], the remnants of this virus are still visible in most professional careers. Study has shown an increase in virtual meetings [8], an increase that is expected to be structural [17]. The increase in the amount of virtual meetings has also introduced the concept of 'Zoom-fatigue' [4], where Zoom can be used interchangeably with other virtual conferencing platforms. It is thus important to assess how to combat this 'Zoom fatigue'. Research suggests to foster engagement as a way to combat this 'Zoom fatigue' [13]. The goal of this paper is to find whether lexical alignment is associated with higher group involvement and thus engagement. The results of this paper could then contribute to organizing more effective virtual meetings.

Involvement can be referred to as the level a participant is interacting with other participant(s). Alignment is the support of an ongoing activity, with lexical alignment seen as a way to show involvement and to establish interpersonal involvement [20]. Earlier research shows that lexical alignment is correlated to involvement [3]. This is however only researched in a dyadic setting and not in a group setting, which is therefore what this research will look into. This will be done by looking at the other-repetitions, which is when one participant repeats words used by another participant, within a certain amount of time. The amount of other-repetitions is then analysed together with annotated group involvement data using linear regression to look for the correlation between the two concepts.

Two research questions are posed to look at the correlation between lexical alignment and group involvement.

- Does lexical alignment have a positive influence on group involvement in online meetings?
- Can lexical alignment predict group involvement in online meetings?

The hypothesis for the first question is that more lexical alignment means more group involvement. The answer to the second research question would then be that a linear regression model would be able to predict the group involvement.

The structure of the rest of the paper will be as follows. Section 2 will show the literary background with previous research that form the groundwork for this research. In section 3 the corpus will be discussed, the annotation of the data will also be discussed and finally the methods for analysing the data will be described. In section 4 the results retrieved using these methods will be reviewed. Then section 5 will discuss how responsible research has played a part in this research. Section 6 will discuss the results and the methods that were used, including some recommendations for future work. Section 7 will state the conclusions from this research and finally section 8 will state some acknowledgements to people that have helped during the course of this research.

2 Background

Involvement, interchangeably used with engagement, can be defined as the importance a participant in an interaction places on the desire to remain with the other participant(s) and carry on the interaction [12]. Previous studies have looked at the role of involvement, both between human-human [10] and human-agent interaction [15]. As these interactions also occur within groups, research has also focused on group engagement, which was described by [11] as "a group variable that is calculated as the average of the degree to which individual people in a group are engaged in spontaneous, non-task-directed conversations." Research by Kamaludin and Mohd [7] also looked into the significance of involvement in online collaborative settings. They investigated the effects of many elements on involvement and collaborative results in an online learning setting.

Alignment can be defined as the support of an ongoing activity [6]. Lexical alignment in conversation occurs when individuals repeatedly use the same terms to refer to an object. This phenomenon is influenced by historical factors, including past references, shared conceptualizations, and the gradual simplification or abandonment of conceptual pacts [2]. In previous studies, lexical alignment specifically has been correlated with the conveying of emotional stance [18]. Furthermore, a research has suggested positive correlation between lexical alignment and task performance in neurotypical individuals, concluding that interventions promoting shared vocabulary could enhance communication abilities [16].

Previous research has utilized a multimodal system in groups to predict involvement [9]. In this research the lexical alignment modality is chosen as that is not commonly used in multimodal systems to see if it can be used in addition to the modalities already used in multimodal systems.

Tannen tells in [20] that repetition gets used as a way to establish interpersonal involvement. Research has been done in which lexical alignment is correlated with involvement [3]. There is however no use of annotation of involvement.

The research gap in current literature is the fact that lexical alignment in relation to involvement has not yet been looked at in a group setting, which is therefore what this research looks at.

3 Methodology

To answer the research questions we need a couple of things and as this will be a long section it is useful to split it into three substantial parts. The first subsection will introduce the MEMO corpus used in this research and will explain why it is relevant. The second subsection will discuss how the MEMO corpus was annotated and why it was annotated that way. The last subsection will talk about which statistics are used to analyse the data, why they were chosen and how they were implemented.

3.1 The Corpus

The MEMO corpus is a collection of audio and video recordings from online meetings consisting of four to six people. One of these people is the moderator who helps keep the conversation going. There are a total of fifteen groups with three group sessions each. There are also validated transcripts available in Video Text Track (VTT) format. These were obtained using automatic speech recognition methods and corrected by hand where needed. This is all part of the corpus. Unfortunately there are four transcripts missing, of which three are of the same group.

For this research the MEMO corpus is interesting because it contains small groups in online meetings, which is a natural setting. The data was collected during and after the coronavirus, which means most people were most likely already familiar with online meetings. The conversations are also reasonably natural, which makes it a good corpus to analyse for involvement. There was no comparison to other corpora as it was handed to us by our supervisor and responsible professor. This meant this corpus was inherent to this research project.

It is also good to think about the biases that are already ingrained in this dataset, these are discussed in section 5.

3.2 Annotations

As a research group we have made some decisions on how we wanted to annotate this dataset. To start annotating we first needed a definition of involvement that we wanted to use for our annotation of group involvement. After some discussion we settled on the following definition for group involvement: “The perceived degree of interest or involvement of the majority of the group.” [5]. As this still contained the word involvement we also chose to define conversational involvement as follows: “The process by which interactors start, maintain, and end their perceived connections to each other during an interaction.” [15].

Using this definition of engagement every member of the group annotated a part of the data. This was done using a Likert scale from 1-5. This was done to have a good layer of granularity while also not giving too much range, which would make the data a lot harder to use, as there are four people annotating. We also decided upon 5 second segments to annotate as this gives enough insight into the conversation to get an idea of the group involvement while still not averaging out too much.

The annotation of first group was skipped, because there was a mistake in the recording, which made not the whole group visible in the first session. This made it impossible to annotate group involvement and as multiple people needed data per group it was decided that all sessions of this group were dropped. This gave more annotations for the other groups and therefore a better view of the involvement in those groups.

The next important decision was concerning the division and stratification of the data. We came to the conclusion that we wanted every member of the group to annotate every session of every group, but for it to not cost an extraordinary amount of time, so that we still had enough time to do our data exploration and analysis. We therefore decided that everyone would get random five second time-slices in every session adding up to about five minutes of annotated time per session. We also made sure to have at least ten percent overlap between annotators to make sure we can compute inter-rater reliability. It was chosen to use the ICC3k method described in [14] to look at each pair of annotators. This was chosen because there was a fixed set of raters, which were considered a fixed effect. As there were four raters the k-rater version was chosen.

3.3 Statistics

The collected data was analyzed using linear regression analysis, as the hypothesis for the research question is a linear correlation.

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \tag{1}$$

The variables β_0 and β_1 were contrasted in which β_0 was the variable defined for the amount of repetitions within a 5 second time slice. The variable β_1 defines the level of group involvement for that same 5 second time slice on a Likert scale of 1-5. Performing this linear regression analysis results in a dependent variable Y_i , which should be an indicator for high group involvement moments, according to the hypothesis to the research question.

Experimental Setup

The first step was to process the transcript data. The first necessity was to remove all stop-words (according to the nltk stopwords package) and garbage words. The next step was to POS-tag and lemmatize every word that was left. This was done using the NLTK package in Python. Both of these previous steps were done because [3] also follows these steps. These steps are important because they remove garbage words and words that fill up sentences, but do not contain any value. Furthermore, the POS-tagging and lemmatizing are important, because they bring the word back to its base word. This is important, because this way if the word gets repeated in another form (e.g. past tense instead of present tense for a verb) it still gets picked up as an other-repetition.

Then the other-repetitions needed to be processed. The data that was gathered during this process is as follows: The word that was repeated, the speaker, the previous speaker using the word, the five second window the word was said in, the time between the start of the two timeframes the word and its previous occurrence were said in and the group and session the word was used in. Occurences where the previous speaker is the same as the current speaker were dropped as these are not other-repetitions, but just repetitions and therefore not of interest. After that, for each timeslice it is counted how many repetitions there are and the according annotation data is retrieved. If no annotation data was available the timeslice was dropped from the dataset. If more than one annotation was available the annotations were averaged. The amount of repetitions has been limited by a maximum amount of time between the use of the words. It was chosen to do this in a range from ten to sixty seconds, in steps of ten seconds. So the repetitions were counted for amount of repetitions within ten seconds, twenty seconds, thirty seconds and so on up to sixty seconds. This was chosen as no other methods of systematically analysing other-repetitions was found in literature. This method of analysing captures repetitions even if someone else spoke in between the original speaker of the word and the repeater, of which there is a good possibility in online group meetings. It also gives a view into if fast repetitions have a different impact on the group involvement than slower repetitions.

4 Results

First off, after computing the other-repetitions, 1 shows how many other-repetitions each session contains.

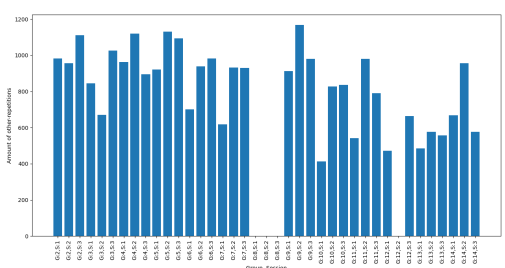

Figure 1: Division of the other-repetitions over the different groups and sessions

After this, six different analyses were performed, in which the distinction between the analyses are the maximum amount of time between the other-repetitions which are counted. The other-repetitions were counted for other-repetitions within ten seconds of its original saying, twenty seconds and so on up to sixty seconds.

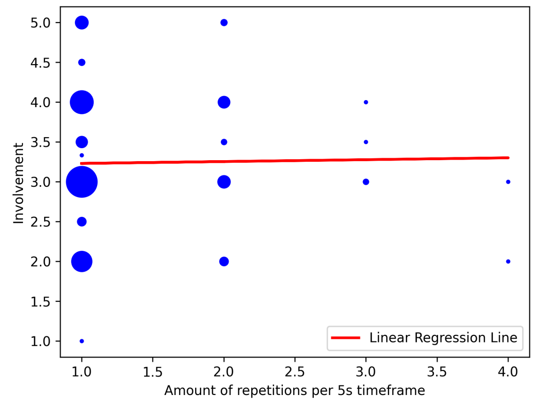

Figure 2: Repetitions within 10 seconds to involvement

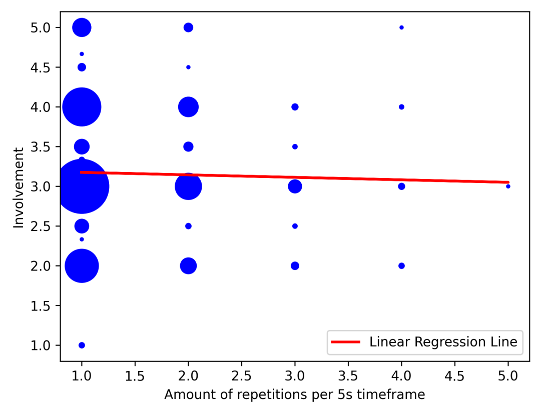

Figure 3: Repetitions within 20 seconds to involvement

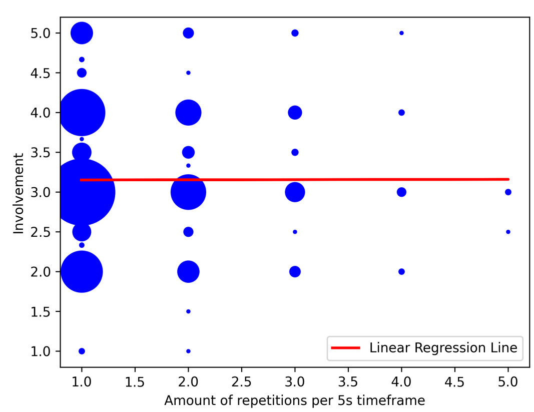

Figure 4: Repetitions within 30 seconds to involvement

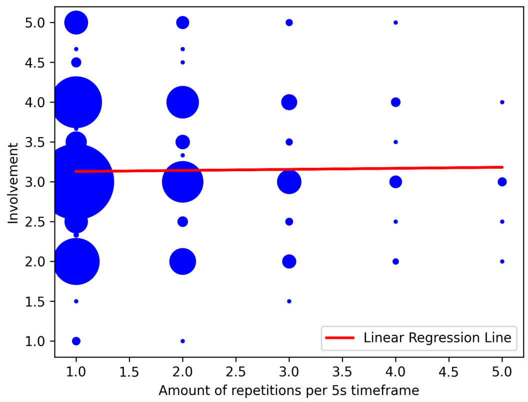

Figure 5: Repetitions within 40 seconds to involvement

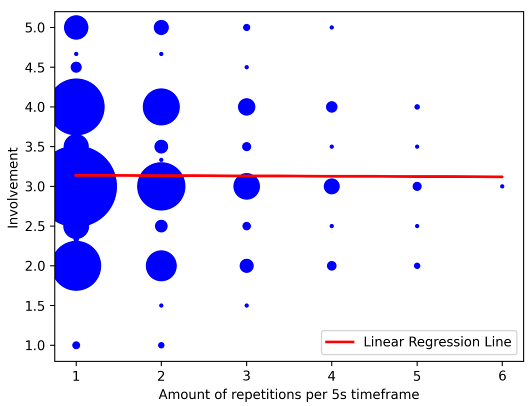

Figure 6: Repetitions within 50 seconds to involvement

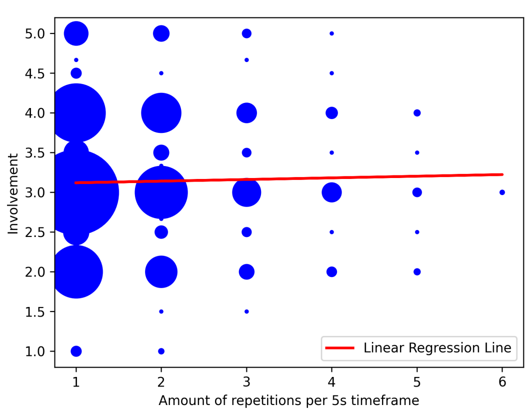

Figure 7: Repetitions within 60 seconds to involvement

After performing the linear regression analysis results show that there is no clear linear correlation between the amount of repetitions within a five second time interval and the involvement of the group.

The results of the linear regression are as follows:

Time in between repetitions (s)	**LR Coefficient**	**MSE**	**R2**
10	0.0234136	4.35	-17.80
20	-0.03123905	3.90	-10.04
30	0.00200735	3.67	-7.53
40	0.01328554	3.49	-5.77
50	-0.00388725	3.44	-5.06
60	0.02068029	3.29	-4.21

Table 1: Results for the different linear regression analyses for the different maximum times between repetitions

As the R2 values are all negative this suggests that the mean is a better predictor of the involvement than the linear model. The high Mean Square Error numbers are also quite high for a scale of 1 to 5. The figures also visually show the fact that there is no linear correlation between these variables.

All the graphs show that a higher number of other-repetitions does not correspond to high involvement. It however also does not correspond to low involvement.

The results of the performed analysis do not answer the first research question: ”Does lexical alignment have a positive influence on group involvement?” as there was no statistically significant correlation found using linear regression. As this was the outcome of the first question, the second question, , also gets a negative answer. The fact that there is no linear correlation between lexical alignment and group involvement makes it not possible to predict involvement.

5 Responsible Research

The first thing to be very careful with is biases. These can come up in different ways, like with the annotation of the data by four different people or by having biases in the dataset already. How we as a research group tried to mitigate these was by always bringing them up when decisions had to be made. This way we always kept the biases in the back of our mind when making choices.

The first origin of bias to think about is the bias in the original data. As this is a small amount of people (about 75) only a small sample of the population is represented. There might also be biases surrounding the subject which is discussed as the coronavirus might be a sensitive topic.

The next origin of bias is the annotation of the data, which is an inherently subjective process. First of all every annotator has a different level of empathy, which makes them interpret the people differently, making them give different scores. Furthermore there is also a bias in myself and at least one other group member which makes us calibrate to each group, leveling out our annotations over the group. This means that a group that might be on average higher on involvement still gets an average score because a ’3’ just means that for that group the involvement is average. Our group also consisted of 3 males and only 1 female, which is also a possible source of bias.

Lastly there might be a bias in the analysis of the data. This can be mostly prevented by choosing the right tools to analyse the data. The methods used are not in itself prone to bias, but the way they are applied are. Although no statistically significant correlation is found it is still important to think about the biases that could be at play. However, as the statistical method used does not seem to be the correct one, this can be considered a bias that is influencing the outcome of this research.

Another significant thing we as a research group had to think about was the fact that we had 4 sets of annotations which we wanted to combine. This is a hard thing to do as all of us have our own intrinsic biases while annotating making the data hard to conglomerate. Because we wanted to do this responsibly we had to invest significant research time into this, only to be still left with only reasonably reliability. This is unfortunate, but it would be good to invest more time into this, should this be done again.

It is also good to think about the origin of the data. As we have raw video data and we are doing the annotations ourselves, there is not really anything to question about this. It is however good to think about how the data was gathered. At the start the moderator asks for consent to be their moderator and people have also signed a consent form. There are however the transcripts, which are subject to automatic speech recognition and manual correction. From the parts that were looked at this is definitely not perfect, but it was felt that this does not impact the data enough to warrant spending hours to check and correct all of the transcripts.

Lastly it is also important to think about the reproducibility of the experiment. The Experiment Setup section gives a very clear account of what exactly was done with the data, so that it is not hard to reproduce this experiment if the reader has basic programming knowledge and access to the data.

6 Discussion

This research looked into the relation between lexical alignment and group involvement using two research questions.

The first research question is: "Does lexical alignment have a positive influence on group involvement in online meetings?". The Mean Squared Error and R2 values in 1 show that there is no linear correlation between these two parameters. Thus, more other-repetitions does not mean more group involvement. This is unexpected as [19] suggests repeating other people's words does show people's involvement. The reasons this does not align may be one of the following. First of all, it could be that more repetition does not mean more group involvement and that a few other-repetitions could also be enough to show involvement, so that there does not exist a linear correlation between the two concepts. It could also be that people do this in dyadic conversations, but not or a lot less in a group. Another possibility is that if one person is showing involvement this does not necessarily mean that the group is involved, this is supported by the fact that there is research into how individual engagement relates to group involvement [10][9].

The second research question that was posed is: "Can lexical alignment predict group involvement in online meetings?" As the answer to the previous research question was no, the answer to this question is also no within the limits of this research. It might be that it is possible to predict group involvement in online meetings using lexical alignment, but not using the methods used in this research.

As all data is treated equally, while it may not be, this also needs to be thought about. In 1 you can see that the data is not completely equally distributed. This might be correlated to things like personalities of the people in the group, which is also part of the parallel research done by someone else in the same research group as the author. Future work could look into what might be the cause of this.

It is also good to look at the limitations of this research, because this can also help future research. First of all, the annotation data we have produced as a project group was hard to use. This is because we did not align our annotation strategy enough as a group. If we would have tried to align our annotation strategies more this might have also lead to a better definition of conversational involvement, as the one we used (The process by which interactors start, maintain, and end their perceived connections to each other during an interaction." [15]) was not very usable with our annotation division of random stratification over time throughout the session. The result of this was a not very good Intraclass Correlation score (ICC). Using the methods given in [14] and looking at the ICC3k values (see 2) shows that the data has reasonable overlap, but not enough to say that each rater is an extension of the other raters, which made it hard to trust the data. It was however still chosen to use this data as the ICC values are acceptable, but it made it harder to look into for example the length of peaks and troughs in involvement.

Raters	ICC3k value
1 & 2	0.584122
1 & 3	0.759784
1 & 4	0.526881
2 & 3	0.595724
2 & 4	0.620876
3 & 4	0.653286

Table 2: ICC3k values between each pair of raters

As the proposed method for analysing other-repetitions was not taken from other research this could also be a part of the fact that no correlation was found. It might therefore be interesting for future work to only look at other-repetitions in the immediate reply to another participant.

Lexical alignment is not only other-repetitions, but also shared vocabulary. A definition for this is given in [3], but this is only for the shared vocabulary between two people. Interesting future work that can be done is to develop a definition for a group of people so this can also be analysed.

Other future work that can be done is looking into if repetition or use of certain words (e.g. emotions) can be correlated to involvement. Another detail that was not looked at in this research that might be of interest is how far into the meeting or in which of the three sessions of the group the other-repetitions took place. It could for example be the case that multiple other-repetitions in short succession increases involvement throughout these other-repetitions, therefore this is possible future work.

What becomes clear through this is that just looking at other-repetitions one can not predict involvement. After this research has been conducted it also has become clear that this is quite a big step from the available literature with a lot of gaps trying to be filled in at the same time, the jump from dyadic to group meetings, the use of annotated data to analyse data on a large scale and the creation of a new way of trying to analyse the data. This is definitely a point of attention and something the author can improve upon in future research.

7 Conclusion

The goal of this research paper was to research whether the analysis of lexical alignment can be seen to have a function in improving involvement of interaction in groups within online meetings. After the conducted linear regression analysis, results showed that other-repetitions do not correlate to higher involvement within groups in online meetings. This was shown by the Mean Squared Error (MSE) and R2 values, as the MSE values were high and the R2 values were negative. It can therefore also be concluded that other-repetitions can not be used to predict involvement in online group meetings.

8 Acknowledgements

The author wants to thank a couple of people that were essential during this research. First of all, my project group. They were crucial as we did all the annotation together, but Andy Li and Sebas Holtreman also built the script for the annotation stratification. Ana Hobai also did the ICC value calculations which were very insightful. Every time we is used throughout this paper, that is speaking about the project group. Someone who was also very important is Masha Tsfasman for giving useful feedback and pushing in the right direction. The last person to thank is the reponsible professor, Catholijn Jonker, who was very clear in what she wanted from us, to learn the scientific method and that we help each other.

References

[1] Michela Biancolella, Vito Luigi Colona, Ruty Mehrian-Shai, Jessica Lee Watt, Lucio Luzzatto, Giuseppe Novelli, and Juergen K.V. Reichardt. Covid-19 2022 update: transition of the pandemic to the endemic phase. volume 16. BioMed Central Ltd, 12 2022.

[2] Clark H. H. Brennan, S. E. Conceptual pacts and lexical choice in conversation. journal of experimental psychology. *Learning, memory, and cognition*, 22:1482–1493, 06 1996.

[3] Sabrina Campano, Jessica Durand, and Chloé Clavel. Comparative analysis of verbal alignment in human-human and human-agent interactions, 2014.

[4] G. Fauville, M. Luo, A.C.M. Queiroz, J.N. Bailenson, and J. Hancock. Zoom exhaustion fatigue scale. *Computers in Human Behavior Reports*, 4:100119, 2021.

[5] Daniel Gatica-Perez, Lain McCowan, Dong Zhang, and Samy Bengio. Detecting group interest-level in meetings. volume I. Institute of Electrical and Electronics Engineers Inc., 2005.

[6] Jan Gorisch, Bill Wells, and Guy J Brown. Pitch contour matching and interactional alignment across turns: An acoustic investigation. *Language and Speech*, 55(1):57–76, 2012.

[7] Puteri Kamaludin, Suhaili Mohd Yusof, Sofwah Md Nawi, Nor Afifa Nordin, Nursyafiqah Zabidin, and Noridah Sain. Group online engagement: An analysis from tuckman model. *International Journal of Academic Research in Business and Social Sciences*, 12, 09 2022.

[8] Katherine A. Karl, Joy V. Peluchette, and Navid Aghakhani. Virtual work meetings during the covid-19 pandemic: The good, bad, and ugly. *Small Group Research*, 53(3):343–365, 2022.

[9] Catharine Oertel. Towards developing a model for group involvement and individual engagement. pages 349–352, 2013.

[10] Catharine Oertel and Giampiero Salvi. A gaze-based method for relating group involvement to individual engagement in multimodal multiparty dialogue. pages 99–106, 2013. our individual engagement annotations are based on the rankings of the participants themselves and we are using a predefined annotation scheme rather than relying solely on the third-party annotators' intuitions.

[11] Catharine Oertel, Stefan Scherer, and Nick Campbell. On the use of multimodal cues for the prediction of degrees of involvement in spontaneous conversation. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, pages 1541–1544, 08 2011.

[12] Isabella Poggi. Mind, hands, face and body : a goal and belief view of multimodal communication, 2007.

[13] Luc Rubinger, Aaron Gazendam, Seper Ekhtiari, Nicholas Nucci, Abbey Payne, Herman Johal, Vikas Khanduja, and Mohit Bhandari. Maximizing virtual meetings and conferences: a review of best practices. *International Orthopaedics*, 44:1461–1466, 8 2020.

[14] Patrick E Shrout and Joseph L Fleiss. Intraclass correlations : Uses in assessing rater reliability, 1979.

[15] Candace Sidner and Neal Lesh. Engagement when looking: behaviors for robots when collaborating with people, 2003.

[16] Eigsti I. M. Stabile, M. Lexical alignment and communicative success in autism spectrum disorder. *Journal of speech, language, and hearing research*, 65:4300–4305, 11 2022.

[17] Willem Standaert, Steve Muylle, and Amit Basu. How shall we meet? understanding the importance of meeting mode capabilities for different meeting objectives. *Information Management*, 58(1):103393, 2021.

[18] Jan Svennevig. Other-repetition as display of hearing, understanding and emotional stance. *Discourse studies*, 6(4):489–516, 2004.

[19] Deborah Tannen. How men and women use language differently in their lives and in the classroom. *The Education Digest*, 57(6):3, 1992.

[20] Deborah Tannen. *Talking voices: Repetition, dialogue, and imagery in conversational discourse*, volume 26. Cambridge University Press, 2007.