The MEET Corpus: Collocated, Distant and Hybrid Three-party Meetings with a Ranking Task

Ghazaleh Esfandiari-Baiat, Jens Edlund
Speech, Music and Hearing, KTH
Stockholm, Sweden
geb@kth.se, edlund@speech.kth.se

Abstract

We introduce the MEET corpus. The corpus was collected with the aim of systematically studying the effects of collocated (physical), remote (digital) and hybrid work meetings on collaborative decision-making. It consists of 10 sessions, where each session contains three recordings: a collocated, a remote and a hybrid meeting between three participants. The participants are working on a different survival ranking task during each meeting. The duration of each meeting ranges from 10 to 18 minutes, resulting in 380 minutes of conversation altogether. We also present the annotation scheme designed specifically to target our research questions. The recordings are currently being transcribed and annotated in accordance with this scheme.

Keywords: meetings, multimodal corpora, annotation scheme

1. Introduction

The declaration of COVID-19 as a global pandemic led to widespread implementation of social distancing measures, resulting in a shift of various human social activities from offline to online. In other words, the enforced social isolations in the physical world significantly increased humans' social interactions in the cyber world (Yan, 2020). In a professional context, this was most noticeable as a shift from collocated to remote meetings that changed the work environment. The shift is backed by staggering numbers. As an example, Zoom added >2 million active users monthly during 2020 (Video Conferencing Market Size, Share & Covid-19 Impact Analysis, 2021).

The change subsists in the post-pandemic era, and remote meetings have become a prevalent part of modern work culture. and it is safe to assume that we currently have more people that are well-versed in the art of remote meetings than ever before.

Notwithstanding, important questions have not been adequately addressed, such as:

- How is the structure and dynamics of remote meetings best described?
- Are they as effective as collocated ones?
- How do they differ?
- How, for that matter, is effectiveness evaluated?

The actualization of these questions has motivated us to study remote meetings, with the aim of describing, analyzing and comparing spoken interaction behaviors and the resulting efficiency in collocated, remote and hybrid meetings.

In this paper, we focus on the corpus construction explaining the recording phase (setting, participants and the tasks) and also the annotation phase providing a detailed description of the used annotation scheme.

1.1 Scope

According to Merriam-Webster, the term "meeting" simply refers to "the act of coming together." Other definitions, such as Google's English Dictionary by Oxford Languages, specify a more deliberate gathering, defined as "an assembly of people for a specific purpose." Here, we are concerned with this latter kind of purposeful meeting - the kind that Goffman considers “the natural unit of social organisation 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 to allow the inclusion of remote and hybrid meetings. Other constraints typically associated with the term meeting include synchroneity 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 labelled “knowledge workers”. The term was first used by Drucker around 1960 (Drucker, 1959, 1961). Although it is not a particularly well-defined concept (Scarbrough, 1999), it commonly includes occupations such as doctors, lawyers, scientists and academics. Around the turn of the century, Drucker explicitly included what he labelled “knowledge technologists”, exemplified by computer technicians, lab analysts, paralegals,

ISA-20 Proceedings @LREC-COLING-2024, pages 1–7

software designers, into the group (Drucker, 1999).

2. Background and Related Work

2.1 Studies of Meetings

The study of meetings has received attention in several disciplines. A key body of work is what we may term the social psychology of small groups, a field that reached a peak in the 70s. (Davis et al., 1976) presents a comprehensive overview that is particularly relevant. The focus here is not as much on meetings as it is on the dynamics of work in small groups in general, and although the two are clearly associated, they are not the same. Goffman points to several reasons to hold the two concepts apart, with the strongest being that the crucial meeting attribute of "maintenance of continuous engrossment in the official focus of activity" is "not a property of social groups in general" (Goffman, 1961).

In literature more directly focused on meetings there are several directions that are worth specific mention here. The study of the effects of meeting facilitators blossomed in the 1990s, in part because of the increase of team-based organizations, but also because "the advent of group support technologies" (e.g., audio-video conferencing) created "a special demand for facilitation" (Niederman & Volkema, 1999; here you will also find a brief overview of the field). Directly related to the same technology development are studies of the effects of distance (e.g. in audio-video conferencing). (Fulk & Collin-Jarvis, 2001) provides a comprehensive overview of 20th century work in this field.

2.2 Meeting Types

McGrath (1984) takes off from Hackman's three classes (e.g. Hackman & Morris, 1975; Morris, 1966), where "production" and "problem-solving" becomes generated (ideas and plans, respectively) and "discussion" becomes choose or negotiate depending on the situation. He then adds "execute" as a fourth alternative. Each of these four basic "quadrants" is then divided into two using features from several other classifications. This results in eight task types: planning tasks, creativity tasks, intellective tasks, decision making tasks, cognitive conflict tasks, mixed-motive tasks, competitive tasks, and psycho-motor tasks. This classification - the circumplex model - has been quite influential. In this terminology, our main interest is in the "choose" quadrant, and more specifically in "decision making tasks".

2.3 Mediated Meetings

A great deal of theoretical work on mediated meetings took place quite some time ago. There is relevant work in the group decision support systems (GDSS) field, although it targets groups rather than meetings. DeSanctis & Gallupe, (1987) proposed a division of electronic support systems (for group decisions) into three levels. Level 1 contains "technical features aimed at removing common communication barriers". This is the most relevant level for the present work, as it contains audio/video conferencing.

Fulk & Collin-Jarvis (2001) makes a three-way distinction between group support systems (GSS, which do not seem to differentiate from GDSS, and which refers to all three levels of DeSanctis & Gallupe). Here, their notion of "teleconferencing", which refers to "meetings held through audio-conferencing and video-conferencing systems" is the main area of interest from our perspective. Review papers on teleconferencing started appearing as early as the 1970s (Williams, 1977). We note, however, that even though audio and video conferencing technology has improved by leaps and bounds since its infancy, acceptance of distant meetings may not have increased at the same rate (Blenke et al., 2017), at least not before the pandemic. We also note that the main issue may not be acceptance but rather that video conferencing and face-to-face meetings simply work differently (Denstadli et al., 2012) and that attitudes vary with the type of video conferencing system used (Julsrud et al., 2012).

Face-to-face interaction is another research field that has taken a keen interest in video conferencing, targeting its presumed inability to faithfully transfer communicative cues and the resulting deterioration on quality of interaction. Various complex technical video solutions have been proposed from near the dawn of video conferencing until the present (e.g. Adeboye, 2020; Nguyen & Canny, 2007; Okada et al., 1994; Sellen et al., 1992), as well as considerably more complex solutions involving avatars in order to achieve telepresence over low bandwidth (Al Moubayed et al., 2012; Beskow et al., 2009)

Directly related to the same technology development are studies of the effects of distance (e.g. in audio-video conferencing). Fulk & Collin-Jarvis (2001) provides a comprehensive overview of 20th century work in this field.

For more direct comparisons of remote and collocated communication, digital interaction has been shown to reduce perceived social presence between communicators, potentially hindering relationship-building among collaborators, and leading to a stronger focus on self (their personal goals) and less on their interaction partners (Scholl et al., 2020). Collocated interactions involve richer visual, auditory, tactile, and contextual information, helping people pick up important social cues and share intentions and emotions resulting in feelings of social closeness (Newson et al., 2021). More generally, *media richness theory* posits that interactions held through "richer" communication media (i.e. media

that involve more cues) lead to better communication. Nesher Shoshan & Wehrt (2022) showed that meetings held through video conferences cause more exhaustion, indicating that so-called *Zoom fatigue* may objectively exist. Moreover, participants involved in remote meetings described difficulties in reading social cues of others, while perceiving pressure to provide such cues themselves (ibid).

Concerning efficiency, Denstadli et al. (2012) showed that while remote meetings save time (both in planning and in the duration of the meeting itself) they are not suitable for participants who do not know one another beforehand, and it makes developing contacts difficult. From a more organizational point of view, collocated meetings are preferred because of the desire to develop social relations and social capital and to handle tasks with high ambiguity (ibid). Similarly, Alge et al., (2003) examined the effect of teams' past experiences on their ability to communicate in collocated and remote contexts. Results indicate that teams without knowledge-building experience (no shared past) communicating collocated reported higher openness/trust and shared more unique information than remote teams communicating through a synchronous computer-mediated medium.

3. Method

3.1 Corpus Collection

3.1.1 Participants

The corpus consists entirely of three-party conversations. 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. They were compensated by four cinema tickets. 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 (three individuals per session having three types of meetings consecutively, no individual took part in more than one session).

3.1.2 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). In all three tasks an imaginary situation is explained during which participants must find a way to survive. There is a list of items (10 to 15 items dependent on the task) which could help them in their survival. Participants were asked to rank these items from one to fifteen according to their importance for their survival. The aim of this type of task is to arrive at a group consensus by the end of the meeting. The reason for choosing these survival tasks was that they were well studied and vastly used in the literature. During each meeting the groups had to complete one task and the order in which the tasks were used was randomised.

3.1.3 Setting & Equipment

All meetings took place in the Division of Speech, Music and Hearing (THM) at KTH. Meetings were performed and recorded (both audio and video) in three different settings. The collocated 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 a 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. For the digital and 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.

3.1.4 Process

In each session, participants in groups of three, joined three consecutive meetings (collocated, remote, and hybrid) while working on one of the survival tasks in each meeting. The order in which the meetings took place was randomized for each group. Before the start of each session recording, participants were provided with instructions and asked to sign a GDPR consent form 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 individual preferred order of items. They were given 5 minutes to complete this. 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) they were again given 5 minutes to review their initial individual ranking and modify it if necessary. Each session was completed within 2 hours.

3.2 Corpus Annotation

3.2.1 Data Processing & Annotation Tool

The recordings have been segmented and annotated on various levels using ELAN 6.3 multimodal annotation tool (Sloetjes &

Wittenburg, 2008). With ELAN a user can add an unlimited number of textual annotations to audio and/or video recordings. Annotations can be created on multiple layers, called tiers. Tiers can be hierarchically interconnected (child and parents tier). An annotation can either be time-aligned to the media or it can refer to other existing annotations. The content of annotations consists of Unicode text and annotation documents are stored in an XML format (EAF).

3.2.2 Annotation Scheme

The conversations are annotated on various levels using separate tiers for each layer. On the first level, conversations are manually decomposed into **TURN UNITS**. These units are defined as stretches of speech produced by one participant who occupies the speaker role, bounded by periods of inactivity (i.e. silence) of that speaker. (Brady, 1968; Bunt et al., 2020; Heldner & Edlund, 2010). An annotation segment on a tier starts with the start of the vocalisation by a participant and ends with its end. The minimum silence from a participant required to end aTurnUnit was 500ms. The TurnUnits related to each participant were annotated on separate tiers (**TurnUnit_A, TurnUnit_B and TurnUnit_C**). If there were any other vocalisations by anyone that isn't one of the participants, it is annotated on a separate tier (**TurnUnit_Other**).

On the second level of annotation, 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 **FOCUS** tier has three child tiers (**ITEMS**, **RANK** and **SPEAKER**). Parent and child tiers are linked in such a way that some changes made on a parent tier will also affect its child tires (child tiers are shown with the same color, see Figure1).

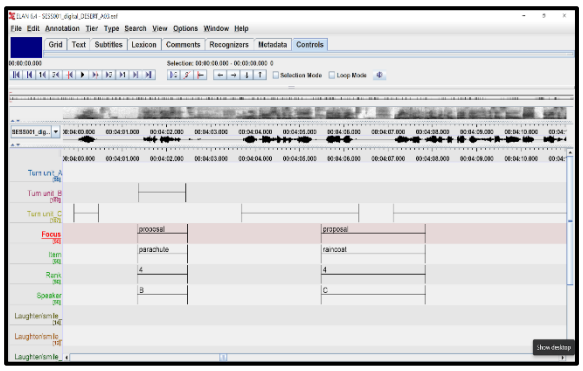

Figure 1: Screenshot of the annotation in ELAN.

Survival tasks only allow 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. The **ITEMS** tier Shows which item from the list is under discussion (the items are different in different tasks). Focus item changes are defined by a simplistic rule: any mention of an item sets that item in focus. In other words, the mention of a list item sets the item focus to that item. In a similar manner to items, the **RANKS** tier shows which position on the list is considered. In the NASA moon survival and the Desert survival task there are 15 items while in the Boys scout survival task there are only 10 items to rank. And finally, the **SPEAKER** tier shows which speaker made the contribution (is talking).

Furthermore, using controlled vocabulary (CV) in ELAN, we annotate focus-changes as one of ***proposal***, ***question***, ***decision*** or ***decision-repeat***. When a certain linguistic type with a limited number of annotation values is frequently used it might be a good idea to associate it with a 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.

- A ***proposal*** leaves both Item and Rank set. For example, a participant says: "*I think map (Item) should be in position 2 (Rank)*" 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 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).
- A ***decision*** marks the point where the group announces the final consensus on an item and its rank:" Okay*, we put map on fifth position".*
- A ***decision-repeat*** marks the instances where a ranking is repeated during the conversation after the decision is made.

Note that in this version of the annotation, we do not annotate grounding and repetitions before the decision at all. The only time we include repetitions (of an already focused Item or Rank, or of a proposed mapping between the two), is when it is a repetition of a decision already made (see decision-repeat above). Other repetitions are simply left unannotated.

LAUGHTER/SMILE are annotated on separate tiers for each speaker when it is audible or visible (Laughter/Smile_A, Laughter/Smile_B and Laughter/Smile_C). The start of the laughter or

smile is marked as the starting point of the annotation segment, and its end is the end of the segment. Laughter is distinguished from smile by a simple token: the former produces an audible sound while the latter does not.

BREATHING (*In_breath* and ***out_breath)*** were also marked on separate tiers (Breathing_A, Breathing_B and Breathing_C) for each speaker and are annotated only when it's audible. The starting point of the annotated segment is when the breathing begins, and it ends when the breathing ends.

Acoustic **SILENCE** (***SIL***) was defined as a segment in which no participant vocalizes, flanked by segments in which some participant vocalizes. Silences were not annotated explicitly but found by extracting segments with no TurnUnits.

Although annotators were instructed to produce adjacent TurnUnits when no silence could be heard, there were a few mistakes of this sort in the original annotations. In a semi-automatic post-processing step, we removed any within-speaker silences of less than 500 ms (which is the minimum duration of gaps to be annotated according to the annotation scheme). A total of 4% of the automatically extracted silences were removed by this process.

In addition, we removed any between-speaker silences of 50 ms or less. The reasoning here is that these acoustic silences are not perceivable as silences, or gaps, in the terminology of Sacks et al (Sacks et al., 1974). On average, the group decision threshold for perceivable acoustic silence between speakers is considerably longer - 120 ms, but as some listeners perceive gaps robustly at as little as 58 ms of acoustic silence, we opted for a conservative threshold of 50 ms (Heldner, 2011). Whenever a silence was removed, the adjacent TurnUnits were corrected so that they become adjacent to each other, by growing the larger of the TurnUnits. Removing all between-speaker silences below the group decision threshold of ~120 ms, another 4 % of the between-speaker silences were removed.

Finally, a new entity was added: the **Unbroken Speech Sequences *(USS)***. This is a continuous sequence during which at least one participant vocalises at each moment, flanked by silence on both sides. Large proportions of overlap, high intensity, and long TurnUnits all contribute to long USSs, whereas large numbers of pauses, short utterances and general inactivity contribute to low USS durations.

3.3 General Statistical Observation

In total, the corpus consists of 6 hours and 20 minutes. The average total active meeting time in a session ranged from 21 to 46 minutes with an average and median of 38 and 39 minutes, respectively. The average single meeting duration ranged from five to 20 minutes with both average and median at 13 minutes.

In terms of TurnUnits, the corpus contains a total of 8149 TurnUnits. The number of TurnUnits produced by a group (in one session) ranged from 587 to 1120 with an average of 815 and a median of 816 TurnUnits. The number of TurnUnits in a single meeting ranged from 117 to 411 with an average of 272 and a median of 277 per meeting.

The total number of unbroken speech sequences (USS) is 5293, with a range from 328 to 689 in a single session, an average of 530 and a median of 567. That means that a typical USS contained 1.5 TurnUnits (average and median), with a highest session TurnUnits/USS at 2.7 and a lowest of 1.3 (note that the floor is 1, here, as each USS holds at least one TurnUnit). The number of USSs in a single meeting ranged from 79 to 270, with an average of 176 and a median of 180. A USS contains on average 1.6 TurnUnits, while the corresponding median was 1.7, with a lowest observation of 1.1 and a highest of 4.1 TurnUnits/USS.

The median TurnUnit duration in the entire corpus was 1.5 s, and the median silence 0.7 s. Medians within sessions range from 1.1 to 2 s for TurnUnits and from 0.6 to 1 s for silences.

4. Summary & Future Work

We have presented the MEET meeting corpus and its annotation scheme. Ten three-person groups were recorded, each in a single session consisting of three separate meetings, each with a different task and condition, for a total of 30 meetings.

This corpus was constructed with the aim of systematically describing, analysing and modelling interaction patterns during different types of meetings plus evaluating the outcome of these different meeting setups. We wanted to know in which setting the participants were more cooperative and had the highest influence on the group consensus.

Although the current legislation does not permit us to share the corpus recordings, we plan to release the interaction models together with a detailed description of how they were derived. As for future work, we will release tools that facilitate work with and analysis of the kind of interaction model we have created. We will also include more of the annotation, for example filled pause annotation (any spoken sound or word used to fill gaps in speech) in the model. Currently, these are manually segmented for all the meetings in the corpus, but not included since they have not been validated. We also plan to have the corpus transcribed. Currently a section of it is transcribed using whisper ASR. It is however unclear how much of the transcriptions can be shared freely.

5. References

Adeboye, D. (2020). Exploring the effect of spatial faithfulness on group decision-making (Master Thesis UCAM-CL-TR-952; p. 64). University of Cambridge.

Al Moubayed, S., Skantze, G., & Beskow, J. (2012). Lip-reading: Furhat audio visual intelligibility of a back projected animated face. In Y. Nakano, M. Neff, A. Paiva, & M. Walker (Eds.), Intelligent Virtual Agents (pp. 196–203). Springer. https://doi.org/10.1007/978-3-642-33197-8_20

Alge, B. J., Wiethoff, C., & Klein, H. J. (2003). When does the medium matter? Knowledge-building experiences and opportunities in decision-making teams. Organizational Behavior and Human Decision Processes, 91(1), 26–37. https://doi.org/10.1016/S0749-5978(02)00524-1

Beskow, J., Salvi, G., & Moubayed, S. A. (2009). SynFace—Verbal and non-verbal face animation from audio. The International Conference on Auditory-Visual Speech Processing AVSP'09.

Blenke, L. R., Gosavi, A., & Daughton, W. (2017). Attitudes towards face-to-face meetings in virtual engineering teams: Perceptions from a survey of defence projects. International Journal of Project Organisation and Management. https://www.inderscienceonline.com/doi/abs/10.1504/IJPOM.2017.085284

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., 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

Davis, J. H., Laughlin, P. R., & Komorita, S. S. (1976). The social psychology of small groups: Cooperative and mixed-motive interaction. Annual Review of Psychology, 27(1), 501–541. https://doi.org/10.1146/annurev.ps.27.020176.002441

Denstadli, J. M., Julsrud, T. E., & Hjorthol, R. J. (2012). Videoconferencing as a mode of communication: A comparative study of the use of videoconferencing and face-to-face meetings. Journal of Business and Technical Communication, 26(1), 65–91. https://doi.org/10.1177/1050651911421125

DeSanctis, G., & Gallupe, R. B. (1987). A foundation for the study of group decision support systems. Management Science, 33(5), 589–609. https://doi.org/10.1287/mnsc.33.5.589

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.

Hackman, J. R., & Morris, C. G. (1975). Group tasks, group interaction process, and group performance effectiveness: A review and proposed integration. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 8, pp. 45–99). Elsevier. https://doi.org/10.1016/S0065-2601(08)60248-8

Hall, E. J. (Jay), & Watson, W. H. (1970). The effects of a normative intervention on group decision-making performance. Human Relations, 23(4), 299–317. https://doi.org/10.1177/001872677002300404

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. (2011). Detection thresholds for gaps, overlaps, and no-gap-no-overlaps. The Journal of the Acoustical Society of America, 130(1), 508–513. https://doi.org/10.1121/1.3598457

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

Julsrud, T. E., Hjorthol, R., & Denstadli, J. M. (2012). Business meetings: Do new videoconferencing technologies change communication patterns? Journal of Transport Geography, 24, 396–403. https://doi.org/10.1016/j.jtrangeo.2012.04.009

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

McGrath, J. E. (1984). Groups: Interaction and performance. Prentice-Hall. https://www.jstor.org/stable/2393041?origin=crossref

Morris, C. G. (1966). Task effects on group interaction. Journal of Personality and Social Psychology, 4(5), 545–554. https://doi.org/10.1037/h0023897

Nesher Shoshan, H., & Wehrt, W. (2022). Understanding "Zoom fatigue": A mixed-method approach. Applied Psychology, 71(3), 827–852. https://doi.org/10.1111/apps.12360

Newson, M., Zhao, Y., Zein, M. E., Sulik, J., Dezecache, G., Deroy, O., & Tunçgenç, B. (2021). Digital contact does not promote wellbeing, but face-to-face contact does: A cross-national survey during the COVID-19 pandemic. New Media & Society, 14614448211062164. https://doi.org/10.1177/14614448211062164

Nguyen, D. T., & Canny, J. (2007). Multiview: Improving trust in group video conferencing through spatial faithfulness. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1465–1474. https://doi.org/10.1145/1240624.1240846

Niederman, F., & Volkema, R. J. (1999). The effects of facilitator characteristics on meeting preparation, set up, and implementation. Small Group Research, 30(3), 330–360. https://doi.org/10.1177/104649649903000304

Okada, K.-I., Maeda, F., Ichikawaa, Y., & Matsushita, Y. (1994). Multiparty videoconferencing at virtual social distance: MAJIC design. Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work, 385–393. https://doi.org/10.1145/192844.193054

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. Language, 50(4), 696–735. https://doi.org/10.2307/412243

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

Scholl, A., Sassenberg, K., Zapf, B., & Pummerer, L. (2020). Out of sight, out of mind: Power-holders feel responsible when anticipating face-to-face, but not digital contact with others. Computers in Human Behavior, 112, 106472. https://doi.org/10.1016/j.chb.2020.106472

Sellen, A., Buxton, B., & Arnott, J. (1992). Using spatial cues to improve videoconferencing. CHI'92, 651–652. https://doi.org/10.1145/142750.143070

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 (FBI100293; Fortune Business Insights, p. 160). (2021). Furtune Business Insights. https://www.fortunebusinessinsights.com/industry-reports/video-conferencing-market-100293

Williams, E. (1977). Experimental comparisons of face-to-face and mediated communication: A review. Psychological Bulletin, 84(5), 963–976. https://doi.org/10.1037/0033-2909.84.5.963

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