Winder: Linking Speech and Visual Objects to Support Communication in Asynchronous Collaboration

Tae Soo Kim
School of Computing, KAIST
Daejeon, Republic of Korea
taesoo.kim@kaist.ac.kr

Seungsu Kim
School of Computing, KAIST
Daejeon, Republic of Korea
seungsu0407@kaist.ac.kr

Yoonseo Choi
School of Computing, KAIST
Daejeon, Republic of Korea
yoonseo.choi@kaist.ac.kr

Juho Kim
School of Computing, KAIST
Daejeon, Republic of Korea
juhokim@kaist.ac.kr

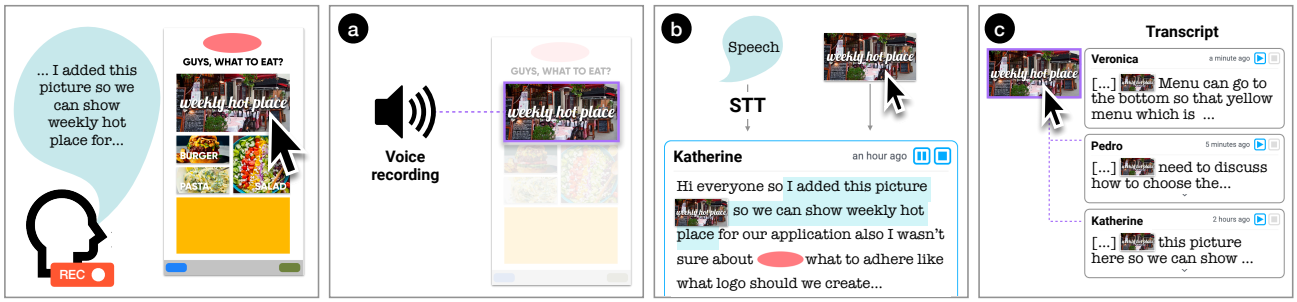

Figure 1: *Winder* allows for communication through linked tapes—multimodal recordings of voice and clicks on document objects. Winder supports three features for tape understanding and navigation: (a) highlighting objects on playback of voice recordings, (b) inline thumbnail images of objects on automatic transcripts, and (c) search of recordings based on objects.

ABSTRACT

Team members commonly collaborate on visual documents remotely and asynchronously. Particularly, students are frequently restricted to this setting as they often do not share work schedules or physical workspaces. As communication in this setting has delays and limits the main modality to text, members exert more effort to reference document objects and understand others' intentions. We propose *Winder*, a Figma plugin that addresses these challenges through *linked tapes*—multimodal comments of clicks and voice. Bidirectional links between the clicked-on objects and voice recordings facilitate understanding tapes: selecting objects retrieves relevant recordings, and playing recordings highlights related objects. By periodically prompting users to produce tapes, Winder preemptively obtains information to satisfy potential communication needs. Through a five-day study with eight teams of three, we evaluated the system's impact on teams asynchronously designing graphical user interfaces. Our findings revealed that producing linked tapes could be as lightweight as face-to-face (F2F) interactions while transmitting intentions more precisely than text. Furthermore, with preempted tapes, teammates coordinated tasks and invited members to build on each others' work.

CCS CONCEPTS

• **Human-centered computing** → **Collaborative interaction**; **Collaborative and social computing systems and tools**; **Empirical studies in collaborative and social computing**; **Empirical studies in interaction design**.

KEYWORDS

Team Collaboration; Asynchronous Communication; Speech; Voice; Multimodal Input; Visual Document; User Interface Design.

ACM Reference Format:
Tae Soo Kim, Seungsu Kim, Yoonseo Choi, and Juho Kim. 2021. Winder: Linking Speech and Visual Objects to Support Communication in Asynchronous Collaboration. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3411764.3445686

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

ACM ISBN 978-1-4503-8096-6/21/05…$15.00
https://doi.org/10.1145/3411764.3445686

1 INTRODUCTION

Teams commonly collaborate around a shared visual document (e.g., user interface (UI) design or presentation slides). This is an essential process in both the workplace and academic settings. For productive and successful collaboration, frequent communication is necessary to develop a shared understanding between team members [25]. However, for teams of students, communicating effectively and efficiently can be challenging. Students are generally novices who lack the design knowledge necessary to express themselves effectively when discussing their design. Additionally, differences in course schedules and the lack of shared office space frequently restrict them to collaborate remotely and asynchronously [42].

Communication-related burdens introduced by an asynchronous setting can discourage students from discussing with their teams. Most asynchronous communication channels rely on text (e.g., text messaging and Google Docs' comments [19]), but typing involves high physical and cognitive demand [28]. Also, although effective referencing is essential in developing a shared understanding [22], text makes referring to visual objects in the document challenging—as pointing while typing is impossible. Unfortunately, students may lack the knowledge needed to overcome the restrictions of text [51].

Furthermore, regular scheduling issues between students [29] may prevent them from frequently checking on their team's messages and documents. This introduces a significant delay in communication and team members' communicative needs may not be satisfied in a timely manner. In addition, when students do check on messages, they may have to look through a large volume of messages [56], and references might not be adequate as the objects referenced may have changed after the messages were sent [52].

As shown by our formative study, feeling uncertain about receiving a response on time, and the aforementioned burdens when producing and consuming messages may lead to communication breakdowns in student teams [34]. Although professional teams have organizational support to handle these breakdowns [25], instructors may be ill-prepared or time-constrained to adequately provide similar support to students [41].

In this paper, we propose a novel way of communication with multimodal messages of voice and clicks—referred to as *linked tapes*—as an alternative form of asynchronous communication. Linked tapes are multimodal messages created by simply speaking while *pointing* at relevant visual objects through clicks, interactions which could require less effort when compared to typing text messages. When a tape is produced, the current version of the document is stored to preserve the temporal context of the tape to enable *change awareness* [44]. Additionally, bidirectional links between objects and voice snippets are automatically generated by temporally mapping the modalities onto each other (Fig. 1). With the voice-to-object link, playing back a voice message can display relevant objects to allow the receiver to effortlessly understand the references in the message. With the object-to-voice link, selecting an object of interest can filter through numerous voice messages to retrieve only those relevant to the object and facilitate the receiver's navigation. The ease of producing linked tapes with the multimodal input and the support provided by the bidirectional links can vitalize communication in asynchronous teams and improve a shared understanding.

We actualized communication based on linked tapes in *Winder*, a plugin for the collaborative UI design tool Figma. The plugin leverages the linked tapes' bidirectional links to implement three main features: (1) highlighting on playback (Fig. 1a), (2) inline thumbnails on transcripts (Fig. 1b), and (3) object-based search (Fig. 1c). In addition, Winder also aims to tackle the problem of communication delays. The plugin periodically prompts the user to produce linked tapes with the goal of preemptively obtaining information which may be needed by team members in the future. Tape-based communication could lessen the burden imposed by this approach—the effort of producing and consuming many messages. Thus, the drawbacks can be outweighed by the potential gains in shared understanding.

To investigate the effect of Winder on the collaboration process of asynchronous teams, we conducted a five-day study with eight teams of three students (N=24). Teams were tasked with designing the UI of a mobile application, which helps friends decide on what and where to eat. On the first day, they discussed ideas as teams and, on subsequent days, each team member worked on the design on different time slots. Our findings showed that the participant teams produced an average of 13.13 tapes and that the average tape length was 53.27 seconds. Analysis of survey responses revealed that, when compared to text messages, participants felt less burdened producing linked tapes due to the ease of speaking and clicking, and felt more confident that their messages would not be misunderstood. Participants also expressed that bidirectional links facilitated navigation through and within tapes, as well as their understanding of these tapes. Furthermore, the study results suggest that tapes recorded preemptively could allow for communication at hand without having team members at hand.

Our work contributes a novel multimodal asynchronous communication tool, *Winder*. Through lightweight interactions in production (i.e., click and voice) and bidirectional links for consumption, the system advances work in asynchronous communication by simultaneously decreasing burden for both senders and receivers—previous work facilitated either but not both. Furthermore, reducing at-the-moment burdens allows for an approach to tackle communication delays that would previously be overly burdensome: prompting users for preemptive recordings to satisfy future communication needs. As a secondary contribution, we present empirical findings that demonstrate the potential of Winder to reduce bilateral communication burden and overcome the detriment of delays in student teams.

2 RELATED WORK

We first review work on incorporating multiple modalities into asynchronous communication, then on communication anchored on documents, and finally on handling communication delays.

2.1 Multimodal Asynchronous Communication

Through modalities that are complementary, multimodal interaction is able to be more efficient and robust (i.e., less error-prone) than unimodal alternatives [37]. Alongside the accepted value of non-verbal methods for building common ground [22], this has inspired a rich body of work to integrate multiple modalities into

asynchronous channels. To enhance expressiveness in asynchronous ideation, SketchComm [31] provides a shared canvas onto which designers can freely interweave audio, photo, video, and hand-drawn sketches to express their ideas. RichReview [54] and RichReview++ [55] record voice annotations alongside ink and pointing gestures on a tablet. The combination of these modalities allows asynchronous collaborators to discuss around a digital text document as they would with a physical document if they were co-present. Other work such as Video Threads [4] relies on video and audio as the main components of communication by allowing users to create threads of video messages. In a more social and emotional context, FamilyStories [24] incorporates voice and physical actions to kindle a feeling of togetherness between family members distributed in place and time. As seen, previous approaches to multimodal asynchronous communication rely on combinations of sketching, gesturing, or video with voice. In our work, we instead focus on clicks to support deictic referencing—an interaction proven to be essential in remote settings [32]—while being simple enough to reduce the effort of message production.

Like previous work, on the other hand, we also incorporate voice or speech as it is faster than typing and rich due to expressive nuances it possesses, such as intonation and loudness [7]. These merits have led to the incorporation of "voice comments" in a variety of popular applications such as Microsoft Word [23] or Google Docs, through third-party plugins [46]. However, despite being faster in production, voice recordings are slow to consume and browse, which has impeded the widespread use of this modality [21]—the feature was discontinued in recent versions of Word. A common approach to tackling this challenge is to automatically generate transcripts through speech recognition. Transcribing facilitates consumption as this allows for keyword search, like in Otter.ai [36], or creating automatic summaries [8, 16, 49]. Speech recognition, however, is limited and may be inaccurate, which has led to the development of systems like TypeTalker [2] that allow the user to correct these errors. Considering the problems with automatic transcripts and the manual effort needed to resolve them, our work takes a different approach: automatically linking snippets in the voice recordings to document objects. With these links, we facilitate message browsing by allowing the user to filter out messages irrelevant to an object of interest. In addition, we provide thumbnail images of the objects inline with automatic transcripts of voice recordings to allow the user to navigate to moments in a recording when potentially interesting objects were selected and discussed.

2.2 Contextual References and Anchored Communication

While referencing items or sections of a document can be as simple as pointing in a F2F setting, this task becomes challenging and complex in online situations [10], even more so if team members are asynchronous. The user must provide detailed descriptions or rely on workarounds such as taking screenshots to adequately express the context and prevent confusion. To reduce the cost of creating contextual references, several systems have been designed to facilitate this process in diverse application scenarios. For discussions surrounding multimedia, Korero [14] supports referencing through linking to multiple portions of a video and, on the click of a button, Snapstream [53] instantaneously creates annotatable snapshots of a live stream. In a different domain, systems like chat.codes [35] and Callisto [52] have also been developed to support communication between programmers by enabling 'pointing' to code segments in chat interfaces.

Beyond incorporating the context of a document into the communication channel through references, substantial work has also explored anchoring the communication on the context itself. For example, the tools by both Churchill et al. [15] and Zyto et al. [57] allow for anchoring discussions on specific locations in text documents. Similarly, LemonAid [12] anchors question-and-answering communication on UI components of a web application to allow the answerers to provide more contextually adequate help. This work builds on these previous approaches on referencing and anchoring. Our proposed approach of *linked tapes* allows for multiple references to specific visual objects in one message—prior work supported referencing that was either singular [12], to general visual frames [14, 53], or to textual content [35, 52]. Additionally, anchored communication is possible as the user can select an object to retrieve recordings in which that object was clicked and record their own comments by talking while selecting the same object.

2.3 Approaches to Handling Communication Delays

Delays in communication are an inevitable aspect of asynchronous collaboration. While working on a document, a member of a team may need to communicate with their team; however, the other members may not always be attentive to the communication channel to respond immediately. Previous work has demonstrated that, aside from hindering the overall productivity of the team, these delays can also have social ramifications such as team members more negatively judging their fellow team members [43] and the overall task [26].

The detrimental consequences of delays in communication has motivated multiple researchers to design interventions to mitigate these. For example, Avrahami and Hudson [3] devised a notification system which distinguishes messages that require the user's immediate attention, and Pielot et al. [39] identified features that could predict a user's attentiveness to text messages, which could help manage expectations regarding response times. However, if the message receiver is certainly unavailable, these approaches will not sufficiently address existing challenges. A distinct approach explored in the domain of collaborative software development is to rely on external assistance. Codeon [11] and MicroMentor [27] connect a developer with remote helpers who can provide assistance when colleagues are not available. However, this type of support incurs a financial cost which may not be practical for all teams. Additionally, while these external assistants can provide technical help, they will not be able to aid with needs specific to a team's collaboration—e.g., understanding why a certain team member performed a certain change in the document.

As an alternative, we suggest preemptively obtaining team members' explanations of actions and intentions by prompting them while they are working. Our approach is inspired by think-aloud protocols used to capture participants' cognitive processes during studies on human subjects [50]. Although this type of information

does not satisfy all communicative needs, a shared understanding of team members' activities and goals allows for team coordination and is thus consequential to the team's success [48].

3 FORMATIVE STUDY

To understand the challenges in communication between team members collaborating on a visual document in an asynchronous setting, we conducted semi-structured interviews with 10 undergraduate students. We focused our investigation on student teams as they frequently collaborate asynchronously [42] and their collaboration experiences are meaningful as they allow for the learning of collaboration skills crucial in the workplace [5]. Additionally, student teams have little or no support available to handle the challenges of asynchrony—unlike teams in the workplace that may have workflows (e.g., Scrum [40]) or managers [25] in place to facilitate communication.

3.1 Interviews

We recruited a total of 10 undergraduate students (seven females and three males) at a technical university in South Korea. All participants had participated in at least one team project in their most recent semester in which the team members collaborated asynchronously on a visual document (e.g., UI design or presentation slides). We conducted one-hour long interviews in which participants were asked to reflect on their asynchronous collaboration experiences by freely looking through the previous interactions they had with their teams (e.g., chat logs or documents created). The interview questions mainly focused on three aspects: (1) what communication needs did students have while working asynchronously; (2) how the asynchronous setting affected their achievement of these needs; and (3) how failing to achieve their needs could impact their collaboration process.

3.1.1 Loss in Shared Understanding Led to Reworking or Subpar Outcomes. Initially, interviewees met synchronously with their team members through video conferencing tools—due to the COVID-19 pandemic—to discuss goals, tasks, and the assignment of these tasks. These discussions usually lasted approximately one hour and served to establish a shared understanding within the teams. After these discussions, each team member worked on their assigned tasks on their own time while communicating through familiar text messaging applications (e.g., Facebook Messenger). As the state of the visual document evolved with each member's contributions, interviewees frequently needed communication within their teams to understand what had changed and why. However, due to issues related to the team's asynchronous setting—which we discuss below—this communication would frequently not take place, which deteriorated the shared understanding of teams. As a consequence, work on the visual document would progress with misunderstandings remaining, which meant some team members had to redo their work later on or the team's outcome would be unsatisfactory. This finding parallels insights on remote work by Olson and Olson [34]—when building common ground, effort is required to resolve misunderstanding and, if this effort is too high, people may proceed without resolving them.

3.1.2 Burden When Producing and Sending Messages. Team members maintained a shared understanding by sharing progress updates or by asking questions about each other's work, but the process of producing and sending these messages was burdensome. Particularly, typing text messages required significant effort, especially when the message was referring to an object in the visual document. To prevent confusing team members, interviewees dedicated additional time thinking of how to write these messages and, in some cases, expended effort taking screenshots of the objects. In addition, interviewees also considered the perspective of team members on the receiving end when sending messages. Interviewees did not want to disrupt their team members and they also recalled on the burden they themselves had previously felt when receiving messages. Social costs related to the effort of producing messages and consideration of others when sending have also been identified in online question and answering [33]. Due to the social costs or burdens, interviewees frequently hesitated to produce and send a message, or even completely withheld from doing so. However, unlike the professional teams studied by Bjørn et al. [6] that faced similar challenges, students had no managerial practices to encourage them to communicate despite the burden they felt.

3.1.3 Burden when Receiving and Consuming Messages. As mentioned previously, interviewees also felt burdened when they were on the receiving end of a message. Due to their schedules and priorities, interviewees were not able to frequently check their team's messages. This meant that, once they were ready to work on the visual document and check messages, they would be faced with a large amount of messages. Significant effort was required to read through all these messages and to understand them, if the messages were not clearly written or sufficiently detailed. If they were unable to read through all the messages, interviewees tried navigating through them to find useful information. But, as all messages are presented in the same way in text messaging applications, distinguishing important information within these messages was challenging. These findings are corroborated with those reported by Zhang and Cranshaw [56]. Thus, collaborating asynchronously placed communicative burdens on not only the sender but also the receiver.

3.1.4 Waiting or Working with Limited Understanding due to Communication Delays. Even if the interviewees overcame the burden of producing and sending a message and asked their team members about their work, team members may not be available to provide an answer. As the team members were working asynchronously, immediate responses to messages were the exception and not the rule. In these situations, some interviewees waited for a response, which could be frustrating and stall the team's progress. Other interviewees worked on the document with a limited understanding, which at times led them to redo their work later as it had been completed with a misunderstanding of the team's goals. Therefore, communication delays impacted the overall productivity of teams as they either halted the effort of team members or caused effort to be wasted.

Our findings expand knowledge on the challenges of remote and/or asynchronous collaboration by revealing difficulties specific to the context of students collaborating asynchronously on a visual document. We gained the insight that team members refrain from

communicating in an asynchronous setting due to their own burden and the possibility of burdening their team members. Therefore, to develop and maintain a shared understanding in teams, reducing the effort of producing and consuming messages is crucial. With these communicative burdens reduced, we hypothesize that preemptively asking team members to explain their work while they are working can increase team productivity despite communication delays. If the explanations are created in advance, team members would not be affected by other members being unavailable to provide explanations as the explanations would already be at hand.

In accordance to the insights from our formative study, we set the following three design goals:

- DG1: Facilitate the production of explanations and referencing of objects in the visual document.
- DG2: Support navigation to required explanations, and the understanding of the explanations and their context.
- DG3: Preemptively obtain the user's explanations of their work and decisions while they are working on the visual document.

4 WINDER

To instantiate our design goals, we first establish a concrete context regarding the type of team and task we aim to support. We consider temporary teams in which members have vastly different schedules—e.g., student teams or cross-institutional research teams. These teams have no set communication practices, which can lead to unclear and infrequent communication, and are unable to allocate substantial blocks of time to work synchronously. In terms of the task, we focus on UI design. Unlike other visual documents like presentation slides that may incorporate significant amounts of text and have inherent structures (e.g., slides are in chronological order and usually have titles at the top), UI designs are highly visual and open-ended in terms of structure. While we believe in the generalization of our design goals, setting a concrete context allows us to design more effective support.

Based on this context and our design goals, we present *Winder* (Fig. 2), a system to support the asynchronous communication of team members collaborating on a UI design document. Winder is a plugin built on top of the collaborative interface design tool Figma. Figma [17] was chosen as the base for our plugin as it is a free service that sees widespread use, and for its flexibility with regards to plugin development.

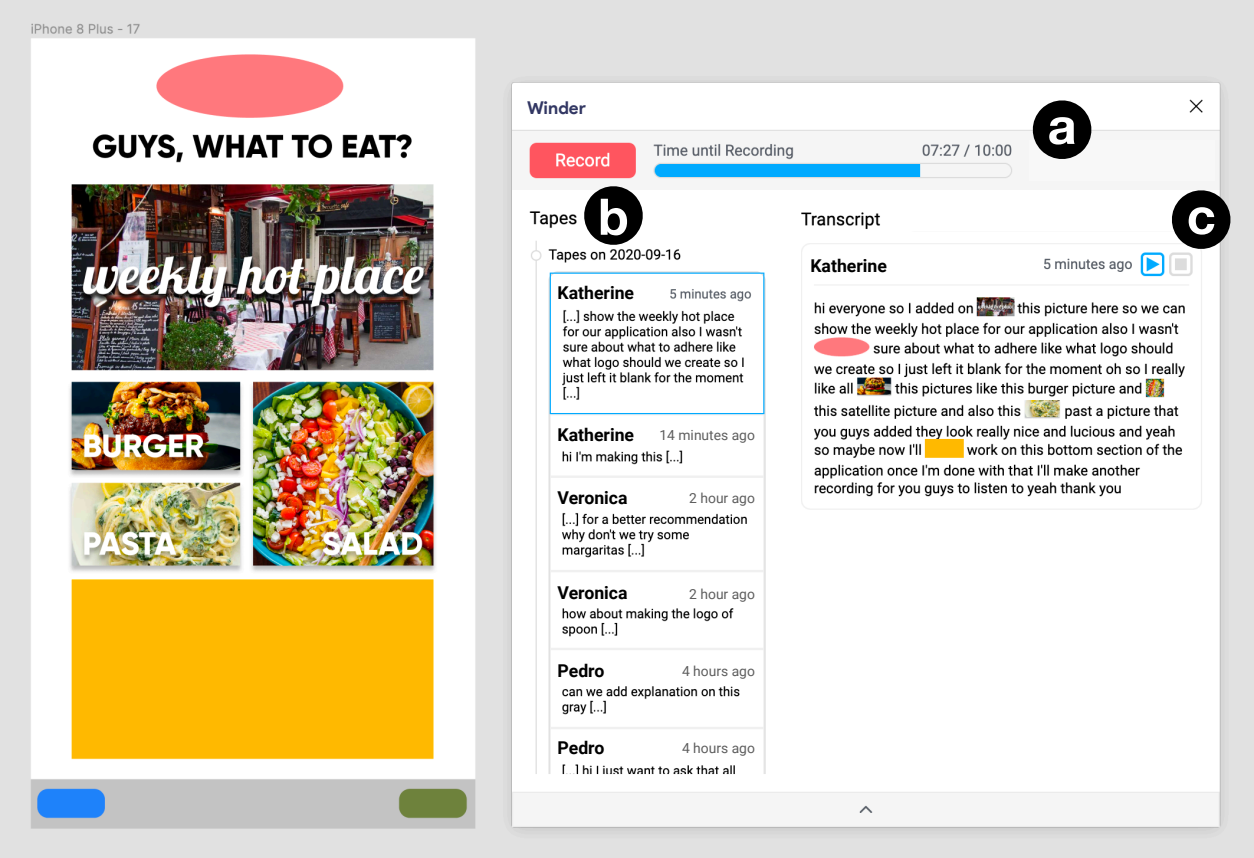

Figure 2: Winder (right) is shown on top of the Figma UI design document. The main components of the plugin's interface are (a) the top bar, (b) the list of linked tapes, and (c) the transcript space.

4.1 Overview of the Plugin

Winder supports asynchronous communication through *linked tapes*—multimodal recordings of voice comments and clicks on UI design objects. The plugin's screen consists of three main components: (a) the top bar, (b) the list of linked tapes, and (c) the transcript space. The top-bar (Fig. 2a) contains the "Record" button, which the user clicks to start a recording, and a timer, which alternates between states—i.e., "Time until Recording", "Recording" or "Playing". The list of linked tapes (Fig. 2b) shows all the tapes recorded by a team, grouped by the days in which they were recorded. Each tape entry displays the user that recorded it, how long ago it was recorded, a "NEW" label if the user has not yet played that tape, and a short fragment from the transcript of the tape's voice recording. The fragment is extracted from the 10 seconds with the highest average amplitude in the voice recording. Clicking on a tape on the list displays the tape's full transcript on the transcript space (Fig. 2c). The top-right of the displayed transcript shows media control buttons to allow for playing, pausing, and stopping of a tape.

4.2 Recording Linked Tapes

Although the user can freely decide when to record a tape, they are also prompted to do so every 10 minutes (DG3). While they are working, the timer on the top-bar counts down from 10 minutes—this interval was chosen based on pilot studies—and, once the timer runs out, the user receives a notification at the bottom of Figma asking them to record a tape. To record a tape, the user clicks on the "Record" button. This stores the current version of the document, and starts recording audio and clicks. The user can then simply comment on their design through their voice and make references to objects in the UI design document by simply clicking on them (DG1). Inspired by the work by Joshi et al. [27] which showed that imposing a 3-minute limit on help sessions could lessen the burden on the communicating parties, we also limit the length of tape recordings to 3 minutes. The amount of time that has progressed since the start of the recording is shown in the top-bar timer. Additionally, our pilot studies revealed that 1 minute is usually sufficient time for a recording. To further encourage the user to be even more concise in their recordings, the timer also begins to blink after 1 minute of recording.

4.3 Playing and Navigating Through Linked Tapes

For each linked tape, Winder generates an automatic transcript of the voice recording. In addition, Winder identifies all the snippets in the tape's voice recording where an object in the UI document had been clicked on and remained selected. By identifying these snippets, the system generates bidirectional links between voice snippets and objects. These links serve as the basis for three features which facilitate the understanding of tapes and the identification of desired information (DG2): (1) highlighting on playback, (2) inline thumbnails on transcripts, and (3) object-based search.

4.3.1 Object Highlighting on Voice Playback (Fig. 3). To play a linked tape, the user can click on the "Play" button on the top-right of a transcript. This shows to the user the version of the document when the tape was recorded, and starts playing the audio of the voice recording. As the audio of the voice recording is playing, Winder highlights the design objects that the creator of the tape had clicked on as they had been clicked during the tape's recording. Also, the object remains highlighted for the duration that it remained selected during the recording. This allows the user to easily distinguish, as they are listening to the voice explanation, what objects the creator of the tape had clicked on and was referring to. Highlighting is performed by making all other objects in the UI design more transparent, such that the highlighted object is more prominent. At any time, the user can pause the tape to pause the audio and highlighting to freely explore the previous version of the document. They can also stop the tape and go back to the current version.

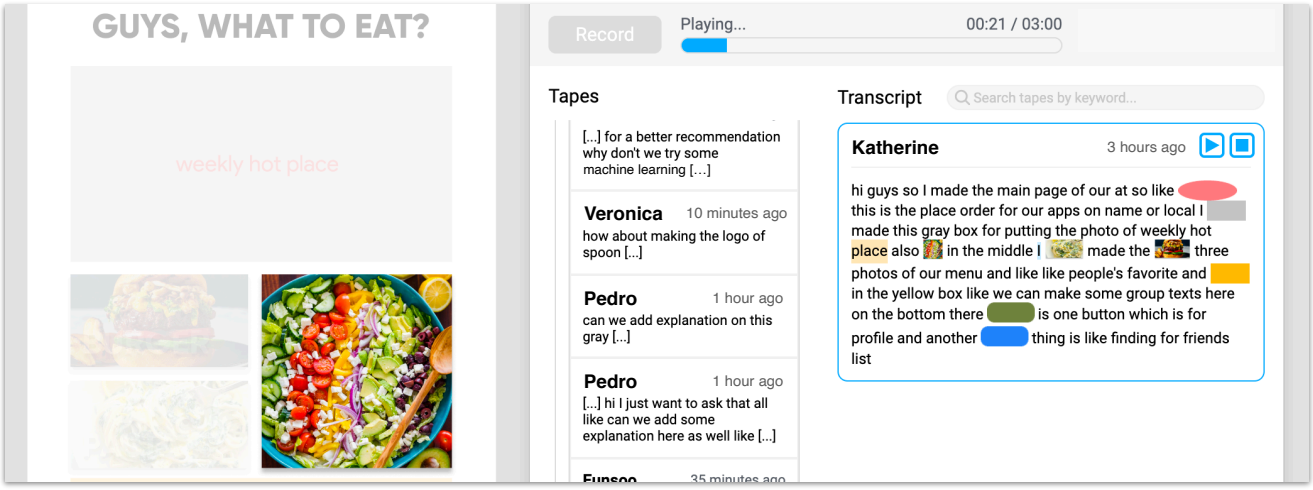

Figure 3: Winder is playing a linked tape, which plays the audio of the voice recording and highlights objects in Figma at the moments they were clicked and selected during the tape's recording (thus the picture of a salad is currently highlighted).

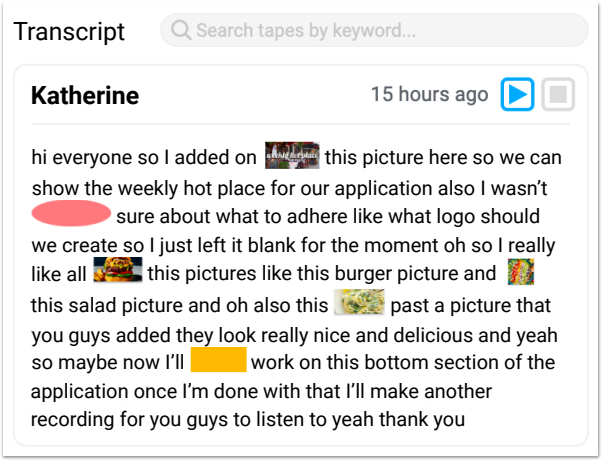

Figure 4: Winder presents an automatically generated transcript for each tape. Thumbnail images of the objects clicked on during the recording of the tape are shown in line with the transcript text to show when they were clicked.

4.3.2 Inline Thumbnail Images on Voice Transcripts (Fig. 4). For each linked tape, Winder embeds thumbnail images of the design objects into the transcript of the voice recording. These thumbnail images are obtained through the Figma API and are shown inline with the words of the transcript on the moments the objects were clicked during recording. With the thumbnail images, the user can see all the objects that were clicked during a tape's recording and also see when they were clicked. The user can quickly browse through tapes by looking at the thumbnail images in each tape. Also, the user can navigate to points in the voice recording by clicking on sections of the transcript. This allows the user to use the thumbnails as anchors for their navigation within a tape.

4.3.3 Object-based Search of Voice Recordings (Fig. 5). Winder also allows the user to search for voice recordings based on design objects of interest. Clicking on an object in the UI document retrieves all the tapes in which that object was clicked on during recording. The plugin then displays the retrieved tapes as a list in the transcript space. Each entry in the list shows the segment of the tape's transcript during which the object had been clicked on and remained selected. The user can skim through this list to quickly get a grasp of all the information related to that object. Also, if the user finds a tape with interesting information, they can display the full transcript by clicking on the expand button at the bottom of the entry or play the tape with the media controls on the top-right.

4.4 Implementation

We implemented the interface of Winder with TypeScript, ReactJS, and CSS. For the backend of the system, we used Node.js for the server and MongoDB for the database. Due to the fact that microphone access is not allowed from within the frame that houses the plugin in Figma, we used a separate browser window to record audio and deliver it to the server. Socket.io [1] was used to allow the plugin to communicate with the browser window used for microphone access. The Google Cloud Speech-to-Text [2] service was used to generate the transcripts from the voice recordings.

[1]https://socket.io/

5 METHODOLOGY

We designed a five-day user study with 24 students assigned into teams of three to investigate how Winder affected the communication within teams asynchronously collaborating on a UI design document. Our study aimed to examine the usage of Winder, as a whole, within the context of close-to-real team dynamics. Thus, the study was not comparative as a valid comparison would require controlled lab studies evaluating each system component in isolation. Through the study, we aimed to answer the following questions:

(1) How does multimodal communication based on voice and clicks affect the burden of sending messages when compared to typing text?
(2) How do bidirectional links between the visual objects and voice recordings support the identification and understanding of information?
(3) What type of content do preemptively recorded linked tapes contain and how do these impact team collaboration?

5.1 Participants

We recruited 24 participants (age M=22.25 and SD=2.98, 9 females and 15 males) who all reported to have no previous experience in UI design. Participants included 20 undergraduate students, 2 graduate students, and 2 industry workers. The industry workers were allowed to participate in the study as they worked in fields unrelated to design and expressed an interest in learning design in their spare time. Each participant was compensated KRW 70,000 ($59.00) for participating for approximately one hour every day for five days (total of five hours). Participants were randomly assigned into teams of three to produce eight teams in total. We formed teams of three as the formative study showed that this was a common size for student teams in course projects. The assignments were adjusted to ensure that none of the members in a team had previously known each other. Recruitment was carried out through online forums. In addition, participants were selected based on their fluency in English—judged through participants' backgrounds or examination scores—to ensure that participant teams did not face communication challenges due to language barriers and the accuracy of speech recognition.

5.2 Study Procedure

Due to the ongoing COVID-19 pandemic, the entire study was conducted remotely through Zoom [3]. On Day 1 of the study, participants were invited to a video call and introduced to the overall study procedure and the task. This task was to design the UI for a mobile application, which solved the following problem: "deciding on what and where to eat with friends is difficult and time-consuming." This problem was chosen from a list of past student projects in an "Introduction to HCI" course at our institution as we believed that it could be a relatable problem for participants and motivate engagement.

[2]https://cloud.google.com/speech-to-text
[3]https://zoom.us

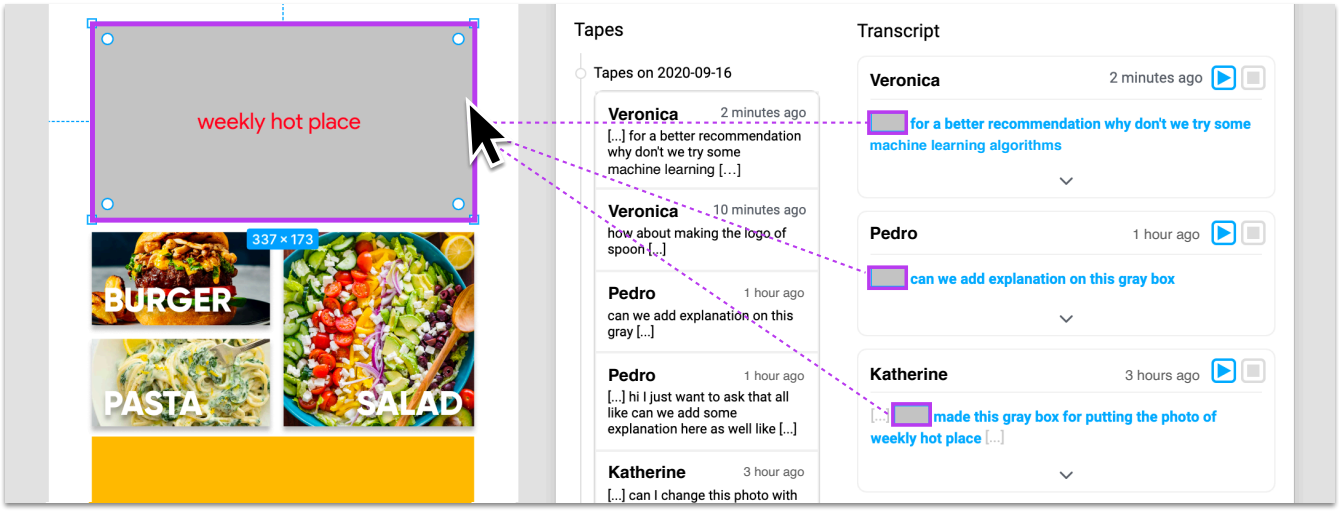

Figure 5: The user clicked on the gray rectangle object in Figma which displayed in Winder all the tapes in which that rectangle was clicked-on during the recording. In the list of tapes, each tape entry shows a transcript snippet of the voice recording during which the gray rectangle was selected.

To further motivate them, participants were also informed that the team with the best design—in terms of usefulness, uniqueness, simplicity, consistency, and completeness—would receive an additional KRW 5,000. Then, each team was sent to a separate video call room in which members were asked to first introduce themselves and then discuss about their design. A shared document was provided to each team which laid out discussion steps for the team to follow—brainstorm ideas, decide on the application's main screens, and distribute tasks. Team members could also make notes of their discussion in that shared document.

On Days 2 to 5 of the study, participants worked on their team's shared Figma document on different time slots to those of their team members. On Day 2, participants were provided with a short tutorial on Figma. Each study day consisted of 45 minutes of design work (including the duration of the tutorials) and 15 minutes of completing a post-survey. The survey asked participants about what they worked on that day and their perceptions on the communication they had with their team members. On their final day of the study (Day 5), participants also responded to an additional survey, which asked them about their overall collaboration and communication experiences. Throughout the four days, participants were allowed to communicate with their team members through the Google Docs document provided on Day 1, the comment feature in Figma, or a chatroom in KakaoTalk—a locally popular messaging application in South Korea. Other team members—those not working at that moment—were allowed to respond through these channels, but were not expected or required to do so.

Since our participants had no previous experiences in UI design collaboration against which they could evaluate their experiences with our system, we aimed to provide them with traditional collaborative experiences during the study. For half of the designing days, they were not allowed to use Winder and were limited to traditional communication tools. Since communication patterns in early and late design stages can differ [30] and this can affect usage of Winder, we aimed to understand system usage throughout the whole design process by allowing half of the participant teams to use the system only during early stages and allowing the other half of teams to use it during later stages. Thus, half of the teams were allowed to use Winder during Days 2 and 3 of the study, while the other half used it only during Days 4 and 5. On the day they would start using the plugin (Day 2 or 4), participants were provided with a tutorial on Winder.

5.3 Measures

Various types of data were collected during the study, including: survey responses, the content of linked tapes produced, chat messages exchanged, text added to the Google Docs, and comments left in Figma. Survey questions asked participants about how and why they sent messages, how and why they tried to understand their team members' actions and intentions, and why either of these processes were easy or difficult. Participants' responses to these questions were used to determine their perceptions on burden during their collaborative experiences. For the linked tapes, the voice recordings were manually transcribed due to the inaccuracies in automatic transcription. For the qualitative analysis, two of the authors conducted a thematic analysis of all the survey responses to derive the main findings of the study. The other types of data were used to verify and supplement these findings.

Also, the same two authors open coded the content of the linked tapes to categorize them based on their purposes. The open coding process involved developing and revising categories by looking through the data, and individual coding. The individual coding resulted in an average Cohen's kappa of 0.56 which indicates moderate agreement—the lowest kappa value was 0.43 and the highest was 0.69 (all within the moderate agreement range). As the initial agreement between coders was not significantly high, the coders

met to discuss, revise the categories' definitions, and resolve conflicts to reach a complete agreement on the codes.

6 RESULTS

Due to difficulties in objectively measuring burden or success in collaboration, our findings instead focus on the experiences and perspectives of our participants. Findings from our study revealed how voice and clicks, and the bidirectional links in *linked tapes* could facilitate asynchronous communication. Additionally, prompting the user to record tapes was found to potentially satisfy some future communication needs, clear misunderstandings, and even allow team member to better coordinate their design work. We provide a summary of our findings with mappings to research questions and system features at the end of the section in Table 4.

6.1 Linked Tapes Statistics and Categories

A summary of the statistics for the linked tapes produced during the study is presented in Table 1. Participants recorded a total of 107 tapes but 2 were not recorded properly due to technical issues with the participant's computers. These 2 tapes were not included in the analysis. The table shows that, on average, tapes were recorded around every nine minutes and their length was under one minute. This indicates that participants generally followed the system's suggestions for tape length and frequency. The linked tape categories derived through our open coding process are shown in Table 2, with details on each category. As seen in the table, describing, explaining one's own work, and coordinating tasks to complete ("Describe", "Justify", and "Coordinate" in Table 2) were the three most common purposes behind tapes during the study. However, tapes were also used for purposes beyond these. For example, some participants (T4M1, member 1 in team 4, and T2M2) mentioned leaving feedback on their other members' design through tapes which is also evidenced by the "Feedback" category. Further, as shown by the "Feedback" and "Clarify" categories, participants also engaged in back-and-forth communication by responding to each other's tapes. For instance, T6M1 mentioned: *"I left messages to answer back to the questions left by my teammates [and say] how I tried to solve their design concerns."* Interestingly, the "Clarify" category was the least common which could possibly be due to misunderstandings being preemptively clarified by the "Describe" or "Justify" tapes.

Statistics on Tapes	Mean	STD	Max	Min
Tapes by Member (number)	4.38	2.04	9	2
Tapes by Team (number)	13.13	3.14	18	10
Objects Clicked (number)	3.14	3.39	17	0
Length (seconds)	53.27	32.98	167	9
Interval between (seconds)	529.85	379.02	1827	72

Table 1: Statistics for the number of tapes created by members or teams, number of objects clicked on in each tape, length of tapes, and intervals between tape recordings.

6.2 Evaluating the Modalities in Winder

Participants felt ease in communicating with the two modalities, clicks and voice, supported in Winder but also pointed at several drawbacks related to these.

6.2.1 Pointing through Clicks. Most participants expressed feeling ease when pointing at or referring to UI objects with clicks. For example, T5M3 felt comfort when referencing design objects through clicks instead of writing words: *"It's pretty convenient to just click and talk rather than using vague terms [in text messaging]."* When discussing burdens related to referencing, several participants (T2M3, T4M1, T4M3, T7M1, T8M1, and T8M3) considered the perspective of the message receiver. All of these participants mentioned how misunderstanding references is possible with text but, by using clicks, they felt more assured that receivers would know what the messages are referencing. For example, T7M1 expressed: *"[Clicking] was the best way to deliver explanations without [confusing team members]."* Thus, due to the lower perceived effort and perceived uncertainty, pointing through clicks can potentially decrease feelings of burden when referencing. However, as G1P2 mentioned, referencing objects with clicks could be "unfamiliar". This led to several participants to hover on objects—instead of clicking—which made the resulting tapes more confusing as deictic references (e.g., "this" or "that") would not be accompanied by an object.

6.2.2 Expressing through Voice. Participants generally reported feeling less burdened when speaking a message instead of typing. T2M3 mentioned how recording linked tapes was easy for her as she could just *"talk instead of typing"*, and T4M1 likened voice recording to having a *"casual conversation"* with her team members. While most participants found voice easier than typing, participants were polarized in terms of which they preferred. To illustrate, T7M3 expressed that, for him, *"audio messages are almost always better than text messages"* while T5M3 mentioned how he *"prefers text over talking"* due to the awkwardness of recording his voice.

This polarization could be attributed to speaking-related burdens. A couple of participants felt conscious of their pronunciation as English was not their first language: *"My pronunciation is really bad, so I do not think that the system [will transmit] my message clearly."* (T8M2). Pressure to speak coherently was also felt by multiple participants, with some feeling that they had to improve on this: *"I am not sure if I explained everything well... I might have to improve on the flow of my recordings."* Previous work [55] also reported on self-consciousness due to pronunciation, but not on that due to coherence. This could be attributed to the time limit on voice recordings imposed by our system. Despite the initial burdens, participants reported gradually becoming accustomed to using their voice. In particular, for example, T8M3 continued sending voice recordings in his team's chatroom on the days he could not use the plugin.

6.2.3 Interweaving Clicks and Voice. Overall, communicating through both clicks and voice was unfamiliar for various participants, and simply familiarizing themselves to this was time-consuming: *"It took me really long to [...] get used to the plugin."* (T3M1). Nonetheless, several participants reported becoming accustomed to this form of communication within one study day (less than one hour). For instance, T2M1 mentioned how he became *"less anxious about*

Tape Categories	Description	Example Snippet	Percentage (%)
Describe	Describes what the explainer added, modified, or deleted.	**T1M3:** *"These two buttons I am modifying it for a better image."*	83.8
Justify	Explains the design rationale behind changes.	**T2M3:** *"I thought that feature is only necessary for this page only and you don't need that for results and reservation page so I got rid of those."*	46.6
Coordinate	Assigns tasks to others or reports on future tasks to be completed.	**T4M2:** *"I think it might be a good idea if you change the text for these to what you did here if you have the time to do it."*	40
Build on others	Invites others to work on one's own work or mentions having worked on another's work.	**T6M1:** *"[T6M2] told me that [they] added this button here what I actually erased it so you can't see."*	16.19
Feedback	Provides or asks for feedback on work.	**T7M3:** *"Instead of leaving the comments, it [would be] better if there's just the picture, the star rating, and the user comments."*	14.28
Social	Expresses social or emotional comments (e.g., praise, concern).	**T3M3:** *"I feel bad [as] I'm leaving too much work for [T3M2] but my time ran out."*	12.38
Clarify	Provides or asks for clarifications on completed work.	**T7M1:** *"My intention [of making] this page, this week's hot place, is to show some list[s] of some new restaurants and fancy restaurant."*	8.57

Table 2: The categories of tapes by the purpose of their content. These categories were not mutually exclusive—a tape could belong to multiple categories. Each entry in the table includes the name and description of a category, an example transcript snippet from a tape in that category, and the percentage of tapes which were coded with that category out of all tapes.

[creating] messages". One participant (T5M2) became so accustomed to the modalities that, once they could not rely on them in the second half of the study, they felt more alone without them: *"It felt more like I was working on my own now that I couldn't leave detailed voice messages."* Also, it is possible that the perceived lower burden in communicating with the two modalities encouraged all participant teams, apart from T3, to mostly send information through linked tapes (see "Total" in Table 3). Our analysis in Table 3 focuses on the number of words, instead of the number of messages/tapes, due to the high variance in their amount of content—the shortest message/tape contained one word while the longest contained more than 100. Thus, measuring the number of words could better represent the amount of information transferred within teams [45].

6.3 Bidirectional Links for Navigation and Understanding

The bidirectional links in Winder's linked tapes support three main features: (1) highlighting on playback, (2) inline thumbnails on transcripts, and (3) object-based search.

6.3.1 Facilitating Understanding with Object Highlighting. Without the plugin, one participant (T3M1) explained how he had to manually map discussions onto the related objects in the UI design: *"I went through the chat room discussions one by one and tried to match them with the corresponding [objects in the] UI design."* On the other hand, with the highlighting of objects when playing back recordings, participants noted how understanding what members were talking about felt easier: *"I could [easily] see what they were talking about so I quickly understood their [messages]."* (T6M2). Several participants (T2M2, T4M1, T4M3, T5M2, and T8M1) also felt that the highlighting allowed them to more precisely pinpoint what they should focus their attention on and track document changes: *"The recordings were useful to highlight exactly which parts had changed."* Additionally, some participants (T3M1 and T6M1) expressed that the highlighting made them feel, at least momentarily, that their team members were co-present: *"With the voice recordings and the feature that showed what the users clicked on as they talked, it was as if we were working together."*

6.3.2 Thumbnail Images to Support Navigation Through and In Recordings. Our findings reveal that some participants were indeed able to use the additional information provided by these thumbnails when navigating to points of interest in voice recordings: *"That was helpful because it allowed me to listen to what I wanted to hear [and reduce] time lost [..] listening to all the boring [recordings]."* (T1M2). When navigating through different tapes, T3M1 explained how the thumbnails gave him an overall sense of what the recording was about at a glance: *"[You could] look right away at the [UI objects] that were clicked on in one transcript."* Beyond simply perceiving what was clicked on, T5M3 indicated how he could also gain a high-level view of his team's design with the thumbnail images: *"It also gave me a general sense of [our team's UI] themes."* However, as noted by T5M2 and T7M3, the shapes of the objects affected the usability of the thumbnails as the images were resized to match the height of the text.

6.3.3 Potential to Enhance Productivity and Prevent Conflicts Through Object-based Search. Participants T5M1, T8M2, and T8M3 mentioned how object-based search could potentially increase their productivity by helping them quickly find the design rationale behind objects or identify what object-specific tasks needed to be completed. T7M1 and T7M3 used the feature to have a conversation anchored on a set of icons to clarify the purpose of those icons.

Group	Day 2		Day 3		Day 4		Day 5		Total	
	Tapes	Other	Tapes	Other	Tapes	Other	Tapes	Other	Tapes	Other
T1	739	0	1016	0	-	185	-	81	1755	266
T2	-	0	-	165	959	0	1275	0	2234	165
T3	-	757	-	854	702	294	655	305	1357	2210
T4	863	28	1154	164	-	426	-	310	2017	928
T5	361	135	398	193	-	50	-	188	759	566
T6	-	50	-	87	577	35	928	0	1505	172
T7	-	0	-	49	365	0	615	0	980	49
T8	255	0	551	46	-	184	-	209	806	439
								Average	1426.63	599.38

Table 3: Total word counts in the transcripts of tapes recorded and other communication channels for each team and study day. Days in which each team used Winder are filled in green or yellow. The average total number of words communicated by a team through linked tapes was 1426.63 (max=2234, min=759, SD=554.15) and through other channels was 599.50 (max=2211, min=49, SD=708.88).

Participant T8M2 also noted that the feature could increase his productivity as it helped him quickly find information on an incomplete object and he later completed it based on that information. Besides these deliberate uses, participant T2M3 mentioned how the feature unexpectedly showed her a recording, which allowed her to recall information: *"When I was changing [a group of four icons], I went back to listen to the message that was recorded previously."* Additionally, as pointed out by T1M3, object-based search allowed him to prevent a potential conflict in his team: *"There were certain parts that I wanted to modify and, if I clicked on those parts and there were recorded comments about [related future plans], I would not delete or edit those features in my own way."* This object-based search, however, was not used frequently by participants. Participants T1M2, T5M3, and T7M3 all mentioned not using the object-based search as they only had to listen to a small number of tapes, but saw how it could be useful if their designs were more complex.

6.4 Content and Effect of Preemptively Recording Linked Tapes

Our study results showed that teams used linked tapes for diverse purposes and that they could reduce some of their needs for direct communication. In turn, this could help with the coordination within teams and could also encourage team members to cooperate by working on top of each other's work.

6.4.1 Satisfying Communication Needs. A couple of participants (T7M2 and T1M1) explicitly stated that, when they had tapes recorded in advance by team members, they did not have to directly communicate with their team. T1M1 explained that the preempted tapes helped her gain an understanding of her team's work *"without chatting directly"* with her team members. Similarly, T6M3 noted that a team member assigned him a task on a tape so he could work without talking to his team: *"It helped me understand what I had to do. For example, one of our members left a message saying that I had to create a new page containing user information."* In addition, for some participants, listening to preempted tapes prevented misunderstandings and the potential back-and-forth in communication needed to resolve these confusions: *"The recordings left behind by my group members helped clarify some of the misunderstandings or confusions that I had."* (T2M1). Furthermore, we observed that, on days in which they had the plugin, all eight participant teams relied mostly on the linked tapes to communicate, with three teams (T1, T2, and T7) relying solely on the tapes (see Table 3).

6.4.2 Impact on Team Collaboration. Preemptively recorded linked tapes were used by some teams to coordinate their efforts. Half of the participants detailed how the tapes allowed them to identify remaining tasks and decide on new ones. Participant T2M3 mentioned: *"Understanding [my teammates'] intentions really helped me guideline what I needed to work on and how I could improve the [UI screens]."* Some participants also coordinated what "territories" [47]—sections of the document—they should or should not modify. Particularly, T1M3 kept himself from modifying his team members' work if it had not been allowed explicitly: *"It prevented me from deleting/editing [my team members'] work without their consent."* On the other hand, as seen by the "Build on others" category in Table 2, participants also welcomed team members into their own "territories" so that they could cooperatively iterate on specific parts of the design.

Beyond supporting their low-level task coordination, several teams used preempted tapes to support their collaboration at a higher level. For example, as noted by T4M3 and T5M1, listening

to team members' tapes, allowed the team to maintain design consistency. As the task was designing a UI, it required maintaining a shared direction on what the application does in addition to what the design looks like. For this purpose, preempted tapes were also useful. T6M1 left a tape in which she clicked on design objects to illustrate the application's user flow: *"Using the [plugin] it was easy to show the decision-making journey, because I can click on icons in chronological order and explain the journey step by step."* The content of recorded tapes showed that more than a third of participants (N=9) also illustrated user scenarios in their tapes.

6.4.3 Influence on the Social Factors of Teams. Beyond supporting work-related aspects, preemptively recorded tapes were also used to positively affect social factors—such as confidence, motivation, and trust. One participant (T8M2) mentioned how the nuances of voice helped him gain a sense of how confident his team members felt: *"By listening to recordings on the plugins, you can figure out [...] what they feel confident about, based on their tones."* Three participants (T1M2, T2M1, and T8M3) expressed feeling more motivated to work on their UI design after listening to team members' tapes: *"Understanding [my team members] actions and intentions was fun somehow and made me work harder."* (T1M2). Also, by increasing understanding of team members' intentions, preemptively recorded tapes could also increase trust: *"Thankfully, I've understood that they all had their own plans, which made me trust in them."* (T1M3).

6.4.4 Burden of Prompting and Preempted Tapes. While preemptively recording tapes was beneficial, several participants expressed feeling burdened by these due to several reasons. Firstly, the timer, which notified members to record, pressured some of the participants (T2M1, T3M2, T5M3, T7M1, and T8M2). T8M2, for instance, mentioned: *"Even though the time limit does not have to be obeyed, it still made me feel a huge pressure to [record] any type of comment."* Similarly, T5M3 suggested having an option to personally change the timer length: *"An option to change the timer [length] would be helpful because the work segments everyday are not going to be same length."* In addition, three participants (T3M2, T4M3, and T5M2) felt concerned that their recorded tapes would burden their team members due to their content: reminding on how much work is left (T5M2), assigning tasks (T3M2), or providing feedback (T4M3). Some participants (T1M3, T2M1, and T4M2) also felt pressured after listening to the tapes left by their team members. For example, T4M2 expressed that listening to his teammates *"preciously"* describe their design burdened him to work harder. However, other participants (T1M3, T4M3, and T8M1) also reflected positively on pressure as they argued that it is needed for successful collaboration.

7 CASE STUDIES

We observed that each team in the study showed unique patterns of communication as well as collaboration outcomes. In this section, we present in-depth case studies of two representative teams from the study. These teams were selected as their patterns of collaboration and communication revealed Winder's strengths and weaknesses. Through a grounded analysis, we investigated each team's cases along two dimensions: (1) working patterns, and (2) communication uses (with a focus on their use of Winder).

7.1 Team 1: Narrow Use of Winder Limits Improvement but Nonetheless Boosts Productivity

Team 1's initial discussion (Day 1) seemed to impact their communication quantity and purposes during the rest of the study. All the team members expressed satisfaction with that discussion in their survey responses as *"each feature was thoroughly discussed"* (T1M3) and they divided their individual tasks *"absolutely"* (T1M2). Members mostly kept to their own assigned screens, and most of their tapes focused on explaining edits and announcing future tasks they planned to complete. This is reflected by the categories of their tapes: all tapes (100% out of 18 tapes) were of the "Describe Work" category and 44.44% of the "Coordinate Tasks" category (both percentages were higher than the averages shown in Table 2). Team 1's members rarely used the tapes to provide each other with feedback (11.11% of "Feedback" category) or to invite others to work on top of their own work (11.11% of "Build on Others" category). As T1M3 mentioned in frustration, there was *"no sense of teamwork"*. Even though Team 1 did not use Winder for diverse purposes, its unavailability on Days 4 and 5 impacted their productivity. Without the plugin, communication virtually halted—aside from two conversations initiated by T1M3 (see row "T1", columns "Day 4" and "Day 5" in Table 3). As mentioned by T1M1, this made it "hard to know" what else she could do. Overall, due to the success of the discussion on Day 1 and the initial availability of the plugin, Team 1 still designed a UI which all the team members felt highly satisfied about. However, as noted by T1M3 on his last day, their narrow use of Winder prevented the team from further developing their original ideas—they only designed features that were discussed on the first day.

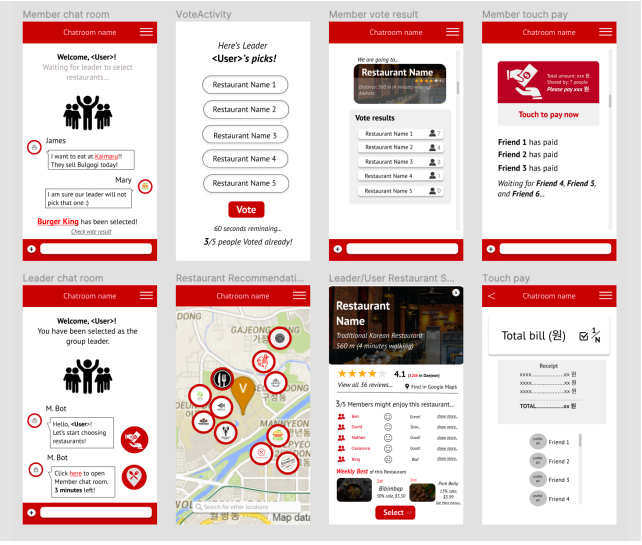

Figure 6: The final UI of the application designed by Team 1. They created an app named 'Meets' which allows groups to create specialized chatrooms for searching restaurants, voting, and splitting the bill.

Research question	System feature	Potential	Challenges
RQ 1: How does **multimodal communication** based on voice and clicks affect the **burden of sending messages** when compared to typing text?	Clicks	• Less perceived effort when referencing compared to text. • Greater feelings of assurance that references will not be misunderstood.	• Unfamiliarity led to hovering on objects and greater confusion.
	Voice	• Less perceived effort in talking compared to typing. • Users generally became accustomed to voice as a modality within two study days.	• Self-consciousness about pronunciation. • Burden to speak coherently during recording.
RQ 2: How do **bidirectional links** between the visual objects and voice recordings support the **identification and understanding of information**?	Object highlighting on voice playback	• Reduced perceived effort of mapping messages to referred to objects. • Promotes sensation of co-presence despite asynchrony.	-
	Inline thumbnail images on voice transcripts	• Can support navigation to points of interest within transcripts. • Can act as overviews of tape content and design themes.	• Variance in shape and size of objects can reduce readability of transcripts.
	Object-based search of voice recordings	• Can facilitate identification of design rationale for parts of interest. • Can incite spontaneous recalling of previously recorded information. • Potential to reduce future conflicts by preventing modification undesired by other members.	• Not needed with small-scale or simple designs.
RQ 3: What **type of content** do preemptively recorded linked tapes contain and how do these **impact team collaboration**?	Preempted tapes	• Preemption can reduce some needs for direct or back-and-forth communication. • Tapes used to coordinate tasks and work "territories". • Tapes used to coordinate high-level design decisions. • Can promote confidence, trust, and motivation.	• Presence of a prompting timer pressured users. • Inflexible timer length led to recording of meaningless tapes. • Burden to work harder due to others' preemptive tapes. • Concern about burdening others with content of one's own tapes.

Table 4: Summary of the study findings, mapped based on research questions, relevant system features, and the potential and challenges.

7.2 Team 3: Frequent Tape Recording could Remedy the Detriment of Unequal Contribution

In contrast to Team 1, Team 3 frequently communicated through various channels (e.g., Google Docs, chatroom, and Winder), but still had problems reaching a shared understanding. T3M2 was the team's 'de-facto leader': she set the design layout and decided on tasks. The two other members—T3M1 and T3M3—worked based on her design. To communicate, T3M2 mainly relied on the team's Google Docs to write detailed updates. This accounted for Team 3 having the highest usage of other channels among participant teams (see "Total" in Table 3). Both T3M1 and T3M3 expended a significant amount pf time understanding T3M2's updates and were unable to work much during the limited session time. This problem was partially remedied once Winder was provided as T3M1 and T3M3 could easily understand T3M2's updates in the context of the document. However, T3M2 preferred *"writing memos to voice messages"* and kept using Google Docs—she had the lowest number of tapes recorded (N=2) among all study participants. Due to T3M2's reluctance to communicate through tapes and the time it took to understand her written comments, T3M1 and T3M3 had to frequently use the tapes to ask T3M2 to complete tasks for them. This is reflected by their most common tape category being "Coordinate Work" (85.71% out of 14 tapes). Overall, Team 3's case reveals that Winder can facilitate shared understanding in teams with largely unequal contribution levels. However, the effectiveness is dependent on whether the largest contributor records frequently or not.

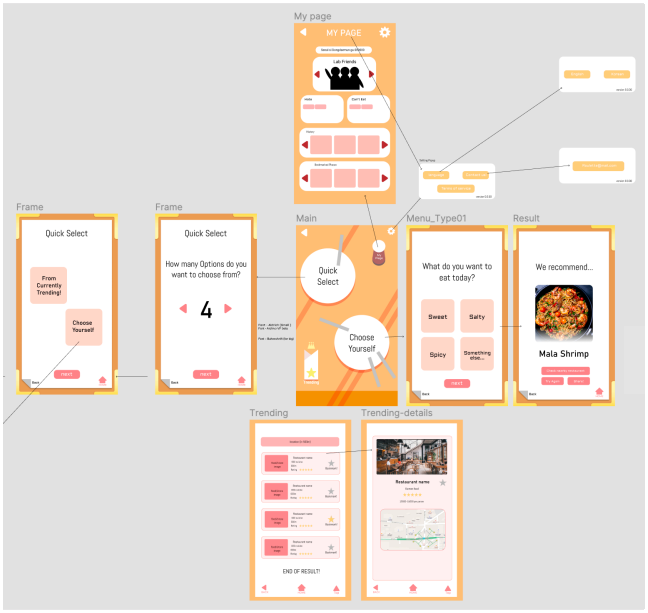

Figure 7: The final UI of the application designed by Team 3. They created a unique menu decision-making interface based on a roulette of restaurants, and screens to explore preferences and trends. They marked connections between the screens to show the user flow.

8 DISCUSSION

In this paper, we propose linked tapes, a novel form of asynchronous communication that integrates multimodal input and visual referencing. Our evaluation of Winder, a system which instantiates linked tapes, showed that tapes could lessen the perceived effort of producing comments and references to visual objects, while also facilitating navigation and understanding of others' comments. The study also showed that preempted tapes could potentially satisfy future communication needs. In this section, we provide a summary of what worked well and what did not with Winder. Then, we discuss how linked tape communication can be used in practice and generalized to other tasks. Finally, we discuss the implications of our work on asynchronous communication in general.

8.1 Winder Increases Shared Understanding and Augments Collaboration

Our evaluation focused on a holistic evaluation of Winder. As the features are interrelated, it is challenging to discern the particular effect of individual features. However, based on participants' comments regarding the features, we interpret the possible relationships between the features and how they could have led to the observed effects in our study.

Winder can potentially encourage students to communicate more frequently with their team members. This observed effect could be attributed to both the prompting and users perceiving less effort when producing messages with voice and clicks. While the greater transfer of information between team members can benefit shared understanding, it can also lead to greater burden. Despite participants generally considering production to be easier with Winder's modalities, the effort is not negligible, the use of voice can be awkward, and prompting was seen as a source of burden. Also, more messages produced entails greater burden on receivers. Thus, while Winder reduced feelings of burden related to typing, it could also introduce other types of burden and, at times, with no benefit for the teams as tapes could contain meaningless information—some participants recorded tapes simply because they were prompted.

Referencing visual objects through clicks appeared to reassure students that their messages would not be misunderstood. Our participants' comments regarding the perceived clarity of references—through the object highlighting—infers that this reassurance was justified. Aside from supporting clarity in references, processing clicks and voice also allowed for inline thumbnails and object-based search, which supported students navigation through and within tapes. While participants expressed that this support could reduce their perceived effort in consumption, most of them still listened to all of their team members' tapes to gain a complete understanding. Several participants described this process as "tedious", showing that the bidirectional links had a limited effect on reducing consumption burden. To resolve this, future work could leverage bidirectional links and the document history to generate automatic summaries to allow for a general understanding of changes. These summaries could be similar to the *workflow histories* by Grossman et al. [20] but enriched with the information shared in the voice recordings.

By prompting users to record linked tapes as they are working, Winder obtains preemptively recorded tapes, which could, to a certain extent, substitute direct communication. Preempted tapes were used for diverse purposes—such as coordinating tasks and encouraging cooperation on the same UI screens. During the study, however, most of the tapes involved simple descriptions of completed work. As seen by the case study of Team 1, only providing these low-level comments can limit the benefits of preempted tapes. As tape recording resembles reflection—thinking and talking about one's own actions—providing guides based on reflection literature could result in high-level reflections that are more beneficial to teams' collaborative processes. For example, future work could adopt the levels of reflection proposed by Fleck and Fitzpatrick [18].

8.2 Integrating Linked Tapes into Teams' Application Ecosystems

While our study revealed that certain teams communicated exclusively through linked tapes, in practical scenarios, student teams may already rely on several familiar channels for communication (e.g., chat). Stacking linked tapes on top of these channels can be more detrimental than beneficial. As seen from our study, there is effort involved in keeping track of this additional channel and students may feel dissatisfied due to its unfamiliarity. Thus, for the success of Winder and linked tapes, it is crucial to establish how they fit in students' existing ecosystems. Course instructors could recommend ideal arrangements students could follow to use Winder alongside other channels. A partially successful arrangement which was observed in the study is to use a shared document to track high-level tasks, chat for semi-synchronous discussions on goals and direction, and linked tapes for more detailed and document-centric comments. In addition, Winder could be connected to communication channels already used by students to lower their barriers to entry [1]. For example, the system could send transcript snippets and object thumbnails from linked tapes to social chat applications already used by students.

8.3 Generalizability of Winder and Linked Tapes

While our investigation focused on UI design, we believe that Winder could be modified for other types of document-related tasks performed in online education settings. In particular, tasks in which referencing is common and students' frequently communicate asynchronously could be benefited by a Winder-like system. Two potential tasks are collaboration on presentation slides and discussion on lecture videos. For asynchronous collaboration on presentation slides, adapting Winder would be straightforward—presentations contain discrete objects (e.g., slides, text boxes, and shapes). However, the system might not be as beneficial in this task as presentations have an inherent structure—e.g., chronological order of slides—which could make referencing through text less challenging. For asynchronous discussions about lecture videos [13], computer vision techniques, such as video object segmentation [38], could be used to extract discrete objects from the video. Then, students could click on these objects to discuss the lecture content.

8.4 Implications for Asynchronous Communication

Our work advances research in asynchronous communication through linked tapes, which bidirectionally integrate the communication channel and the document context. Previous work has primarily explored integration in a single direction—only referencing, which brings the context into the discussion, or only anchored communication, which incorporates the discussion into the context. As our study suggested, however, bidirectional integration allows for both discussions 'across' the document (higher-level and considering the whole picture) and 'into' objects (in-depth and specific to details). We suggest future researchers to consider bidirectional integration to support asynchronous collaboration. Additionally, our initial investigation into the impact of preemptive communication on asynchronous collaboration showed potential for future development. By prompting team members to communicate in advance to a system, it could allow a future user to have communication at hand without waiting for the availability of their team members. While our approach was naive—fixed intervals in which the user was softly notified—future work can explore this concept in more sophisticated ways. For example, prompting can be more contextual—based on intermediate steps detected in the user's workflow [9].

9 LIMITATIONS

Our work has limitations which we address in this section.

First, as we evaluated Winder in a controlled study instead of a deployment study, participants' communicative behaviors might have been affected by the setting. It is possible that they were pressured to record tapes as they were being observed by researchers. On the contrary, participants might have communicated less as the monetary award for the team with the best design might have not been enough incentive. Future work could explore the longitudinal use of Winder through a deployment study in a university project course.

Second, as evaluating the quality of designs is subjective and quality might be largely dependent on the characteristics of the members, we did not analyze how the use of tapes affected the quality of the outcomes. Therefore, we were unable to concretely measure how Winder impacted the quality of teams' collaborations.

Third, we conducted our study with participants who had no previous experience in UI design and were assigned into teams of three. The experiences of team members and the size of teams can significantly affect collaboration and communication behaviors. For example, professional designers may have established communication practices (e.g., routines and terminology) which can lead to different usage patterns of Winder and reactions to the system compared to those observed in our study. Future work could also investigate how Winder can be used by domain experts and bigger team sizes.

Finally, our study was carried out with a sample size of 24 participants, or 8 teams. More samples are needed to gain more conclusive findings. Also, our study focused on investigating participants' perceptions on burden when using Winder, as a whole, in close-to-real settings. For more conclusive findings on the effect of our system's features on burden, controlled studies quantitatively evaluating

each feature against baselines are needed. For example, a possible setup could be measuring the time taken to transmit a certain amount of information with voice and clicks and comparing that to the time taken through typing.

10 CONCLUSION

This paper presents *Winder*, a novel system that supports asynchronous communication between students in UI design collaboration. Winder provides communication through linked tapes—multimodal recordings of voice and clicks that contain bidirectional links between the comments and document objects. Additionally, by prompting the user to record tapes, Winder preemptively obtains information that can substitute direct communication when satisfying team members' needs, thus reducing the impact of communication delays. A five-day user study showed the effectiveness of linked tapes and preemptive recording in the collaborative processes of students. Finally, we discussed how Winder can be used in practice and generalized to other contexts, and the implications of linked tapes on general asynchronous communication.

11 ACKNOWLEDGMENTS

This research was supported by the KAIST UP Program. The authors would like to thank the members of KIXLAB for their support and feedback. We also thank our study participants for their time. Finally, we thank the reviewers for helping us improve our paper through their guidance and feedback.

REFERENCES

[1] Paige Abe and Nickolas A Jordan. 2013. Integrating social media into the classroom curriculum. *About Campus* 18, 1 (2013), 16–20.

[2] Ian Arawjo, Dongwook Yoon, and François Guimbretière. 2017. TypeTalker: A Speech Synthesis-Based Multi-Modal Commenting System. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 1970–1981. https://doi.org/10.1145/2998181.2998260

[3] Daniel Avrahami and Scott E. Hudson. 2004. QnA: Augmenting an Instant Messaging Client to Balance User Responsiveness and Performance. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work* (Chicago, Illinois, USA) *(CSCW '04)*. Association for Computing Machinery, New York, NY, USA, 515–518. https://doi.org/10.1145/1031607.1031692

[4] 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

[5] Stephanie Bell. 2010. Project-based learning for the 21st century: Skills for the future. *The clearing house* 83, 2 (2010), 39–43.

[6] Pernille Bjørn, Morten Esbensen, Rasmus Eskild Jensen, and Stina Matthiesen. 2014. Does Distance Still Matter? Revisiting the CSCW Fundamentals on Distributed Collaboration. *ACM Trans. Comput.-Hum. Interact.* 21, 5, Article 27 (Nov. 2014), 26 pages. https://doi.org/10.1145/2670534

[7] Barbara L. Chalfonte, Robert S. Fish, and Robert E. Kraut. 1991. Expressive Richness: A Comparison of Speech and Text as Media for Revision. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New Orleans, Louisiana, USA) *(CHI '91)*. Association for Computing Machinery, New York, NY, USA, 21–26. https://doi.org/10.1145/108844.108848

[8] Senthil Chandrasegaran, Chris Bryan, Hidekazu Shidara, Tung-Yen Chuang, and Kwan-Liu Ma. 2019. TalkTraces: Real-Time Capture and Visualization of Verbal Content in Meetings. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300807

[9] Minsuk Chang, Ben Lafreniere, Juho Kim, George Fitzmaurice, and Tovi Grossman. 2020. Workflow Graphs: A Computational Model of Collective Task Strategies for 3D Design Software. In *Proceedings of Graphics Interface 2020* (University of Toronto) *(GI 2020)*. Canadian Human-Computer Communications Society / Société canadienne du dialogue humain-machine, Toronto, 114 – 124. https://doi.org/10.20380/GI2020.13

[10] Yuan-Chia Chang, Hao-Chuan Wang, Hung-kuo Chu, Shung-Ying Lin, and Shuo-Ping Wang. 2017. AlphaRead: Support Unambiguous Referencing in Remote Collaboration with Readable Object Annotation. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 2246–2259. https://doi.org/10.1145/2998181.2998258

[11] Yan Chen, Sang Won Lee, Yin Xie, YiWei Yang, Walter S. Lasecki, and Steve Oney. 2017. Codeon: On-Demand Software Development Assistance. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 6220–6231. https://doi.org/10.1145/3025453.3025972

[12] Parmit K. Chilana, Andrew J. Ko, and Jacob O. Wobbrock. 2012. LemonAid: Selection-Based Crowdsourced Contextual Help for Web Applications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 1549–1558. https://doi.org/10.1145/2207676.2208620

[13] Gayle Christensen, Andrew Steinmetz, Brandon Alcorn, Amy Bennett, Deirdre Woods, and Ezekiel Emanuel. 2013. *The MOOC phenomenon: who takes massive open online courses and why?* Available at SSRN 2350964. https://dx.doi.org/10.2139/ssrn.2350964

[14] Soon Hau Chua, Toni-Jan Keith Palma Monserrat, Dongwook Yoon, Juho Kim, and Shengdong Zhao. 2017. Korero: Facilitating Complex Referencing of Visual Materials in Asynchronous Discussion Interface. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 34 (Dec. 2017), 19 pages. https://doi.org/10.1145/3134669

[15] Elizabeth F. Churchill, Jonathan Trevor, Sara Bly, Les Nelson, and Davor Cubranic. 2000. Anchored Conversations: Chatting in the Context of a Document. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (The Hague, The Netherlands) *(CHI '00)*. Association for Computing Machinery, New York, NY, USA, 454–461. https://doi.org/10.1145/332040.332475

[16] Patrick Ehlen, Matthew Purver, John Niekrasz, Kari Lee, and Stanley Peters. 2008. Meeting Adjourned: Off-Line Learning Interfaces for Automatic Meeting Understanding. In *Proceedings of the 13th International Conference on Intelligent User Interfaces* (Gran Canaria, Spain) *(IUI '08)*. Association for Computing Machinery, New York, NY, USA, 276–284. https://doi.org/10.1145/1378773.1378810

[17] Figma. 2019. Figma: the collaborative interface design tool. Retrieved September 13, 2020 from https://www.figma.com/

[18] Rowanne Fleck and Geraldine Fitzpatrick. 2010. Reflecting on Reflection: Framing a Design Landscape. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction* (Brisbane, Australia) *(OZCHI '10)*. Association for Computing Machinery, New York, NY, USA, 216–223. https://doi.org/10.1145/1952222.1952269

[19] Google. 2016. Use comments & action items - Computer - Docs Editors Help. Retrieved September 13, 2020 from https://support.google.com/docs/answer/65129

[20] Tovi Grossman, Justin Matejka, and George Fitzmaurice. 2010. Chronicle: Capture, Exploration, and Playback of Document Workflow Histories. In *Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology* (New York, New York, USA) *(UIST '10)*. Association for Computing Machinery, New York, NY, USA, 143–152. https://doi.org/10.1145/1866029.1866054

[21] Jonathan Grudin. 1988. Why CSCW Applications Fail: Problems in the Design and Evaluationof Organizational Interfaces. In *Proceedings of the 1988 ACM Conference on Computer-Supported Cooperative Work* (Portland, Oregon, USA) *(CSCW '88)*. Association for Computing Machinery, New York, NY, USA, 85–93. https://doi.org/10.1145/62266.62273

[22] Carl Gutwin and Saul Greenberg. 2002. A descriptive framework of workspace awareness for real-time groupware. *Computer supported cooperative work* 11, 3-4 (2002), 411–446.

[23] Susan Harkins. 2009. Insert voice comments into a Word document. https://www.techrepublic.com/blog/microsoft-office/insert-voice-comments-into-a-word-document/

[24] Yasamin Heshmat, Carman Neustaedter, Kyle McCaffrey, William Odom, Ron Wakkary, and Zikun Yang. 2020. FamilyStories: Asynchronous Audio Storytelling for Family Members Across Time Zones. 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, 1–14. https://doi.org/10.1145/3313831.3376486

[25] Pamela J Hinds and Suzanne P Weisband. 2003. Knowledge sharing and shared understanding in virtual teams. In *Virtual teams that work: Creating conditions for virtual team effectiveness*, C.B. Gibson and S.G Cohen (Eds.). Jossey-Bass, San Francisco, CA, USA, 21–36.

[26] Sun Young Hwang, Negar Khojasteh, and Susan R. Fussell. 2019. When Delayed in a Hurry: Interpretations of Response Delays in Time-Sensitive Instant Messaging. *Proc. ACM Hum.-Comput. Interact.* 3, GROUP, Article 234 (Dec. 2019), 20 pages. https://doi.org/10.1145/3361115

[27] Nikhita Joshi, Justin Matejka, Fraser Anderson, Tovi Grossman, and George Fitzmaurice. 2020. MicroMentor: Peer-to-Peer Software Help Sessions in Three Minutes or Less. 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, 1–13. https://doi.org/10.1145/3313831.3376230

[28] Barry M Kroll. 1978. Cognitive egocentrism and the problem of audience awareness in written discourse. *Research in the Teaching of English* 12, 3 (1978), 269–281.

[29] Heng-Yu Ku, Hung Wei Tseng, and Chatchada Akarasriworn. 2013. Collaboration factors, teamwork satisfaction, and student attitudes toward online collaborative learning. *Computers in human Behavior* 29, 3 (2013), 922–929.

[30] James A. Landay and Brad A. Myers. 1995. Interactive Sketching for the Early Stages of User Interface Design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '95)*. ACM Press/Addison-Wesley Publishing Co., USA, 43–50. https://doi.org/10.1145/223904.223910

[31] Guang Li, Xiang Cao, Sergio Paolantonio, and Feng Tian. 2012. SketchComm: A Tool to Support Rich and Flexible Asynchronous Communication of Early Design Ideas. 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, 359–368. https://doi.org/10.1145/2145204.2145261

[32] Paul Luff, Christian Heath, Hideaki Kuzuoka, Keiichi Yamazaki, and Jun Yamashita. 2006. Handling Documents and Discriminating Objects in Hybrid Spaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Montréal, Québec, Canada) *(CHI '06)*. Association for Computing Machinery, New York, NY, USA, 561–570. https://doi.org/10.1145/1124772.1124858

[33] Haiwei Ma, Bowen Yu, Hao Fei Cheng, and Haiyi Zhu. 2019. Understanding Social Costs in Online Question Asking. 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/10.1145/3290607.3313042

[34] Gary M Olson and Judith S Olson. 2000. Distance matters. *Human–computer interaction* 15, 2-3 (2000), 139–178.

[35] Steve Oney, Christopher Brooks, and Paul Resnick. 2018. Creating Guided Code Explanations with Chat.Codes. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 131 (Nov. 2018), 20 pages. https://doi.org/10.1145/3274400

[36] Otter.ai. 2019. Otter.ai: Otter Voice Meeting Notes. Retrieved September 13, 2020 from https://otter.ai/

[37] Sharon Oviatt. 1999. Ten Myths of Multimodal Interaction. *Commun. ACM* 42, 11 (Nov. 1999), 74–81. https://doi.org/10.1145/319382.319398

[38] Federico Perazzi, Anna Khoreva, Rodrigo Benenson, Bernt Schiele, and Alexander Sorkine-Hornung. 2017. Learning video object segmentation from static images. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (Holulu, HI, USA). IEEE, New York, NY, USA, 2663–2672.

[39] Martin Pielot, Rodrigo de Oliveira, Haewoon Kwak, and Nuria Oliver. 2014. Didn't You See My Message? Predicting Attentiveness to Mobile Instant Messages. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 3319–3328. https://doi.org/10.1145/2556288.2556973

[40] Lene Pries-Heje and Jan Pries-Heje. 2011. Why Scrum works: A case study from an agile distributed project in Denmark and India. In *2011 Agile Conference*. IEEE, New York, NY, USA, 20–28.

[41] Linda Riebe, Antonia Girardi, and Craig Whitsed. 2016. A systematic literature review of teamwork pedagogy in higher education. *Small Group Research* 47, 6 (2016), 619–664.

[42] Chiara Rossitto, Cristian Bogdan, and Kerstin Severinson-Eklundh. 2014. Understanding constellations of technologies in use in a collaborative nomadic setting. *Computer Supported Cooperative Work (CSCW)* 23, 2 (2014), 137–161.

[43] Oliver J Sheldon, Melissa C Thomas-Hunt, and Chad A Proell. 2006. When timeliness matters: The effect of status on reactions to perceived time delay within distributed collaboration. *Journal of Applied Psychology* 91, 6 (2006), 1385.

[44] James Tam and Saul Greenberg. 2004. A framework for asynchronous change awareness in collaboratively-constructed documents. In *International Conference on Collaboration and Technology*. Springer, Berlin, Germany, 67–83.

[45] Yla R. Tausczik and James W. Pennebaker. 2013. Improving Teamwork Using Real-Time Language Feedback. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 459–468. https://doi.org/10.1145/2470654.2470720

[46] Texthelp. 2015. Read&Write Literacy Support Software | Texthelp. Retrieved December 17, 2020 from https://www.texthelp.com/en-gb/products/read-write

[47] Jennifer Thom-Santelli, Dan R. Cosley, and Geri Gay. 2009. What's Mine is Mine: Territoriality in Collaborative Authoring. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 1481–1484. https://doi.org/10.1145/1518701.1518925

[48] James D Thompson. 2003. *Organizations in action: Social science bases of administrative theory*. Transaction publishers, Piscataway, NJ, United States.

[49] 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.

[50] MW Van Someren, YF Barnard, and JAC Sandberg. 1994. *The think aloud method: a practical approach to modelling cognitive*. Citeseer, London.

[51] Alonso H. Vera, Thomas Kvan, Robert L. West, and Simon Lai. 1998. Expertise, Collaboration and Bandwidth. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Los Angeles, California, USA) *(CHI '98)*. ACM Press/Addison-Wesley Publishing Co., USA, 503–510. https://doi.org/10.1145/274644.274712

[52] April Yi Wang, Zihan Wu, Christopher Brooks, and Steve Oney. 2020. Callisto: Capturing the "Why" by Connecting Conversations with Computational Narratives. 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, 1–13. https://doi.org/10.1145/3313831.3376740

[53] Saelyne Yang, Changyoon Lee, Hijung Valentina Shin, and Juho Kim. 2020. Snapstream: Snapshot-Based Interaction in Live Streaming for Visual Art. 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/10.1145/3313831.3376390

[54] Dongwook Yoon, Nicholas Chen, François Guimbretière, and Abigail Sellen. 2014. RichReview: Blending Ink, Speech, and Gesture to Support Collaborative Document Review. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (Honolulu, Hawaii, USA) *(UIST '14)*. Association for Computing Machinery, New York, NY, USA, 481–490. https://doi.org/10.1145/2642918.2647390

[55] Dongwook Yoon, Nicholas Chen, Bernie Randles, Amy Cheatle, Corinna E. Löckenhoff, Steven J. Jackson, Abigail Sellen, and François Guimbretière. 2016. RichReview++: Deployment of a Collaborative Multi-Modal Annotation System for Instructor Feedback and Peer Discussion. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work and Social Computing* (San Francisco, California, USA) *(CSCW '16)*. Association for Computing Machinery, New York, NY, USA, 195–205. https://doi.org/10.1145/2818048.2819951

[56] Amy X. Zhang and Justin Cranshaw. 2018. Making Sense of Group Chat through Collaborative Tagging and Summarization. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 196 (Nov. 2018), 27 pages. https://doi.org/10.1145/3274465

[57] Sacha Zyto, David Karger, Mark Ackerman, and Sanjoy Mahajan. 2012. Successful Classroom Deployment of a Social Document Annotation System. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 1883–1892. https://doi.org/10.1145/2207676.2208326