

MEET: a corpus for studies of meeting characteristics

Ghazaleh Esfandiari & Jens Edlund

KTH Speech, Music & Hearing

Stockholm, Sweden

edlund@speech.kth.se, geb@kth.se

Abstract

We present the MEET corpus. The corpus was collected with the aim of systematically studying the effects of physical, digital or hybrid work meetings on collaborative decision-making meetings. It consists of 10 sessions, where each session contains three recordings: a physical, a digital and a hybrid meeting between three participants. The participants are working on a different survival ranking task during each meeting. Each meeting is around 15 minutes, resulting in 450 minutes of conversation altogether. We also present the first version of an annotation scheme designed specifically to target our research questions. The recordings are currently being transcribed and annotated in accordance with this scheme.

1. Introduction

As COVID-19 was declared a global pandemic, widely implemented social distancing measures were taken to control the pandemic. Various types of human social interactions (e.g., shopping, banking, learning, meeting, and entertaining) were shifted from dominantly offline to dominantly online. In other words, the enforced social isolations in the physical world significantly increased humans' social interactions in the cyber world (Yan, 2020).

In 2020, Zoom added >2 million active users monthly (*Video Conferencing Market Size, Share & Covid-19 Impact Analysis*, 2021). Having all its down sides, it provided a golden opportunity for researchers to study

this shift from different angles and perspectives. The working environments changed as distant meetings quickly replaced face-to-face meetings everywhere. But how can we describe these distant meetings? Are they as effective as the physical meetings? In what way are they different? How can we evaluate the effectiveness of such meetings?

As researchers interested in face-to-face communication, we were motivated by such questions to study distant meetings. The aim was to describe, analyze and compare the structure of human interaction in physical, digital and hybrid meetings and also to develop a method by which meetings could be objectively compared with respect to their outcome. For this purpose the MEET corpus was constructed. It consists of 10 full sessions. Each session contains three recordings of physical, digital and hybrid meetings between three participants. The participants are working on survival ranking tasks during the meetings. Each meeting is around 15 minutes, resulting in 450 minutes of conversation altogether. The recordings are currently being transcribed and annotated on various levels.

2. Background

2.1. Meetings

Merriam-Webster defines “meeting” as “an act or process of coming together”¹. Others include a measure of deliberation, such as Google’s English Dictionary by Oxford Languages²: “an assembly of people for a particular purpose”. Here, we are concerned with this latter kind of purposeful meeting - the type of gathering Goffman describes as “the natural unit of social organization in which focused interaction occurs”, where focused interaction is “when people effectively agree to sustain for a time a single focus of cognitive and visual attention, as in a conversation, a board game, or a joint task sustained by a close face-to-face circle of contributors” (Goffman, 1961). We will however interpret “close face-to-face circle” loosely in order to allow the inclusion of distant and hybrid meetings. Other constraints typically associated with the term meeting include synchronicity among the participants and a limitation in time, denoted by a beginning and an end (Fulk & Collin-Jarvis, 2001). We further limit the scope to professional meetings, and more specifically those that occur in the segment of the workforce that has been labeled “knowledge workers”. The term was first used by Drucker around 1960 (Drucker, 1959, 1961). Although it is not a particularly well-defined concept (Scarborough, 1999), it commonly includes occupations such as doctors, lawyers, scientists and academics. Around the turn of the century, Drucker explicitly included what he labeled “knowledge technologists”, exemplified by computer technicians, lab analysts, paralegals, software designers, into the group (Drucker, 1999).

¹

<https://www.merriam-webster.com/dictionary/meeting>

² <https://languages.oup.com/google-dictionary-en/>

2.2 Meeting evaluation

Notwithstanding the voluminous literature across many disciplines, several authors point to methodological shortcomings concerning the assessment and evaluation of meetings. Davenport et al. found “an astounding lack of knowledge in what actually improves performance” and that “measures, controls, hypotheses and even the recording of lessons learned” were often lacking. Instead, “fad, fashion and faith” drove decisions about new work environments for knowledge workers (Davenport et al., 2002). In general, we note that an overwhelming majority of studies of meeting quality uses some manner of subjective, survey based (post-meeting questionnaire) perceived meeting quality as the quality metric. Often, this seems to be accepted as a necessary state of affairs. We are aiming for comparable and objective assessments of meeting outcomes. Consequently, we have opted for a method in which the meetings are experiments rather than real-world meetings, and where the meeting task structure is predefined as part of the general experiment design, but where little or no other restrictions are imposed on meeting design.

3. Method

Here, we describe the meeting tasks, the recordings, and the first version of the annotation scheme.

3.1 Tasks

Three different ranking tasks were used during the meetings: NASA moon survival (Hall & Watson, 1970; Littlepage et al., 1995), Desert survival (Lafferty & Pond, 1974; Littlepage et al., 1995) and the Camping game survival (Hare, 1952) task. The reason for choosing these survival tasks was that they were well studied and vastly used in the literature. The order in which the

tasks were used was randomized. In all three tasks participants were asked to rank items according to their importance to their survival. The aim of this type of task is to arrive at a consensus by the end of the discussion.

3.2 Participants

Altogether thirty individuals (13 females and 17 males) participated in this study. They were mainly recruited through the Accindi digital platform where researchers and study participants can interact. Participants were between 23 and 48 years old. They were all fluent English speakers and had no hearing problems. They formed groups of three while participating in meetings.

3.3 Setting & equipment

All meetings took place in the Division of Speech, Music and Hearing (THM) at KTH. The meetings were performed and recorded (both audio and video) in three different settings. The physical meeting took place in the seminar room at TMH, where participants gathered around a table working on their task. Their meeting was recorded using the meeting owl pro (360-degree camera, mic, and speaker) which was placed at the center of the table and connected to a host computer. In addition, separate Xoom voice recorders were used to capture audio of each individual. As for the digital/hybrid meetings, ZOOM video conferencing software was used and participants were placed in separate booths while connecting over zoom. They were asked to use full screen mode while selecting the gallery view and “Hide self” in the gallery options. The meetings were recorded both through zoom and voice recorders. During the hybrid meeting, two participants were sharing the same room while the third participant was connected through zoom.

3.4 Process

In each session, participants in groups of three, joined three consecutive meetings (physical, digital and hybrid) while working on the above mentioned survival tasks. The order in which the meetings took place was randomized for each group. Before the start of the sessions, participants were asked to sign a GDPR and fill out a demographic form. Each meeting, regardless of the setting, consisted of three phases: the pre-meeting, the in-meeting and the post-meeting phase. Before the meeting (the pre-meeting phase) participants were asked to work on the given task individually and write down their preferred order of items. They had 5 minutes to complete it. During the meeting (the in-meeting phase) they had 15 minutes to discuss the same task with their group mates and come up with a group consensus. After the meeting (the post-meeting phase) again they had 5 minutes to review their initial ranking individually and had the chance to modify it if needed. The whole session was completed within 2 hours.

4. Results

4.1 Data processing

The recordings have been segmented and annotated on various levels using ELAN 6.3 multimodal annotation tool. ELAN is a professional tool for the creation of complex annotation on video and audio resources. Annotations can be created on multiple layers, called tiers. Tiers can be hierarchically interconnected. An annotation can either be time-aligned to the media or it can refer to other existing annotations (Sloetjes & Wittenburg, 2008).

4.2 Annotation scheme

At the first level of annotation, the conversations were segmented into units

defined as stretches of communicative activity produced by one participant who occupies the speaker role, bounded by periods of inactivity (i.e. silence) from that speaker (Brady, 1968; Bunt, 2015; Heldner & Edlund, 2010). We borrow the term Turn Unit from 24617-2 (Bunt et al., 2020). The turn units for each participant were annotated on separate tiers (**TurnUnit_A**, **TurnUnit_B**, **TurnUnit_C**). If there were any vocalizations by anyone who wasn't one of the participants, it was annotated on a separate tier (**TurnUnit_Other**). The minimum silence from a participant required to end a turn unit was 500ms. The **Focus** tier tracks entities currently under discussion in the conversation. This can be seen as a linear, incremental and simplified version of the *questions under discussion* concept (Ginzburg, 2001; Larsson, 2002).

The survival tasks only allows for two types of task internal entities to focus on: the **Items** on the list, the **Rank** on which each item belongs. Task externally, we are also interested in which **speaker** is behind an utterance. Anything else is either not immediately related to the task, or an attribution, argument, etc. that is associated with one or more of these three entities. In our model, each time an item or a rank is mentioned, the entity becomes the focus of its kind.

Furthermore, we annotate focus-changes as one of *proposal*, *question*, *decision* or *decision-repeat*. A *proposal* leaves both Item and Rank set. For example, a participant says: 'I think map (Item) should be in position 2 (Position)', or responds "In the second position" to the question "Where should we place the map?". A *question* sets one of Rank or Item and in effect vacates the other: 'what do you think

TurnUnit_A	none	none
TurnUnit_B		
TurnUnit_C		
Focus	Item Rank Speaker	Proposal Question Decision Decision-repeat
Item	none	The list of available items in each task(10 or 15 items, depending on the task)
Rank	none	Items were ranked from 1 to 10 or from 1 to 15, depending on the task.
Speaker	none	A, B or C
Laughter / smile_A	none	Laughter Smile
Laughter / smiler_B		
Laughter / smile_C		
Breathing_A	none	In-breath Out-breath
Breathing_B		
Breathing_C		

Table 1 : Annotation scheme, MEET annotation manual v1.0.0-1.0.0

should go first' (sets Rank to 1 and Item empty) or 'Where do you think map should be placed?' (sets Item to map and Rank to empty).

Tiers	Child tiers	Controlled vocabulary
--------------	--------------------	------------------------------

A *decision* marks the point where the group announces the final decision on an item and its rank. A *decision-repeat* marks the instances where a ranking is repeated after the decision is made.

Laughter and *smile* are annotated on a separate tier (**Laughter/smile**) for each speaker when it is audible or visible. Laughter is distinguished from smile by a simple token: the former produces an audible sound while the latter does not. *In_breath* and *out_breath* were marked on a separate tier (**Breathing**) for each speaker. Table 1 shows a summary of the annotation scheme.

As seen in Table 1, the FOCUS tier has child tiers. Parent and child tiers are linked in such a way that some changes made on a parent tier will also affect its child tiers (child tiers are shown with the same color, see Figure 1). When a certain linguistic type with a limited number of annotation values are frequently used it might be a good idea to associate it with a Controlled Vocabulary (CV). Such a CV consists of a number of predefined values that a user can choose from when editing an annotation, in order to make the task of the annotators less error-prone. Table 1 shows the CVs used in this annotation scheme.

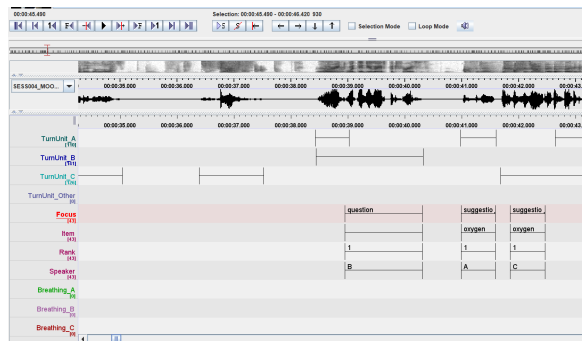


Figure 1: Screenshot of the annotation in ELAN.

As seen in Figure 1, each annotation is entered on a tier and assigned to a time interval. As the annotation is progressing the annotation scheme is also being updated

based on the annotator's feedback and discussions. A metadata file and also an annotation manual accompany the corpus.

5. Summary and future work

This study presents the MEET corpus. This corpus was developed and designed for studying and evaluating meetings in different forms (physical, digital and hybrid). The data was collected during 10 sessions of recorded meetings, altogether providing 8 hours of multi-party conversations. The participants are working on survival ranking tasks and their aim is to arrive at a group consensus during their meeting. Based on our research questions, the recordings were transcribed and annotated on various levels using the ELAN annotation tool. The annotation work and also the annotation scheme are still ongoing work.

References

- Brady, P. T. (1968). A statistical analysis of on-off patterns in 16 conversations. *The Bell System Technical Journal*, 47(1), 73–91.
<https://doi.org/10.1002/j.1538-7305.1968.tb00031.x>
- Bunt, H. (2015, April). On the principles of semantic annotation. *Proceedings of the 11th Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA-11)*.
<https://aclanthology.org/W15-0201>
- Bunt, H., Petukhova, V., Gilmartin, E., Pelachaud, C., Fang, A., Keizer, S., & Prévot, L. (2020). The ISO standard for dialogue act annotation, second edition. *Proceedings of the Twelfth Language Resources and Evaluation Conference*, 549–558.
<https://aclanthology.org/2020.lrec-1.69>
- Davenport, T. H., Thomas, R. J., & Cantrell, S. (2002). *The mysterious art*

- and science of knowledge-worker performance. *4*(1), 22–30.
- Drucker, P. F. (1959). *Landmarks of tomorrow*. Harper.
- Drucker, P. F. (1961). Fifty years of management—A look back and a look forward. *Journal of Engineering for Industry*, *83*(3), 366–370.
<https://doi.org/10.1115/1.3664530>
- Drucker, P. F. (1999). Knowledge-worker productivity: The biggest challenge. *California Management Review*, *41*(2), 79–94.
<https://doi.org/10.2307/41165987>
- Fulk, J., & Collin-Jarvis, L. (2001). Wired meetings: Technological mediation of organizational gatherings. In F. Jablin & L. Putnam (Eds.), *The new handbook of organizational communication* (pp. 625–663). SAGE Publications, Inc.
<https://doi.org/10.4135/9781412986243>
- Ginzburg, J. (2001). Interrogatives: Questions, facts and dialogue. In *The Handbook of Contemporary Semantic Theory*. Blackwell.
- Goffman, E. (1961). *Encounters: Two studies in the sociology of interaction* (p. 152). Bobbs-Merrill.
- Hare, A. P. (1952). A study of interaction and consensus in different sized groups. *American Sociological Review*, *17*(3), 261–267.
<https://doi.org/10.2307/2088071>
- Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, *38*(4), 555–568.
<https://doi.org/10.1016/j.wocn.2010.08.002>
- Lafferty, J. C., & Pond, A. W. (1974). *The desert survival situation: A group decision making experience for examining and increasing individual and team effectiveness*. Human Synergistics.
- Larsson, S. (2002). *Issue-based dialogue management* [PhD Thesis]. Department of Linguistics, Göteborg University.
- Littlepage, G. E., Schmidt, G. W., Whisler, E. W., & Frost, A. G. (1995). An input-process-output analysis of influence and performance in problem-solving groups. *Journal of Personality and Social Psychology*, *69*(5), 877–889.
<https://doi.org/10.1037/0022-3514.69.5.877>
- Scarbrough, H. (1999). Knowledge as work: Conflicts in the management of knowledge workers. *Technology Analysis & Strategic Management*, *11*(1), 5–16.
<https://doi.org/10.1080/095373299107546>
- Sloetjes, H., & Wittenburg, P. (2008, May). Annotation by category: ELAN and ISO DCR. *Procs. of LREC'08*. The Sixth International Conference on Language Resources and Evaluation (LREC'08), Marrakech, Morocco.
http://www.lrec-conf.org/proceedings/lrec2008/pdf/208_paper.pdf
- Video conferencing market size, share & Covid-19 impact analysis* (No. FBI100293; Furtune Business Insights, p. 160). (2021). Furtune Business Insights.
<https://www.fortunebusinessinsights.com/industry-reports/video-conferencing-market-100293>
- Yan, Z. (2020). Unprecedented pandemic, unprecedented shift, and unprecedented opportunity. *Human Behavior and Emerging Technologies*, *2*(2), 110–112.
<https://doi.org/10.1002/hbe2.192>