MeetingCoach: An Intelligent Dashboard for Supporting Effective & Inclusive Meetings

Samiha Samrose[1], Daniel McDuff[2], Robert Sim[2], Jina Suh[2], Kael Rowan[2], Javier Hernandez[2], Sean Rintel[2], Kevin Moynihan[2], Mary Czerwinski[2]
[1]University of Rochester, [2]Microsoft Research
ssamrose@cs.rochester.edu,{damcduff,rsim,jinasuh,kael.rowan,javierh,serintel,kevinmo,marycz}@microsoft.com

ABSTRACT

Video-conferencing is essential for many companies, but its limitations in conveying social cues can lead to ineffective meetings. We present MeetingCoach, an intelligent post-meeting feedback dashboard that summarizes contextual and behavioral meeting information. Through an exploratory survey (N=120), we identified important signals (e.g., turn taking, sentiment) and used these insights to create a wireframe dashboard. The design was evaluated with in situ participants (N=16) who helped identify the components they would prefer in a post-meeting dashboard. After recording video-conferencing meetings of eight teams over four weeks, we developed an AI system to quantify the meeting features and created personalized dashboards for each participant. Through interviews and surveys (N=23), we found that reviewing the dashboard helped improve attendees' awareness of meeting dynamics, with implications for improved effectiveness and inclusivity. Based on our findings, we provide suggestions for future feedback system designs of video-conferencing meetings.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in collaborative and social computing**.

KEYWORDS

group, feedback, meeting, sensing, video-conferencing

ACM Reference Format:
Samiha Samrose[1], Daniel McDuff[2], Robert Sim[2], Jina Suh[2], Kael Rowan[2], Javier Hernandez[2], Sean Rintel[2], Kevin Moynihan[2], Mary Czerwinski[2]. 2021. MeetingCoach: An Intelligent Dashboard for Supporting Effective & Inclusive Meetings. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3411764.3445615

1 INTRODUCTION

Meetings are a primary mode of work [58], but many employees find them frustrating and even counter-productive when good meeting practices are lacking or violated [1, 46]. The violations of general meeting norms and disrespectful behaviors have been shown to be negatively correlated with meeting effectiveness and satisfaction [45]. In 2020, the "stay at home" orders and travel restrictions of the COVID-19 pandemic dramatically accelerated the use of the video-conferencing meetings for work. By March 2020, daily usage of video-conferencing services such as Zoom and Microsoft Teams had increased by 300% and 775% respectively, and video-conferencing apps jumped to the top of the Apple app store.[1] Although video-conferencing has the potential to reduce the cost and effort behind organizing travel, space, and scheduling of in-person meetings[2], the "fractured ecologies" [37] of video-conferencing can aggravate negative outcomes and marginalize some members of the meeting [38]. Video-conferencing has consistently presented communicative challenges [19, 52] and introduced distractions [61] which can lead to ineffective and non-inclusive meetings [23, 30]. A primary goal of remote collaboration tools should be to support the most effective meetings possible for all participants. Cutler et al. [9] conducted a large scale e-mail survey on remote meeting effectiveness (N=4,425) at a technology company (pre-COVID-19) and showed that online meeting effectiveness correlates with meeting inclusiveness, participation, and comfort in contributing. They identify a large potential financial benefit to companies that can achieve these goals, and an opportunity to establish and maintain a workplace culture in which everyone feels free to contribute. There are clearly opportunities for technological solutions that assist attendees in feeling more included and improving meeting effectiveness by helping them understand their own and others' core meeting dynamics.

This paper reports on an exploratory study with in situ work teams to identify the challenges they face in video-conferencing based meetings, and proposes a post-meeting feedback system to address the issues. In particular, we aimed to provide insights on the following research questions: (1) What aspects of meetings do video-conferencing attendees need help with?; (2) How can we leverage AI systems to make video-conferencing meetings more inclusive and effective?; (3) How should AI-extracted meeting features (including content, behavioral measurements, and sentiment) be categorized and visualized in a feedback dashboard?; and (4) What concerns exist regarding data privacy and accuracy for such systems?

Our work addresses these research questions through a series of user studies and design iterations. Via an initial exploratory requirement analysis survey of 120 information workers, we identified the communicative signals (e.g., participation, facial sentiment) which are important in improving meeting effectiveness and inclusion.

CHI '21, May 8–13, 2021, Yokohama, Japan

ACM ISBN 978-1-4503-8096-6/21/05…$15.00
https://doi.org/10.1145/3411764.3445615

[1]https://www.marketwatch.com/story/zoom-microsoft-cloud-usage-are-rocketing-during-coronavirus-pandemic-new-data-show-2020-03-30
[2]https://financesonline.com/video-web-conferencing-statistics/

We conducted a longitudinal user study to record in situ video-conferencing meetings from eight teams over a four-week period. We used the insights from the requirement analysis survey to create a wireframe prototype of a post-meeting dashboard and 16 participants from the user study teams evaluated the design and helped us further refine the components. Finally, we developed an AI sensing system to quantify these meeting dynamics features and created personalized, interactive, post-meeting dashboards for the participants. Through surveys (N=23) and interviews (N=9), we found that participants were able to become more aware of the group dynamics by reviewing the dashboard. Our study shed light on the privacy concerns participants had regarding such insights within the dashboard. The dashboard also helped participants identify and recollect important events of the past meetings. Our findings also showed that participants perceived that the dashboard would improve meeting effectiveness and inclusivity. In addition, participants expressed that actionable suggestions were more helpful than data visualizations alone. The main contributions of this work are as follows:

- We developed MeetingCoach, an AI-driven dashboard that provides both contextual and actionable insights based on meeting behaviors.
- We implemented both behavioral (e.g., *participation*) and topical (e.g., *questions*) meeting dynamics features in our feedback system using state-of-the-art AI.
- We identified and implemented shared and private design approaches for different feature components based on users' preferences from two design iterations and evaluations.
- We demonstrated that MeetingCoach helped improve behavioral awareness and recollection of past meeting events, and bears the potential to improve perceived effectiveness and inclusivity.
- We proposed design guidelines explaining the need for actionable suggestions, reminders or highlights based on timing, and multi-modal feature implementations to be adopted for future video-conferencing feedback systems.

2 RELATED WORK

2.1 Factors in Meeting Dynamics

Meeting effectiveness includes both task processing and interaction efficiency by a team [10, 14, 31, 47, 53]. Dickinson and McIntyre [10] emphasized the importance of goal specification, entailing identification and prioritization of tasks and sub-tasks, in agendas and other meeting resources. Even with clear goals, though, interaction efficiency has a clear impact on both outcomes and satisfaction. Balanced, active, and equal participation have been found to improve team performance [13, 32]. Depending on the type of meeting, equal participation may not always be applicable or feasible, but in a collaborative decision-making discussion, participation from all members ensures at least the exchange of opinions and a sense of "being heard", which ultimately improves team satisfaction [32, 50]. Turn-taking patterns also influence team performance and satisfaction, as some members may dominate the discussion without realizing they are doing so, reducing time for other members to voice their opinion or expertise [14]. Lawford [31] found that rapport building through verbal and non-verbal signals was an important factor in effective and inclusive discussions. To ensure coordination and rapport, affect management has been found to play an important role in a team's success [4, 8]. Barsade [4] has shown that a member's positivity can improve the mood of the whole team, making it more inclusive and improving the quality of decision-making. Cannon-Bowers et al. [8] discussed the importance of effective strategy formulation to consider alternative courses of action in case of disagreement or task failure. Non-verbal gestures, through head nodding and shaking, indicate signs of agreement or disagreement, and levels of interest, acknowledgement, or understanding [22, 34].

While the face-to-face views of video-conferencing intuitively seem to support the above, they have been found to constrain attention to the non-verbal signals and the overall progress of the meeting [19, 23]. Sellen [51] showed that having video did not improve the interruption or the turn-taking rate for video-conferencing meeting participants compared to audio-only ones. This implies that even though video is important in online meetings, it cannot fully resemble in-person meeting dynamics. Especially during long meetings, additional support for monitoring meeting progress and participation may be needed. Taking into consideration these concerns, we designed and developed an automated feedback system to summarize meeting content and attendee behaviors, with the goal toward improving meeting dynamics over time.

2.2 Feedback Systems for Videoconference Meetings

Researchers demonstrated the impact of feedback systems on meeting dynamics and discussion outcomes for in-person, text chat, and video-conferencing meeting setups [7, 11, 24, 27, 35, 43, 48, 56]. Feedback on participation [48], turn-taking [13], interruption [49], agreement [25], and valence [15], have effectively improved group discussion dynamics. These studies show that the timing (e.g., real-time, post-meeting) and the design (e.g., number of features, visualization strategies) of the feedback are important in effectively modulating collaboration behaviors in a group discussion.

Researchers explored feedback systems with affective, behavioral and topical group discussion features. Dimicco et al. [11] presented a real-time, shared-display feedback system measuring speaking contributions from audio recordings, visualized as bar graphs during a co-located meeting. They showed that effective visualization can help improve the group discussion, even though shared visualization can also motivate behavioral changes due to social pressure [39]. Nowak et al. [44] provided feedback on vocal arousal and explored the impact on oneself and one's partners behavior during a negotiation-based task conducted over the phone. They found that real-time feedback can be difficult to process during an ongoing task and can negatively impact user performance.

Therefore, even though real-time feedback has been found to be effective in modulating behaviors during a discussion, it can also be distracting and cognitively demanding [54, 56]. Samrose et al. [48] presented *CoCo*, an automated post-meeting feedback system providing summarized feedback on talk-time, turn-taking, speech overlap and sentiment through a chatbot for a video-conference group discussion. They showed that post-meeting feedback can effectively make successive discussion more balanced. Through a longitudinal study in a video-conference learning environment, *EMODASH* [15], an interactive dashboard providing feedback on affective meeting features, improved behavioral awareness over time. Instead of showing numeric or categorical feedback on linguistic features, Tausczik and Pennebaker [56] provided real-time and individualized actionable suggestions in text chat group discussions. Their findings showed that individualized suggestions helped teams shape their behavior; however, too much information in the feedback can be cognitively taxing. Suggestion-oriented feedback has been found effective in behavior modulation [54], and could be useful post-meeting. As we will later explain, our design incorporated an individualized and suggestion-oriented approach to a post-meeting feedback system. While identifying the feedback features for our dashboard during our requirement analysis, we prioritized features from these prior systems, such as talking time and turn-taking, but with an eye toward reducing cognitive load.

Beyond the meeting context, researchers have developed a number of interfaces that allow users to view emotional or affective signals captured by self-report, diaries or sensor-based systems. These have been used in several domains, such as self-reflection [21, 36, 41], data exploration and information retrieval [18, 59, 60], stress management [2, 20], and studying interpersonal dynamics [28, 29]. Data portraits can help people understand more about themselves and other people [12]. However, there is still a lot left to be understood about how to best represent complex and subjective data, such as emotions or group dynamics. Affective data is often highly dimensional, multi-modal, and continuous, all difficult when designing useful visualizations. There are also important privacy concerns raised by creating digital systems and artifacts that encode highly personal information [12].

The feedback needs of organizational teams conducting videoconferencing meetings require special attention, as these teams are relatively stable over time and the members need additional support in tracking the progress and the outcomes of their meetings [8, 26, 33]. As video-conferencing discussions can be prone to distraction and multitasking [30], we hypothesize that a meeting feedback system that helps members reflect on meeting goals and progress could be a useful tool. Feedback on the non-verbal behaviors can also help with effective and inclusive videoconference meetings. The Matrix to a Model of Coordinated Action (MoCA) presented by Lee and Paine [33] is an elaborate framework that explains a complex collaborative environment by seven dimensions. Within the context of MoCA, the post-meeting feedback dashboard that we propose is characterized as a long-term asynchronous periodic event for small teams placed in different locations.

In this study, we observe the meeting challenges and the needs of several in situ work teams, and propose and test technological solutions for them. Meetings have evolved from engaging inperson, to fully computer-mediated (all members join remotely via audio/video/chat), and hybrid (some members join remotely to group/s who are together in person), and each have their distinct character [17, 19, 58]. This study focuses on the fully-computer-mediated meetings of teams in which all members were remote due to COVID-19's mandatory requirement to work from home. All teams used the same video-conferencing system. We followed an iterative, human-centered design process to address our research questions. Our approach was comprised of two main phases. In the first phase, we performed a preliminary investigation via survey to understand the current challenges and needs for online meetings, and gathered design requirements for our technology probe. Informed by our findings from the preliminary study, in phase 2 we conducted two design iterations through a longitudinal study of actual remote, recurring team meetings. In the following sections, we describe the details of the requirements analysis and the longitudinal studies.

3 PHASE 1: REQUIREMENT ANALYSIS SURVEY STUDY

In the first phase of our work, we aimed to identify the challenges and the expected solutions through a survey study. Our goal is to understand how participation and inclusivity during meetings affect meeting effectiveness, what social signals are relevant in the context of meeting effectiveness, and what challenges people face in videoconferences. The findings from our survey were used to gather requirements and inform the design of our AI-assisted feedback system that we discuss later (Section 4). The survey was approved by the Microsoft Research Institutional Review Board (IRB).

3.1 Survey Design

The topics of our survey questions spanned the frequency of meetings, meeting effectiveness, challenges during meetings, and useful information and signals from meetings. We asked the participants to self-assess the meeting effectiveness, perceived inclusivity, and participation in the most recent effective and ineffective meeting that they had. We asked the participants to provide quantitative and qualitative feedback on the importance of a variety of information for meeting effectiveness, such as social signals, meeting summary, participation of the attendees (both time spent talking and turn-taking), tone of the meeting, etc. Participants were asked to reflect on when this information would be useful and how it could be presented to them (e.g., personalized view, highlight reel).

3.2 Participants

We recruited survey participants through an e-mail advertisement at a large technology company. A total of 120 completed survey responses were collected. While participating in the survey, participants were working from home and joining meetings via videoconferencing due to COVID-19 mandatory requirements. Participants reported as working in the roles of a program manager (25%), developer (23%), manager (17%), researcher (3%), and administrative assistant (1%). 58% of participants reported that they organized 1-5 meetings per week, and 40% of participants reported that they attended more than 15 meetings per week.

3.3 Analysis & Findings

We used a chi-square test on quantitative responses to investigate differences between effective and ineffective meetings. We also performed a thematic analysis [6] on open-ended responses to derive themes around topics of interest (e.g., challenges, information needs). We then categorized the responses into various themes and quantified their occurrence. Below, we highlight three findings from our analysis: (1) factors that influence meeting effectiveness, (2) challenges participants faced in online meetings, (3) solutions that participants proposed or desired, and (4) privacy and trust related concerns.

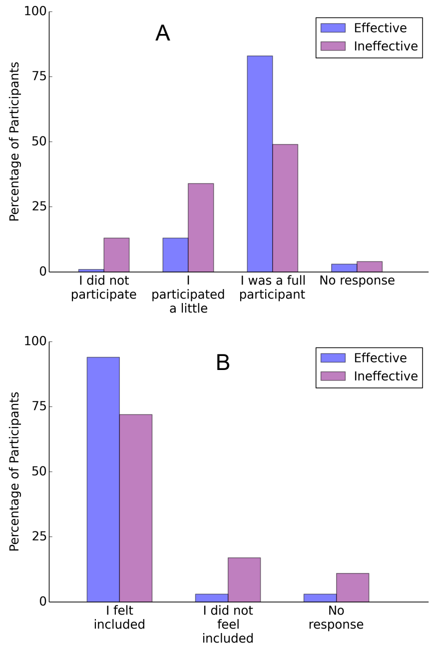

Figure 1: Relationship between effective & ineffective meetings with respect to participation & inclusivity. (A) Full or equal participation prevailed more in effective meetings ($\chi^2(1, N = 231) = 34.89, p < 0.001$). (B) The feeling of inclusivity was more prominent in effective meetings ($\chi^2(1, N = 240) = 21.78, p < 0.001$).

3.3.1 ***Meeting Effectiveness & Inclusivity.*** From the meetings that the participants reported as effective or ineffective, we compared the differences in self-assessed participation or feeling of being included. We found that full participation was significantly more prevalent in effective meetings in comparison with ineffective ones ($\chi^2(1, N = 231) = 34.89, p < 0.001.$). We also found that the feeling of being included was significantly more prevalent in effective meetings in comparison with ineffective ones ($\chi^2(1, N = 240) = 21.78, p < 0.001$). Fig. 1 shows the comparison of effective and ineffective meetings across participation and inclusivity.

3.3.2 ***Challenges in Online Meetings.*** The top two challenges faced in online meetings that participants reported were related to the connection issues (69.3%) and the meeting agenda (or lack of) (64.91%), as can be seen in Fig 2. Excluding tech issues, responses indicate that attendees faced difficulties with maintaining agenda items (64.91%), repetition (46.49%), reaching goals (30.7%), being heard (self: 28.95%, others: 27.19%), negative tone (14.91%). A total of 78% of the participants reported understanding social signals to be very important in meetings and 55% of the participants expressed that social signals are more difficult to understand in online meetings, even with audio/video, compared to face-to-face meetings.

3.3.3 ***Proposed Solutions.*** Through thematic analysis on the responses, we identified five main categories of solutions brought up by the participants to address the aforementioned challenges. (1) *Preparing for the meeting*: 37% of participants wanted to have a clear agenda before the meeting and a summary of goal/outcome after the meeting. (2) *Ensuring productive meeting dynamics*: 30% of participants mentioned specific features that could assist in maintaining productive meeting dynamics – 26% were participation or turn-taking related (e.g., attendees speaking as per their role in the agenda, who are participating and for how long, whether someone is getting a chance to speak), and others included having enough pauses for smoother turns, what questions are being asked, what the agreement level is. (3) *Improving the usability of the tools*: 22% of participants asked to improve general hardware/software/usability related components of the video-conferencing platform. (4) *Conveying social cues*: 14% of participants suggested keeping attendees' cameras on to understand each other's social cues. (5) *Moderating the meeting*: 8% of participants requested some form of effective moderation whereas 4% brought up the need to have a dashboard that summarizes the meeting events or notes, and the other 4% suggested having a meeting moderator who can keep the discussion on track and productive during the meeting. Nine out of ten participants expressed interest in a post-meeting feedback tool with 56% asking for combined highlights of audio, video, and transcripts.

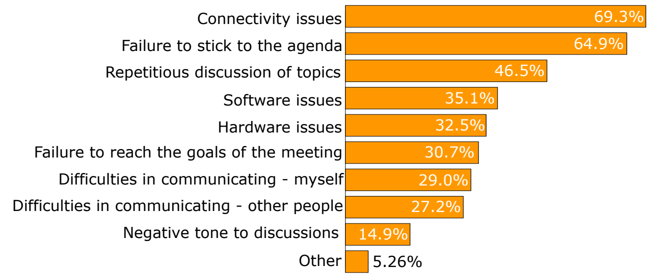

Figure 2: Percentage of participants in our survey that reported experiencing challenges in workplace meetings.

3.3.4 ***Privacy and Trust Concerns.*** Participants brought up concerns regarding with whom the meeting data would be shared. They did not want sentiment as a behavioral feature to be included in the post-meeting tool for four reasons: (1) trust of AI: suspicion regarding how sentiment computation works, its accuracy, what the outcome would be if the computation is wrong, etc. (*"I don't feel that AI is good enough at sentiment."*), (2) privacy concerns: participants felt uncomfortable sharing their own or accessing others' sentiment data (*"I think it would be overly intrusive to have this data from other participants"*), (3) applicability: participants were not certain how they would use sentiment information (*"I'm not sure if I'd find that useful, or how."*), and (4) subjectivity: sentiment might be context-dependent and negative/neutral tone might not make a meeting ineffective (*"I have some doubts on whether this can be done accurately, given how subjective some of this information would be. I would have to see an example."*). People were intrigued to see how the real data would look (*"As long as the intelligence could truly determine negativity or positivity it would be good documentation and a way to think about how to flip a negative meeting to the positive with the same group."*)

Participants expressed challenges and opportunities regarding behavioral (e.g., participation, turn-taking) and contextual (e.g., agreement, questions) issues faced in online meetings. Their interest in a post-meeting tool with highlighted meeting events was also prominent. Based on these, we identified the features and designed a post-feedback tool, which we discuss in Section 4.

4 PHASE 2: LONGITUDINAL USER STUDY

The second phase of our work involved a 4-week longitudinal user study. We collected in situ meeting data to develop an AI system that extracts salient and desired signals identified in our requirement study (Section 3). The collected data were used to conduct two design iterations and evaluations of a meeting feedback system with the participants, first with mocked data (Section 4.2) and second with real data from meetings captured during the user study (Section 4.3).

4.1 Longitudinal Study Design

Our longitudinal user study involved capturing audio and video streams from regularly scheduled weekly meetings of teams. We targeted recurring meetings of teams rather than one-off meetings because the members of recurring meetings have more opportunities to invest in improving their meeting effectiveness. We collected separate audio-video stream from each participant of the meeting. We captured the meetings of in situ work teams to ensure that the data used for our system and the meeting feedback we provided were contextually appropriate and relevant. In addition, each member of the participating teams provided feedback on our initial wireframe design (Section 4.2) and evaluated their own, personalized interactive dashboards (Section 4.3).

The participants were compensated $10 per meeting hour for taking part in the data collection, and an additional $10 for participating in an interview after the study. We recruited a total of 8 teams from the same company (49 participants, with an average of 6 participants per team ($min = 3$, $max = 10$)), who consented for us to record their weekly project meetings over 4 weeks. This second phrase study was also approved by the IRB. All teams conducted their meetings via the Microsoft Teams video-conferencing platform, closely mirroring their current workplace meeting practices: at the time of the study, all meetings were conducted remotely due to the COVID-19 work-from-home policy. We recorded a total of 28 meetings with an average duration of 32 minutes ($min = 11$, $max = 62$).

4.2 Design and Evaluation of Meeting Feedback Wireframe

Our findings from Section 3 informed the initial design of the meeting feedback system. We first enumerated specific feedback features that could support challenges and solutions expressed by our survey participants. We then designed meeting feedback wireframe that included those feedback features. Using the prototype wireframe, we conducted a survey study with our longitudinal study participants at the end of their 2^{nd} week of the 4-week longitudinal study. The goal of the survey was to evaluate the usability and effectiveness of each of these features and to inform which features or components should appear in the next design iteration. We collected a total of 16 completed evaluations from the participants. Here, we describe the details of the design of the meeting feedback wireframe and the results of the evaluation survey of our first design.

4.2.1 ***Design Strategies & Feature Definitions.*** As our previous requirement analysis survey study revealed the need for a post-meeting feedback tool with meeting highlights, we focused on designing a wireframe prototype with the meeting dynamics features. First, we categorized the feedback features into two classes: (1) behavioral (e.g., participation, sentiment), and (2) topical (e.g., transcription, questions about the meeting content). Second, we identified that participants felt more comfortable sharing the topical contents with others, whereas they wanted most behavioral features to be kept private. We incorporated those priorities in the wireframe dashboard design. Finally, participants expressed interest in a combination of context-based summary and action items. As such, we organized the feature visualization into three ways: (1) summarized: showed a quick snippet of the average measurement of a feature; (2) suggestive: provided actionable suggestions related to the individual's feature outcomes; and (3) temporal: presented meeting events in a contextualized timeline manner. For all temporal features, the idea was that upon clicking on any of the event components, the corresponding video and transcript portions would be

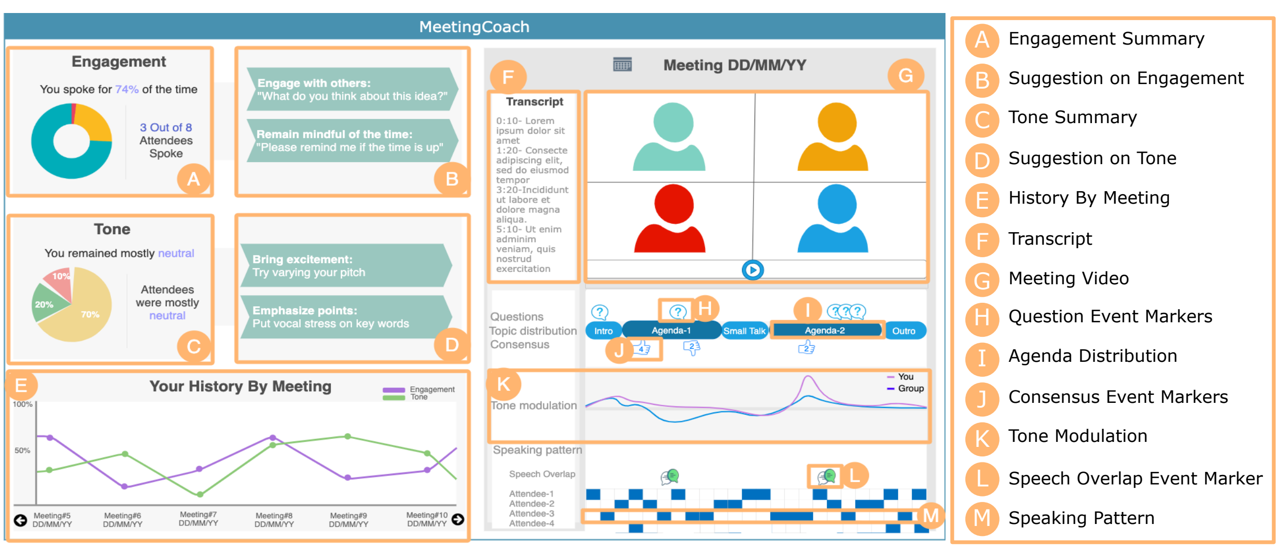

Figure 3: Visualization of our wireframe dashboard design and its components.

Feature Definition	Type	Viz.	Privacy
(A) Engagement Summary: The percentage of time one spoken during the whole meeting. It also showed the number of attendees who spoke at least once. The intention was to help everyone feel included and heard.	Behavioral	Summarized	Shared
(B) Suggestion on Engagement: Tips on improving meeting engagement, which were updated based on whether the "Engagement" metric was "too high", "too low", or "equal". The intention was to provide specific guidance to improve group engagement in future.	Behavioral	Suggestive	Private
(C) Tone Summary: The percentage of time one's sentiment remained positive/neutral/negative. The sentiment was measured from facial signals. The intention was to grow awareness about perceived tone.	Behavioral	Summarized	Private
(D) Suggestion on Tone: Tips on improving one's meeting sentiment related to facial expression. The intention was to provide specific guidance to improve individual sentiment in future.	Behavioral	Suggestive	Private
(E) History by Meeting: Average speaking time and sentiment in past meetings. The intention was to track behavior modulation over the time.	Behavioral	Summarized	Private
(F) Transcript: Timestamped text of what was spoken during the meeting by which participant. The intention was to allow reading the script for connecting context to the other temporally visualized features.	Topical	Temporal	Shared
(G) Meeting Video: The recorded audio-video feed of the meeting. The intention was to allow replaying the video for connecting context to the other temporally visualized features.	Topical	Temporal	Shared
(H) Question Event Markers: The moments when and what questions were asked by which attendee.	Topical	Temporal	Shared
(I) Agenda Distribution: Identification of which agenda items were discussed during the meeting and when.	Topical	Temporal	Shared
(J) Consensus Event Markers: The moments when attendees agreed or disagreed. It was measured by a combination of head-nods and head-shakes. The number represented the count of members contributing to that agreement/disagreement.	Topical	Temporal	Shared
(K) Tone Modulation: Facial sentiment of the group and the individual throughout the meeting. The intention was to be able to connect affect-rich moments with other meeting attributes, such as agenda.	Behavioral	Temporal	Private
(L) Speech Overlap Event Markers: The moment when two or more members spoke at the same time. The intention was to observe speaker's floor transfer, interruption, etc.	Behavioral	Temporal	Shared
(M) Speaking Pattern: Showed which member spoke during what portion of the meeting. The intention was to review turn-taking, speaking duration, agenda-wise speaker selection, etc.	Behavioral	Temporal	Shared

Table 1: Definitions of the features included in our wireframe dashboard. Each feature was associated with behavioral or topical information, some were summarized states, some temporal graphical visualizations, and others suggested actions. Those that were only shown to the individual are labeled as private.

played/highlighted. We organized the left panel to have the summarized and suggestive features, whereas the right panel presented all the temporal features in their timelines. The wireframe prototype design is shown in Fig 3. We provide our primary feature definitions and the corresponding constraints in Table 1.

4.2.2 ***Results & Next Steps.*** The responses from the evaluation survey showed that participants perceived the system prototype to be important (Fig. 4A, $M = 4.5, SD = 1.62$) and capable of impacting meeting inclusivity (Fig. 4B, $M = 4.63, SD = 1.32$). Notably, as the wireframe version was not interactive, the contextual information tied to the temporal feedback could not be fully explored. Participants were interested in using the interactive version of MeetingCoach and discussed the potential of the AI sensing capabilities:

> P2: *"I like the dashboard and the insights it provides on engagement and recommendation. The effectiveness and impact of the dashboard really depend on the definition (AI algorithm) of the metrics on the dashboard."*

> P8: *"This would be great if you could hover over the components to see what the questions, comments, and topics were."*

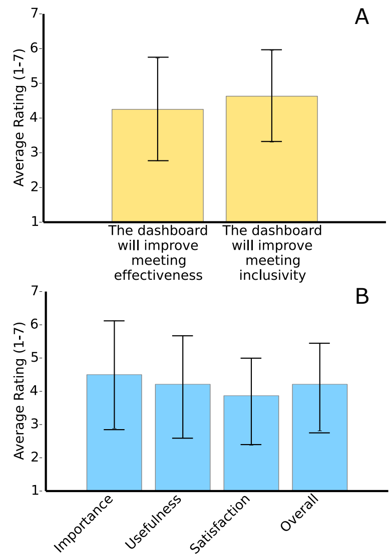

Figure 4: Survey responses evaluating the wireframe dashboard. (A) Perceived agreement on the impact of the dashboard on improving meeting effectiveness and inclusivity. (B) Rating of the potential usability of the wireframe.

Based on these responses, we modified the implementation of the interactive dashboard features: (1) how each feature was defined (e.g., hovering over the term *consensus* revealed that it was captured from visual signals with a combination of head-nods and head-shakes), and (2) what inherent information each feature holds (e.g., hovering over each *consensus event* revealed the exact distribution of head-nod/head-shake leading to the overall consensus, and upon clicking on an event marker, the corresponding portion of the meeting video was played).

Among the features, *speaking turn* ($M = 5.8, SD = 0.98$), *engagement* ($M = 5.53, SD = 1.32$), *and topic distribution* ($M = 5.47, SD = 1.15$) were perceived as the most useful features. The usefulness of the wireframe feature components are shown in Fig. 7 (yellow bars). Participants valued the idea of quickly traversing through specific portions of the meeting and reviewing the discussion (*"I would personally see more use for it (speaking turn) to jump to specific sections in a meeting recording where a specific speaker is talking"*). Some participants also asked for suggestions based on turn-taking patterns, whereas some mentioned that it might be too context dependent. Temporal sentiment modulation was rated the least useful ($M = 3.73, SD = 1.24$). Participants found it interesting but were unsure how to utilize the information (*"This might be useful, but not sure how to take action on it"*). However, some also expressed an interest in evaluating it with their own real data. Participants also suggested some modifications to the representations of the components:

> P7: *"This (speech overlap) should be integrated into the speaking turn graphic if it is meant to overlap."*
> P2: *"Is it (average tone) accessible and how can I be sure the information is private and confidential to me?"*

In the next step, we updated the design to integrate speaking patterns and speech overlap into one representation. We updated terms such as *engagement* to *participation* as used in some previous literature, *tone* to *your sentiment* to make its scope unambiguous. Accurately identifying different agendas of the meeting and precisely marking the timestamp for each of them were difficult to implement. Another major challenge was differentiating between small talk and a short agenda discussion. Therefore, we exclude *topic distribution* from the final dashboard. The next section discusses the interactive dashboard in detail.

4.3 Design and Evaluation of Interactive Feedback Dashboard

Our evaluation of the prototype wireframe revealed the need for several design changes. We carried out these changes in a second design iteration phase, extracted the proposed features from actual meeting data collected from the study, and implemented the interactive dashboard. At the end of the 4-week study, we evaluated the interactive dashboard through a mix of semi-structured interviews and a survey study with our study participants.

First, we conducted a 30-minute, semi-structured interview session with nine randomly selected participants, one at a time, to capture their interaction behaviors with the dashboard. We conducted the interviews first so that all participants were interviewed about their first impression of the dashboard. The interview adopted 'think aloud' or cognitive walk-through approach to understand the interaction patterns while answering the interview questions. The semi-structured questions asked the interviewees to find out who spoke most and least in the meeting, how meeting participation changed from week to week, when meeting sentiment changed and whether they agreed, when a question was asked by attendee-X and whether it was answered, who interacted the most, how likely it was that they would continue using the system, what meeting inclusivity meant to them and whether the dashboard supported it, and what further modification the dashboard needed.

Second, the participants in the study received an individualized feedback dashboard populated with their actual meeting data. Each participant could see only their own participation history over multiple meetings, average and temporal sentiment, etc., and meeting specific shared information, such as the meeting transcript, temporal group sentiment, etc. After interacting with the dashboard, the participants filled out a survey evaluating the system's ability to improve meeting awareness, effectiveness, and inclusivity. Here, we describe the details of the system design, feature implementation, and the evaluation results of our interactive dashboard.

4.3.1 ***System Development.*** We used Microsoft Teams meetings for our data collection. A customized video recording bot[3] was created to record the video and audio data from each participant in the meeting separately. This bot was added to all the meetings in our study and the recordings stored on a secure server. The recordings were then processed with a series of software pipelines (shown in Fig. 5) to automatically extract features related to the verbal and non-verbal content of the meeting. After extracting the feature signals and processing them into categorized feedback metrics, we implemented a web-based interactive dashboard by using HTML,

[3]https://docs.microsoft.com/en-us/microsoftteams/platform/bots/calls-and-meetings/real-time-media-concepts

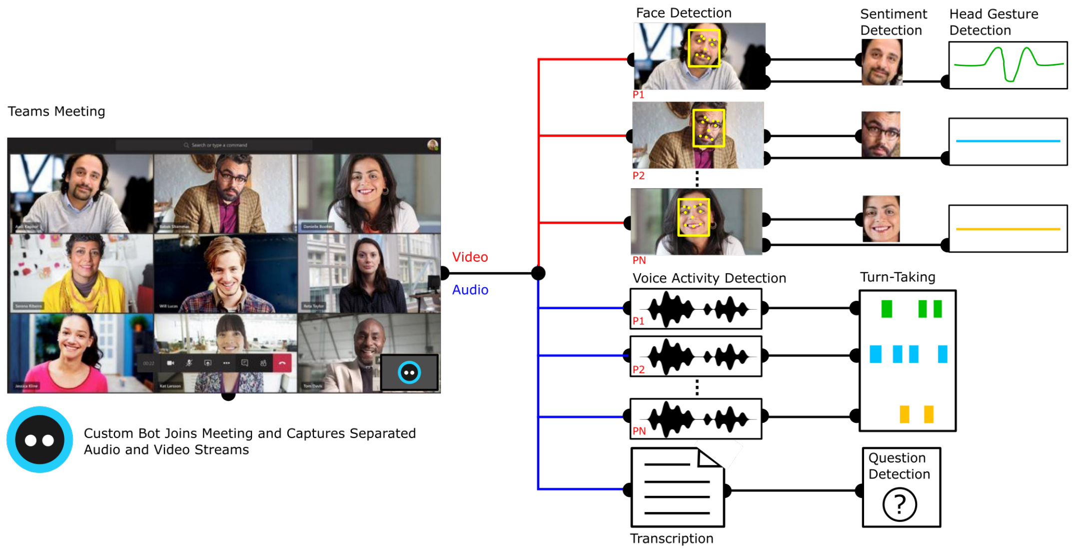

Figure 5: We developed a custom Microsoft Teams Bot to record and analyze the independent audio and video streams from each member of an online meeting. Each video stream was processed to detect the face of the participant and then analyze expressions and head gestures. Each audio stream was processed to detect voice activity of the participant and then determine turn-taking. The audio was transcribed and questions detected from the resulting text.

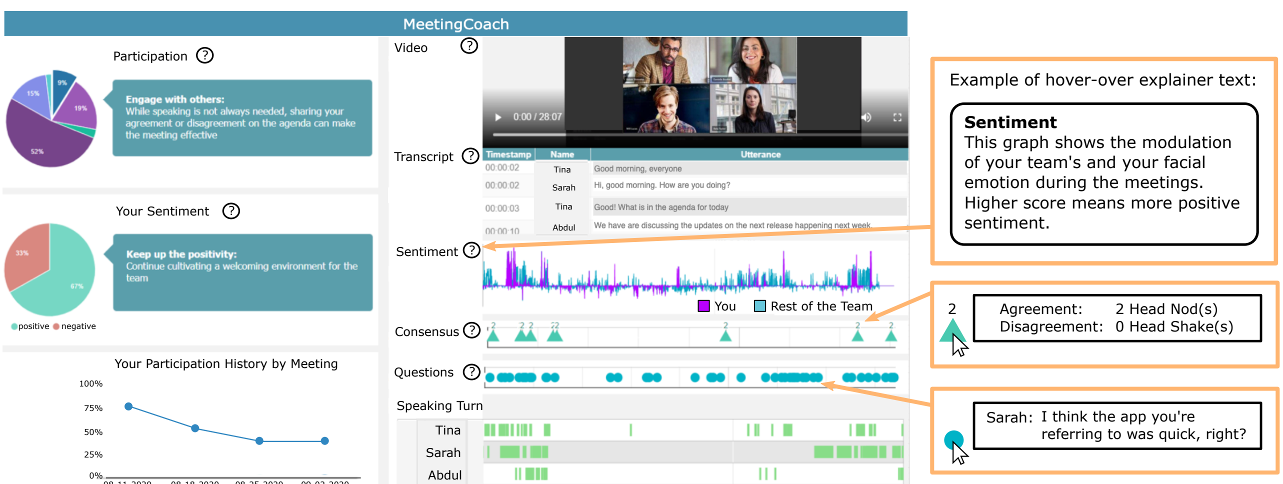

Figure 6: A screenshot of our interactive web dashboard on the left. Hover-over functionalities are shown on the right.

jQuery, D3.js, and Google Charts Developer Tool. The dashboard was hosted in Microsoft Azure so that participants could interact with it from anywhere. Participants received unique access IDs to view their individualized dashboards. Fig 6 shows the interface of the interactive dashboard. Right side of the corresponding figure shows the hover functionalities. Upon clicking a temporal event, the corresponding recorded meeting video was played.

*4.3.2 **Feature Implementation**.* After collecting the recorded meeting audio-video feed, we used a multi-modal sensing pipeline (shown in Fig. 5) to extract the group dynamics related signals from the collected feed and process them into feedback features:

Transcription and Questions Detection. We used the Microsoft conversation transcription API[4] to extract the transcripts of the meeting recordings. Given that the video and audio feeds of each attendee were recorded separately, we used the speech-to-text system to extract the transcript for each attendee and then synchronized them with time stamps.

Participation & Turn-taking Detection. Our participation metric was based on the duration for which each meeting attendee spoke during the meeting. The audio received from the recordings was sampled at 16kHz and passed through the Microsoft Windows

[4]https://docs.microsoft.com/en-us/azure/cognitive-services/speech-service/conversation-transcription

Voice Activity Detector (VAD) [55] which provided a Boolean value for every second of audio. For each attendee of a meeting, *participant* feature was computed as the percent average talk-time. To implement *speaking-turn*, the individually separated audio for each attendee was analyzed and then the metrics were synchronized using the video timestamps.

Face and Facial Landmark Detection. We used the Microsoft Face API[5] to detect the faces in each of the video frames and applied a landmark detector to identify the eyes, nose, and mouth. These data are used for the downstream components of head gesture detection and facial sentiment classification.

Agreement/Disagreement and Consensus Detection. We used a Hidden Markov Model (HMM) to calculate the probabilities of the head nod and head shake gestures. The HMM used the head yaw rotation value over time to detect head shakes, and the head Y-position values over time to detect head nods. Consensus was measured whenever two or more attendees asserted either head nod or head shake signals in the synchronized timeline.

Sentiment Classification. The faces were cropped from the video frames using the bounding box information provided by the face detector, and the resulting image-patches were sent to a facial expression detection algorithm. The facial expression detector returned eight probabilities, one for each of the following basic emotional expressions: anger, disgust, fear, happiness, sadness, surprise and neutral. This is a frequently employed categorization of facial expressions; however, it is not without critics, as displays of emotion are not uni-modal or necessarily universal [3, 16]. We used the publicly available perceived EmotionAPI[6], allowing other researchers to replicate our method. The emotion detection algorithm is a Convolutional Neural Network (CNN) based on VGG-13 [5] which has been externally validated in prior work [40, 42].

Suggestion Construction. We prepared a list of suggestions to appear based on the different levels of participation and sentiment. Example of suggestions included "Before moving onto the next agenda, consider checking whether everyone is on the same page" (more than average participation), "While speaking is not always needed, sharing your agreement or disagreement on the agenda can make the meeting effective" (less than average participation), "Continue cultivating a welcoming environment for the team" (more positive sentiment), "Consider expressing any concern you may have about the agenda" (more negative sentiment), "Your webcam was off during the meeting. Turning on video can help your team engage better" (no sentiment data due to no video feed), etc.

4.3.3 ***Results.*** The average responses of the survey questions are presented in Fig 8. In general, on a scale of 1-7 in which the higher the better, participants found the system to be important ($M = 4.91, SD = 1.41$) and useful ($M = 5.23$, $SD = 1.38$). They agreed that the dashboard would improve their meeting awareness ($M = 5.45$,$SD = 0.86$). They perceived that the dashboard would improve meeting effectiveness ($M = 4.45$, $SD = 1.47$) and inclusivity ($M = 5.23$, $SD = 1.23$). Five participants specifically commented that they found the system to be useful enough to utilize across other meeting series (*"I really like the types of data that are presented in the dashboard, and I would be curious to use it for more meetings to help drive inclusivity and engagement."*).

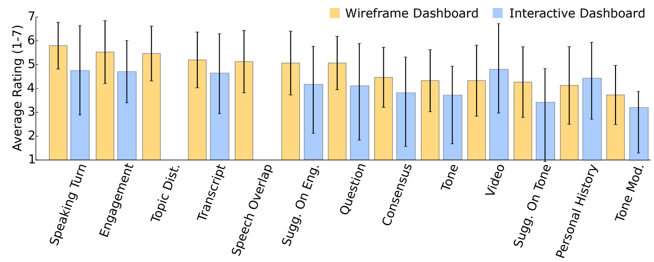

Figure 7: Comparison of the usefulness ratings of the feature components in our iterative evaluation phases. Yellow and blue bars represent the ratings participants provided to each component after reviewing the wireframe and the interactive dashboards, respectively. Notably, *Topic Distribution* and *Speech Overlap* ratings are null for interactive dashboard as the features are excluded during iteration.

After interacting with the dashboard and reviewing their own meeting data, participants were able to draw insights from the meeting ($M = 4.68, SD = 1.43$). However, the impact varied across different features. For example, participants could easily determine at which point each attendee spoke and for how long from the *turn-taking* feature ($M = 6.00$, $SD = 0.82$), and could deduce from the *consensus* feature who generally contributed to any decisions made ($M = 3.45$, $SD = 1.22$). Although the *consensus* feature visualized when an agreement/disagreement may have occurred, inspections of the meeting video and transcript were still needed to understand exact details, such as whether the group made a decision and precisely who contributed to it.

In terms of the usefulness of the feature components, *video*, *speaking turn*, and *participation* were highly rated (shown in Fig 7 (blue bars)). Participants appreciated having the *meeting video* to providing detailed context for the event highlights. During the semi-structured interview, participants expressed that having the video and the transcript helped them verify the accuracy of the features and understand the context of that event snippet. *Sentiment* related features received lower ratings. The interview participants expressed that because of the nature of any workplace meeting the sentiment features remained mostly neutral, therefore most of the time there was not any specific events to focus on. However, participants also expressed that feedback might be only useful in case of a major shift towards negative sentiment.

> P13: *"I'd be worried about the accuracy of the sentiment measurements. Maybe a better understanding of what "negative" sentiment means, together with some examples, would help alleviate my worry."*
> P23: *"People just might be serious in the meeting, which might translate as negative, and that would not be appropriate."*

Participants expressed how comfortable they were with the privacy strategy adopted in our system. On a scale of 1-5 ("Not at all" to "A great deal"), participants felt comfortable knowing their participation ($M = 4.41, SD = 0.67$) and turn-taking ($M = 3.59, SD =$

[5] https://azure.microsoft.com/en-us/services/cognitive-services/face/
[6] https://docs.microsoft.com/en-us/xamarin/xamarin-forms/data-cloud/azure-cognitive-services/emotion-recognition

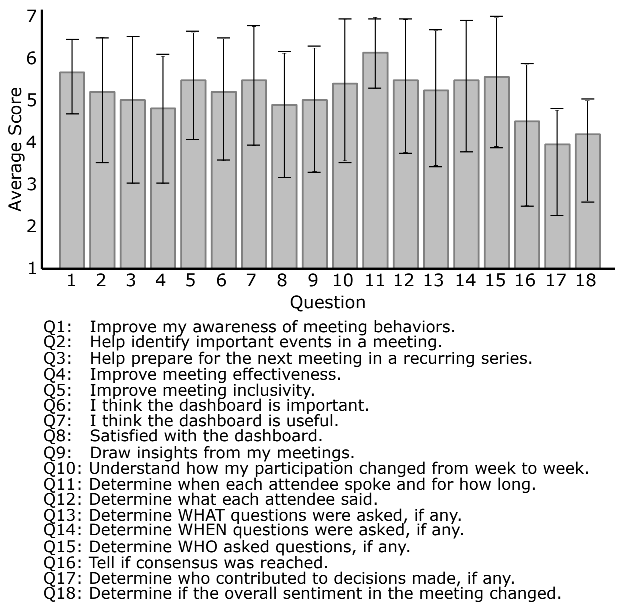

Figure 8: User ratings evaluating the interactive dashboard

1.18) information were shared and sentiment information was private ($M = 3.36, SD = 1.26$).

Participants also provided recommendations for improving the dashboard. They suggested having fewer and/or customizable features: *"The dashboard should be customizable so that individuals can see at a glance the things they most care about."* Participants requested feature refinements: *" It's hard, for example, to identify important events of a meeting because there are so many events. On the questions view, for example, some events just say "Yes?" and some are paragraphs of text long."* New features were also suggested: *"I'd love to see stats on interruptions as well. Maybe also stats on aggressive behavior, e.g., raising my voice."* To improve trust in the system, users mentioned that it could be more transparent: *"I feel further work is required to improve transparency, information design and layout, and trust with the user using the system."*

For most feature components, participants expressed the need for actionable suggestions that could help them effectively modify their dynamics in future, if needed. In our current dashboard, only two summarized features on the left panel (participation and sentiment) had explicit suggestions. The participants thought that those brief suggestions regarding how to improve meeting behaviors in the future made the feedback more impactful.

> P2: *"I really enjoy the actionable recommendations. I think it's great to have these recommendations on the screen. The recommendations were actually the first thing I looked at."*
> P3: *"At the end of a meeting you don't want to sit and study data. You want one nudge. Good nudges would be ... giving hints on what you could do next time to make the meeting more inclusive."*

Our dashboard was a combination of event highlights (e.g., what questions were asked) and behavior patterns (e.g., length of each speaking turn). The findings showed that based on when feedback was being delivered, the information need might be different. Participants mentioned that actionable suggestions would be useful to have as a reminder right *before* the next meeting, whereas the event highlights would be useful to review after the meeting.

> P11: *"I feel there's lots of interesting ideas in this 'one dashboard' that I'd rather see in different places rather than all in one overview. I would rather receive an email or a reminder just before a meeting about the information in the left column to serve as a reminder on how to improve my upcoming meeting. The right-column looks very useful in Teams as maybe a tab inside the meeting chat-channel? I'd personally split these two up into separate components."*
> P13: *"I feel like this would be more valuable to managers and new members on the team if they're trying to figure out how they fit in. I would probably not use this for every meeting, but I would use it to reflect to meetings that I want to reflect on, either if I was feeling a meeting didn't go well that I wanted it to, or if I wanted to look back and remember what was said during the meeting."*

5 DISCUSSION

Our study reveals a unanimous agreement that attendees of videoconferencing meetings are in the need of feedback assistance to make their meetings more effective and inclusive. The inability to understand the social cues, in association with hardware- and software-related challenges, make online meetings difficult for workplace meeting attendees. However, currently no major videoconferencing platforms incorporate such elements. There is an opportunity to help attendees improve their meeting dynamics by leveraging the audio-video feed within such platforms.

The complexity of meeting dynamics makes it challenging to generate contextually appropriate feedback for the attendees. We explored the usefulness of various behavioral and topical features and provided the context through a temporal visualization. Participants found it useful to have the recorded meeting video and the transcript included with the dashboard, as these enabled them to quickly review the context of different feature events.

Group-based features require data from participating team members. If not everyone in the meeting agrees to be recorded and included in the feedback tool, these features may not fully describe the meeting dynamics. Although our work studied real workplace meetings, we only recruited teams whose entire team agreed to participate in the study. We found that potential participants carried concerns regarding how the data would be captured and shared, who would have access to the data and the dashboard, how would sensitive meeting contents be excluded from external analysis, etc. In addition to our potential participants' concerns, team members may feel coerced into participating in the group's effort to improve its meeting effectiveness and inclusivity, even when one is not comfortable with sharing their video and audio feeds. Another concern is in using this data for performance evaluations, with or without the employees' direct knowledge. Making sure that the data capture and the dashboard visualization addresses such concerns of all the attendees is key to deploying such a system. Future research should study how to appropriately capture consent and communicate the role of the feedback tool.

5.1 Implications for Design

During the iterative process of the study, the participants mentioned suggestions, concerns, and expectations they had regarding a post-meeting dashboard. Extrapolating from those suggestions, we suggest some key design improvements and future possibilities.

5.1.1 ***Provide actionable suggestions.*** Being able to dive into the temporal features is valuable for examining specific events and cross-matching those with the meeting context. However, analyzing so many features is potentially time-consuming and overwhelming. Based on our findings, we propose that feedback systems, especially with a high number of features, need to provide more actionable suggestions that can help direct the behavior nudges.

From an implementation perspective, providing suggestions for every meeting dynamic is challenging, as the suggestions will by necessity be context dependent. For example, *turn-taking* patterns will greatly vary based on the type of the meeting (scrum, planning meeting, presentation, etc.). To understand the meeting context, AI systems need more information about the meeting type and the team culture, beyond simply capturing the meeting audio-video signals.

5.1.2 ***Modify feedback design based on its delivery timing.*** Our iterative study revealed that meeting attendees have different needs before, after, and during the meeting. Therefore, feedback should be modified based on the timing. We categorize the prospective delivery or timing of feedback information as below:

(a) Reminders before Meetings: Right before attending a meeting, quickly reviewing any actionable suggestions or feedback from the past meetings can help maintain effectiveness and inclusivity. We find that participants wanted actionable suggestion as a reminder immediately before meetings. These notifications can help individual attendees remember behavioral goals. Therefore, we propose that summarized suggestions from the last meeting could be sent as a reminder before the next meeting of the series.

(b) Highlights after Meetings: Participants mentioned that the temporal information, especially regarding the topical features (e.g., transcript, questions) would be more useful when someone missed the meeting, or needed to recollect particular information from the previous meeting. They also mentioned that exploring the temporal behavior features (e.g., consensus) after the meeting would be useful to evaluate what actions needed to be taken in the next meeting. Thus, feedback systems built with the intent to help formulate goals for the next meeting should incorporate event highlights from the previous meetings.

5.1.3 ***Provide training opportunities based on feedback from a series of meetings.*** The current implementation focuses on per-meeting feedback, but participant interest in splitting feedback into pre-meeting and post-meeting time periods indicates that there is the potential for further valuable longitudinal use of the feedback. Aggregating personal and team feedback across multiple meetings might enable deeper learning and foster good habits for the long-term. Further, feedback from such a system could also be used to personalize meeting training programs for both managers and individuals who are seeking to measurably improve their effectiveness and inclusion.

5.1.4 ***Incorporate multi-modal signals to provide rich feedback.*** Our findings show that to get deeper insights of a feature of group discussion, analyzing a single modality might not be enough. Multi-modal analysis of a feature may not only provide a more refined and accurate metric, but also may achieve more trust from the user. For example, in our implementation we measured *consensus* by using head nod and shake visual signals. Combining this with language properties to capture whether words related to agreement or disagreement were used co-temporally with head nods and shakes could improve the understanding of consensus. Previous work [56] has also used agreement analysis from a single modality by applying Language Style Matching. Based on our findings, we propose that future systems should incorporate multi-modal signals to implement features for capturing group dynamics.

5.2 Limitations

The meetings we recorded, while diverse in terms of gender and job role, were all from the employees of a single large technology company during COVID-19 mandatory working from home restrictions. Future work will be needed to validate whether they generalize beyond that sector and context. In addition, the team meetings were recurring, which meant that the study participants all knew one another to some degree. This may also have influenced the perceived effectiveness and inclusivity of the meetings, but also the comfort level with sharing the tone and participation information about the meetings. All participants were remote in this study, but enabling such a system for hybrid meetings will add a significant level of extra complexity, both technically and in the way that people can use the system. Being recorded for a study might have caused a Hawthorne effect impacting how participants behave. However, we presume the Hawthorne effect to be minimal because our participants were already accustomed to being recorded during meetings for documentation purposes, and the meetings we recorded over 4 weeks were recurrent meetings with already familiar colleagues. In our study, the participants interacted with the dashboard once, but frequent and repeated usage over time can provide better insights on the impact of such a dashboard on meeting dynamics. In our future work, we would explore the dashboard usage over a longer period of time.

6 CONCLUSION

Video-conferencing meetings are a standard feature of modern work, even though meeting effectiveness and inclusivity can be diminished due to the constraints on the availability of social signals. We conducted an iterative study with two phases, including a requirements analysis and interactive system design (wireframe and interactive dashboards) to understand the effectiveness of these tools for video-conferencing meeting attendees. Our requirement analysis phase showed that, even though equal participation is not always expected, participation to some degree can increase the perceived effectiveness and inclusivity of a meeting. It also revealed that, in online meetings, attendees are painfully aware of not having access to each other's social cues, and that there is a need for a post-meeting dashboard to summarize meeting dynamic highlights. While some video-conferencing platforms provide some real-time features to make up for their constraints (e.g., hand raises, thumbs up, clapping, etc.), no major commercial platform leverages affective signals to provide actionable meeting metrics.

In our iterative system design, we included behavioral and contextual features in a summarized, temporal, and suggestive user interface dashboard. Based on the privacy concerns expressed by the participants, we kept sentiment features private to the individual, while including other features shared just among the meeting attendees (exactly as would be true of a meeting recording). The evaluation showed that reviewing behavior history over time can improve attendee's awareness to the meeting dynamics. Participants expressed an interest in comparing the meeting dynamics over different meeting series and expected more actionable suggestions that they could use in future meetings. Based on the feedback from our participants, we proposed having meeting reminders with suggestive feedback before a meeting, and event highlights with detailed insights after the meeting.

The kind of sensing explored in this study looks forward to a future of human-AI partnerships. We are at the point where we can influence the nature of those partnerships. On the one hand, such systems can act as prosthetics, enabling people to do things they otherwise could not but also setting up a relationship of dependency. On the other hand, such systems could be empowering and change the skills and capabilities of *users* over time [57]. Meetings, even when mediated, are valuable precisely because they leverage the immediacy and intimacy of human social connection to achieve what would be difficult to achieve by other means. We hope future systems incorporating similar kinds of affective and actionable highlights will enable people to be more inclusive and effective during meetings and, in the long-run, improve their comfort and productivity.

REFERENCES

[1] Joseph A Allen, Nale Lehmann-Willenbrock, and Steven G Rogelberg. 2015. *The Cambridge handbook of meeting science.* Cambridge University Press.

[2] Yadid Ayzenberg, Javier Hernandez, and Rosalind W. Picard. 2012. FEEL: frequent EDA and event logging - a mobile social interaction stress monitoring system. In *Extended Abstracts of Human Factors in Computing Systems.* 2357–2362.

[3] Lisa Feldman Barrett, Ralph Adolphs, Stacy Marsella, Aleix M Martinez, and Seth D Pollak. 2019. Emotional expressions reconsidered: challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest* 20, 1 (2019), 1–68.

[4] Sigal G Barsade. 2002. The ripple effect: Emotional contagion and its influence on group behavior. *Administrative science quarterly* 47, 4 (2002), 644–675.

[5] Emad Barsoum, Cha Zhang, Cristian Canton Ferrer, and Zhengyou Zhang. 2016. Training deep networks for facial expression recognition with crowd-sourced label distribution. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction.* 279–283.

[6] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[7] Dan Calacci, Oren Lederman, David Shrier, and Alex 'Sandy' Pentland. 2016. Breakout: An open measurement and intervention tool for distributed peer learning groups. *arXiv preprint arXiv:1607.01443* (2016).

[8] Janis A Cannon-Bowers, Scott I Tannenbaum, Eduardo Salas, and Catherine E Volpe. 1995. Defining competencies and establishing team training requirements. *Team effectiveness and decision making in organizations* 333 (1995), 380.

[9] Ross Cutler, Yasaman Hosseinkashi, Jamie Pool, Senja Filipi, Robert Aichner, and Johannes Gehrke. 2020. Meeting Effectiveness and Inclusiveness in Remote Collaboration. *arXiv preprint arXiv: To Appear* (2020).

[10] Terry L Dickinson and Robert M McIntyre. 1997. A conceptual framework for teamwork measurement. *Team performance assessment and measurement* (1997), 19–43.

[11] Joan Morris DiMicco, Katherine J Hollenbach, Anna Pandolfo, and Walter Bender. 2007. The impact of increased awareness while face-to-face. *Human–Computer Interaction* 22, 1-2 (2007), 47–96.

[12] Judith Donath, Alex Dragulescu, Aaron Zinman, Fernanda Viégas, and Rebecca Xiong. 2010. Data portraits. In *ACM SIGGRAPH 2010 Art Gallery.* 375–383.

[13] Wen Dong, Bruno Lepri, Taemie Kim, Fabio Pianesi, and Alex Sandy Pentland. 2012. Modeling conversational dynamics and performance in a social dilemma task. In *2012 5th International Symposium on Communications, Control and Signal Processing.* IEEE, 1–4.

[14] Norah E Dunbar and Judee K Burgoon. 2005. Perceptions of power and interactional dominance in interpersonal relationships. *Journal of Social and Personal Relationships* 22, 2 (2005), 207–233.

[15] Mohamed Ez-Zaouia, Aurélien Tabard, and Elise Lavoué. 2020. Emodash: A dashboard supporting retrospective awareness of emotions in online learning. *International Journal of Human-Computer Studies* 139 (2020), 102411.

[16] José-Miguel Fernández-Dols and Maria-Angeles Ruiz-Belda. 1997. Spontaneous facial behavior during intense emotional episodes: Artistic truth and optical truth. In *The psychology of facial expression*, James A Russell and José-Miguel Fernández-Dols (Eds.). Cambridge University Press, New York, NY, 255–274.

[17] C Marlene Fiol and Edward J O'Connor. 2005. Identification in face-to-face, hybrid, and pure virtual teams: Untangling the contradictions. *Organization science* 16, 1 (2005), 19–32.

[18] Michelle Gregory, Nancy Chinchor, Paul Whitney, Richard Carter, Elizabeth Hetzler, and Alan Turner. 2006. User-directed sentiment analysis: Visualizing the affective content of documents. In *Proceedings of the Workshop on Sentiment and Subjectivity in Text.* 23–30.

[19] Steve Harrison. 2009. *Media space 20+ years of mediated life.* Springer Science & Business Media.

[20] J. Hernandez, X. Benavides, P. Maes, D. McDuff, J. Amores, and R.W. Picard. 2014. AutoEmotive: Bringing empathy to the driving experience to manage stress. In *Proceedings of the Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, DIS.* https://doi.org/10.1145/2598784.2602780

[21] J. Hernandez, D. McDuff, R. Fletcher, and R.W. Picard. 2013. Inside-out: Reflecting on your inner state. In *IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops 2013.* https://doi.org/10.1109/PerComW.2013.6529507

[22] Dirk Heylen. 2005. Challenges ahead: Head movements and other social acts in conversations. *Virtual Social Agents* (2005), 45–52.

[23] Ellen A Isaacs and John C Tang. 1994. What video can and cannot do for collaboration: a case study. *Multimedia systems* 2, 2 (1994), 63–73.

[24] Alejandro Jaimes, Takeshi Nagamine, Jianyi Liu, Kengo Omura, and Nicu Sebe. 2005. Affective meeting video analysis. In *2005 IEEE International Conference on Multimedia and Expo.* IEEE, 1412–1415.

[25] Jeroen Janssen, Gijsbert Erkens, and Gellof Kanselaar. 2007. Visualization of agreement and discussion processes during computer-supported collaborative learning. *Computers in Human Behavior* 23, 3 (2007), 1105–1125.

[26] Florian Jentsch, John Barnett, Clint A Bowers, and Eduardo Salas. 1999. Who is flying this plane anyway? What mishaps tell us about crew member role assignment and air crew situation awareness. *Human Factors* 41, 1 (1999), 1–14.

[27] Taemie Kim, Agnes Chang, Lindsey Holland, and Alex Sandy Pentland. 2008. Meeting mediator: enhancing group collaborationusing sociometric feedback. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work.* 457–466.

[28] Uroš Krčadinac, Jelena Jovanović, and Vladan Devedžić. 2012. Visualizing the affective structure of students interaction. In *International Conference on Hybrid Learning.* Springer, 23–34.

[29] Kostiantyn Kucher, Daniel Cernea, and Andreas Kerren. 2016. Visualizing excitement of individuals and groups. In *Proceedings of the 2016 EmoVis Conference on Emotion and Visualization.* 15–22.

[30] Anastasia Kuzminykh and Sean Rintel. 2020. Classification of Functional Attention in Video Meetings. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* 1–13.

[31] G Ross Lawford. 2003. Beyond success: Achieving synergy in teamwork. *The Journal for Quality and Participation* 26, 3 (2003), 23.

[32] Harold J Leavitt. 1951. Some effects of certain communication patterns on group performance. *The Journal of Abnormal and Social Psychology* 46, 1 (1951), 38.

[33] Charlotte P Lee and Drew Paine. 2015. From The Matrix to a Model of Coordinated Action (MoCA) A Conceptual Framework of and for CSCW. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing.* 179–194.

[34] Jina Lee and Stacy C Marsella. 2010. Predicting speaker head nods and the effects of affective information. *IEEE Transactions on Multimedia* 12, 6 (2010), 552–562.

[35] Gilly Leshed, Diego Perez, Jeffrey T Hancock, Dan Cosley, Jeremy Birnholtz, Soyoung Lee, Poppy L McLeod, and Geri Gay. 2009. Visualizing real-time language-based feedback on teamwork behavior in computer-mediated groups. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 537–546.

[36] Madelene Lindström, Anna Ståhl, Kristina Höök, Petra Sundström, Jarmo Laaksolathi, Marco Combetto, Alex Taylor, and Roberto Bresin. 2006. Affective diary: designing for bodily expressiveness and self-reflection. In *CHI'06 extended abstracts on Human factors in computing systems.* 1037–1042.

[37] Paul Luff, Christian Heath, Hideaki Kuzuoka, Jon Hindmarsh, Keiichi Yamazaki, and Shinya Oyama. 2003. Fractured ecologies: creating environments for collaboration. *Human–Computer Interaction* 18, 1-2 (2003), 51–84.

[38] Sandi Mann and Lynn Holdsworth. 2003. The psychological impact of teleworking: stress, emotions and health. *New Technology, Work and Employment* 18, 3 (2003), 196–211.

[39] Rob McCarney, James Warner, Steve Iliffe, Robbert Van Haselen, Mark Griffin, and Peter Fisher. 2007. The Hawthorne Effect: a randomised, controlled trial. *BMC medical research methodology* 7, 1 (2007), 30.

[40] Daniel McDuff, Eunice Jun, Kael Rowan, and Mary Czerwinski. 2019. Longitudinal Observational Evidence of the Impact of Emotion Regulation Strategies on Affective Expression. *IEEE Transactions on Affective Computing* (2019).

[41] Daniel McDuff, Amy Karlson, Ashish Kapoor, Asta Roseway, and Mary Czerwinski. 2012. AffectAura: an intelligent system for emotional memory. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 849–858.

[42] Daniel McDuff, Kael Rowan, Piali Choudhury, Jessica Wolk, ThuVan Pham, and Mary Czerwinski. 2019. A Multimodal Emotion Sensing Platform for Building Emotion-Aware Applications. *arXiv preprint arXiv:1903.12133* (2019).

[43] Poppy Lauretta McLeod, Jeffrey K Liker, and Sharon A Lobel. 1992. Process feedback in task groups: An application of goal setting. *The Journal of applied behavioral science* 28, 1 (1992), 15–41.

[44] Michael Nowak, Juho Kim, Nam Wook Kim, and Clifford Nass. 2012. Social Visualization and Negotiation: Effects of Feedback Configuration and Status. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (Seattle, Washington, USA) *(CSCW '12)*. ACM, New York, NY, USA, 1081–1090. https://doi.org/10.1145/2145204.2145365

[45] Isabelle Odermatt, Cornelius J König, Martin Kleinmann, Maria Bachmann, Heiko Röder, and Patricia Schmitz. 2018. Incivility in meetings: Predictors and outcomes. *Journal of Business and Psychology* 33, 2 (2018), 263–282.

[46] Steven G Rogelberg. 2018. *The surprising science of meetings: How you can lead your team to peak performance*. Oxford University Press, USA.

[47] Marcello Russo. 2012. Diversity in goal orientation, team performance, and internal team environment. *Equality, Diversity and Inclusion: An International Journal* (2012).

[48] Samiha Samrose, Ru Zhao, Jeffery White, Vivian Li, Luis Nova, Yichen Lu, Mohammad Rafayet Ali, and Mohammed Ehsan Hoque. 2018. Coco: Collaboration coach for understanding team dynamics during video conferencing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 1–24.

[49] Sanat Sarda, Martin Constable, Justin Dauwels, Shoko Dauwels, Mohamed Elgendi, Zhou Mengyu, Umer Rasheed, Yasir Tahir, Daniel Thalmann, and Nadia Magnenat-Thalmann. 2014. Real-time feedback system for monitoring and facilitating discussions. In *Natural Interaction with Robots, Knowbots and Smartphones*. Springer, 375–387.

[50] Mark Schittekatte and Alain Van Hiel. 1996. Effects of partially shared information and awareness of unshared information on information sampling. *Small Group Research* 27, 3 (1996), 431–449.

[51] Abigail J Sellen. 1995. Remote conversations: The effects of mediating talk with technology. *Human-computer interaction* 10, 4 (1995), 401–444.

[52] John Short, Ederyn Williams, and Bruce Christie. 1976. *The social psychology of telecommunications*. John Wiley & Sons.

[53] Tony L Simons and Randall S Peterson. 2000. Task conflict and relationship conflict in top management teams: The pivotal role of intragroup trust. *Journal of applied psychology* 85, 1 (2000), 102.

[54] M.I. Tanveer, E. Lin, and M.E. Hoque. 2015. Rhema: A Real-time in-situ Intelligent Interface to Help People with Public Speaking. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*. 286–295.

[55] Ivan J Tashev. 2015. Offline voice activity detector using speech supergaussianity. In *2015 Information Theory and Applications Workshop (ITA)*. IEEE, 214–219.

[56] Yla R Tausczik and James W Pennebaker. 2013. Improving teamwork using real-time language feedback. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 459–468.

[57] Anja Thieme, Ed Cutrell, Cecily Morrison, Alex Taylor, and Abigail Sellen. 2020. Interpretability as a dynamic of human-AI interaction. *Interactions* 27, 5 (2020), 40–45.

[58] Wilbert van Vree. 2019. Formalisation and Informalisation of Meeting Manners. In *Civilisation and Informalisation*. Springer, 291–313.

[59] Shiliang Zhang, Qingming Huang, Shuqiang Jiang, Wen Gao, and Qi Tian. 2010. Affective visualization and retrieval for music video. *IEEE Transactions on Multimedia* 12, 6 (2010), 510–522.

[60] M. Zisook, J. Hernandez, M.S. Goodwin, and W. Picard R. 2013. Enabling Visual Exploration of Long-term Physiological Data. In *IEEE Conference on Visual Analytics Science and Technology*.

[61] Annuska Zolyomi, Andrew Begel, Jennifer Frances Waldern, John Tang, Michael Barnett, Edward Cutrell, Daniel McDuff, Sean Andrist, and Meredith Ringel Morris. 2019. Managing Stress: The Needs of Autistic Adults in Video Calling. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–29.