

# It is Okay to be Distracted: How Real-time Transcriptions Facilitate Online Meeting with Distraction

Seoyun Son School of Computing, KAIST Daejeon, Republic of Korea seoyun.son@kaist.ac.kr Junyoug Choi School of Computing, KAIST Daejeon, Republic of Korea joonchoi518@kaist.ac.kr Sunjae Lee School of Computing, KAIST Daejeon, Republic of Korea sunjae1294@kaist.ac.kr

Jean Y. Song Information and Communication Engineering, DGIST Daegu, Republic of Korea jeansong@dgist.ac.kr Insik Shin School of Computing, KAIST Daejeon, Republic of Korea ishin@kaist.ac.kr

Germany. ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/ 3544548.3580742

### ABSTRACT

Online meetings are indispensable in collaborative remote work environments, but they are vulnerable to distractions due to their distributed and location-agnostic nature. While distraction often leads to a decrease in online meeting quality due to loss of engagement and context, natural multitasking has positive tradeoff effects, such as increased productivity within a given time unit. In this study, we investigate the impact of real-time transcriptions (i.e., full-transcripts, summaries, and keywords) as a solution to help facilitate online meetings during distracting moments while still preserving multitasking behaviors. Through two rounds of controlled user studies, we qualitatively and quantitatively show that people can better catch up with the meeting flow and feel less interfered with when using real-time transcriptions. The benefits of real-time transcriptions were more pronounced after distracting activities. Furthermore, we reveal additional impacts of real-time transcriptions (e.g., supporting recalling contents) and suggest design implications for future online meeting platforms where these could be adaptively provided to users with different purposes.

#### **CCS CONCEPTS**

• Human-centered computing  $\rightarrow$  Interactive systems and tools; Collaborative and social computing systems and tools.

### **KEYWORDS**

Video-conferencing, multitasking, distraction, online meeting, realtime transcriptions

#### **ACM Reference Format:**

Seoyun Son, Junyoug Choi, Sunjae Lee, Jean Y. Song, and Insik Shin. 2023. It is Okay to be Distracted: How Real-time Transcriptions Facilitate Online Meeting with Distraction. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg,* 

CHI '23, April 23-28, 2023, Hamburg, Germany

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM. ACM ISBN 978-1-4503-9421-5/23/04...\$15.00 https://doi.org/10.1145/3544548.3580742

#### **1** INTRODUCTION

Online meetings are becoming essential as a substitute for faceto-face meetings and in-person communications. In particular, the demand for online meetings has increased dramatically since the COVID-19 pandemic. For instance, the percentage of Americans working from home increased from 15% to 50.2% immediately after the pandemic [13], and on Zoom, one of the leading online meeting platforms, customers with ten or more employees increased by about 620% from 2019 to 2021 [18, 19]. As many organizations actively adopt telecommuting to benefit both individuals and the organization [17, 57, 58], the importance of online meetings will likely grow, even after the pandemic [6, 40, 47, 60].

Nevertheless, online meetings are well known for their own problems, which include decreased concentration [31], fatigue [11, 26], and a lack of non-verbal interaction [35, 54, 56]. Moreover, they are vulnerable to technical issues [39, 42, 67]. These problems induce distractions both directly [39] and indirectly [15] during an online meeting, making it difficult for participants to focus on the meeting context. That is, people are more prone to being distracted during remote online meetings, and this can negatively affect engagement, attention, and the overall quality of the online meeting [15].

However, one key observation that many people overlook is that not all distractions are harmful to the meeting experience or must be avoided. It is natural for us to be distracted during a long meeting, and often, such distractions can even benefit the individual's productivity during the meeting. For instance, when working from home, one may have to take care of children suddenly when they cry or require supervision. Such incidents are inevitable and are not subject to be resolved or avoided. Moreover, multitasking, one of the most common factors of distraction, has advantages such as increased productivity and creativity by allowing people to complete more tasks in a given time [2, 15], keeping arousal levels high [1], or creating a synergistic effect with different related tasks [10].

Many previous works have attempted to enhance the online meeting experience by eliminating the cause of the distraction itself [41, 65], encouraging participants to avoid distractions [3, 20], and reducing the negative impact of one individual's distractions

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

#### CHI '23, April 23-28, 2023, Hamburg, Germany



Figure 1: OPARTS features: (A) live video of each participant; (B) meeting topic and task announcement of the user study experiment; (C) control buttons for the audio stream, video stream, and multitasking user study task; (D) display of popular keywords; (E) toggle button to switch between the summary mode and full-transcript mode; (F) display of keywords of each utterance; (G) real-time full-transcript of the speech; and (H) real-time summary of the speech.

on other participants [4, 49]. However, few studies have investigated how to overcome the shortcomings of distraction while still embracing the benefits and inevitability of distractions.

The high-level insight that underpins our study is as follows: what is equally important with minimizing distractions is *to enable individuals quickly to catch up on the meeting context after they are distracted*. To this end, this paper explores how online meeting tools can help participants retain their productivity, even in distracting situations. Specifically, we leverage three different levels of realtime transcriptions (i.e., full-transcript, summary, and keywords) to help online meeting participants quickly to catch up on the meeting context after being distracted. In earlier works, transcriptions were widely used to help participants recall or catch up on missing parts *after* the meeting is over [37, 69, 77]. To the best of our knowledge, this study is the first to explore the effectiveness of *real-time* transcriptions in assisting people to be engaged in meetings, even during distracting situations.

To understand the usefulness of real-time transcriptions during online meetings, we design and implement OPARTS (**O**nlinemeeting **P**latform with **A**I-based **R**eal-time **T**ranscription**S**), a webbased online meeting platform that provides three different levels of real-time transcriptions: *i*) full-transcripts, *ii*) summaries of each utterance, and *iii*) keywords of each utterance and popular keywords. Figure 1 shows the features implemented in OPARTS. In addition, OPARTS is designed to log various user behaviors, including which application the user is currently viewing and how actively each feature of the system is being used — in order to analyze both the conscious and unconscious effects of real-time transcriptions.

We conducted two independent controlled experiments with OPARTS, where the first experiment (n=43) was to observe *how* 

*participants experienced* OPARTS during the meeting, especially after a distraction, and the second experiment (n=28) was to *quantitatively measure* whether participants indeed understood the meeting content better when using OPARTS. Through the two experiments, we confirmed both qualitatively and quantitatively that real-time transcriptions help online meeting participants quickly catch up and actively engage in the meetings. Also, through an in-depth log analysis of user behaviors, we confirmed that participants more actively use real-time transcriptions immediately after being distracted. Furthermore, we report various key findings and observations (Table 1), which can be utilized to improve the online meeting experiences, including participants' preferences toward different levels of real-time transcriptions, the characteristics of the different levels of real-time transcriptions, and various effects of real-time transcriptions in online meetings.

Below we summarize the three main contributions of this paper:

- We designed and implemented an online meeting tool, OPARTS, which provides different levels of real-time transcriptions– full-transcripts, summaries, and keywords– to mitigate the adverse effects of a distraction without restricting users from experiencing such distractions.
- We designed and conducted two rounds of controlled studies (n=43 and 28) to evaluate not only the *subjective opinions* of the participants about the helpfulness of real-time transcriptions but also the *actual usefulness* of real-time transcriptions through a quantitative comparison with a baseline condition.
- We report the key findings from our experiments, which are summarized in Table 1, and provide guidance for enhancing the online meeting experience with real-time transcriptions.

Son, et al.

Benefits of using real-time transcriptions in the online meeting where distractions also exist (Section 5.3)				
<b>Retaining Context</b>	Participants can catch up on the missed gist more rapidly.			
Better Understanding	Participants can understand the meeting context well.			
Users' tendencies toward different levels of real-time transcriptions (Section 5.6)				
Observation 1	Users prefer full-transcripts and summaries over keywords. Users find keywords not very useful in online meetings.			
Observation 2	There is no significant superiority between a full-transcript and a summary. The numbers of participants who considered full-transcripts as more useful, and vise versa, were well balanced.			
Observation 3	Participants have no clear preference between full-transcripts and summaries. Rather, most of them use the default transcription mode (randomly assigned) (76.7%).			
Perceived characteristics of the different levels of real-time transcriptions (Section 5.6)				
Full-transcript	Accurate and detailed. Allows users to confirm a specific part of a past conversation and check the exact utterance in real time.			
Summary	Concise and clear. Easy to get the main point quickly after a distraction.			
Keyword	Suitably represents the flow of the meeting. Easy to understand the flow of the ongoing topic after concentration is lost.			
Additional benefits of using real-time transcriptions in an online meeting (Section 6.1)				
<b>Recalling Contents</b>	Participants can check on forgotten contents or missed key points in real time.			
Multi-modal Listening	Participants can read the transcription in real time when audio is unstable or different intonations make speech difficult to understand.			

Table 1: Summary of the findings of providing real-time transcriptions in online meetings.

#### 2 RELATED WORK

#### 2.1 Distractions During Online Meetings

Distraction Factors in Online Meetings. It is common for peo-2.1.1 ple to experience distractions during meetings. Especially in an online meeting environment, because participants are isolated in their own spaces, and considering that their surroundings may not be suitable for work, it is even more challenging not to be distracted. In fact, several studies have found that working from home complicates workers' efforts to maintain a concentration-friendly working environment, especially when sharing their household with pets, children, or other family members [39, 42, 73, 74]. In particular, multitasking is one of the most common factors inducing distractions during meetings. Multitasking activities during online meetings can be divided into meeting-related and non-meeting-related activities. Meeting-related activities include sharing meeting materials or taking notes [2, 64], while non-meeting-related activities include light tasks such as emailing, communicating with colleagues, eating, and cleaning [15, 39]. Unintentional interruptions by other people or pop-ups of mobile applications can also lead to non-meeting-related multitasking.

2.1.2 *Pros and Cons of Multitasking.* Multitasking refers to carrying out a subtask simultaneously to one's main task, such as watching a video while doing an assignment or texting friends while talking with colleagues. Generally, multitasking behavior is known

to affect content memorization [7, 34], psychological cost [5], and task performance negatively [51]. However, it has also been found that multitasking can increase creativity [38] and improve decisionmaking performance on a simple task [66]. A recent study by Cao et al. [15] showed that while some people experience negative effects of multitasking, such as missing important information and feeling fatigue during an online meeting, others experience increased productivity when using meeting-related multitasking and relieve fatigue by focusing less on less important content. Other works also showed that multitasking during meetings can increase task performance [64, 72] and help with efficient time management [2]. Meanwhile, some studies have reported that the benefits of multitasking are based on tradeoffs. Zijlstra et al. [79] and Mark et al. [48] found that people can increase their productivity by multitasking, but doing so introduces side effects such as stress, higher frustration, time pressure, and the requirement for extra effort. Therefore, to minimize the adverse effects of multitasking while preserving its benefits, we focus on helping participants quickly to catch up with a meeting without intervening in their multitasking behavior.

# 2.2 Tools for Solving Problems Caused by Distraction

Many existing studies have proposed online meeting tools that can mitigate the shortcomings of distractions during online meetings. One approach is to encourage participants to improve their attention and engage more during meetings so that participants can spontaneously avoid distractions. CoCo [63] encourages participants to speak more during meetings by providing post-meeting feedback on participants' speaking time and adopting a turn-taking methodology. MeetCues [3] encourages participants to engage more actively during a meeting by providing real-time and post-meeting feedback. However, this approach does not take into consideration the benefits and inevitability of distractions.

Another approach is to reduce the negative emotions that participants experience as a result of another participant's distraction. Marlow et al. [49] found that the multitasking behaviors of a single participant can induce unpleasant emotions in other participants, and possibly lead to less engagement of other participants as well. To solve this problem, Avrahami et al. [4] proposed a switched camera view that switches between multiple cameras whenever a participant turns her head to multitask in a dual-monitor environment to hide one's multitasking behaviors from others. However, although such approach can preserve the benefits of multitasking, it does not directly address the problems that individual participants face as a result of their own distractions. Moreover, because both studies evaluated their results based on participants observing the recorded video of an online meeting between two participants, their results did not fully reflect the real-world online meeting experience.

As such, studies to solve the problems associated with distractions during online meetings have yet to be conducted. In particular, we lack studies on how to solve the problems that individual participants face due to their own distractions while preserving the benefits of multitasking. Therefore, in accordance with previous studies that reported that helping distracted people get back into the context is critical to increasing the effectiveness and experience of online meetings [2, 42], our study aims to explore how real-time transcriptions affect participants in an online meeting environment where various distractions occur.

#### 2.3 Automated Support Tools for Use in Online Meeting

2.3.1 Adopting the Advantages of Face-to-face Meetings. Several studies have attempted to bring the strength of face-to-face meetings into online meetings. Unscheduled ad hoc conversations before and after a meeting can encourage creative and productive ideas but these are not likely to occur if participants are not physically together. Mingler [65] addressed this problem with a video conferencing system that allows private ad hoc conversations. Also, various studies have attempted to extract non-verbal signals during an online meetings to make meetings more effective and inclusive. Most notably, they analyzed videos containing participants' faces to extract non-verbal signals such as those related to consensus, tone [62], or emotion [25].

2.3.2 Providing Additional Information. Researchers have also tried to support meeting participants by providing additional information after the meeting or in real time. Some studies modified videos recorded during video conferencing events by adding text annotations [53] or links to previous meeting videos [27, 28]. Recorded videos help participants recall or catch up on the contents, even

if they did not participate in the meeting [8, 75]. Some systems generated key phrases [23] or annotated the contributions of [24] participants' speeches to enrich the meeting experience.

Other studies focused on increasing productivity during online meetings by adding various features. Lee et al. [43] proved that providing the stage of the current discussion and allowing participants to use system-recommended opinions can reduce the load of the moderator in a meeting. Popescu-Belis et al. [59] implemented a system that finds backup materials for the ongoing meetings with real-time transcripts so the meeting contents can be enriched. Tur et al. [71] implemented a system that provides an analysis after a meeting, including summaries and decisions made during the meeting.

2.3.3 Improve the Accessibility. Various support tools for improving the limited accessibility of those who are deaf and hard of hearing (DHH) are also being developed. Kafle et al. [36] proved that highlighted subtitles could support DHH groups with better readability, enhanced understandability, and a lower task load. Another study demonstrated the need to convey emotions for DHH groups in a voice-only meeting by providing emojis that can express emotions along with subtitles [55]. There was also a study about reducing visual dispersion by introducing a word cloud in each participant's background and a separate caption window [33].

### 2.4 Utilizing Transcription as an Online Meeting Tool

Whittaker et al. [76] reported that reading a transcript is more efficient than watching/listening to recorded video/audio because participants can skim and find their desired information faster when using a transcript. Also, allowing participants to mark or highlight the transcript during the meeting not only helps participants better recall the meeting content but also offers a lower cognitive load compared to traditional note-taking methods [37]. Similarly, providing a summary during a meeting is proven to be effective when participants attempt to recall and catch up on the meeting. Zhang et al. [77] showed that providing a summary of conversations in chatbased meetings can effectively assist participants in their efforts to catch up on or recall the meeting while also making it easier to document the meeting content. Likewise, Tucker et al. [69] showed that providing an audio summary of a meeting helps latecomers to catch up on the ongoing meeting efficiently.

Based on the findings of previous studies, we leverage transcriptions as a tool to assist participants in their efforts quickly to catch up on a meeting after a distracting event. More specifically, we adopt three levels of transcriptions (i.e., full-transcripts, summaries, keywords) to investigate comprehensively how people use each different transcription type. Our study differs from previous studies in an important way. First, while previous works either provide transcriptions *after* the meeting [62, 71], conduct an evaluation with recorded feeds of past meetings [4, 69], or suggest the use of real-time transcriptions as a future work [15], we provide transcriptions in *real time* and conduct an experiment in an actual meeting environment.

In addition, the effectiveness of a real-time summary and keywords in an online meeting environment is under-explored. While there exist online meeting tools that provide real-time full-transcripts [29, Server every five seconds. If the Silence Detection Server captures 50], to the best of our knowledge, there are no studies or commercial products that provide summaries or keywords in real time during an online meeting. Server every five seconds. If the Silence Detection Server captures a period silence longer than 4.5 seconds from the recorded audio file, it automatically responds with a speech end recognition signal to the OPARTS server. For efficiency, the Silence Detector scans

# 3 OPARTS: ONLINE-MEETING PLATFORM WITH AI-BASED REAL-TIME TRANSCRIPTIONS

To investigate the effectiveness of real-time transcriptions provided during online meetings and observe how people utilize different levels of transcriptions, we present OPARTS, a system that provides different levels of granularity of transcriptions and logs various user behaviors.

#### 3.1 Real-time Transcriptions

Figure 1 shows the proposed interface of OPARTS with the fulltranscript, summary, and keyword features. Because we attempted to observe patterns when participants used the three types of transcriptions, not intending to compare each transcription in independent environments, we designed the OPARTS UI to allow participants to use all types of transcriptions freely. Participants can switch between the full-transcript and summary modes using the mode toggle button (Figure 1E). Note that for fairness, we balanced which mode is displayed as the default mode. In full-transcript mode, the real-time full-transcript of each utterance is displayed in the message box (Figure 1G), and in summary mode, a real-time summary of each utterance is shown (Figure 1H). Keywords are provided regardless of the mode selected. At most five keywords from each utterance are displayed under the corresponding fulltranscript or summary (Figure 1F). When certain keywords appear frequently, we show them as popular keywords at the top of the OPARTS UI (Figure 1D).

Figure 2 shows the system architecture of OPARTS. The OPARTS server runs a speech-to-text (STT) module, a Summarizer module, and a Keyword module, which in turn support the OPARTS features mentioned above. Below we describe the technical challenges we resolved when implementing OPARTS.

3.1.1 STT Module. We used Microsoft Azure's STT API to generate participants' full-transcripts. The STT API gives a "speech end recognition" response when it detects the end of a speech from the user's audio stream. Based on this signal, we designed the initial system to recognize whether a speech is over and send a summary generation request. However, from multiple internal pilot studies, we found that when the user did not turn off their microphone themselves, it took about 18 - 19 seconds from the time that the user ended the speech until the system received the speech end recognition signal. Therefore, it took more than 20 seconds for the participant to check the summary after someone ended their speech, including three seconds of summary generation latency. To overcome this latency, we additionally implemented what is referred to as a Silence Detection Server to improve the performance of the STT API by internally determining the speech end detection time. When a participant speaks, OPARTS's server sends an audio stream to the STT API and saves an audio file continuously. Simultaneously, it sends a silence detection request to the Silence Detection

Server every five seconds. If the Silence Detection Server captures a period silence longer than 4.5 seconds from the recorded audio file, it automatically responds with a speech end recognition signal to the OPARTS server. For efficiency, the Silence Detector scans only the last ten seconds of the audio file because it is the most likely section to include the end of a speech. After receiving the speech end recognition signal, the server proceeds with a summary request. As a result, the speech end detection latency was reduced to 7.68 seconds on average. The additional latency from the silence detection server was trivial because this latency was only about 0.41 seconds.

3.1.2 Summarizer Module. We used the BERT [21] and BART [44] models to generate a real-time summary of each speech. As mentioned earlier, the summary is analyzed when the full-transcript generation step is completed. After OPARTS's server receives the summary, it shows participants the summary and sends a keyword generation request to the keyword module. Prior research attempted to increase the quality of the summary in meeting environment by implementing a model that handled ill-formed sentences [78] and utilized videos of the users to construct the summary [45, 52]. Unfortunately, this process was difficult to apply in our experiment given that we summarized each speech in real time. Instead, we combined the BERT and BART models with a short process time of less than three seconds. OPARTS compares the summaries from BERT and BART and selects the most proper summary by using a scoring strategy we established. The strategy is based on the Rouge Score [46] and Universal Sentence Encoder concepts [16]. Rouge Score evaluates the summary from 0 to 1 by comparing text sequences in the summary and the original fulltranscript. Universal Sentence Encoder computes the embeddings that represent sentences, making it possible to compute the semantic similarity between two sentences. In this way, we determined whether the summary has a similar meaning to the original fulltranscript, assigning a score from 0 (not similar at all) to 1 (identical). Our scoring strategy combined these scores by calculating their mean value and chose the summaries with higher scores between the BERT and BART results.

3.1.3 Keyword Module. We used the RAKE [61] algorithm to generate the keywords of each utterance and to select popular keywords. The keyword module generates the keywords from each utterance and applies the results to popular keywords. We compared various keyword extraction algorithms (e.g., YAKE [14], keyBERT [30]), focusing on the processing time and accuracy. We selected the RAKE algorithm because it is at least 100 times faster than other algorithms and extracts keywords with high accuracy. The keyword module selects at most five keywords among the keywords from the full-transcript and the summary of each speech. With these keywords, it computes popular keywords that most commonly appear in the speeches, putting more weight on more recent keywords.

#### 3.2 Loggers

We implemented various loggers to gain a better understanding of the results of the user responses and to analyze the usefulness of the real-time transcriptions provided during the online meetings quantitatively.



Figure 2: OPARTS system architecture: the OPARTS server generates full-transcripts, summaries, and keywords in real time using the STT, Summarizer, and Keyword modules, respectively. The STT module produces full-transcripts based on users' speeches and periods of silence. With the full-transcripts, the Summarizer module generates summaries. Finally, the Keyword module produces keywords from the full-transcripts and summaries. This process takes less than ten seconds per utterance.

3.2.1 User Focus Logger. In order to observe whether participants focus on the online meetings or are distracted by some other tasks, the User Focus Logger records which applications are receiving user interactions (e.g., keyboard input, scroll inputs, click inputs). Participants are considered 'on-focus' while OPARTS is on top of the screen and 'out-of-focus' otherwise. Short out-of-focus durations (shorter than ten seconds) are counted as on-focus, taking into account that they would not interfere with the following of the meeting context. For example, if a participant is on-focus for 20 seconds and then becomes out-of-focus for the next five seconds, and then turns back to on-focus for the next 15 seconds, the entire time of 40 seconds is considered as on-focus.

3.2.2 User Activity Logger. In order to track how participants use OPARTS, the User Activity Logger records their interaction with OPARTS, including window scrolls, mode toggles, audio inputs, and button clicks. For instance, participants can be considered as actively using the transcriptions of OPARTS when they scroll the transcription window to read the contents of past conversations. To characterize such user interaction behaviors on the key features of OPARTS, we divide the entire meeting period into ten-second intervals. Then, any ten-second interval is considered as 'active' if any of the above activities occurred in the interval, and the corresponding '*degree of activeness*' is measured in terms of the number of activities occurring during the interval.

#### 4 STUDY DESIGN

To understand how real-time transcriptions help people to keep up with meetings even during distractions, we formulated three research hypotheses:

H1: Real-time transcriptions can make it easier for online meeting participants to keep up with a meeting when an *explicit-distraction* (i.e., unwanted distraction) occurs.H2: Real-time transcriptions can make it easier for online meeting participants to keep up with a meeting when an *implicit-distraction* (i.e., spontaneous distraction) occurs.

H3: Participants' reliance on real-time transcriptions will increase after distracting events.

To verify the hypotheses and investigate the usability of real-time transcriptions, we conducted two independent controlled experiments with OPARTS. In both experiments, four conditions were compared through a 2 × 2 experimental setting. The four conditions are: OE (OPARTS Explicit-distraction), OI (OPARTS Implicit-distraction), BE (Baseline Explicit-distraction), and BI (Baseline Implicit-distraction), where the first letter represents whether OPARTS was used or not (OPARTS vs. Baseline) and the second letter represents the type of the distractions (Explicit-distraction vs. Implicit-distraction). Explicit distractions include spontaneous multitasking behaviors plus additional unwanted multitasking tasks such as answering simple routine questions or typing non-meeting related text, while implicit distractions include only spontaneous multitasking as any activity other than participating in the meeting.

The first study was conducted in a real-world online meeting environment, where six to eight participants in a group debated a specific topic through an online meeting. After the meeting, the participants were asked how they utilized OPARTS during the meeting and how useful OPARTS was under distracting situations. In the second study, individual participants participated in a simulated meeting environment as a silent participant (i.e., only listened). After the meeting, they answered questions related to the meeting contents, where the score quantitatively shows whether the participants actually understood the meeting. Below, we explain the detailed design of our two experiments.

# 4.1 Study 1: Online Group Meeting Using OPARTS and Participants' Responses

4.1.1 Participants. We recruited 53 participants from 18 different countries through the student community sites of two universities (ten participants were excluded from the analysis for clarity of the result. See Section 5.1.1). The participants consisted of people without hearing disabilities. Each participant received 30 USD for

participating in the experiment. Before the recruiting process, participants were told that they would participate in a group meeting conducted in English with five to seven other participants and use a video conferencing system developed by the experimental team. All of them participated in the meeting in their personal space (e.g., home) online and participated in the experiment using a personal laptop or desktop with a camera, audio, and microphone working.

4.1.2 *Conditions.* Table 2 shows the four conditions used in the experiments: **B**aseline Implicit-distraction (BI), **B**aseline Explicit-distraction (BE), **O**PARTS Implicit-distraction (OI), and **O**PARTS Explicit-distraction (OE).

**Implicit-distraction condition vs. Explicit-distraction condition** As mentioned earlier, we aim to assist people in catching up with what they missed in a meeting while retaining their natural multitasking abilities. Hence, we investigated the effects of transcription by differentiating the intensity of the distraction rather than limiting it.

Because we did not control participants' surrounding environment and given that they joined the meeting remotely from an individual place such as a house or office, we justifiably infer that all participants are in an implicit-distraction situation where they are vulnerable to spontaneous multitasking and distractions by external stimuli factors. From the pre-questionnaire conducted at the time of recruiting, 94% of participants reported that they multi-tasked in previous online meetings. Regarding the types of multitasking, 58% of the participants reported that they had experience with meetingrelated multitasking (i.e., note-taking, searching for information), and 74% reported that they have engaged in non-meeting-related multitasking (i.e., text messaging, assignments, web surfing, email, using social network services). Accordingly, we consider that the implicit-distraction condition includes spontaneous multitasking (i.e., meeting-related multitasking, non-meeting-related multitasking) and other unwanted distractions such as their phone ringing or disturbances in the surrounding environment.

In the explicit-distraction condition, we intentionally increased the distraction intensity by explicitly assigning additional multitasking tasks to each participant. The multitasking task was to answer a questionnaire with nine questions, including four short-answer questions and five essay questions. We assigned an email-like lightweight questionnaire as the multitasking task, referring to the fact that email is a common multitasking behavior [15].

**Baseline condition vs. OPARTS condition** Each participant is assigned to either an implicit-distraction or an explicit-distraction condition and participates in two meetings: one in the OPARTS condition and the other in the baseline condition. While the overall functions of the OPARTS system were provided in the OPARTS condition, the transcriptions were not provided (i.e., only video and audio streams were provided) in the baseline condition. Therefore, each participant experienced a meeting with transcriptions in the OPARTS condition and a generic meeting without transcriptions in the baseline condition.

4.1.3 Tasks and Procedures. We divide the participants into eight groups with six to eight participants each, where four groups experience an explicit-distraction condition, and the other four are assigned an implicit-distraction condition. Each group experiences two types of meetings (OPARTS condition and baseline condition)

and two types of discussion topics (1) *Is a college education worth it?* and (2) *Should we regulate youth games?*. The orders of the conditions and the meeting topics were counterbalanced such that every group had a different meeting scenario. For example, group A is assigned the BI condition and topic 2 in the first meeting and the OI condition and topic 1 in the second meeting. In contrast, group B is assigned the OI condition and topic 2 in the first meeting and the BI condition and topic 1 in the second meeting.

Participants in explicit-distraction conditions were guided to complete additional tasks autonomously during the meeting. The questions of the assigned multitasking task were structured so as not to overlap in the two meetings (e.g, "What did you eat yesterday for dinner?", "What did you eat yesterday for lunch?"). We sent a text message in the middle of the meeting (when 15, 10, and 5 minutes were left) to remind them of their additional tasks; otherwise, we did not intervene in the meeting. For other implicit distraction behaviors (e.g., web surfing, texting, note taking), we did not ask the participants to refrain from performing spontaneous distractions. Rather, we guided participants to participate in the meeting as they would in their usual meeting environment.

A 30-minute tutorial was provided in advance so that the participants could familiarize themselves with the functions of OPARTS, the multitasking tasks, and the meeting topics before the experiment. The experiment was conducted in the following order: a 30-minute ice-breaking session, a 30-minute first meeting, a 30minute post-survey including a break time, a second 30-minute meeting, and a 30-minute post-survey and final survey. We minimized the effects of long experiments by adding a break time and counterbalancing each condition. Each post-survey asks about the meeting conducted immediately before, and the final survey asks about the overall meeting experience.

4.1.4 *Measures and Analysis.* We measured the effectiveness of the transcriptions through post-survey questions (i.e., 5-point Likert scale, multi-select, and narrative questions) and user behavior logging.

In the post-survey questions, to measure how much the realtime transcriptions affected participants during the meeting, the participants rated the following:

- (1) How much did the distraction interfere with you?,
- (2) How difficult was it to catch up on the meeting after the distraction?,
- (3) How well did you understand the content of the meeting?, and
- (4) Which methods did you use to catch up on the meeting after the distraction?.

To measure which factor participants felt was more helpful among the full-transcript, summary, and keywords provided during the meeting, participants in the OPARTS conditions rated the following:

- To what degree was each transcription useful during the meeting?,
- (2) How much did each transcription help you to catch up after the distraction?,
- (3) Which methods did you use to catch up on the meeting after the distraction?, and

	Implicit-distraction Explicit-distraction	
Baseline	BI	BE
OPARTS	OI	OE



(4) Between the two inter-switchable transcription modes (i.e., fulltranscript mode or summary mode), which mode did you use more often?.

In addition, we collected log data from OPARTS's User Focus Logger and User Activity Logger to analyze the user behaviors from a more fine-grained perspective. For a statistical analysis, we conducted a pairwise comparison using a one-tailed Wilcoxon signed-rank test.

# 4.2 Study 2: Simulated Online Meeting to Quantitatively Measure Participants' Understanding

4.2.1 Participants. For the second study, we recruited 38 participants from 12 different countries through the student community sites in five universities (ten participants were excluded from the analysis to ensure a rigorous analysis of the results. See Section 5.1.2). Each participant received 11.5 USD for participating in the experiment. The participants also were given information about the meeting format (i.e., participating in a meeting conducted in English as a silent participant) and the meeting tool (the video conferencing system developed by the experimental team) before joining the experiment. They all participated in the meeting in their personal space (e.g., home) online using a personal laptop or desktop with audio working.

4.2.2 Conditions. We used the same  $2 \times 2$  experimental setting used in Study 1: BI, BE, OI, and OE (see Section 4.1.2). However, note-taking was prohibited in all conditions to more accurately measure the effectiveness of only the real-time transcriptions. In addition, we kept the duration of the explicit distraction (i.e., multitasking task) equal (1.5 min) among the participants to improve the accuracy of the between-subject comparison. We also instructed participants to conduct the same multitasking task (i.e., typing phrases from the novel *Harry Potter and the Philosopher's Stone*). We double-checked whether the given multitasking task was performed correctly by asking participants' for summaries of what they typed and collecting the sentences that they typed.

4.2.3 Tasks and Procedures. We randomly assigned ten participants to each OE and BE condition and nine to the OI and BI condition. Participants listened to a prerecorded discussion on the topic *Is a college education worth it?* for 20 minutes and solved the true/false questions. They were guided to participate in a *simulated* online meeting as a *silent participant* in the group discussion. For the prerecorded discussion, we selected one discussion from Study 1 and used the first 20 minutes of it. A 15-minute tutorial was provided in advance so that the participants could familiarize themselves with the functions of OPARTS and the multitasking tasks. In explicit distraction conditions (i.e., OE and BE), we assigned a total of four

multitasking tasks, one every five minutes. To prevent participants from predicting when the tasks would start or from sloppily performing the multitasking tasks, we informed them that the tasks would start randomly during the meeting and that they could finish the experiment earlier if they worked diligently on the multitasking tasks.

4.2.4 Measures and Analysis. We quantitatively measured how well each participant understood the meeting by asking them 12 true/false questions after they completed the online meeting simulation. The questions were the same throughout all conditions. The difficulty of the questions was adjusted to be neither too challenging (i.e., including overly detailed information) nor too easy (i.e., can be easily inferred from the overall meeting context). In addition, we added an *I'm not sure* option to prevent false positive answers.

The questions were from three different intervals of the explicit distraction meeting conditions–a fully engaged interval (2 minutes), a distraction interval (1.5 minutes), and an RfD interval (1.5 minutes)–, with four questions from each interval. The RfD (recovery from distraction) interval is the initial part of the on-focus interval when the users start to re-focus on the meeting and try to recover from the distraction. The fully engaged interval is the remainder of the on-focus interval, during which the users are expected to focus entirely on the meeting, and the distraction interval is when participants are not focusing on the meeting (i.e., out-offocus). The participants performed given explicit multitasking tasks during the distraction interval.

To support the quantitative results, we asked participants in the OPARTS conditions to answer the following questions:

- (1) How much did the transcriptions help you catch up after multitasking? and
- (2) How much of the meeting did you miss while reading the transcription?.

We also analyzed log data from OPARTS's User Focus Logger to confirm the correctness of the test conditions. We conducted a onetailed Mann-Whitney U test on the scores of true/false questions for the statistical analysis.

# 5 RESULTS & EVALUATION

We analyzed the results of our two user studies and the logs collected from OPARTS to shed light on 1) whether real-time transcriptions are effective in helping participants recover from a distraction and 2) the usefulness of different types of real-time transcriptions. In addition, from the user comments and statistical results, we elicit a few insights and discussion points that could potentially serve as guidance on how real-time transcriptions can be adopted in online meeting tools. Our results suggest that real-time transcriptions are effective in both explicit and implicit distraction conditions, confirming all three hypotheses; meanwhile, the additional distraction of reading the transcriptions does not interfere with meeting participation. Additionally, users have varying preferences with regard to the full-transcripts and summaries, while keywords are less helpful than the other two types.

#### 5.1 Filtering

5.1.1 Study 1. Prior to the analysis, we compared each participant's user study response with the OPARTS logs to filter the data of participants who were clearly misguided. When asked "Which mode did you use most frequently?", seven out of the 53 participants selected a mode that actually turned out to have never been used. In addition, three participants from the explicit-distraction group did not perform the assigned multitasking tasks. As a result, ten participants in total were excluded from the analysis, resulting in 20 participants for the explicit-distraction condition (i.e., BE and OE) and 23 participants for the implicit-distraction condition (i.e., BI and OI).

*5.1.2 Study 2.* We filtered participants in the explicit-distraction condition to analyze only participants who were sufficiently distracted. We included participants whose total number of lines typed during the multitasking exceeded the median value in each OPARTS and baseline condition. The median value of the typed number of lines was 12.5 in the OE condition and 11.5 in the BE condition. As a result, ten participants were excluded from the analysis.

# 5.2 Verifying the Suitability of the Test Conditions

Before discussing the results of our evaluation, we present a statistical analysis (one-tailed Mann-Whitney U test) of whether our test conditions functioned as intended. Through a log analysis, we confirmed that the participants in the explicit-distraction conditions had longer out-of-focus durations and a greater number of out-of-focus events compared to those in the implicit-distraction conditions. In Study 1, the participants in the explicit-distraction conditions (i.e., OE and BE) were out-of-focus for 8.81 (SD=3.97) minutes, while the participants in the implicit-distraction conditions (i.e., OI and BI) were out-of-focus for 1.91 (SD=2.69) minutes on average. We confirmed that there was a statistically significant difference between the results from the two conditions (U=1700.0and p<.001). Likewise, the participants in the explicit-distraction conditions experienced out-of-focus 8.25 (SD=5.40) times, while the participants in the implicit-distraction conditions were out-offocus 4.35 (SD=5.95) times. The statistical analysis suggests that there was a significant difference between the results (U=1378.5and p<.001). We confirmed that Study 2 had the same patterns in the out-of-focus duration and the number of out-of-focus events between the different distraction conditions as Study 1. Note that the out-of-focus duration and number of events were conservatively measured given that we did not count multitasking behaviors that did not involve the device used for OPARTS, such as the use of a smartphone or daydreaming during the meeting.

# 5.3 H1: Real-time transcriptions help participants to keep up with the meeting when an explicit-distraction occurs.

5.3.1 Study 1. We analyzed the data collected from the explicitdistraction group (i.e., OE and BE) to explore the effectiveness of real-time transcriptions under the meeting condition where distractions are more common and inevitable. From the statistical analysis result of the post-survey questions, we found that the participants were significantly less disturbed (W=59.0 and p<.01) and felt that it was easier to catch up on the meeting (W=82.0 and p<.005) when using OPARTS. The average scores on the question "How much did the distraction interfere with you?" (lower is better) were 3.20 (SD=1.11) for the OE condition and 3.85 (SD=0.875) for the BE condition (Figure 3a). Similarly, the average scores for the question "How difficult was it to catch up on the meeting after the distraction?" (lower is better) were 2.35 (SD=0.875) for the OE condition and 3.10 (SD=1.25) for the BE condition (Figure 3b). There was no significant difference in how much the participants thought that they understood the meeting contents (W=54.5 and p>.05; Figure 3c). We reason this to be a result of the low complexity of the meeting topics, as all participants responded that they understood well, and participants with OPARTS actually understood better than those without, which will be covered in section 5.3.2. The average scores on the question "How well did you understand the meeting?" (higher is better) was 4.10 (SD=0.852) for the OE condition and 3.95 (SD=1.10) for the BE condition.

We also found that participants actively utilized the real-time transcriptions. Table 3 shows how the participants responded to the multi-select question "Which methods did you use to catch up on the meeting after the distraction?". When real-time transcriptions were not given (i.e., BE), participants mainly used infer from the conversation to make up for the missing parts of the conversation. All 20 participants reported that they used this method, whereas only two participants responded that they used other methods as well (i.e., asking other participants for the missing parts). This indicates that inferring from the ongoing conversation is the traditional and the most dominant method for users to catch up on the meeting when they were distracted. On the other hand, when real-time transcriptions were given (i.e., OE), only 13 participants responded that they used infer from the conversation, and all participants responded that they used OPARTS, of which seven participants only used OPARTS. This shows that the real-time transcriptions are effective enough, to the extent that users rely more on OPARTS than the traditional method to which they are already accustomed.

In line with the findings from the user responses, the logs collected from OPARTS also suggest that the majority of the participants actively utilized real-time transcriptions. From the user activity analysis, we found that out of 20 participants who reported that they used OPARTS, 17 injected active scroll events into OPARTS, scrolling up and down the transcriptions to make up the missing parts of the meeting.

*5.3.2* **Study 2**. We compared the true/false scores from the explicitdistraction groups (i.e., OE and BE) to quantitatively verify H1. For the overall score (0~1), participants using OPARTS scored significantly higher than those who did not (U=22 and p<.05). The CHI '23, April 23-28, 2023, Hamburg, Germany



with you?" (lower is better)



(b) "How difficult was it to catch up on the meeting after the distraction" (lower is better)



(c) "How well did you understand the meeting?" (higher is better)

Figure 3: The number of participants who assigned a score to each item on the survey in the OE/BE conditions (N=20). According to the survey result, in the explicit-distraction conditions, (a) participants felt significantly less interfered with due to the distraction and (b) found it easier to catch up after the distraction in the OPARTS condition compared to the baseline. Additionally, (c) participants thought that they understood meeting contents better when the OPARTS condition is given, but there was no significant difference.

Method Type	Method	# of answers in	
		OE condition	BE condition
Traditional Methods	Infer from the conversation	13	20
	Ask other participants	1	2
OPARTS Features	Summary	14	
	Full-transcript	11	
	Keyword	7	IN/A

Table 3: Response to a multi-select question "Which methods did you use to catch up on the meeting after the distraction?" in the OE/BE conditions (N=20). Participants in the BE condition mostly inferred from the conversation, while participants who were given OPARTS mainly used OPARTS's real-time transcriptions.

average overall scores were 0.62 (SD=0.0456) for the OE group and 0.45 (SD=0.151) for the BE group (Figure 4a). Similarly, for the distraction interval only, participants from the OE group scored higher than those from the BE group. The average score of the OE group was 0.55 (SD=0.209), while the average score of the BE group was 0.40 (SD=0.285) (Figure 4b). Although we found no statistically significant difference between the two scores (U=16.5 and p>.05), we conjecture that the difference will be more significant if the complexity of the conversation increases, as the post-survey question "How much did the transcriptions help you catch up after the multitasking?" was scored 4.4 (SD=0.548) out of 5.

We also tested for a ripple effect of providing transcriptions by comparing the scores of questions from the RfD interval (i.e., estimated time interval during which participants are expected to use reading the transcriptions after the distraction). Contrary to our expectation that the OE group's score would be slightly lower due to the time spent reading the transcriptions, participants in the OE condition showed a significantly higher average score compared to those in the BE condition (U=22.0 and p<.05). The average scores were 0.8 (SD=0.209) for the OE condition and 0.45 (SD=0.209) for the BE condition (Figure 4c). In line with the quantitative result, the average score on the post-survey question was 2.00 (SD=1.00) out of 5, confirming that the distraction caused by reading the transcription was not large enough to hamper the meeting experience. Even if a negative ripple effect arises due to the transcriptions, real-time transcriptions can help participants recover from it.

# 5.4 H2: Real-time transcriptions help participants to keep up with the meeting when an implicit-distraction occurs.

*5.4.1* **Study 1**. To analyze the effectiveness of real-time transcriptions when no explicit distractions are given (i.e., only implicit distractions), we analyzed the data collected from the OI and BI conditions. In these conditions, although we did not deliberately inject any distractions or instruct participants to engage in multitasking, the participants spontaneously lost focus during the meeting (see Figure 7b). OPARTS's user focus logger reported that 15 out of 23 participants experienced a distraction at least once.

Son, et al.

CHI '23, April 23-28, 2023, Hamburg, Germany



Figure 4: The average score of the true/false questions from the OE/BE conditions (N=5). In the explicit-distraction conditions, participants who used OPARTS scored significantly higher than those who did not. They also scored higher with questions from the distraction interval, but there was no significant difference. Additionally, participants in the OE condition scored significantly higher with the questions from the RfD interval.

The statistical analysis results of the post-survey questions conveyed that the participants in the implicit-distraction group (i.e., OI and BI) also found OPARTS useful for catching up on the meeting after a distraction (W=73.0 and p<.005). The average scores on the question "How difficult was it to catch up on the meeting after the distraction?" (lower is better) were 1.61 (SD=0.941) for the OI condition and 2.74 (SD=1.48) for the BI condition (Figure 5b). However, we found no significant difference (W=57.0 and p>.05) between the two conditions (i.e., OI and BI) on the question "How much did the distraction interfere with you?" (Figure 5a), as all distractions were spontaneous. The average scores were 2.48 (SD=1.34) for the OI condition and 2.70 (SD=1.30). For the question "How well did you understand the meeting?", we found that the participants thought that they understood meeting significantly better when using OPARTS (*W*=57.0 and *p*<.05). The average scores were 4.61 (SD=0.499) for the OI condition and 4.09 (SD=0.848) for the BI condition.

Similar to the findings from **H1**, participants under implicit distractions also utilized real-time transcriptions actively. Table 4 shows how participants responded to "Which methods did you use to catch up on the meeting after the distraction?" The results convey a trend similar to that of **H1**. When real-time transcriptions were not given (i.e., BI), participants mainly used *infer from the conversation* to catch up on the meeting, whereas when real-time transcriptions were given (i.e., OI), participants mainly used OPARTS. One interesting observation is that the percentage of participants who relied heavily on OPARTS (i.e., participants who only used OPARTS) (56%) was greater than that of the participants under an explicit distraction (35%). This indicates that participants found real-time transcriptions effective enough even in the implicit-distraction condition, where only spontaneous distractions existed.

The user activity logger of OPARTS also suggested that the majority of the participants utilized OPARTS's transcription features. According to the user activity analysis, 19 out of 23 participants scrolled the transcriptions at least once during the meeting. In addition, the average amount of scroll events during the meeting differed only by 9% between the implicit group and the explicit group, indicating that the participants in the implicit-distraction condition utilized real-time transcription as much as those in the explicit-distraction condition.

5.4.2 Study 2. We compared the true/false scores from the implicitdistraction group (i.e., OI and BI) to quantitatively validate H2. Because the implicit-distraction condition does not include fixed distractions or an RfD interval, we only compared the overall scores. The average overall scores were 0.58 (SD=0.118) for the OI condition and 0.54 (SD=0.167) for the BI condition (Figure 6). However, there was no significant difference in the scores between the two conditions (U=48.5 and p>.05). Our conjecture with regard to this result is as follows: because the meetings were short and the topics were easy, participants had no trouble understanding the overall meeting contents. We believe that real-time transcriptions will contribute more to people's understanding of meetings if the topics require more effort to follow, as difficult meetings tend to cause more distractions [15] and require more of a cognitive load to catch up on the missing parts. We discuss this matter in more detail in the Limitations and Future Work section (see Section 6.3).

# 5.5 H3: Participants' reliance on real-time transcriptions increases after distracting events.

Lastly, to verify whether participants utilized real-time transcriptions more actively after the distraction, we conducted an in-depth analysis of the logs collected from OPARTS. First, we plotted each participant's degree of activeness with respect to the time axis to observe at which interval users more frequently use real-time transcriptions. Then, to grasp on the high-level view of this phenomenon, we divided the users' on-focus intervals into two sections: 1) recovery from a distraction interval (RfD) <sup>1</sup> and 2) a fully engaged interval <sup>2</sup>. Figure 7 shows the resulting graph of two of our participants. According to our statistical analysis, the RfD interval showed a significantly denser degree of activeness compared to the fully engaged interval (W=369.0 and p<.005). The average degrees were 0.55 (SD=0.0589) for the RfD interval and 0.35 (SD=0.0365) for the fully engaged interval. This indicates that participants used

<sup>&</sup>lt;sup>1</sup>A time interval where users start to re-focus on the meeting and try to catch up on the meeting after a distraction–each interval has a maximum length of 90 seconds. <sup>2</sup>The remainder of the on-focus interval, during which users are expected to be fully focused on the meeting.

CHI '23, April 23-28, 2023, Hamburg, Germany



with you?" (lower is better)



(b) "How difficult was it to catch up on the meeting after the distraction?" (lower is better)



20

(c) "How well did you understand the meeting?" (higher is better)

Figure 5: The number of participants who reported each score on the survey in the OI/BI conditions (N = 23). According to the survey result, (b) participants who used OPARTS felt it to be less difficult to catch up after a distraction compared to those who did not. (a) There was no significant difference in the degree to which the distractions distracted. (c) Participants responded that they thought that they understood the meeting better when using OPARTS, but we found this result unreliable.

Method Type	Method	# of answers in	
		OI condition	BI condition
Traditional Methods	Infer from the conversation	9	20
	Ask other participants	2	1
	Catch up not needed	0	2
OPARTS Features	Full-transcript	14	
	Summary	12	- N/A
	Keyword	7	

Table 4: Responses to the multi-select question "Which methods did you use to catch up on the meeting after the distraction?" in the OI/BI conditions (N=23). Participants in the BI condition mostly inferred from the conversation, while participants who were given OPARTS mainly used OPARTS's real-time transcriptions.



Figure 6: The average score of the overall true/false questions from the OI/BI conditions (N=9). In the implicit-distraction conditions, participants who used OPARTS scored higher than those who did not, though without statistical significance.

real-time transcriptions more actively after being distracted and that our hypothesis thus holds true.

# 5.6 Usefulness of Different Types of Real-time Transcriptions

On top of the effectiveness of real-time transcriptions as a whole, we also analyzed the usefulness of each individual type of realtime transcription (i.e., full-transcripts, summaries, and keywords). First, we present statistical results derived from user responses and logs collected from OPARTSin Study 1 to verify which type of real-time transcriptions was more useful and more actively used during the experiment. Then, we identify the characteristics of each transcription type based on the user comments collected from the user study.

**Statistical Results.** One observation from the user responses is that the full-transcript and the summary are more useful than the keyword type. As shown in the table 3 and 4, the number of participants who reported that they used the full-transcript (avg N=12.5) and summary (avg N=13) was significantly higher than those who reported that they used the keyword (avg N=7). In addition, on the questions *"How useful was each function during the meeting?" (higher is better)* and *"How much did each function help you to catch up after* 

Son, et al.





Figure 7: Example user activity analysis for a recovery from a distraction. The degree of activeness shows the window scroll count in the transcription area in units of ten seconds. There are areas indicating periods when the participant is out-of-focus and areas referred to as Recovery from Distraction (RfD) intervals, which is immediately after the participant became on-focus. The participant actively used the transcription features by scrolling the transcription area, especially in the RfD interval.

the meeting?" (higher is better), the full-transcript and the summary scored significantly higher than the keyword type (W=487.0 and p<.05, W=358.0 and p<.05 for the full-transcript and the keyword types, respectively; W=444.0 and p<.005, W=361.0 and p<.001 for the summary and the keyword types, respectively; see Figure 8). The average scores for the questions "How useful was each function during the meeting?" (higher is better) were 3.63 (SD=1.27) for the full-transcript, 3.67 (SD=1.19) for the summary, and 2.93 (SD=1.32) for the keyword types. The average scores for "How much did each function help you to catch up after the meeting?" (higher is better) were 3.47 (SD=1.39) for the full-transcript, 3.74 (SD=1.24) for the summary, and 2.84 (SD=1.48) for the keyword types.

Meanwhile, one interesting observation is that although we expected the summary to be more useful than the full-transcript given that the summary provides denser information and it is easier to read, we found no significant difference between the scores assigned these two types in both questions (W=328.0 and p>.1, W=338.5

and *p*>.1). Similarly, in the post-survey question, where we asked "Between two inter-switchable transcription modes (i.e., full-transcript mode or the summary mode), which mode did you use more often?", 41.86% of the participants reported that they mainly used the full-transcript mode, while 58.14% of the participants reported that they used the summary mode mainly. There was no clear difference as to which mode users thought was more useful.

Another interesting observation is, contrary to our expectation, that users would switch between modes from time to time as needed (e.g., summary after a long distraction and full-transcript after a short distraction); most participants tended to use only one of the modes regardless of the duration of the distraction or the meeting condition. Figure 9 shows a distribution graph of the participants with respect to the ratio at which they used the two modes, demonstrating that most participants resided at either end of the distribution. This reveals that the majority of the participants stuck to a single mode throughout the meeting, whereas only a few participants used both modes equally. Moreover, Figure 10 shows that 25.6% of the participants never switched their mode at all, and even when they did, most participants (51.15%) returned to the default mode and used it mainly throughout the meeting. This shows that most of the participants did not have a clear preference between the two modes but rather tended mainly to use the mode given as the default.

Characteristics of different types of real-time transcriptions. Although participants showed no clear preferences towards different types of transcriptions, the user comments collected from the survey suggest that each type has different usage purposes and characteristics.

**Full-transcript.** Participants found the full-transcript particularly useful when they did not want to miss any details and when ensuring that they were correctly following the context: "Because there was a case where I couldn't understand the specific part of the other person's opinion. So, instead of reading the summary of his opinion, I needed to confirm some specific part. Therefore, full-transcript mode was more useful in my experience [BG13]," "I like to get the detailed view even when I don't read the whole of it [BG32]," "Because I do not want to miss any detail [AG23]." Participants also commented on how the real-time full-transcript is useful after a short distraction, as it generates a transcription word by word in real time: "I could follow up even if I got distracted for a moment because the full-transcript immediately generated what others were saying [BG28]."

Summary. On the other hand, participants preferred the summary because it provides more understandable text, makes it easy to catch the gist, and is shorter: "The full-transcript can be too long so summary mode is easier to scroll and glance through just to get the general gist [AG36]," "Summarizing keeps stuff simple. Since I also had to listen to the discussion, having a short summary was easier than to read the entrire[sic] thing [BG34]." In addition, it is easier to identify the gist from the summary owing to its length and the characteristic that it omits unnecessary phonetic expressions such as 'uh' or 'um': "Some speakers (especially me) had a lot of utters (uh, I, I think, um, so, and). Some speakers beat around the bush. Summaries provide a shorter and more comprehendible[sic] text [AG23]." However, we also noticed that a few participants felt awkward with regard to the quality of the summary: "Sometimes summary has a little inaccuracy so I was referring to the full-transcript for clarification [AG38]." Therefore, there is a possibility that the preference tendency in Section 5.6 will change if the summary quality is improved.

**Keyword.** Despite the analysis results showing that keywords are less helpful than the full-transcript and the summary, several participants felt keywords were useful for reading and understanding the topic quickly: "It was useful as it helps understand the whole points made by the speaker by just looking at few words [BG17]," "It helped to understand the trend of the discussion. Especially after I came back from doing subtask [AG33]."

On the other hand, participants also commented that the keywords did not provide sufficient information due to the short length of the list: "It quite a deal to understand a speaker's points by keyqords[sic] only. It's better to use the real-time summary [AG48]," "Not as useful because I am aware of the keywords while listening, even if a little distracted [AG36]."

#### 5.7 Confidence of Participants in Their Memory

One interesting observation from Study 2 is that participants in the OPARTS condition were more confident with their true/false answers. Participants in OPARTS conditions (i.e., OE and OI) selected *I'm not sure* significantly less than participants in the baseline conditions (i.e., BE and BI) (U=309.5 and p<.005). The average numbers of *I'm not sure* were 1.05 (SD=0.911) for OPARTS conditions and 2.14 (SD=1.28) for the baseline conditions. This indicates that participants who used OPARTS felt more focused during the meeting and thus had more confidence in their memory.

#### 6 DISCUSSION

#### 6.1 Additional Benefits of Using Real-time Transcriptions in Online Meeting

Sections 6.1.1 and 6.1.2 describe the additional benefits of real-time transcriptions as commonly noticed by a number of participants(12 out of 43 in Section 6.1.1 and 6.1.2, respectively), and Section 6.1.3 summarizes potential benefits worthy of further study as noticed by a few participants.

6.1.1 Support Recalling Contents. Twelve out of 43 (27.9%) of the participants from Study 1 commented that the benefit of the realtime transcription is not limited to distracted situations. Functions in OPARTS can be helpful in situations where speech is difficult to understand. According to the participants' comments, OPARTS helped them catch up with the flow when too many people spoke simultaneously: "There were some times when multiple people were speaking, it was a bit burden to catch up all the ideas. However, thanks to full-transcript and summary section, I could reread while people were arguing and obtain more defensive arguments [AG18]." Other participants also commented that skimming the transcription helped them catch the point when the speech was long: "However, there was one instance where one of the speakers had gone on a particularly lengthy speech and I had forgotten what their first point was. To rectify this, I simply glanced at the full-transcript panel and saw a few 'buzz words' that reminded me [AG44]."

6.1.2 Support Multi-modal Listening. Twelve out of 43 (27.9%) participants of Study 1 also commented that real-time transcriptions are useful for supporting hearing problems. In online meetings, it is sometimes challenging to have a seamless conversation due to noise or technical issues [39, 42, 67]. Many participants noted that realtime transcriptions were useful when meetings were interrupted due to temporal network issues or the surrounding environment because real-time transcriptions can assist or replace the audio stream: "It captures almost everything. So we can go back to read at anytime, especially when you cannot hear the speaker well or when you are distracted [AG23]," "When the voice was not clear, it was a great help in understanding the content [BG37]." In addition, there were comments that it was helpful to solve communication problems that could occur because of different accents from the various nationalities of the participants: "There is a pronunciation that I cannot understand, but when the person who used that pronunciation talked, I used the full-transcript usefully [AG26]." From the comments above, we can confirm that the quality of transcriptions is sufficient to capture what people said, even when they had various accents.



(a) "To what degree was each (b) "How much did each function help you function useful during the meet-to catch up after the meeting?"(higher is ing?"(higher is better) better)

Figure 8: The average score of each transcription types on the survey questions (N=43). Participants felt that the full-transcript and the summary were more effective than the keyword, while there was no significant difference found between the scores assigned to the full-transcript and the summary.



Figure 9: Distribution of the participants with respect to the ratio at which participants used two modes. The graph shows that participants tended to use only one of the two modes.



Figure 10: Mode usage tendency for participants. Participants tend to use default mode more frequently, regardless of which mode was given as the default mode.

*6.1.3 Potential benefits.* Through a post-survey, many participants shared detailed comments describing their experiences and expectations with OPARTS. Based on their comments, we summarize the potential benefits of real-time transcriptions, which are worthy of further study.

(a) Reduce Mental Fatigue. We find that transcriptions can alleviate mental fatigue, one of the well-known problems associated with online meetings. Some participants commented that the burden of attending meetings was reduced: *"I felt much more comfortable since the script is written in real-time* [BG36]." We expect that reducing mental fatigue can make people better engage in meetings and benefit from a virtuous cycle that reduces multitasking caused by mental fatigue.

(b) Allow People to Participate Less in Unnecessary Parts of a Meeting. We anticipate that real-time transcriptions will allow people to follow the ongoing topic of a meeting while remaining less engaged in the less important parts of the meeting. Participants noted that the provided transcriptions reduced the sense of obligation to listen to all the contents of the meeting: "I was just waiting for the summary to come instead of listening for what they were saying [AG38]," "It also gave me the impression that I could skip listening and find what I need in the text [BG32]." This also aligns with the future work suggested in an earlier study: capturing topics from real-time transcriptions would help people to skip unnecessary parts of a meeting [15].

(c) Enhance Meeting Performance. Providing a real-time transcription is expected to help with performance degradation problems by, for instance, decreasing the uniqueness of ideas [2] by reducing the cognitive load required to participate in the meeting after a distraction. In fact, there were comments that the real-time summary helped to improve the discussion quality: *"Having fast real-time summary helped me to better catch the important points and come up with more defensive and good counterarguments [AG18]."* 

(d) Support Documenting. We also find comments that align with previous studies about note-taking assistants. Kalnikaite et al. [37] showed that providing real-time automatic speech recognition and a simple annotation tool can help people to reduce their cognitive load when taking notes. In addition, one participant noted that the functions in OPARTS helped with note taking: "... and the full-transcript was even more useful because I can take note of specific points and arguments [AG35]." Since meeting-related multitasking can cause a decrease in performance regarding the conversation pace and interaction [2], this suggests the positive aspect that transcription can solve the performance degradation experienced by people who engage in meeting-related multitasking.

(e) Assist Deaf or Hard-of-Hearing Participants. From the results, we find the possibility that the real-time transcription functions provided by OPARTS can help improve the accessibility for DHH individuals. Previous works showed a diverse preference of captioning from the DHH community. To satisfy the preferences

#### CHI '23, April 23-28, 2023, Hamburg, Germany

of DHH users, they suggested providing options to select the style of the captions desired by each person [12] or providing different types of captions according to the degree of hearing [55]. Therefore, OPARTS, which provides various transcription levels, is expected to be an excellent option to help DHH users participate in real-time online meetings.

(f) Apply in Remote Online Lectures. Some participants mentioned that the functions of OPARTS would be helpful in a remote lecture because the full-transcript provides details in real time: "*T think using this in real-time lecture is really good for students, because sometimes we fell asleep during the lecture or gets easily distracted to use phone. So, these different features are really important to get back the flow of the lecture* [*AG11*]." According to previous studies, much multitasking occurs in the classroom environment. Students who multitask during lectures are less focused [7] and remember less content [32]. Also, in non-face-to-face lectures, internet connection problems often cause one to miss the content [9, 22]. Transcriptions can help students catch up on what they missed and become more engaged. However, unlike online meetings, lectures are usually one-way affairs; thus, different effects of transcription depending on the type of video conferencing should be considered.

## 6.2 Design Guideline for Real-time Transcriptions in Online Meeting Tools.

6.2.1 Adequate Amount of Information to Provide in Real-time Online Meetings. While most participants found OPARTS useful enough, one participant reported that the real-time transcriptions in OPARTS induced distractions to her: *"Too much text distracts my attention from what people are talking about [AG12]."* Studies have shown that providing too much information during a realtime meeting is cognitively burdensome and distracting [68, 70]. To maximize the benefits of real-time transcription, we should also consider the amount of distraction caused by the given functions. Although real-time transcriptions in OPARTS did not negatively affect our study, to maintain a positive impact from transcription and reduce distractions in online meeting platforms where various functionalities also exist, future applications should find an appropriate amount of information to display during meetings or should give users a choice to hide the functionalities as desired.

6.2.2 Adequate Design for Displaying Transcriptions. Some participants suggested changing the design of the transcriptions as an improvement. The readability of transcriptions can be increased by displaying them in a more adequate format (e.g., bullet points, paragraph division): "It would be better if the full-transcripts were divided into paragraphs [BG22]," "It would be better if the summaries with shorter lengths were provided in bullet points [BG36]." Some others pointed out the positions of the transcriptions: "The full-transcript mode could be changed to subtitles in the video pane[sic] if possible [BG32]." Indeed, although OPARTS places transcriptions next to the video so that users can freely scroll up and down, other forms of online meeting tools can reposition the transcriptions in the form of subtitles so that users can watch videos and read transcriptions simultaneously.

*6.2.3* Suitable Transcription Type for Real-time Online Meetings. We find that each transcription type has varying characteristics

and that their suitable usages are very distinct (see 5.6). Thus, to maximize the impact of transcriptions in reducing the adverse effects of a distraction, the types of transcriptions provided should be flexible. We expect that the type of effective transcription will differ depending on the meeting characteristics. For example, because the online-meetings used in our study were short and given that the topics did not change often, participants found the keywords relatively unnecessary. However, in longer meetings where the topics change frequently, we expect keywords to be very useful to those seeking quickly to grasp the topic and following the context. Also, when a meeting includes many details, such as numbers, the preference for the full-transcript may increase, and in meetings that include many long utterances where grasping the key sentences is important, the preference for a summary may increase. Likewise, we believe that the usefulness of different levels of transcriptions will differ depending on the characteristics of the meeting as well as the characteristics of the members, such as the frequency of such meetings (i.e., one-time, weekly, monthly), intimacy between participants (i.e., coworkers, first meet), and the level of understanding of each participant's meeting agenda (i.e., beginner, expert).

#### 6.3 Limitations and Future Work

From the user study and its results, we have noticed that our study has several limitations that future works can explore further.

First, we chose discussion topics for which it is easy to share opinions and are universal enough to create an environment where first-timers can easily engage in meetings. From Study 1, the score on the question "how well did you understand the meeting?" showed that most participants could easily understand the contents of the meeting, as intended. From Study 2, participants in the implicitdistraction condition showed high true/false scores regardless of the presence of OPARTS. We reason that this stemmed from the low complexity of the meeting topic. Because the topic was designed to be universal and easily debatable between non-native English speakers, it did not require much concentration to catch up with the meeting contents. As a result, the need for OPARTS diminished when it comes to helping participants understand the meeting and when no explicit distractions were given: "The content is not very complicated. SO[sic] it is easy to infer. I don't think that would be the case if we were discussing a more technical or complicated topic [BG24]." We will further explore the effect of OPARTS on the understanding of meeting contents using more complex and non-universal topics.

Second, our User Focus Logger insufficiently captured the distraction behaviors, missing some. Because the logger reports that the users are 'on-focus' as long as they have OPARTS on the top of the screen, we were not able to capture some spontaneous distractions, such as the use of smartphones, daydreaming, and others. We attempted to resolve this limitation by implementing an eye gazer inside OPARTS. However, many participants did not have a sufficient environment to run the eye gazer on their own devices (e.g., no camera, a low-spec computer, among others), and there was a geographical limitation to providing a separate eye gazer device because the participants were distributed in their own convenient locations. In future work, we will adopt various measures, such as a

lightweight eye gazer, a brain wave detector, and video recognition to calculate the 'on-focus' interval more strictly.

Lastly, our work has the potential to reveal other interesting observations when targeting different meeting characteristics or when used for long-term observations. As mentioned in Section 6.2.3, the usefulness of each transcription is expected to differ due to varying meeting characteristics. In addition, during long-term observations, factors such as the importance, length, and difficulty of the meeting topic are also expected to affect the OPARTS usage patterns of the meeting participants.

#### 7 CONCLUSION

In this study, we explored the various impacts of real-time transcriptions in assisting people to catch up with a meeting flow and helping them feel less interfered with, even after distracting activities. To the best of our knowledge, our study is the first to investigate the role of different real-time transcriptions (i.e., fulltranscript, summary, and keyword) during online meetings with controlled experiments and the first to prove their effectiveness. Our two studies suggest that real-time transcriptions can very effectivly help users recover from distractions (i.e., unwanted distractions and spontaneous distractions) during a meeting. Also, our analysis of various user behavior logs shows that users have no clear preference between the full-transcript and the summary while finding keywords less helpful than the other two. From our discussions of other impacts of real-time transcriptions and their varying characteristics, we anticipate that future online meeting tools will provide users with the best possible online meeting experience with guidance on implementing real-time transcriptions.

#### ACKNOWLEDGMENTS

We are grateful to Hayeon Lee for her contributions to the early design of this study and to Dae R. Jeong for feedback on this paper. We also thank all the participants for lending their time and sharing their experiences and the reviewers for their valuable insights to improve the paper. This work was supported in part by ERC (NRF-2018R1A5A1059921), NRF-2020R1A2C2005479, and the DGIST Start-up Fund Program (2021070007) funded by the Ministry of Science and ICT of the Korea Government.

#### REFERENCES

- Rachel F Adler and Raquel Benbunan-Fich. 2012. Juggling on a high wire: Multitasking effects on performance. *International Journal of Human-Computer Studies* 70, 2 (2012), 156–168.
- [2] Alberta A. Ansah, Yilun Xing, Amudha Varshini Kamaraj, Diana Tosca, Linda Boyle, Shamsi Iqbal, Andrew L. Kun, John D. Lee, Michel Pahud, and Orit Shaer. 2022. "I Need to Respond to This" – Contributions to Group Creativity in Remote Meetings with Distractions. In 2022 Symposium on Human-Computer Interaction for Work (Durham, NH, USA) (CHIWORK 2022). Association for Computing Machinery, New York, NY, USA, Article 7, 12 pages. https://doi.org/10.1145/ 3533406.3533411
- [3] Bon Adriel Aseniero, Marios Constantinides, Sagar Joglekar, Ke Zhou, and Daniele Quercia. 2020. MeetCues: Supporting online meetings experience. In 2020 IEEE Visualization Conference (VIS). IEEE, IEEE, Salt Lake City, UT, USA, 236–240.
- [4] Daniel Avrahami, Eveline van Everdingen, and Jennifer Marlow. 2016. Supporting Multitasking in Video Conferencing Using Gaze Tracking and On-Screen Activity Detection. In Proceedings of the 21st International Conference on Intelligent User Interfaces (Sonoma, California, USA) (IUI '16). Association for Computing Machinery, New York, NY, USA, 130–134. https://doi.org/10.1145/2856767.2856801
- [5] Brian P Bailey and Joseph A Konstan. 2006. On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Computers in human behavior* 22, 4 (2006), 685–708.

- [6] Mary Baker. 2020. Gartner Survey Reveals 82% of Company Leaders Plan to Allow Employees to Work Remotely Some of the Time. Gartner. Retrieved August 3, 2022 from https://www.gartner.com/en/newsroom/press-releases/2020-07-14-gartner-survey-reveals-82-percent-of-company-leaders-plan-to-allowemployees-to-work-remotely-some-of-the-time
- [7] Louise Barkhuus. 2005. "Bring Your Own Laptop Unless You Want to Follow the Lecture": Alternative Communication in the Classroom. In Proceedings of the 2005 ACM International Conference on Supporting Group Work (Sanibel Island, Florida, USA) (GROUP '05). Association for Computing Machinery, New York, NY, USA, 140–143. https://doi.org/10.1145/1099203.1099230
- [8] Jeremy Barksdale, Kori Inkpen, Mary Czerwinski, Aaron Hoff, Paul Johns, Asta Roseway, and Gina Venolia. 2012. Video Threads: Asynchronous Video Sharing for Temporally Distributed Teams. In Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (Seattle, Washington, USA) (CSCW '12). Association for Computing Machinery, New York, NY, USA, 1101–1104. https://doi.org/10.1145/2145204.2145367
- [9] Gabrielle Benabdallah, Sam Bourgault, Nadya Peek, and Jennifer Jacobs. 2021. Remote learners, home makers: How digital fabrication was taught online during a pandemic. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (Yokohama, Japan) (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 350, 14 pages. https://doi.org/10.1145/ 3411764.3445450
- [10] Raquel Benbunan-Fich. 2012. Measurement of multitasking with focus shift analysis. International Conference on Information Systems, ICIS 2012 3 (01 2012), 2230–2243.
- [11] Andrew A Bennett, Emily D Campion, Kathleen R Keeler, and Sheila K Keener. 2021. Videoconference fatigue? Exploring changes in fatigue after videoconference meetings during COVID-19. *Journal of Applied Psychology* 106, 3 (2021), 330.
- [12] Larwan Berke, Khaled Albusays, Matthew Seita, and Matt Huenerfauth. 2019. Preferred Appearance of Captions Generated by Automatic Speech Recognition for Deaf and Hard-of-Hearing Viewers. In Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (Glasgow, Scotland Uk) (CHI EA '19). Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3312921
- [13] Erik Brynjolfsson, John J Horton, Adam Ozimek, Daniel Rock, Garima Sharma, and Hong-Yi TuYe. 2020. COVID-19 and remote work: An early look at US data. Technical Report. National Bureau of Economic Research.
- [14] Ricardo Campos, Vítor Mangaravite, Arian Pasquali, Alípio Jorge, Célia Nunes, and Adam Jatowt. 2020. YAKE! Keyword extraction from single documents using multiple local features. *Information Sciences* 509 (2020), 257–289.
- [15] Hancheng Cao, Chia-Jung Lee, Shamsi Iqbal, Mary Czerwinski, Priscilla N Y Wong, Sean Rintel, Brent Hecht, Jaime Teevan, and Longqi Yang. 2021. Large Scale Analysis of Multitasking Behavior During Remote Meetings. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (Yokohama, Japan) (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 448, 13 pages. https://doi.org/10.1145/3411764.3445243
- [16] Daniel Cer, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St. John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, Yun-Hsuan Sung, Brian Strope, and Ray Kurzweil. 2018. Universal Sentence Encoder. https://doi.org/10.48550/ARXIV.1803.11175
- [17] Mikaela Cohen. 2021. Brex CEO on taking his company virtual-first and the right ways to think about remote work. CNBC. Retrieved August 8, 2022 from https://www.cnbc.com/2021/07/15/brex-ceo-on-the-right-ways-to-thinkabout-remote-work.html
- [18] Zoom Video Communications. 2020. Zoom Video Communications Reports Fourth Quarter and Fiscal Year 2020 Financial Results. zoom. Retrieved September 12, 2022 from https://investors.zoom.us/news-releases/news-release-details/zoomvideo-communications-reports-fourth-quarter-and-fiscal-year
- [19] Zoom Video Communications. 2022. Zoom Video Communications Reports Fourth Quarter and Fiscal Year 2022 Financial Results. zoom. Retrieved September 12, 2022 from https://investors.zoom.us/news-releases/news-release-details/zoomvideo-communications-reports-fourth-quarter-and-fiscal-1
- [20] Snigdha Das, Sandip Chakraborty, and Bivas Mitra. 2022. I Cannot See Students Focusing on My Presentation; Are They Following Me? Continuous Monitoring of Student Engagement through "Stungage". In Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization (Barcelona, Spain) (UMAP '22). Association for Computing Machinery, New York, NY, USA, 243–253. https://doi.org/10.1145/3503252.3531307
- [21] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. https://doi.org/10.48550/ARXIV.1810.04805
- [22] Shivangi Dhawan. 2020. Online learning: A panacea in the time of COVID-19 crisis. Journal of educational technology systems 49, 1 (2020), 5–22.
- [23] Joan Morris DiMicco. 2004. Designing Interfaces That Influence Group Processes. In CHI '04 Extended Abstracts on Human Factors in Computing Systems (Vienna, Austria) (CHI EA '04). Association for Computing Machinery, New York, NY, USA, 1041–1042. https://doi.org/10.1145/985921.985969

- [24] 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.
- [25] 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.
- [26] Liz Fosslien and Mollie West Duffy. 2020. How to combat zoom fatigue. Harvard Business Review 29 (2020), 1–6.
- [27] Andreas Girgensohn, Jennifer Marlow, Frank Shipman, and Lynn Wilcox. 2015. HyperMeeting: Supporting Asynchronous Meetings with Hypervideo. In Proceedings of the 23rd ACM International Conference on Multimedia (Brisbane, Australia) (MM '15). Association for Computing Machinery, New York, NY, USA, 611–620. https://doi.org/10.1145/2733373.2806258
- [28] Andreas Girgensohn, Jennifer Marlow, Frank Shipman, and Lynn Wilcox. 2016. Guiding Users through Asynchronous Meeting Content with Hypervideo Playback Plans. In Proceedings of the 27th ACM Conference on Hypertext and Social Media (Halifax, Nova Scotia, Canada) (HT '16). Association for Computing Machinery, New York, NY, USA, 49–59. https://doi.org/10.1145/2914586.2914597
- [29] Google. 2022. Use captions in a meeting. Retrieved August 10, 2022 from https://support.google.com/meet/answer/9300310?hl=en
- [30] Maarten Grootendorst. 2020. KeyBERT: Minimal keyword extraction with BERT. https://doi.org/10.5281/zenodo.4461265
- [31] Carl Gutwin and Saul Greenberg. 2002. A descriptive framework of workspace awareness for real-time groupware. Computer Supported Cooperative Work (CSCW) 11, 3 (2002), 411–446.
- [32] Helene Hembrooke and Geri Gay. 2003. The laptop and the lecture: The effects of multitasking in learning environments. *Journal of computing in higher education* 15, 1 (2003), 46–64.
- [33] Ryo Iijima, Akihisa Shitara, Sayan Sarcar, and Yoichi Ochiai. 2021. Word Cloud for Meeting: A Visualization System for DHH People in Online Meetings. In Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility (Virtual Event, USA) (ASSETS '21). Association for Computing Machinery, New York, NY, USA, Article 99, 4 pages. https://doi.org/10.1145/ 3441852.3476547
- [34] Shamsi T. Iqbal, Jonathan Grudin, and Eric Horvitz. 2011. Peripheral Computing during Presentations: Perspectives on Costs and Preferences. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Vancouver, BC, Canada) (CHI '11). Association for Computing Machinery, New York, NY, USA, 891–894. https://doi.org/10.1145/1978942.1979073
- [35] 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.
- [36] Sushant Kafle, Peter Yeung, and Matt Huenerfauth. 2019. Evaluating the Benefit of Highlighting Key Words in Captions for People Who Are Deaf or Hard of Hearing. In Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility (Pittsburgh, PA, USA) (ASSETS '19). Association for Computing Machinery, New York, NY, USA, 43–55. https://doi.org/10.1145/3308561.3353781
- [37] Vaiva Kalnikaitundefined, Patrick Ehlen, and Steve Whittaker. 2012. Markup as You Talk: Establishing Effective Memory Cues While Still Contributing to a Meeting. In Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (Seattle, Washington, USA) (CSCW '12). Association for Computing Machinery, New York, NY, USA, 349–358. https://doi.org/10.1145/2145204.2145260
- [38] Chaitali Kapadia and Shimul Melwani. 2021. More tasks, more ideas: The positive spillover effects of multitasking on subsequent creativity. *Journal of Applied Psychology* 106, 4 (2021), 542.
- [39] Katherine A Karl, Joy V Peluchette, and Navid Aghakhani. 2022. Virtual work meetings during the COVID-19 pandemic: The good, bad, and ugly. *Small Group Research* 53, 3 (2022), 343–365.
- [40] Jack Kelly. 2021. Salesforce Says '9-To-5 Workday Is Dead' And Employees Will Only Come Into The Office One To Three Days A Week. Forbes. Retrieved August 10, 2022 from https://www.forbes.com/sites/jackkelly/2021/02/10/salesforcesays-9-to-5-workday-is-dead-and-employees-will-only-come-into-the-officeone-to-three-days-a-week/?sh=4f8952f5d16c
- [41] G Ross Lawford. 2003. Beyond success: Achieving synergy in teamwork. The Journal for Quality and Participation 26, 3 (2003), 23.
- [42] Minha Lee, Wonyoung Park, Sunok Lee, and Sangsu Lee. 2022. Distracting Moments in Videoconferencing: A Look Back at the Pandemic Period. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (New Orleans, LA, USA) (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 141, 21 pages. https://doi.org/10.1145/3491102.3517545
- [43] Sung-Chul Lee, Jaeyoon Song, Eun-Young Ko, Seongho Park, Jihee Kim, and Juho Kim. 2020. SolutionChat: Real-Time Moderator Support for Chat-Based Structured Discussion. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376609
- [44] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2019. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. https://doi.org/10.48550/ARXIV.1910.13461

- [45] Manling Li, Lingyu Zhang, Heng Ji, and Richard J Radke. 2019. Keep meeting summaries on topic: Abstractive multi-modal meeting summarization. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Florence, Italy, 2190–2196. https://doi.org/10.18653/v1/P19-1210
- [46] Chin-Yew Lin. 2004. ROUGE: A Package for Automatic Evaluation of Summaries. In *Text Summarization Branches Out*. Association for Computational Linguistics, Barcelona, Spain, 74–81. https://aclanthology.org/W04-1013
- [47] Susan Lund, Anu Madgavkar, James Manyika, Sven Smit, Kweilin Ellingrud, Mary Meaney, and Olivia Robinson. 2021. The future of work after COVID-19. Retrieved Jan 25, 2023 from https://www.mckinsey.com/featured-insights/futureof-work/the-future-of-work-after-covid-19
- [48] Gloria Mark, Daniela Gudith, and Ulrich Klocke. 2008. The Cost of Interrupted Work: More Speed and Stress. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Florence, Italy) (CHI '08). Association for Computing Machinery, New York, NY, USA, 107–110. https: //doi.org/10.1145/1357054.1357072
- [49] Jennifer Marlow, Eveline van Everdingen, and Daniel Avrahami. 2016. Taking Notes or Playing Games? Understanding Multitasking in Video Communication. In Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (San Francisco, California, USA) (CSCW '16). Association for Computing Machinery, New York, NY, USA, 1726–1737. https://doi.org/10.1145/2818048.2819975
- [50] Microsoft. 2022. View live transcription in a Teams meeting. Microsoft. Retrieved August 10, 2022 from https://support.microsoft.com/en-us/office/view-livetranscription-in-a-teams-meeting-dc1a8f23-2e20-4684-885e-2152e06a4a8b
- [51] Christopher A. Monk, Deborah A. Boehm-Davis, and J. Gregory Trafton. 2002. The Attentional Costs of Interrupting Task Performance at Various Stages. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 46, 22 (2002), 1824–1828. https://doi.org/10.1177/154193120204602210 arXiv:https://doi.org/10.1177/154193120204602210
- [52] Gabriel Murray, Pei-Yun Hsueh, Simon Tucker, Jonathan Kilgour, Jean Carletta, Johanna D. Moore, and Steve Renals. 2007. Automatic Segmentation and Summarization of Meeting Speech. In Proceedings of Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT). Association for Computational Linguistics, Rochester, New York, USA, 9–10. https://aclanthology.org/N07-4005
- [53] Mukesh Nathan, Mercan Topkara, Jennifer Lai, Shimei Pan, Steven Wood, Jeff Boston, and Loren Terveen. 2012. In Case You Missed It: Benefits of Attendee-Shared Annotations for Non-Attendees of Remote Meetings. In Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (Seattle, Washington, USA) (CSCW '12). Association for Computing Machinery, New York, NY, USA, 339–348. https://doi.org/10.1145/2145204.2145259
- [54] Gary M Olson and Judith S Olson. 2000. Distance matters. Human-computer interaction 15, 2-3 (2000), 139–178.
- [55] Kotaro Oomori, Akihisa Shitara, Tatsuya Minagawa, Sayan Sarcar, and Yoichi Ochiai. 2020. A Preliminary Study on Understanding Voice-Only Online Meetings Using Emoji-Based Captioning for Deaf or Hard of Hearing Users. In Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility (Virtual Event, Greece) (ASSETS '20). Association for Computing Machinery, New York, NY, USA, Article 54, 4 pages. https://doi.org/10.1145/3373625.3418032
- [56] Kazuhiro Otsuka, Hiroshi Sawada, and Junji Yamato. 2007. Automatic Inference of Cross-Modal Nonverbal Interactions in Multiparty Conversations: "Who Responds to Whom, When, and How?" From Gaze, Head Gestures, and Utterances. In Proceedings of the 9th International Conference on Multimodal Interfaces (Nagoya, Aichi, Japan) (ICMI '07). Association for Computing Machinery, New York, NY, USA, 255–262. https://doi.org/10.1145/1322192.1322237
- [57] Kim Parker, Juliana Menasce Horowitz, and Rachel Minkin. 2020. How the Coronavirus Outbreak Has – and Hasn't – Changed the Way Americans Work. Pew Research Center. Retrieved Jan 25, 2023 from https://www.pewresearch.org/social-trends/2020/12/09/how-the-coronavirusoutbreak-has-and-hasnt-changed-the-way-americans-work/
- [58] Kim Parker, Juliana Menasce Horowitz, and Rachel Minkin. 2022. COVID-19 pandemic continues to reshape work in America. Pew Research Center. Retrieved Jan 25, 2023 from https://www.pewresearch.org/social-trends/2022/02/16/covid-19-pandemic-continues-to-reshape-work-in-america/
- [59] Andrei Popescu-Belis, Jonathan Kilgour, Alexandre Nanchen, and Peter Poller. 2010. The ACLD: Speech-Based Just-in-Time Retrieval of Meeting Transcripts, Documents and Websites. In *Proceedings of the 2010 International Workshop on Searching Spontaneous Conversational Speech* (Firenze, Italy) (SSCS '10). Association for Computing Machinery, New York, NY, USA, 45–48. https://doi.org/10. 1145/1878101.1878111
- [60] Gartner Research. 2020. COVID-19 Bulletin: Executive Pulse, 3 April 2020. Gartner. Retrieved August 3, 2022 from https://www.gartner.com/en/documents/3982949
- [61] Stuart Rose, Dave Engel, Nick Cramer, and Wendy Cowley. 2010. Automatic keyword extraction from individual documents. *Text mining: applications and theory* 1, 1-20 (2010), 10–1002.

- [63] 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.
- [64] Advait Sarkar, Sean Rintel, Damian Borowiec, Rachel Bergmann, Sharon Gillett, Danielle Bragg, Nancy Baym, and Abigail Sellen. 2021. The Promise and Peril of Parallel Chat in Video Meetings for Work. In Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (Yokohama, Japan) (CHI EA '21). Association for Computing Machinery, New York, NY, USA, Article 260, 8 pages. https://doi.org/10.1145/3411763.3451793
- [65] Jaeyoon Song, Christoph Riedl, and Thomas W. Malone. 2021. Online Mingling: Supporting Ad Hoc, Private Conversations at Virtual Conferences. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (Yokohama, Japan) (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 340, 10 pages. https://doi.org/10.1145/3411764.3445776
- [66] Cheri Speier, Joseph S Valacich, and Iris Vessey. 1999. The influence of task interruption on individual decision making: An information overload perspective. Decision sciences 30, 2 (1999), 337–360.
- [67] Amy Tan and Ahmet M. Kondoz. 2008. Barriers to Virtual Collaboration. In CHI '08 Extended Abstracts on Human Factors in Computing Systems (Florence, Italy) (CHI EA '08). Association for Computing Machinery, New York, NY, USA, 2045–2052. https://doi.org/10.1145/1358628.1358636
- [68] M. Iftekhar Tanveer, Emy Lin, and Mohammed (Ehsan) 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 (Atlanta, Georgia, USA) (IUI '15). Association for Computing Machinery, New York, NY, USA, 286–295. https://doi.org/10.1145/2678025.2701386
- [69] Simon Tucker, Ofer Bergman, Anand Ramamoorthy, and Steve Whittaker. 2010. Catchup: A Useful Application of Time-Travel in Meetings. In Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work (Savannah,

Georgia, USA) (CSCW '10). Association for Computing Machinery, New York, NY, USA, 99–102. https://doi.org/10.1145/1718918.1718937

- [70] Simon Tucker, Ofer Bergman, Anand Ramamoorthy, and Steve Whittaker. 2010. Catchup: A Useful Application of Time-Travel in Meetings. In Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work (Savannah, Georgia, USA) (CSCW '10). Association for Computing Machinery, New York, NY, USA, 99–102. https://doi.org/10.1145/1718918.1718937
- [71] Gokhan Tur, Andreas Stolcke, Lynn Voss, Stanley Peters, Dilek Hakkani-Tur, John Dowding, Benoit Favre, Raquel Fernández, Matthew Frampton, Mike Frandsen, et al. 2010. The CALO meeting assistant system. *IEEE Transactions on Audio, Speech, and Language Processing* 18, 6 (2010), 1601–1611.
- [72] Nathan W Twyman, Jeffrey G Proudfoot, Ann-Frances Cameron, Eric Case, Judee K Burgoon, and Douglas P Twitchell. 2020. Too busy to be manipulated: How multitasking with technology improves deception detection in collaborative teamwork. *Journal of Management Information Systems* 37, 2 (2020), 377–395.
- [73] Lena Waizenegger, Brad McKenna, Wenjie Cai, and Taino Bendz. 2020. An affordance perspective of team collaboration and enforced working from home during COVID-19. European Journal of Information Systems 29, 4 (2020), 429–442.
- [74] Bin Wang, Yukun Liu, Jing Qian, and Sharon K Parker. 2021. Achieving effective remote working during the COVID-19 pandemic: A work design perspective. *Applied psychology* 70, 1 (2021), 16–59.
- [75] Pierre Wellner, Mike Flynn, and Maël Guillemot. 2004. Browsing recorded meetings with Ferret. In International Workshop on Machine Learning for Multimodal Interaction. Springer, Springer-Verlag, Berlin, Heidelberg, 12–21.
- [76] Steve Whittaker, Simon Tucker, Kumutha Swampillai, and Rachel Laban. 2008. Design and evaluation of systems to support interaction capture and retrieval. *Personal and Ubiquitous Computing* 12, 3 (2008), 197–221.
- [77] Amy X Zhang and Justin Cranshaw. 2018. Making sense of group chat through collaborative tagging and summarization. Proceedings of the ACM on Human-Computer Interaction 2, CSCW (2018), 1–27.
- [78] Jiyuan Zheng, Zhou Zhao, Zehan Song, Min Yang, Jun Xiao, and Xiaohui Yan. 2020. Abstractive meeting summarization by hierarchical adaptive segmental network learning with multiple revising steps. *Neurocomputing* 378 (2020), 179– 188.
- [79] Fred RH Zijlstra, Robert A Roe, Anna B Leonora, and Irene Krediet. 1999. Temporal factors in mental work: Effects of interrupted activities. *Journal of Occupational* and Organizational Psychology 72, 2 (1999), 163–185.