Designing Intelligent Systems to Support Workspace Collaboration

by

Ehsan Jahangirzadeh Soure

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Computer Science

Waterloo, Ontario, Canada, 2023

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of contributions

All the research outputs result from collaborative efforts, as no creative work belongs to an individual. Here, I will clarify my contribution to the published works.

Chapter 3: CoUX: Collaborative Visual Analysis of Think-Aloud Usability Test Videos for Digital Interfaces

> Ehsan Jahangirzadeh Soure*, Emily Kuang*, Mingming Fan, Jian Zhao. CoUX: Collaborative Visual Analysis of Think-Aloud Usability Test Videos for Digital Interfaces. IEEE Transactions on Visualization and Computer Graphics (Proceedings of VIS'21), 28(1), pp. 643-653, 2022.

The initial idea came from my supervisor Dr. Jian Zhao and our collaborator Dr. Mingming Fan. We then brainstormed on the system's design and consolidated the final design decisions through a formative study, which was contributed equally by Emily Kuang and me. I contributed to all the implementation, system design work, and study design & analysis. The results were then consolidated by my co-author through the second study. I mainly contributed to the paper writing of the sections that I had more contribution and the majority of the figures in the paper. Emily Kuang and I have contributed equally to this paper, so I used the parts that I contributed significantly in this thesis.

Chapter 4: IVA: An Empirical Investigation and Design of Intelligent Virtual Assistant for Online Meetings

> Ehsan Jahangirzadeh Soure, Ali Neshati, Hanyu Xu, Jian Zhao. IVA: An Empirical Investigation and Design of Intelligent Virtual Assistant for Online Meetings. To be submitted.

The initial idea resulted from numerous brainstorming sessions with Dr. Ali Neshati and my supervisor Dr. Jian Zhao. I contributed to the designing and majority of the tool's implementation, alongside conducting half of the formative study and all of the second study. Dr. Ali Neshati contributed to both study analysis and paper writing. Hanyu Xu contributed to the system backend implementation. My supervisor edited and revised the initial drafts written by Dr. Ali Neshati and me before submission. I was the major contributor to this work's design, implementation, studies, and writing.

---

*These authors contributed equally to this work

Abstract

Complex problems and interprofessional work require more resources to be involved, which has been possible through collaboration. Collaborative work is evolving from physical collaboration to more virtual forms through digital media over time. The evolution of collaboration is not just about how and where it is conducted. With the improvements in artificial intelligence, computers are moving towards playing a collaborative role instead of just supporting human-human collaboration. Empowering collaboration with computer systems can result in effective collaboration. In this thesis, we design intelligent systems to elevate the workspace collaboration in two different dimensions, time and space.

We introduce a collaborative visual analytics tool, CoUX, to facilitate UX evaluators collectively reviewing and discussing think-aloud usability test videos of digital interfaces. We designed CoUX based on a formative study with two UX experts and insights from the literature. CoUX enables collaboration amongst UX evaluators for logging, commenting, and consolidating the discovered problems with a chatbox-like user interface in order to reduce bias and errors. Furthermore, CoUX is the first system in this thesis to integrate synchronous and asynchronous collaboration in the think-aloud review workflow.

We implement IVA, an intelligent virtual assistant that automatically provides tools and services to meeting attendees based on their conversations to facilitate collaboration and multitasking amongst them. To design IVA, we conducted a formative study to investigate users' current practices and concerns for online meetings and how an AI assistant can facilitate collaboration and multitasking. IVA features a set of components, such as a customizable infinite environment where users can add and manipulate any applications besides the main meeting session. To validate our assumptions and design, we evaluated our design by mocking collaborative meetings with our participants using a Wizard-of-Oz implementation. In general, participants found IVA less distracting and more efficient compared to existing online meeting platforms.

Acknowledgements

I thank my supervisor, Professor Jian Zhao, for his support, patience, and guidance throughout my research. I also thank Professors Daniel Vogel and Mark Hancock, who served as readers for this thesis.

Dedication

I dedicate this thesis to my family.

Table of Contents

Author's Declaration **ii**

Statement of contributions **iii**

Abstract **iv**

Acknowledgements **v**

Dedication **vi**

List of Figures **x**

List of Tables **xii**

1 Introduction **1**

1.1 Overview 1

1.2 Activities and Collaboration in User Experience Evaluation 2

1.3 Activities and Collaboration in General Workplace Meetings 3

2 Background & Related Work **5**

2.1 Overview 6

2.2 Collaboration 6

2.3 Usability Test Analysis Tools 7

2.4 AI-Assisted UX Data Analysis 8
2.5 Collaborative Visual Analysis Tools 8
2.6 Tools for Virtual Meetings 9
2.7 Studies about Smart Assistants 10
2.8 Multitasking and Collaboration in Meetings 11

3 CoUX: Collaborative Visual Analysis of Think-Aloud Usability Test Videos for Digital Interfaces 12

3.1 Design of CoUX 14
3.1.1 Design Considerations 15
3.2 CoUX System 16
3.2.1 System Overview 16
3.2.2 Video Analysis and User Feature Extraction 17
3.2.3 CoUX User Interface 19
3.3 User Study 22
3.3.1 Participants and Apparatus 22
3.3.2 Study Videos 23
3.3.3 Task and Design 23
3.3.4 Procedure 24
3.4 Results 24
3.4.1 General User Experience 25
3.4.2 Individual Analysis (RQ1) 26
3.4.3 Collaborative Analysis (RQ2) 27
3.4.4 Session Review Strategies 28
3.4.5 Challenges (RQ3) 30

4 IVA: An Empirical Investigation and Design of Intelligent Virtual Assistant for Online Meetings **32**

4.1 Study I: Investigating the Practices and Needs for Virtual Meetings 34

4.1.1 Interview Design 34

4.1.2 Participants 35

4.1.3 Procedure 35

4.1.4 Results 37

4.2 IVA System 40

4.2.1 System Overview 40

4.2.2 Implemented Components for Tools and Services 42

4.2.3 Infinite Canvas for Collaboration and Multitasking 44

4.2.4 Smart Assistant for Component Recommendations 44

4.3 Study II: Understanding the Effectiveness and Values of IVA 47

4.3.1 Experimental Design 47

4.3.2 Participants 48

4.3.3 Procedure 49

4.3.4 Results 49

4.4 Discussion 54

4.4.1 Overall Findings 54

4.4.2 Take Away Lessons 55

4.4.3 Limitations and Future Work 55

5 Conclusion **59**

References **60**

List of Figures

3.1 CoUX is a collaborative visual analytics tool to support multiple UX evaluators with analyzing think-aloud usability test recordings. From an input video, a video analysis engine extracts various types of features, which are stored on a back-end and presented on a front-end visual interface to facilitate the identification of usability problems among UX evaluators. Moreover, the front-end, consisting of three interactively coordinated panels, communicates with the back-end to support individual problem logging and annotation as well as collaboration amongst a team of UX evaluators. 13

3.2 The CoUX user interface, showing a realistic study session of two UX evaluators (see section 3.3) analyzing a think-aloud video recording of a food delivery mobile app: (A) a Video Player for viewing the video; (B) a Feature Panel for displaying various extracted features to assist the analysis; and (C) a Problem Panel for logging discovered usability problems and discussion. (D) Problem annotation via a dropdown for common heuristics tags (e.g., Nielsen heuristics [74] and Norman principles [79]) and a slider for problem severity rating [75]. (E) A popup panel for adding custom tags. 19

3.3 CoUX supports the collaboration among UX evaluators via a chat thread design: (A) merging two Annotation Cards, (B) merged results, (C) showing the conversation between a pair of UX evaluators, and (B) the Discussion Panel of the selected card. 21

3.4 Participants' questionnaire ratings (Likert 1-5) after the individual phase (Q1-14) and after the collaboration phase (Q15-Q20). 25

3.5 The think-aloud video time visualized over the individual session timeline. The blue dots represent annotations & comments. 29

3.6 The think-aloud video time visualized over the collaborative session timeline. The blue dots represent annotations & comments. The orange dots represent merging two annotations. 31

4.1 IVA is an intelligent virtual assistant helping meeting attendees perform multitasking and collaborate during online meetings. The system provides an Infinite canvas and a set of interactions and components as an environment for multitasking and collaboration. All actions in the system are synchronized through the back-end. The WOZ administrative panel simulates the recommendation of components quickly and accurately. 33

4.2 Window arrangement for three of participants in response to ”Q8.2: Assume you have your online meeting, note taking, drawing applications and your browser. How would you arrange these four windows on one display? Why?”. The Zoom window is minimized to get more screen real estate for other tasks. 36

4.3 The IVA user interface shows all the different modules of the system: (A) the component toolbar button, which opens the toolbar shown with the label C, (B) the Infinite Canvas supports zooming and panning interactions for better multitasking and collaboration, (C) the Component Toolbar has a set of tools mainly used for multitasking, and (D) the Assistant IVA, showing temporarily the shipped Suggested Islands and groups of components in them. 41

4.4 The WOZ administrative panel user interface includes: (A) the Quick Suggestion List to help administrator provide predefined recommendations as quick as a computer would, (B) the JSON Editor, which lets administrators customize their suggestions and ship them with prefilled information, (C) the Component Toolbar hosts all the components shown as buttons with three different states: inactive, active, and selected, easily switchable, and (D) the component shown to user after clicking on the recommendation. 46

4.5 The final canvases after two of the studies. Each color represents one type of component used on the canvas, and some components are used several times. . . 49

4.6 Participants’ questionnaire ratings on a 5-point Likert scale (1 = “strongly disagree” and 5 = “strongly agree”). 50

List of Tables

3.1 Usage statistics of various functions in CoUX. 25

4.1 Formative study interview questions. 57
4.2 Evaluation study questionnaire and interview questions. 58

Chapter 1

Introduction

1.1 Overview

Collaboration is the ”mutual engagement of participants in a coordinated effort to solve a problem together.” Human beings have been collaborating to solve complex problems for a long time. This collaboration has evolved over time into new forms, such as e-collaboration, which encompasses various digital means of collaboration, such as telephones, teleconferencing suits, and computers [56]. Computers have been a tremendous medium for collaboration, both in facilitating collaboration between groups and as a collaborator itself.

Previous work has studied the impact of collaboration in many different domains, like health, learning and education, and research [37]. Generally, when a task is completed collaboratively, all parties achieve better results than if completed individually. For example, in health care, collaboration will lead to better results [115] as a result of a superior mix of skills and greater creativity [12]. The benefits of collaboration have also been studied in business, where collaboration leads to sharing costs and benefits, resulting in reduced risk [85, 118]. Research is another venue that thrives on collaboration where complex and interdisciplinary issues are to be solved. Laal et al. investigate the benefits of collaborative learning in their paper [60]. Although collaboration has been advantageous in numerous domains, there are many domains where it has yet to be applied or has yet to be supported by computers to meet the needs of recent years.

Ongoing improvements and advancements in artificial intelligence (AI) create an excellent opportunity to integrate AI-based support into different domains, like health care, where it can improve accuracy and decision-making. Wang et al., for example, looked at how AutoAI fits into the data science workflow and how professionals feel about working with AI [120]. On the other hand, AI could also be applied to cancer detection tasks, helping physicians, especially less experienced individuals, achieve better results than a physician or AI making the decision alone [115]. There is no constraint on where AI can be beneficial; Rezwana et al. introduce a framework for human-AI co-creativity, explaining the design space for collaboration on creating products and art [88]. However, it is essential to note that for a well-constructed computer supported collaboration it is necessary to integrate the AI support in day-to-day workflows.

This thesis focuses on augmenting workspace collaboration through the integration of intelligent systems in order to improve and empower collaboration. Situated in the domain of UX evaluation, we first propose a system to facilitate human-human collaboration in reviewing think-aloud videos. We tackle the evaluator effect by letting multiple evaluators synchronously and asynchronously collaborate, discuss, and consolidate their results [105]. Afterward, we investigate and design an intelligent virtual assistant (IVA) for general meetings, where collaboration and multitasking happen naturally, to play as a moderator and help meeting attendees manage their meetings. The rationale behind both systems will be introduced in the following sections.

1.2 Activities and Collaboration in User Experience Evaluation

Digital products have become increasingly feature-rich and often require users to navigate through an ever-growing number of onscreen elements, such as pressing a sequence of buttons to place an order on a smartphone. The increasing complexity of digital interfaces makes it challenging to achieve compelling user experience (UX). UX professionals often need to work collaboratively to identify and resolve UX problems via in-depth user evaluations. Of many evaluation approaches, usability testing with *think-aloud protocol* is widely used [26, 68] and considered as the single most useful method [76]. When using think-aloud protocols, participants verbalize their thoughts while performing actions. This allows UX evaluators to gain insights into their thought processes that is inaccessible to mere observations [61].

Despite being useful, analyzing recorded think-aloud videos is tedious, challenging, and time-consuming [16, 26, 30, 78]. First, UX evaluators need to make decisions by attending to multiple behavioral signals in both the visual and audio channels and conducting multiple tasks simultaneously in a fast pace [16], such as observing participants' actions, listening to their verbalized thoughts, inferring usability problems, and taking notes. Moreover, to increase the reliability and completeness of the analysis, UX evaluators are recommended to work collaboratively [30, 31] to overcome the *evaluator effect* [45]—the fact that different UX evaluators may uncover or interpret usability problems differently. Unfortunately, fewer than 30% of UX evaluators have a chance to collaboratively analyze the same usability test session due to practical constraints (e.g., limited company resources [16, 30]).

1.3 Activities and Collaboration in General Workplace Meetings

Online remote meetings have become one of the most common components of people's routines due to the global COVID-19 pandemic, and the number of people using video-communication tools (e.g., Zoom, Google Meet, and MS Teams) is increasing significantly [1]. For example, the number of daily Zoom meetings increased from 10 million in Dec 2019 to 350 million in July 2022 [23]. In addition, many companies and educational institutes, including universities, are planning to continue their remote workflow, even after the global pandemic.

People use Video-conferencing tools for different purposes, including meeting for tutorials and learning [18], work (e.g., managers and employees), conducting research [107], collaborative writing [80], and entertainment [40]. Prior work showed that doing different tasks and activities during an online meeting is very common and can potentially increase people's productivity during online meetings, especially in a remote work environment [14]. Previous research showed that people tend to use a wide range of tools, such as emailing [14], searching [110], and note-taking applications [66] during their online meetings.

Although multitasking can increase individuals' productivity, its effectiveness depends on how the user executes it. A poor multitasking execution can negatively impact users' performance [14, 58]. For instance, if the user keeps minimizing a Zoom meeting to be able to take notes on a word document and then switch back to the Zoom meeting to see the presentation, this is a poor execution of multitasking. However, if the user has access to both the meeting and the document simultaneously (e.g., split screen), this would be a better multitasking execution. This means that for tasks such as note-taking, browsing, and searching (secondary task) during an online meeting (primary task), having an arrangement of the online meeting as well as the required applications might be an advantage. This may also increase the performance and grab the user's attention to the main content in the meeting.

Besides multitasking, collaboration is another key factor in online meetings. For example, classmates developing a prototype for a team project, researchers working on writing a research paper [108], UX designers conducting brainstorming sessions, and manager/employee(s) meetings to discuss the project task assignment, are just a few common examples of collaboration between attendees of virtual meetings. It is common in virtual meetings that one of the attendees acts as the meeting moderator [95]. One of the primary responsibilities of the moderator is to facilitate the process of collaboration [3, 95].

As an example, the moderator starts creating the shared documents (e.g., google Docs, overleaf link, and Miro board) and shares the necessary documents (e.g., pdf files) among other attendees in a meeting. Depending on the context of the meeting, other typical responsibilities of a moderator are scheduling dates (e.g., following meetings and task deadlines), assigning tasks to others, tracking the meeting/presentation time, recording the meeting, monitoring the chat-box, and searching, finding and sharing materials with others (e.g., files, or links). Such amount of responsibilities and effort during an online meeting can be annoying, bothersome, and tedious for the moderator and other attendees.

Chapter 2

Background & Related Work

Preface - *In Chapter 1, we discussed the evolution of collaboration and the challenges of integrating it into daily workflows. An overview of essential topics in the collaboration field will help us better understand these challenges. In this chapter, we provide the relevant technical background and discuss the previous work in this area. To avoid repetition, the related work sections of both chapters 3 & 4 are moved to this chapter.*

2.1 Overview

Our work is inspired and informed by related work in three areas: usability testing analysis, machine learning for user experience research, collaborative visual analysis, virtual meetings tools, and interactions, AI assistants, and multitasking & collaboration.

2.2 Collaboration

Collaboration has been adopted in numerous domains like health care, business, education, and research. Collaboration happens when a group works together towards the same goal. There are no boundaries on where and how this act can happen, which is why it is evolving and covering the digital and virtual worlds [56]. Various studies have been conducted on integrating collaborative workflows in different domains, leading to computer-supported cooperative work [11]. Previous work investigated the benefits and shortcomings of collaboration in health care [83], specifically in cancer recognition [115], bioinformatics research [94], and animal health [62]. Multidisciplinary research collaboration is also studied extensively in past work. Nomaler et al., in their paper, investigate the possible impact of distant collaboration on the paper citation [131]. Another work by Harsanyi et al. covers credit allocation for multi-author works [43]. Collaboration has shown promising results in venues like education [41, 60], business [5, 35], and data science & analytical tasks [120, 125]. Informed and inspired by the previous work, we designed and implemented CoUX, a collaborative tool for think-aloud analysis, facilitating UX evaluators' collaboration [105].

A more recent form of collaboration forms when AI plays as a collaborator to guide humans in different contexts to make better decisions, perform tasks more accurately, and increase efficiency. The research on this part is also similar to the previous part, where researchers apply the concept to various domains to measure the impact and Discover the design space. Health care is one of the top domains that benefit from AI collaboration. Park et al. discover the challenges and opportunities in adding collaborative AI to health care [83]. Similar work focused on the human side of this collaboration and what they need from AI in order to communicate and collaborate effectively [12]. A segment of previous work has been shaped around meetings and teamwork, where AI can help handle cumbersome tasks and lead discussions to achieve greater efficiency [109, 112, 116]. AI can move further and help humans create products, a form of human-AI partnership [55, 88]. For example, letting people collaborate in their neighborhoods by creating street art is one good example of humans and AI co-creating their environments [63]. We applied all the information from the previous works and our formative study to create an integrated meeting system hosting an intelligent virtual assistant to improve collaboration and multitasking during meetings.

2.3 Usability Test Analysis Tools

Numerous commercial tools have been developed to support UX evaluators with conducting usability test and reviewing test session data. The first category is *offline tools* that need to be installed on local machines, such as Morae [113], Noldus Viso [77], and Silverback [102]. These tools allow UX evaluators to review sessions with functionalities like note-taking and marking events on the video progress bar, on top of basic usability test support such as screen recording, survey administration, and results exporting. However, many offline applications have been retired due to the emerging trend of remote and online user testing platforms [72]. These *online platforms* allow for more flexible collaboration, such as UserTesting.com [117] and FullStory [32]. While these tools support a range of user testing and analysis functions, their data analysis capabilities are limited to session playback, note-taking, tagging, and mouse point clouds. In contrast, we design CoUX to meet the increasing demand for online, remote, and collaborative tools that support usability test session review with advanced analysis support.

In the research community, several prototypes have been developed to facilitate UX problem identification. Usability Problem Inspector [4] was designed for UX evaluators to inspect a test session on the fly and was shown to be effective at helping evaluators find important usability problems in an interface design. However, to better understand the user's behavior and interactions, UX evaluators often have to repeatedly review the usability test recording to pinpoint the problems. Skov and Stage conducted an empirical study of a conceptual tool to demonstrate its usefulness for problem identification with a group of usability evaluators [103]. VA2 [8] supports evaluation session analysis by combining multiple sources of information including interaction logs, think-aloud speech, and eye-tracking data. However, unlike CoUX, collaborative features and online remote access are not explored. Several other visual analytics tools support better understanding of users' behaviors based on large interaction logs [22, 89]. However, none of them focus on reviewing think-aloud recordings.

In sum, the above tools primarily provide basic functions for analyzing the content of a test session, such as playback, note-taking, tagging, and some user interaction visualization (e.g., click heatmap), and offer limited collaborative features, such as sharing notes or clips. In contrast, CoUX adopts computational methods to extract rich features from the audio, transcript, and video content of the test session and visualizes these features as auxiliary information to better inform the analysis process. Additionally, CoUX considers the specific collaboration needs among UX evaluators such as discussing and resolving conflicts in detecting UX problems and rating the problem severity.

2.4 AI-Assisted UX Data Analysis

Recently, researchers began to leverage artificial intelligence (AI) to assess the usability of digital interfaces [82] and detect UX problems [38, 42, 54, 84]. For example, user interaction events were utilized to create machine learning (ML) classifiers to detect usability issues of websites [38, 84] and virtual reality applications [42]. In addition, user interaction paths were compared to construct graph-based AI models to detect potential UX problems [54]. Although these automatic methods were promising, they were primarily based on users' interaction logs, which only indirectly reflect some aspects of UX problems and lack a true understanding of the UX problems. In contrast, UX evaluators tend to use multi-modal information from both the acoustic and visual channels of a test session to pinpoint and interpret problems [16].

To address the limitations of AI, VisTA is equipped with AI as an assistant by detecting and highlighting video segments containing potential UX problems [27]. It extracts features such as negative segments and abnormal pitches, which are indicators of UX problems [25, 29]. We employ a similar philosophy to overcome the constraints of AI. We further extract the speech, textual, and visual features from think-aloud usability test recordings and present them to UX evaluators to assist with their analysis in CoUX. Moreover, we take a step further to extract additional features from the video such as scrolling speed.

Unlike VisTA that is designed to support a single UX evaluator, CoUX is able to support both individual analysis and collaboration among UX evaluators.

2.5 Collaborative Visual Analysis Tools

One critical challenge for UX problem detection is the vague evaluation procedures, which can lead to bias or unclear problem criteria [45]. Thus, different UX evaluators could detect different sets of problems when assessing the same interface, known as the *evaluator effect* [45]. Most evaluators perceive this effect when merging their individual findings with teams [46]. Thus, collaboration and involvement amongst UX evaluators are integral to both increasing the reliability [45] and improving completeness of the problems identified [100]. However, few systems have been developed to adequately support collaborative analysis of usability test sessions. When designing CoUX, we strive to support UX evaluators' collaboration for detecting problems, annotating or assessing problem severity using usability heuristics [75], and initiating discussion in one integrated environment.

Moreover, the design of CoUX draws on insights from both co-located (e.g., [52, 65]) and distributed (e.g., [44, 98, 119, 121, 124, 128]) collaborative visualization tools, while these tools do not focus on analyzing think-aloud sessions. In particular, we are inspired by prior work on the support of coordination and synthesis in collaborative analysis activities. Robinson explored the co-located synthesis of findings from paired participants after each had completed an asynchronous individual analysis phase [92]. They found that establishing common ground and role assignment are critical aspects of collaborative synthesis. Mahyar and Tory extended this concept to link common work within a visualization tool to support collaborative sensemaking of documents [65]. CoUX follows these principles by employing both an individual and a collaborative analysis modes, further with the ability to merge problem annotations and severity ratings, helping establish common ground.

Visualizing the analysis history is another strategy for coordination and synthesis, especially in asynchronous collaboration. Sarvghad et al. found that collaborative data analysis can benefit from displaying data dimension coverage of history [98, 99]. Similarly, KTGraph highlights of previously investigated data in a graph visualization to support collaboration [128]. CoUX supports coordination by showing previously annotated UX problems on a video timeline. Also, visual cues of segments of the video timeline are changed based on the state of the problems identified, such as in the uninitiated or in-progress phases.

Furthermore, allowing analysts to use tags and links to organize their comments and identify others' contributions improves final analytic results [124, 127]. Accordingly, CoUX enables user-generated comments and tags for identified problems to explicitly communicate the intent, uncertainty, and progress of their discussion via conversational threads.

2.6 Tools for Virtual Meetings

Meetings were an essential part of everyday workflows in most places, and it is no different when the work is remote, the online meetings play a significant role in synchronizing colleagues and tasks. Previous work investigated virtual meetings' possible positive or negative impacts in different domains. For example, Correia et al. evaluated the quality of education through video conferencing [19]. A branch of research in this field discusses the tools that emerged from the participants' needs during the meetings to shape future meeting platforms. Marlow et .al [66], in their paper, reports information about users sharing habits in meetings and reveal that participants in their study mentioned the need for an improved share screen experience, where two or more share screens can happen in parallel.

The tools for the virtual meeting are either used directly to run the meeting or as a tool used in conjunction with the meeting. Zoom, Microsoft Teams, Google Meet, Skype, and BlueJeans are some of the online meeting platforms. On the other hand, Google docs, Slack, Email clients, Miro, and Calendar are the tools used for note taking, whiteboard collaboration, and scheduling with any of the previously mentioned meeting platforms. Previous work investigated the use of accompanying tools in real-world scenarios like supporting collaboration in the science lab for students [93], group model building (GMB) workshops [123], and documentation [64]. Miro, email, chat application, Google docs, and Google Drive were just a few tools reported in the previous studies in conjunction with virtual meeting apps.

Moreover, previous work worked on habits like multitasking in online meetings and its possible negative or positive impact [14]. Previous work mentions external distractions as a reason for multitasking and suggests that meeting tools should be designed to promote attendees' active contribution. However, focus changing between tools can act against positive multitasking. We take a step further to conduct a formative study to understand users' practices, their challenges, and possible solutions for online meetings. Additionally, the designed system, IVA, supports most of the functionalities and tools discussed in previous work, and all different tools are integrated into the system as components with interfaces similar to users' day-to-day applications.

2.7 Studies about Smart Assistants

AI assistants have been designed and developed for different domains such as home usage, and education [114, 130]. They are widely used as voice assistants or agents in chatbots [13] mostly in a form of Human-AI collaboration [28]. However, in the context of meetings, they are mainly studied as robots being impactful during in-person meetings. Previous research has shown that assistants can positively contribute by providing topics to facilitate the discussion [47, 51, 57]. Hohenstein et al. also talk about adding AI assistants to Human-Human interactions instead of Human-assistant interactions [48]. Besides, previous work focused on the use cases of AI assistants in different domains and their potential impact to play a role beyond their everyday life use cases, like whether they can be effective in education and self regulated learning scenarios or not [97, 114].

On the other hand, a group of papers investigated possible features of assistants, like whether it is active or passive [73], assistant response length [39], built to be proactive [70]. As an example, Rienks et al. investigates the proper form that a an assistant can operate through a set of WOZ experiments [91]. McGregor et al. investigated challenges in adding speech-based technology as is to the group meetings, and reported the shortcomings of these systems for the group meeting setup [69]. However, the previous works have not covered the possibility of adding an assistants, specifically designed for meetings, to modern video conferencing platforms, which motivated our investigation of AI assistants in online meeting scenarios. Apart from that, previous assistants offer limited functionality, like proposing topics during meetings. With the advantages of WOZ implementation, IVA provides a wider range of actions to cover the tasks of a meeting moderator including tool recommendation, scheduling meetings, note taking, and task management. This serves as a vehicle for us to study users' perceptions and behaviors with smart assistants for meetings.

2.8 Multitasking and Collaboration in Meetings

With the Covid 19 pandemic, virtual meetings are considered an essential part of any remote work. In-meeting multitasking is a part of all virtual meetings, where attendees need to perform tasks simultaneously with their meeting [20, 96, 122]. It is also important to remember that meetings are naturally collaborative processes, so multitasking in the meetings impacts self and all other participants [50]. Marlow et al. have shown that to prevent attention drift in the main conversation, multitasking could be more acceptable on a single screen than having multiple screens [66]. Thus, if an online meeting system is correctly implemented, it can support positive multitasking in remote meetings [14] and provide different levels of engagement to its users [59]. However, little research has been done on designing and experimenting with intelligent systems that offer multitasking for online meetings.

On the other hand, meetings are generally collaborative. Previous work focused on making interfaces to help smoother interactions in different domains such as data analysis and knowledge work [7, 10]. Past studies aimed to improve and expand the available resources like screens [7], multiple computing surfaces [10] or simulate the real world experience in the digital world [2]. Numerous tools are built based on extending the screen to help users freely collaborate and use the space they desire. For example, Figma provides an infinite canvas and a collaborative environment for designers to work on design projects together, and Miro provides the same infinite canvas to let users collaborate and share ideas. Thus, it is evident that infinite canvas is being accepted as a collaborative environment. SAGE2 [67] is another tool implemented based on the positive influence of large displays to improve the collaboration between groups of users [111]. Therefore, IVA uses the same infinite canvas to facilitate collaboration through a familiar approach and as an excellent solution to extend the screen space on a limited monitor.

Chapter 3

CoUX: Collaborative Visual Analysis of Think-Aloud Usability Test Videos for Digital Interfaces

Preface - *Our first step towards applying collaboration to a different domain was to develop an integrated tool for think-aloud analysis sessions, where the results of these sessions were highly pruned to the evaluator effect. The solution to this problem is collaboration, and we took the first step to facilitate this collaboration through an integrated system synchronously and asynchronously. This chapter comprises the system design and implementation of the published paper.*

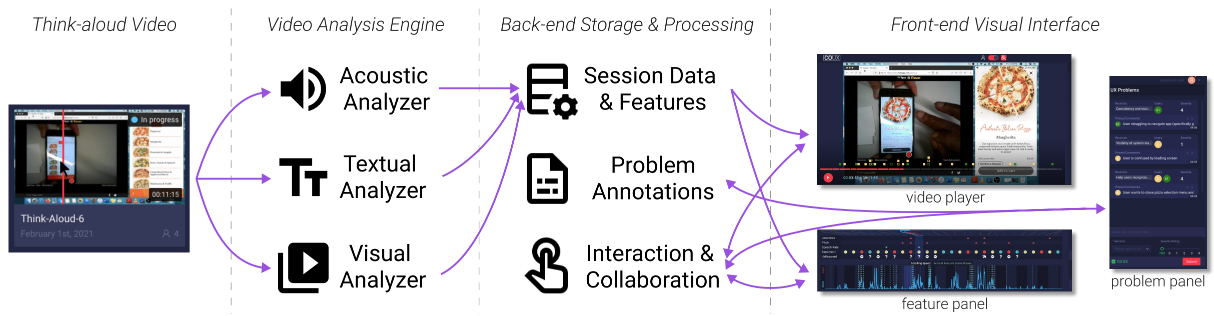


Figure 3.1: CoUX is a collaborative visual analytics tool to support multiple UX evaluators with analyzing think-aloud usability test recordings. From an input video, a video analysis engine extracts various types of features, which are stored on a back-end and presented on a front-end visual interface to facilitate the identification of usability problems among UX evaluators. Moreover, the front-end, consisting of three interactively coordinated panels, communicates with the back-end to support individual problem logging and annotation as well as collaboration amongst a team of UX evaluators.

As mentioned in Chapter 1, reviewing think-aloud sessions is a demanding task, and challenges arise when UX professionals must work collaboratively to reduce the evaluator effect. With the pandemic hit, synchronized collaboration is also challenging to conduct.

To mitigate these challenges, we propose a collaborative visual analytics tool, *CoUX*, to assist a team of UX evaluators with identifying, discussing and consolidating usability problems in think-aloud usability test videos for digital products. Our approach is partially inspired by recent studies extracting acoustic and textual features (e.g., loudness, pitches, and sentiment) from a video to help identify usability problems [24, 25, 27, 29]. We further leverage various machine learning techniques to detect acoustic and textual features directly from the audio (without manual transcripts), as well as user interactions (e.g., scrolling speed and scene breaks) from the video frames. To better support UX evaluators' decision making, CoUX segments a video into meaningful chunks based on the semantics exhibited in the think-aloud audio, extracts various visual, acoustic, and textual features, and visualizes the information collectively on multiple synchronized timelines. This design allows UX evaluators to easily attend to multiple streams of information likely indicating problems, to discover problems that might be otherwise overlooked, and make informed decisions about the occurrence and severity of the problems.

More importantly, CoUX is empowered with a collaborative decision support for discussing and consolidating analysis results among multiple UX evaluators. We draw on insights from studies of collaboration amongst UX evaluators and collaborative visualization (e.g., [30, 44, 119, 121, 124, 128]). CoUX allows UX evaluators to analyze a video independently, and then enter a collaborative mode to discuss and summarize their analyses, minimizing the evaluator effect [45]. In independent analysis, detected usability problems and their severity levels, as well as UX evaluators' reasoning, are automatically organized in a chat box like interface. During collaboration, UX evaluators are enabled with interactive and visual support from CoUX to make decisions collaboratively by discussing their findings in structured conversational threads and consolidating the results, synchronously or asynchronously.

Our design of CoUX is grounded in design considerations derived from the literature and our interviews with two UX professionals. For evaluation, we conducted a user study with six pairs of UX practitioners on collaborative think-aloud video analysis tasks. The results indicate that CoUX helped improve the completeness and reliability of their analyses with an effective support for discovering, discussing, and consolidating UX problems. CoUX allowed them to spot problems that they might otherwise have neglected, and encouraged focused conversations to seek clarification from and respond to their partners.

In summary, we make the following contributions: (1) a video analysis pipeline that extracts multiple acoustic, textual, and visual features from a think-aloud recording to facilitate UX problem identification; (2) a visual analytics tool, CoUX, that supports problem identification, annotation, and collaboration for UX evaluators in an integrated environment; (3) Insights into the results of a user study with six pairs of UX practitioners on collaborative think-aloud video analysis tasks.

3.1 Design of CoUX

Our main goal is to support UX evaluators in making decisions of usability problems and generating reliable annotations via collaboration. Towards this, we conducted 30-minute semi-structured interviews with two experts (E1 and E2) who are experienced in UX research. E1 is an assistant professor in information science at a university, whose research applies mainly qualitative methods. They complete the majority of their data analysis through Google Sheets [36]. E2 is a UX researcher at a start-up company with over four years of experience practicing UX. As part of his daily job, he uses Zoom [129] and Gong.io [34] to conduct and analyze user evaluation sessions. The goal of these interviews was to understand the current practices and challenges of UX evaluators in analyzing video-recorded usability test sessions and assess their needs for a new collaborative decision making and video analysis tool.

3.1.1 Design Considerations

Based on our interview findings and prior work, we derived the design considerations for CoUX.

D1: Leverage various information about the video to enhance the robustness of problem identification. Research has indicated that users tend to verbalize their thoughts with abnormal speech features (e.g., abnormal loudness, pitch, and speech rate) when they encounter problems [25, 29]; Their verbalizations also contain more negative sentiments, questions, and verbal fillers [25, 29]. Moreover, UX evaluators can identify more usability problems when these features are presented during analysis [27]. When discussing her video analysis strategies, E1 said: *"I do analyze the speech features but I don't have a good automatic tool to do so."* E2 also mentioned that he observes *"hesitation and pauses in users' speech"* to decide whether they encounter a usability problem. Furthermore, UX evaluators also correlate these verbalizations with the visual content of the recordings. In an international survey of UX professionals, 95% of them believed that the user's actions (e.g. scrolling on the interface, pressing the wrong button) were helpful in identifying usability problems [26]. CoUX supports these needs for determining UX problems by employing machine learning to automatically extract acoustic, textual, and visual features from the recording, which are then presented collectively on its interface.

D2: Provide an integrated environment for both video review and problem logging to ease the problem annotation. In addition to displaying useful information, it is critical to provide a seamless user interface for both video review and problem annotation. Previous studies have shown that UX evaluators often have to review recordings and take notes in separate applications, such as spreadsheets, text editors, and presentation tools [30]. This finding was echoed by E1 who usually stores all the videos in a separate folder while all the analysis and coding is done on a spreadsheet. As a result, she finds that *"organizing and sorting through the files has been tricky."* E2 experiences a similar problem as he reviews the videos on Zoom cloud recordings but keeps his annotations in a separate document. Using separate applications leads to difficulty when trying to pinpoint specific problems during discussions. E1 said that *"we don't have a way to solve timestamps so we just have to manually track it down and put it on a cell and then when we want to review it, we have to retreat to that specific segment in the video."* E2 mentioned *"sometimes the design lead wants to see exactly how the user reacted so I need an easier way to show her the snippet of the recording."* To address these challenges, CoUX provides an integrated environment with both video reviewing and problem logging functions, allowing UX evaluators to become more organized and efficient during usability test video analysis.

D3: Support collaboration between UX evaluators with both individual and collaborative modes. UX evaluators may have their own biases and limitations when analyzing usability problems, which is known as the "evaluator effect" [45]. Thus, it is important to have multiple evaluators collaborate with each other. Indeed, collaboration amongst evaluators has been found to enhance both the reliability [45] and thoroughness [100] of the problems identified. To serve this purpose, collaboration typically happens among two or more evaluators who first perform independent analysis of the same data [30, 45]. E1 stated that she and at least one other coder would annotate the same video individually by hiding the columns on a spreadsheet. E2 also described reviewing the video individually at first before sharing results with colleagues, which is in line with this best practices process. We aim to design CoUX by following this workflow with two modes: an individual mode for independent problem identification and a collaborative mode for problem merging, decision making, and discussion. This mitigates the confirmation bias since evaluators rely on their own judgment for initial assessments and decisions before seeing others' results.

D4: Allow for both synchronous and asynchronous communication between UX evaluators. Maintaining effective communication between UX evaluators is critical to achieve successful collaboration during the analysis of usability problems. Research has shown that the most frequent form of collaboration is short discussions at the outset of analysis [30]. This was reiterated by E1: *"after we finished coding, we'll highlight the disagreements and then during our meeting time we'll discuss and resolve those highlights."* E2 also mentioned that he discusses the results with the team in short meetings after the session. This type of synchronous communication should be supported by CoUX, e.g., with an instant messaging feature. Further, in the event that a synchronous meeting is not possible, which is not uncommon in practice, E1 and her collaborators would leave comments on the spreadsheet and tag the other person. Thus, asynchronous communication should also be supported to allow the messages to be viewed and discussed at a later time. Thus, we aim to adopt a similar workflow where UX evaluators can discuss and decide both synchronously and asynchronously using comments in a thread and consolidate their opinions using interactive visual support from CoUX.

3.2 CoUX System

3.2.1 System Overview

We developed the CoUX system based on the aforementioned design considerations. As shown in Figure 3.1, CoUX consists of a back-end storage & processing and a front-end visual interface, both of which require data extracted from a video analysis engine.

The video analysis engine contains three modules for extracting different types of features from the session recording, including the *Acoustic*, *Textual*, and *Visual* Analyzers (**D1**). The outputs of the video analysis engine are uploaded into the *Session Data & Features* storage hosted in the back-end. The back-end also contains the *Problem Annotations* and *Interaction & Collaboration* storage. The Problem Annotations storage saves all the inputs from UX evaluators regarding the usability problems, while the Interaction & Collaboration storage supports all the actions that the UX evaluators perform in the front-end.

The front-end is composed of three interactively coordinated views: the *Video Player*, *Feature Panel*, and *Problem Panel*. The Video Player allows UX evaluators to play, pause, and rewind the session recording, as well as view a timeline of their annotations above the video progress bar. The Feature Panel presents all the extracted features and highlights the ones that correspond to the current timestamp of the video. Lastly, the Problem Panel allows UX evaluators to enter descriptions of problems that they identified, the design heuristics or principles violated (e.g., Nielsen's heuristics [74], Norman's principles [79]), custom tags, and their severity ratings [75]. The interface includes a toggle for UX evaluators to switch between *individual* and *collaboration* modes (**D3**). In the individual mode, the Problem Panel displays the comments entered by a single UX evaluator. In the collaboration mode, the Problem Panel also displays the comments of other UX evaluators and allows for both synchronous and asynchronous communication through the chat functionality (**D4**). The three above views together are shown on the same CoUX interface, which provides UX evaluators an integrated environment for both video review and problem annotations (**D2**).

3.2.2 Video Analysis and User Feature Extraction

To assist UX evaluators with thorough identification of usability problems, CoUX analyzes think-aloud videos by segmenting them into small meaningful chunks and extracting various features related to the user in the video (**D1**). The video segments are automatically detected using the Auditok library [101] at periods of silence characterized by the lack of acoustic activity. By doing so, the entire long video is cut into small "bite-size" portions to facilitate UX evaluators' analysis, each of which may correspond to one or few usability problems. Each segment is then transcribed using the Google Speech Recognition API [126]. The audio, transcript, and video of the segments are used to extract three main categories of user features: acoustic, textual, and visual.

- **Pitch:** Users tend to change their pitch when they encounter a problem while thinking aloud [25, 29]. For the corresponding audio of each segment, we computed the frequency of the speech using the "sound to speech function" in the Praat-Parselmouth library [53]. Based on the mean and the standard deviation of the pitch over the entire session, a segment is categorized as containing abnormal pitch if at least 10% of the values are over two standard deviations away from the mean. Thus, it is given one of the three values: 1 for abnormally high, 0 for normal, and -1 for abnormally low.

- **Loudness:** Loudness has been shown as another useful speech feature for analyzing usability test sessions [17]. We utilized the "sound to intensity" function in Praat-Parselmouth [53] to extract the intensity of the sound (in dB). The detection of abnormalities and assigned values are the same as the pitch feature.
- **Speech Rate:** We computed the speech rate by dividing the number of words spoken in a segment by its duration, where the number of words was counted based on the transcript. Only abnormally slow speech is detected based on prior research showing that users slow down when encountering an issue [25, 29]. Thus, each segment is labelled 1 for abnormally slow or 0 for normal.
- **Negations:** Negations in users' think-aloud verbalizations may indicate that they encounter a usability problem [25, 29]. To determine if users said a negation, we applied a keyword-matching to the transcripts to detect the following words: *no*, *not*, *don't*, *doesn't*, *didn't*, *can't* and *never* [24, 25].
- **Questions:** Questions are another type of indicator for usability problems, indicating a user may be in doubt. Similar to negations, we utilized a keyword-matching algorithm containing the following words: *what*, *which*, *why*, *how*, and *where* [24, 25, 29].
- **Verbal Fillers:** Verbal fillers indicate hesitations in the user's speech, which may suggest a problem. We utilized a keyword-matching algorithm containing the words: *um*, *uh*, and *like* [24, 25, 29].
- **Sentiment:** The sentiment of a user's speech (e.g., positive, neutral, or negative) is another source of useful information for problem identification [25, 29, 33, 106]. We used the Valence Aware Dictionary and Sentiment Reasoner library [49] to detect the sentiment based on the transcripts for each video segment. Based on the compound score (between -1 and 1), a segment is labelled as positive ($(0.2, 1]$), negative ($[-1, -0.2)$), or neutral ($[-0.2, 0.2]$).
- **Scrolling Speed:** When using a digital product, the amount of scrolling may reflect a user's confusion. For example, frequently scrolling back and forth on a webpage could indicate that a user has difficulty in understanding the interface [6]. Thus, we extracted the scrolling speed (in the amount of pixel movement per second) for each segment using the dense optical flow algorithm from OpenCV [81], resulting in a continuous time-series.
- **Scene Break:** Frequent switching of views may also indicate that the user has difficulty locating the desired item on a digital interface [6, 38]. We used the OpenCV-based video scene detection library [15], which performs a comparison of sequential frames in a video and detects substantial changes in content. This results in a series of timestamps of these scene breaks.

These features are meant to provide extra information to help UX evaluators review think-aloud sessions of digital products and make decisions regarding usability problems. The features are selected based on our interviews and the literature as mentioned above. However, it remains an open question of whether this feature set is complete.

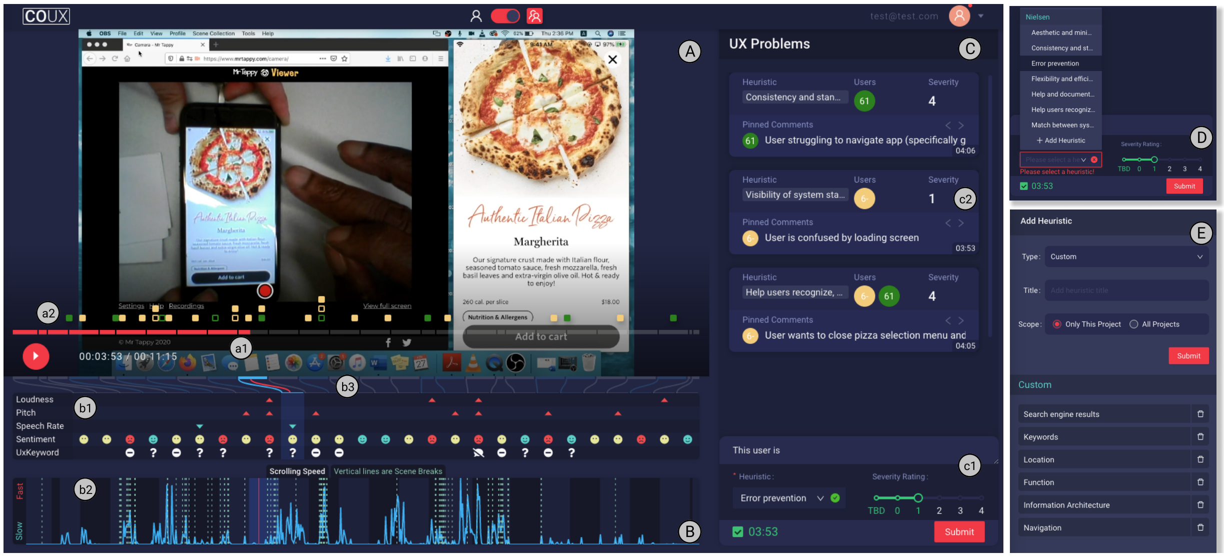


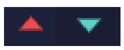
Figure 3.2: The CoUX user interface, showing a realistic study session of two UX evaluators (see section 3.3) analyzing a think-aloud video recording of a food delivery mobile app: (A) a Video Player for viewing the video; (B) a Feature Panel for displaying various extracted features to assist the analysis; and (C) a Problem Panel for logging discovered usability problems and discussion. (D) Problem annotation via a dropdown for common heuristics tags (e.g., Nielsen heuristics [74] and Norman principles [79]) and a slider for problem severity rating [75]. (E) A popup panel for adding custom tags.

3.2.3 CoUX User Interface

For better work organization, CoUX features a project management page showing all the videos that need to be analyzed upon logging into the system. Clicking on any video opens the main CoUX interface. This interface consists of three key components (Figure 3.2): (a) a Video Player for viewing the recorded think-aloud sessions, (b) a Feature Panel for displaying various extracted features based on the analysis in subsection 3.2.2, and (c) a Problem Panel for logging discovered usability problems and discussing them with other UX evaluators.

Problem Identification

Effectively identifying potential UX problems is the key objective of reviewing a think-aloud video. On the left, CoUX comprises all necessary elements for problem identification based on various information extracted from the video (**D1**). First of all, an integrated video player (Figure 3.2-A) is provided to prevent any switching between different tools, which is the largest element on the screen to facilitate the video browsing. The player supports all regular functionalities like play, pause, forward, and rewind. Further, similar to the YouTube chaptered design, the player progress bar shows the automatically-generated segments  (Figure 3.2-a1) that split the video into “bite sizes” (see subsection 3.2.2).

Below the player, visualizations are placed on the Feature Panel (Figure 3.2-B) to facilitate the use of all the extracted features while reviewing the video. CoUX distinguishes discrete and continuous features, and displays them on two sub-panels. First, discrete features (i.e., all the acoustic and textual features) are visualized in the Feature Matrix (Figure 3.2-b1), where rows indicate the features and columns represent the video segments. All values in the matrix are shown as icons and colors instead of text to allow UX evaluators to quickly scan and recognize the feature values that could signify a problem. For example,  represent neutral, negative, and positive sentiments;  represent filler words (e.g., um, uh), negations, and questions; and  represent high and low anomalies. Second, continuous features (i.e., the visual features) are shown in a Feature Chart (Figure 3.2-b2), where the scrolling speed is implemented as a line chart and the scene breaks are represented as vertical green lines.

These features serve as auxiliary data for the video to enhance the thoroughness of problem identification by UX evaluators. While the video is playing, CoUX dynamically highlights the corresponding segments in both the Feature Matrix and Feature Chart, with a lighter blue background. In addition, a red vertical line representing the playhead moves on the Feature Chart while the video is playing. In contrast, the column width of the Feature Matrix does not reflect the time length of each segment (instead, a fixed width). Thus, a Sankey visualization [90] (Figure 3.2-b3) is placed between the player progress bar and the matrix to indicate the correspondence. Similarly, a red curve  is shown on the Sankey to indicate the playhead. This design increases the readability and scalability; if each column width of the Feature Matrix maps exactly to the segment length, some columns could be too narrow to display any readable features whereas others could be very wide, wasting the space. Lastly, all the above visualizations are clickable, which facilitates navigation to different parts of the video.

Problem Annotation

Once an evaluator identifies a UX problem, they can log the problem with the Problem Panel (Figure 3.2-C), integrated seamlessly within CoUX (**D2**). When an evaluator starts to type in the chatbox-like interface at the bottom of the panel (Figure 3.2-c1), the video automatically pauses so that they do not need to manually click the video controls. Annotations can be bound to video playtime by checking the time check box 

. Evaluators can add comments or descriptions

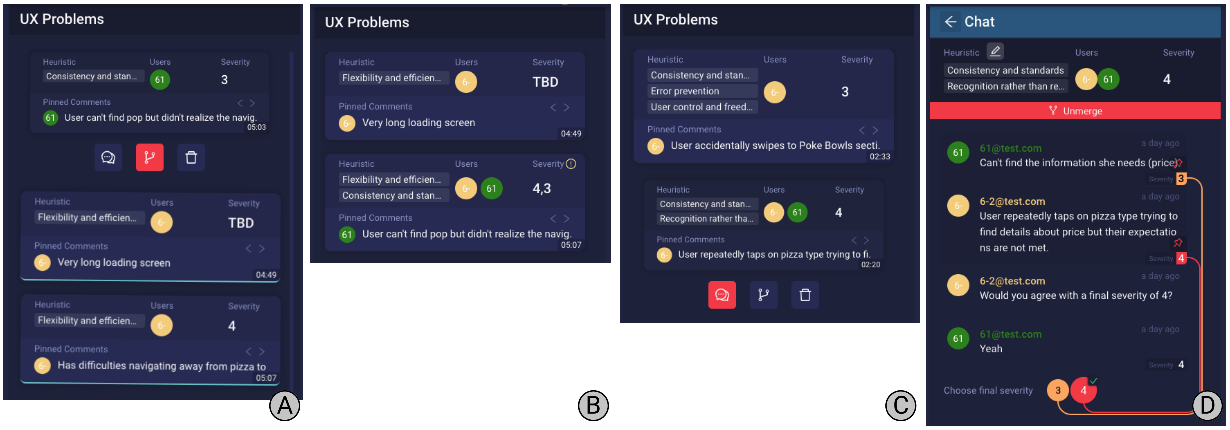


Figure 3.3: CoUX supports the collaboration among UX evaluators via a chat thread design: (A) merging two Annotation Cards, (B) merged results, (C) showing the conversation between a pair of UX evaluators, and (B) the Discussion Panel of the selected card.

for the identified problem, and select predefined heuristic tags from a grouped dropdown list and a severity level (0–4, where 4 indicates the highest severity) [75] with a slider (Figure 3.2-D). CoUX supports common tags including Nielsen's heuristics [74] and Norman's principles [79]. Moreover, evaluators can add their custom tags via a popup panel (Figure 3.2-E). These tags can be created within custom groups and set to either applicable to a specific video or all videos in a project.

After an annotation is submitted, CoUX adds an Annotation Card (Figure 3.2-c2) to the Problem Panel, which displays all the cards bound to the active video segment. Each Annotation Card shows the problem tags, severity, comments/descriptions, and corresponding evaluators. Moreover, the Annotation Timeline (Figure 3.2-a2) updates with a new solid Annotation Square pinned onto the video progress bar, which shows an overview of all problems with colors indicating their creators.

Problem Discussion and Collaboration

CoUX supports an individual mode and a collaborative mode to mitigate the "evaluator effect" (**D3**). In the individual mode, an evaluator can oversee their own Annotation Cards and Squares. When switching to the collaborative mode with the mode button on the top, evaluators can navigate to each other's identified problems by simply clicking the corresponding elements and start a discussion to consolidate their annotations. Evaluators can still create new problem annotations in this collaborative mode. The discussion/consolidation is moderated via chat threads, similar to Slack [104], to support both synchronous and asynchronous collaboration (**D4**).

As it is possible that evaluators created different annotations about the same underlying problem during the individual mode, CoUX allows them to merge the Annotation Cards. To do so, an evaluator first clicks a card and then three buttons pop up: discuss, merge, and delete (Figure 3.3-A). When the "merge" button is clicked, a shaking animation highlights mergeable cards. In addition to a merged Annotation Card (Figure 3.3-B), a new Annotation Square is added on the Annotation Timeline while the previous squares become hollow. Currently, merging is only allowed for problems in the same video segment, but more than two cards can be merged.

Moreover, clicking the "discuss" button enters an in-situ Discussion Panel (Figure 3.3-D) of this Annotation Card (Figure 3.3-C). They can then add new comments and propose a different severity rating for the problem, or discuss the merging if applicable. New and existing comments are displayed based on their timestamps. Evaluators can also pin important comments. A thread visualization helps evaluators review all the proposed severity ratings (Figure 3.3-D). If an annotation has a conflict in the severity rating (i.e., more than one severity ratings are proposed), evaluators are asked to determine the final severity for the annotation; otherwise, this problem remains unresolved, with a warning icon Severity associated with the Annotation Card. Evaluators can also add or remove heuristics by clicking on the edit button Heuristic on the top of the panel. These Annotation Cards on the Problem Panel provide an informative summary about a problem. Each card shows all the tags, severity ratings, participating evaluators, and pinned comments in a carousel view (Figure 3.3-C). For merged cards, evaluators can also unmerge them through a button Unmerge.

3.3 User Study

We conducted a user study to assess the usefulness and effectiveness of CoUX in think-aloud video analysis. Specifically, our exploration was guided by: **RQ1** - How does CoUX support evaluators in analyzing think-aloud sessions? **RQ2** - How do teams work together and communicate during their analysis through CoUX? **RQ3** - What are the general challenges in collaborative UX video analysis?

3.3.1 Participants and Apparatus

We recruited 12 participants (two males, nine females, and one not disclosed, aged 23–32) via social media and mailing lists. They were UX designers ($N = 4$), UX researchers ($N = 4$), and

UX/HCI graduate students ($N = 4$). On average, they had three years of experience in UX ($SD =$ 2.2). Eleven (91.7%) self-reported being very familiar or extremely familiar with identifying usability problems, with one participant being moderately familiar ($M = 4.17, SD = 0.55$). The participants were recruited in pairs. They had all collaborated with their partners before on at least one project. Seven (58.3%) were very or extremely familiar with their partner, with the rest being moderately familiar ($M = 3.83, SD = 0.80$).

Participants completed the study remotely with their own computers while communicating with the moderator through video-conferencing software. Participants were asked to make the application window full screen throughout the study. Participants used the largest screen available. The average display size was 20 inches ($SD = 7.12$).

3.3.2 Study Videos

We collected two recorded usability test sessions in which users were instructed to use digital products with the think-aloud protocol. In the practice video (length: 3 minutes 34 seconds), a user was asked to find a photo of an instruction manual for an early telescope on a Science and Technology Museum's website. In the study video (length: 11 minutes 15 seconds), a user was asked to complete three tasks on a Food Delivery Mobile App, including: (1) find the Wegmans store on Amherst St.; (2) buy 10 bottles of classic Coke and 10 bottles of Sprite, and some full sheet pizzas with any topping while staying under a budget of $100; and (3) change the pick up order to delivery instead. These videos were chosen since they are representative of digital interfaces: one for a desktop website and the other for a smartphone application. There were also numerous usability issues in both videos which promoted discussions between the participants and their partners.

3.3.3 Task and Design

Each pair of participants conducted the study together and was asked to review the study videos and identify usability problems using CoUX. There were two phases in the study session: (1) an *Individual* phase and (2) a *Collaborative* phase. In the individual phase, participants identified usability problems and submitted the annotations of these problems independently. In the CoUX interface, they could only see the problem cards that they had inputted. In the collaborative phase, the problem annotations of both partners were revealed to each other. Then, they were asked to review each other's annotations, merge cards as desired, and discuss the problems before reaching a final decision. Splitting the session into two phases was based on the recommendation that to serve the purpose of improving reliability, collaboration should happen among two or more usability practitioners who first perform independent analysis of the same dataset [30].

3.3.4 Procedure

To begin, each pair was given a short video tutorial about CoUX. Participants were able to ask any questions about the study and the system. They were then introduced to the usability test video review task, and instructed to assume that developers of the products will have limited time to address the problems identified in the session. This assumption resembled the fact that UX practitioners often have limited time to analyze test sessions [26, 78] and allowed for a more realistic evaluation of the extracted features and collaboration support in CoUX.

After the tutorial, the participants completed a practice trial by first analyzing the museum video individually for five minutes, then collaborating with their partner for another five minutes. This allowed them to become familiar with the system and the full procedure of the two-phase task. In the study session, participants were first asked to identify usability problems with the food delivery app individually for 25 minutes and then filled out a short survey based on the 5-point Likert Scale, which sought to understand the usefulness of each feature and the ease of use of the annotation functionality in CoUX. After a short break, they had 15 minutes for the *Collaboration* phase where they discussed each other's problems and tried to consolidate them into a final set. At the end, each pair of participants independently completed another short survey about their collaboration experiences. These survey questions were based on previous findings about collaborative analysis [45, 100]. When performing both the individual and collaborative tasks, participants were asked to communicate only within CoUX. This would allow them to fully explore and use CoUX during the study. We then conducted a semi-structured group interview to collect their feedback about the system. All the interview sessions were video-recorded, and participants' interactions with the system (e.g., clicks, video-playing behaviors) were logged. The study lasted about 90–100 minutes and participants received $25.

3.4 Results

We first present participants' general user experience of CoUX (Sec. 3.4.1) and then how they used the features during their individual analysis (Sec. 3.4.2) and collaboration (Sec. 3.4.3) respectively, based on our RQs. Participant *x* in the study pair *n* is labeled as *Pn-x*.

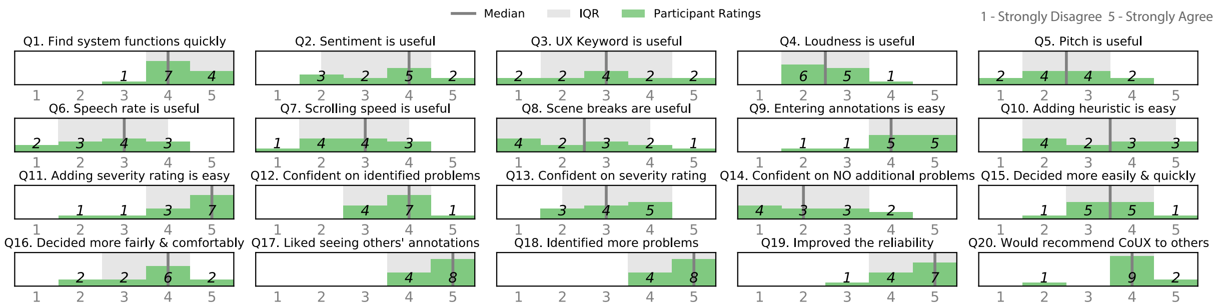


Figure 3.4: Participants' questionnaire ratings (Likert 1-5) after the individual phase (Q1-14) and after the collaboration phase (Q15-Q20).

Table 3.1: Usage statistics of various functions in CoUX.

Function Usage	**Mean (SD)**
Clicks on Feature Matrix (Fig. 3.2-b1)	7.6 (7.1)
Clicks on Feature Chart (Fig. 3.2-b2)	1.2 (2.2)
Number of problem annotations (Fig. 3.2-c1)	18.3 (7.0)
Number of problem merges (Fig. 3.3-A)	1.9 (1.7)
Number of comments per chat thread (Fig. 3.3-D)	2.8 (0.7)
Number of discussed problem annotations	6 (3.6)
Number of additional problems found after collaboration	4.2 (2.7)

3.4.1 General User Experience

Participants found CoUX to be a useful tool for analyzing think-aloud video recordings. They were able to easily find and use all of the functions, including extracted features, problem annotations, and chat threads (Table 3.1). They appreciated that CoUX integrated analytics, collaboration, and communication features in one environment and found it easier to use than previous tools they had used. Most of the participants agreed or strongly agreed to recommend CoUX to others (Figure 3.4).

"Usually we were using Google sheets [36] to coordinate and it was getting quite difficult, because we had to follow up with another person... it was messy and difficult but right now, it seems quite easy [with CoUX]."-P2-1.

3.4.2 Individual Analysis (RQ1)

Problem Identification: Feature Panel

The Feature Panel in CoUX was used by participants as hints or warnings to alert them to potential problems while reviewing a video, as anticipations of problems to help them skip ahead under time pressure, and as checks or anchors to identify areas to revisit in a second pass.

"When I heard... 10 minutes left, for the video that I haven't watched, I picked the more highlighted ones to directly find any problems so that's when I found those markers to be helpful."-P4-1

"The reason I looked at those is more as a hint or warning to see what is coming up. I paid more attention to that segment if there's a red face."-P4-2

Sentiment was rated as the most useful feature, as it was perceived to be accurate and easy to understand. UX keywords were rated as the second-most useful feature and were also perceived to be accurate. Speech features, including loudness, pitch, and speech rate, were rated as less reliable, but participants thought they could be improved with more context and explanation of how they were determined.

"I feel like this is providing useful insight and... the sentiment easily got me to the areas in the video where the user was confused."-P5-1

"The UX keywords matched up with what my impression was while watching the video."-P5-1

"The pitch was interesting, but I feel like I still have to listen to a combination of their tone and the context."-P6-2

Scrolling speed and scene breaks were appreciated by some participants as indicators of confusion or task changes, but were less used because they were difficult to see while watching the video. Some participants felt that the placement of the features on the screen was not ideal and that it was difficult to switch between features while watching the video.

"I actually almost paid all of my attention on the scrolling speed. Compared to the icons, I definitely prefer to look at the visualizations."-P1-2

"I think including this definitely good but... I would like to see the numbers actually I think it will be easier for me to read it"-P3-1

Problem Annotation: Problem Panel

The problem annotation function was particularly popular, with participants entering an average of 18.3 problems (Table 3.1). Participants also appreciated the ability to attach heuristics and severity ratings to each problem description. Some participants had mixed feelings about the requirement to attach a heuristic to every annotation, and suggested that this could be made optional.

"The chat box... is really useful because it's very clear and very easy to use."-P1-1

"I think the heuristics are great, but I don't think it should be mandatory. I usually make notes of activities and those aren't things that I would tie to a heuristic."-P6-2

Participants used the custom tagging function, adding tags that were relevant to their specific experiences and expertise. Overall, the integrated platform provided by CoUX was seen as helpful in assisting with problem annotation and analysis.

"like here there is the older adult and accessibility issues that are more specific than Nielsen's."-P2-1

"We can finish the analysis and make the comments in one screen instead of using lots of applications."-P2-2

3.4.3 Collaborative Analysis (RQ2)

Effects of Collaboration

Overall, collaborative session was perceived as beneficial by the participants. They felt that it helped them make decisions more easily, quickly, fairly, and comfortably. The collaboration also helped improve the completeness and reliability of their results by allowing them to identify more problems and explore different perspectives on the issues (Table 3.1).

"I really like the collaboration part, especially seeing my partner's notes was very helpful."-P3-2

"After I reviewed her opinion and I do think there is a problem, so this collaboration helped us find more problems."-P2-2

Additionally, the collaboration provided a sense of confidence in their ratings and helped ensure unbiased feedback and unique ideas. It appears that the collaborative session was successful in supporting collaboration between UX evaluators and improving the completeness and reliability of their analysis.

"It is good for reducing evaluator bias since we have two people review the same video."-P4-2

"We actually noticed that the same area in the video has problems, but we focused on different aspects of the problem."-P3-2

Usage of Collaboration Support

The Annotation Timeline is a tool that allows participants to view and discuss annotations of usability problems in a collaborative manner. Participants appreciated seeing the annotations of their partners and used the annotation timeline extensively to navigate to each annotation.

"I noticed that... the little blocks were flashing and that meant some changes happened."-P4-1

"In terms of the flashing squares, I would go through and check them for new changes."-P6-2

The merging function was used to have a focused conversation about a certain problem in one place. Participants used the chat threads to document their discussions, seek clarification on problem descriptions, consolidate heuristics and severity ratings, and prepare for video-call or in-person meetings. Participants suggested additional features such as the ability to see the revision history and e-mail notifications of new comments.

"At the three-minute timestamp, there were two cards where we were exactly talking about the same issue of the [user] tapping on 'My Cart' and it was not responsive, so I merged them together."-P6-1

"We can see the process of what we discussed and what we talked about so we won't forget what we say and review the video at the same time."-P1-1

"This is good for me to quickly see what my partner and I agreed on and then we can skip those in the meeting."-P4-2

3.4.4 Session Review Strategies

To understand how the participants reviewed the study video, we adopted a similar approach proposed by Fan et al. [27] to analyzed the pairs of timestamps *(SessionTime, VideoTime)*. The typical strategies were categorized by the number of passes on the video recording and their playback behaviors while going through the pass. Figure 3.5 shows the typical session review strategies.

In general, most participants adopted a *One-pass: Pause-then-Write* strategy. In this case, participants paused the video while entering the problem annotations. This was the most common since we implemented the automatic timestamp functionality where each annotation is associated with a specific timestamp. Thus, participants usually paused the video to type in their annotations so that it will be shown in the segment that contained the problem. Furthermore, the functionality for automatic pausing once participants allowed them to avoid moving their mouse back to the play/pause button. The "pause-and-write" strategy suggests that participants tended to note

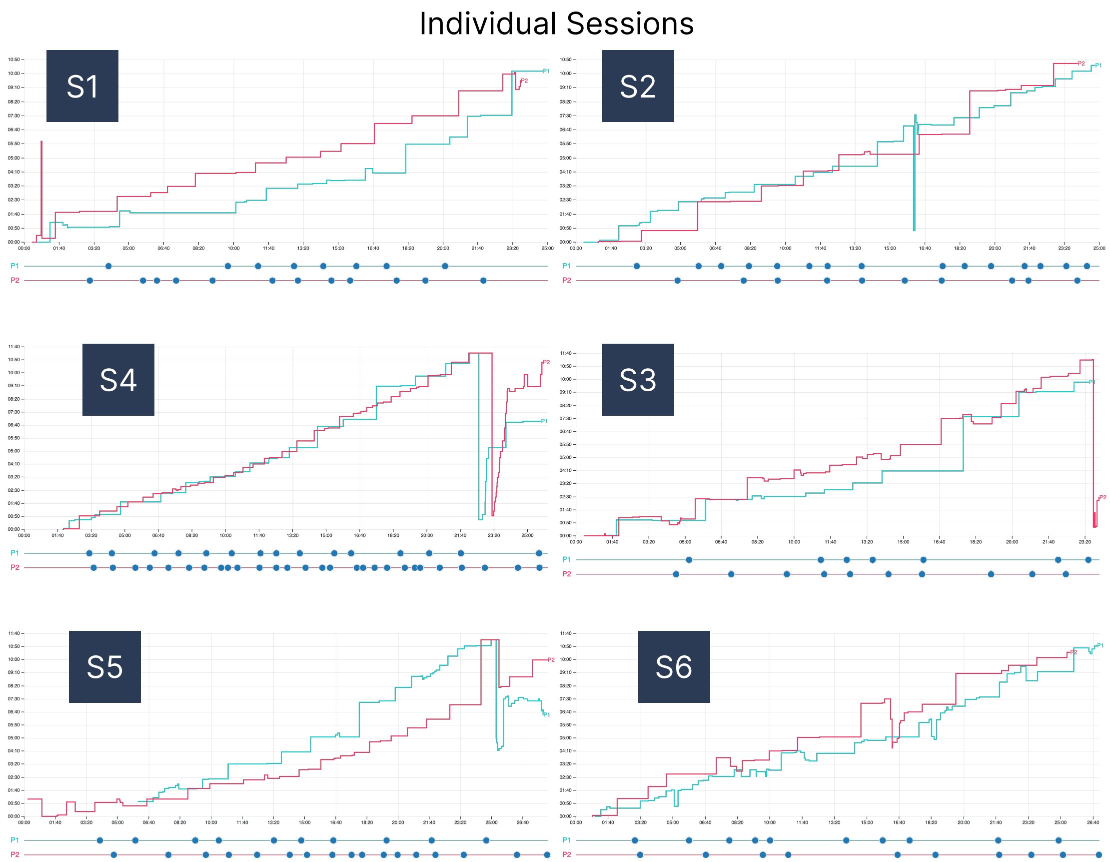


Figure 3.5: The think-aloud video time visualized over the individual session timeline. The blue dots represent annotations & comments.

down their thoughts immediately when they found hints of usability problems. The "one-pass" strategy was partially because some participants did not have enough time to go through the video the second time. Another strategy was the *Two-pass: Pause-then-Write* strategy. In this case, participants went through the entire video and wrote down their annotations. In their second pass, they clicked on their prior annotations using the Annotation Timeline (Fig. 3.2-a2) to add more notes to specific problem descriptions to make sure they didn't miss anything. *P6-2* mentioned that in real life, she would usually *"play it at two times the speed to watch it through once, and then go back to tag and everything"*.

We were interested in understanding how CoUX supports UX evaluators for problem identification and annotation in the individual phase, and how it helps them consolidate their findings in the collaborative phase.

3.4.5 Challenges (RQ3)

Participants highlighted potential challenges in collaborative think-aloud video analysis that could inform future research. These challenges included managing disagreements about the presence and severity of usability problems. Workspace awareness can be an issue when evaluators are working on different parts of the video simultaneously.

"[P4-1] just put she thought that this app is too much trouble, but from my understanding, it's because maybe she's not familiar with the iOS keyboard."-P4-2

"So we disagreed on what the user was trying to do but agreed it is the efficiency of use problem."-P1-1

"It feels like we're not on the same page, because when I work on the first card, she is probably working on the second card, so we cannot get the real-time feedback [on the same card]."-P1-1

Participants suggested that discussions in chat threads could lead to more robust analysis, but also mentioned that video or voice calls could be more efficient for resolving disagreements. In future research, the trade-offs between chat threads and conference calls for communication in collaborative interfaces should be further explored.

"I think communicating via the chat is totally fine, I would rather have a new thread for every issue."-P5-2

"I think it would only be useful to have a call if we actually really disagreed about something and we couldn't come to a consensus in the chat."-P5-2

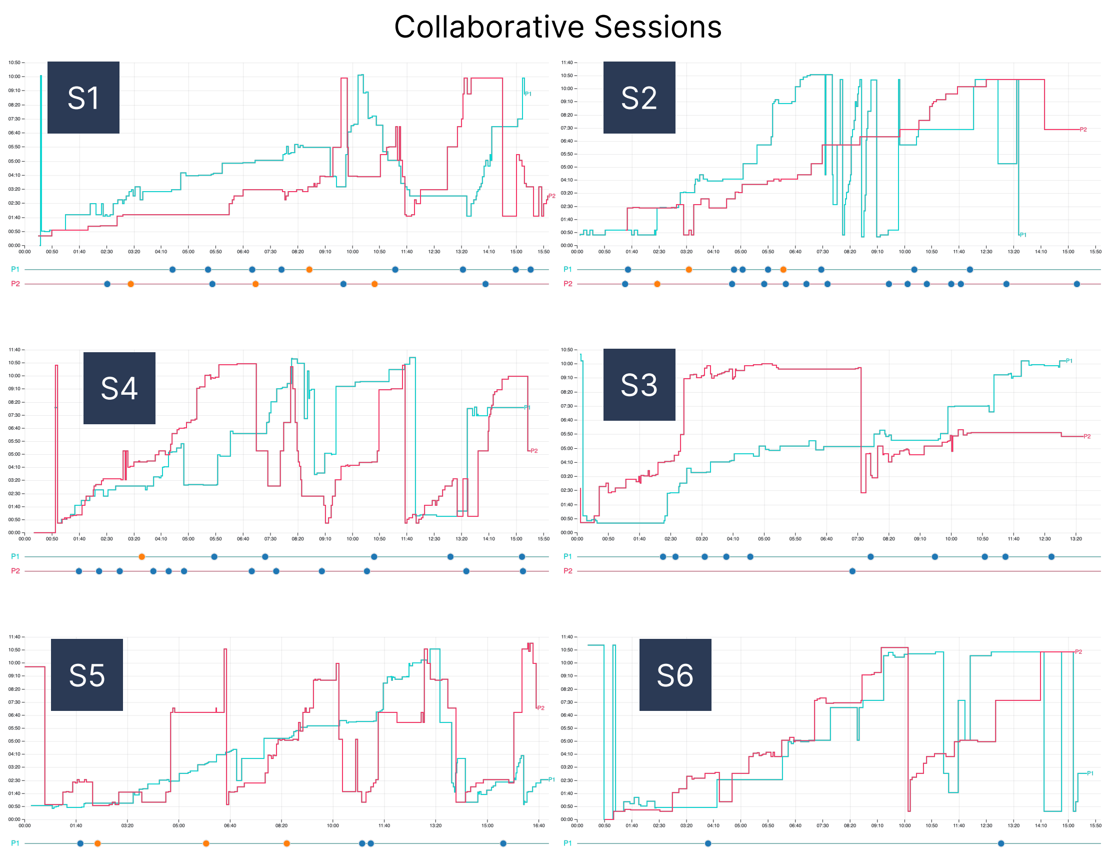


Figure 3.6: The think-aloud video time visualized over the collaborative session timeline. The blue dots represent annotations & comments. The orange dots represent merging two annotations.

Chapter 4

IVA: An Empirical Investigation and Design of Intelligent Virtual Assistant for Online Meetings

Preface - *In chapter 3 we discussed the CoUX system and how it helps the collaboration of multiple UX evaluators. In this chapter, we apply collaboration to meetings, where people collaborate naturally. However, we move a step forward and investigate the impact of having an intelligent virtual assistant as a collaborator in the meetings. We use our learned lessons from the previous paper and integrate the previous works in the online meeting domain alongside a formative study to design a complete system to facilitate meetings of small size. We then explain the implementation and all the design considerations, followed by the results of a study conducted to test the system's usability. Finally, we reflect on the results and the opportunities for future improvements. The background and related work for this chapter has been presented in chapter 2.*

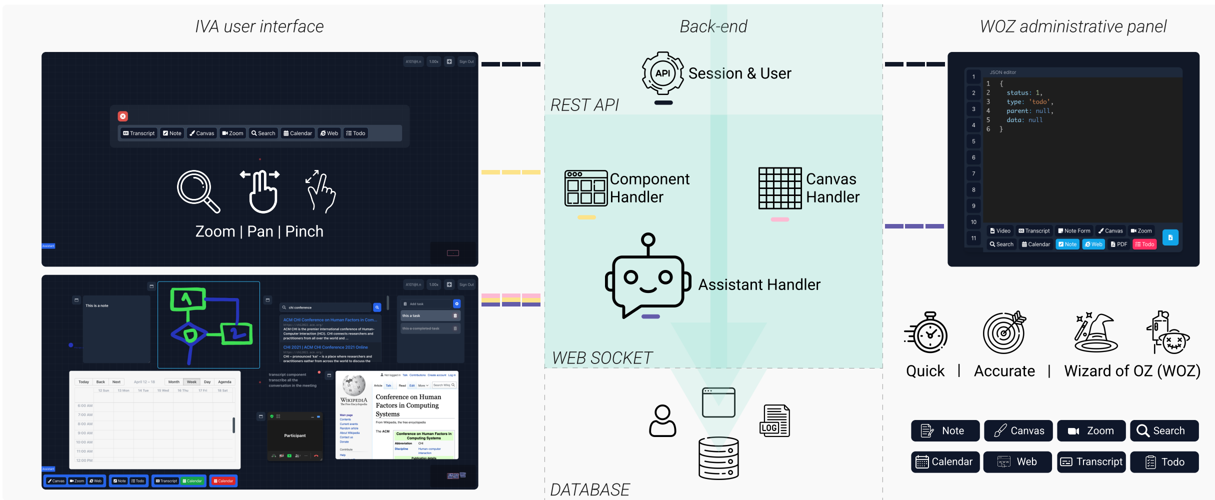


Figure 4.1: IVA is an intelligent virtual assistant helping meeting attendees perform multitasking and collaborate during online meetings. The system provides an Infinite canvas and a set of interactions and components as an environment for multitasking and collaboration. All actions in the system are synchronized through the back-end. The WOZ administrative panel simulates the recommendation of components quickly and accurately.

To address collaboration and multitasking challenges in meetings as discussed in Chapter 1, we propose IVA, an online intelligent meeting assistant capable of automatically providing additional tools and services to the meeting attendance based on the context and content of the online meeting. The IVA system contains two main components. The first component is a customizable infinite canvas where the user can add, relocate, and adjust the arrangement and size of any additional applications, in addition to the main meeting session (e.g., Zoom meeting). This helps the attendees to have an overview of all necessary applications and the meeting simultaneously, which results in improved multitasking performance. The second component is an intelligent assistant 'Iva' that can automatically provide tools and services during online meetings acting as the meeting moderator. Our designed assistant is capable of handling most of the responsibilities of a moderator, which results in increasing the efficiency of the virtual meeting. The assistant is implemented as Wizard-of-OZ (WOZ) for the "best" user experience. The assistant is managed through the WOZ administration panel, which is designed with the accuracy of recommendation, immediateness of action, and easiness of operation in mind.

To design our system, we first conducted a formative study to explore how people attend virtual meetings, the tools and applications they use, and the challenges they may have while attending an online meeting. Understanding when, how, and why people utilize and arrange their tools during an online meeting provides us with deep insights regarding the requirements of our system. We also asked our participants about how an intelligent online assistant can resolve these issues with minimum distraction to enhance the efficiency and performance of meeting attendees. We then designed our system based on the collected data from the formative study. To evaluate IVA, we conducted a user study with twelve users who frequently attend online meetings. They were asked to explore IVA, through a mock up one-on-one virtual meeting with one of the researchers of this paper. The results of this study show that the participants found IVA a practical, helpful, and enjoyable platform that can improve multitasking and collaboration during an online meeting.

Our contributions in this paper are threefold:

- Empirical findings from the formative study regarding the tools people use in an online meeting, challenges they face and possible solutions to these challenges.
- Designing a Wizard-of-Oz-based AI assistant, as a proof of concept, to improve users' performance and efficiency during online meetings by automatically providing tools and services depending on the content of the meeting.
- Insights into the usefulness and design guidelines of AI-based virtual meeting platforms.

4.1 Study I: Investigating the Practices and Needs for Virtual Meetings

Designing our system begins with understanding how people attend virtual meetings, how they use their tools, and what concerns and issues they may experience. We, therefore, conducted a *formative study*, interviewing people who frequently attend remote meetings as part of their daily routine. The results of this study provide us with a deep understanding of the current concerns of online meeting attendees and the tools they require on such remote platforms. Secondly, this study aims to investigate how deploying an intelligent assistant can enhance meeting attendees' performance and efficiency and solve their concerns.

4.1.1 Interview Design

To achieve our goal, we decided to conduct semi-structured one-on-one online interviews with our participants. Each interview session had three sections. In the first section, we asked general questions about participants' current routines and their online meeting tools. The second section was about multitasking and collaboration and how people use multiple tools in addition to meetings. The last set of questions was about participants' concerns and potential improvements in virtual meeting platforms by designing an intelligent assistant to resolve these concerns by automatically providing tools and services.

We designed a set of mainly open-ended questions for each section to encourage our participants to discuss the details and their thoughts. During the remote interview, the researcher (one side of the interview) asked the designed questions. Based on the answers of participants, they could ask further follow-up side questions related to the primary question to clarify participants' responses, as well as get into more details and examples. The three sets of questions are described in (Table 4.1) in details.

The first set of questions were general questions about participants' meetings. The goal of the second set of questions was about the tools that users utilize during their meetings, mainly for collaboration with other attendees and performing multitasking. Understanding the tools and how people use them helps us consider these tools as complementary components of our virtual meeting platform with an embedded intelligent assistant. The last set of questions was related to the design, implementation and requested features and suggestions in our intelligent meeting assistant platform from the participants' point of view. To answer these questions, we asked our participants to be creative and think out of the box, as there was no such platform for virtual meetings.

4.1.2 Participants

We recruited 12 participants (4 females and 8 males; aged 30.3) who had at least three virtual meetings per week for the past four weeks. We added this condition as a pre-screening interview question to ensure our participants have had enough experience with virtual meetings. Our participants contained six students, three UX/UI designers, two university professors, and one software manager. They have a variety of technical backgrounds such as computer science, engineering, design, etc. We compensated our participants with a $15 gift card for their time.

4.1.3 Procedure

Due to the COVID-19 pandemic, we decided to have remote interview sessions with our participants. Participants were asked to pick a quiet place to avoid distractions. Before starting the interview, we explained the objectives of our research and asked our participants to read through

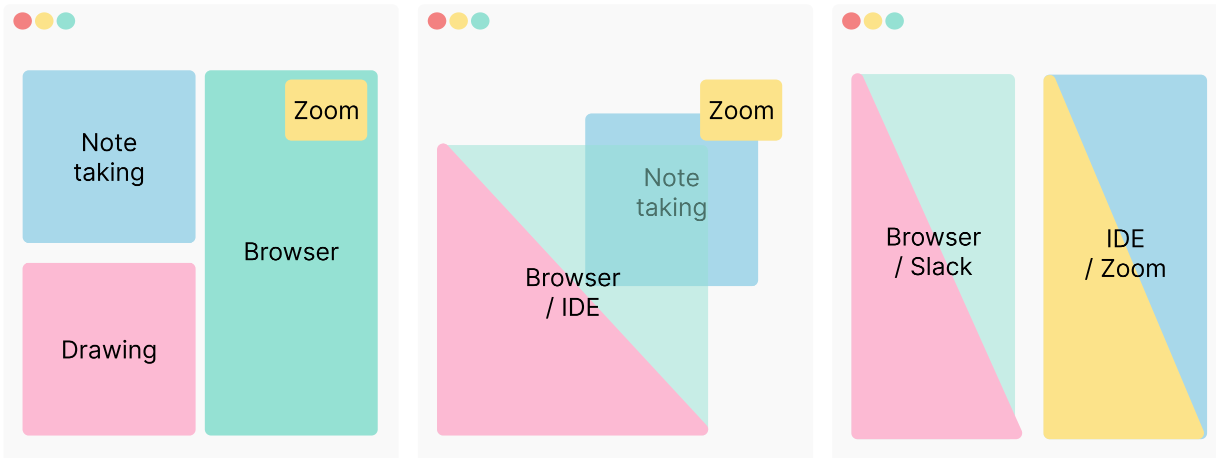


Figure 4.2: Window arrangement for three of participants in response to ”Q8.2: Assume you have your online meeting, note taking, drawing applications and your browser. How would you arrange these four windows on one display? Why?”. The Zoom window is minimized to get more screen real estate for other tasks.

the consent form. We informed our participants that the interview would be recorded for further analysis by the researchers of this paper.

As described earlier, each interview session had three sections. The goal of the first section was to get insights into the current tools participants use, the challenges they face during virtual meetings, and potential solutions to the proposed difficulties. We also asked more general questions about their current meeting routine (e.g., meeting time, topics, number of sessions per week, number of attendees, etc.). The second part of the interview focused on collaboration and multitasking, including questions about how they arrange their tools during an online meeting. For inquiries related to multitasking and windows arrangement, we asked our participants to show us a simple sketch of how they arrange (size and location) their applications while attending virtual meetings. For the last section, the questions were about designing an intelligent assistant embedded into virtual meetings and how it can help users in their meetings. We provided two examples at the beginning of this section to ensure they understood the general idea behind such a system. We asked our participants to think out of the box and be creative. They were also asked to be specific and detailed about scenarios in which an AI assistant can be helpful during virtual meetings.

A thematic analysis was conducted to identify patterns in the interviews. Two researchers coded the interviews independently and then consolidated the themes through a meeting.

4.1.4 Results

In this section, we will report the results of our formative study. First, we will discuss general insights into users' meeting routines. Then, we will report how people perform multitasking, collaborate and how they arrange their tools during their online meetings. Lastly, we will report the result of our users' perspectives regarding the design and functionalities of a meeting-dedicated online assistant.

General Insights into Users' Meeting Routine

Our results show that, on average, our participants attend approximately seven weekly meetings, with a maximum of 15 and a minimum of 3 meetings per week. Most of these meetings are group meetings with more than two attendees. Only two participants reported attending one-on-one meetings more often than group meetings. People who participate in these meetings are primarily researchers (60%), including students, supervisors and research collaborators, and project team members (30%), including software developers, designers, and product managers). In addition, 60% of reported meetings happen in the morning, 30% in the afternoon, and 10% in the evening. No virtual meetings were reported at night or noon.

The majority of participants' meetings are scheduled meetings. Only two out of 12 (16%) participants mentioned that their meetings are primarily unscheduled. This is important as many of our participants commented that attending a regular meeting is more effortless than attending an unscheduled one as they are familiar with the meeting's attendees and the main topic and know what to expect. In this regard, two of our participants mentioned how they prepare their material before scheduled meetings as they know what they may need in the meeting. For instance, P4 said, before my scheduled meetings, I'm going to prepare what I want to do, and P8 commented, I always prepare the questions that I want to ask or the topics I want to discuss before the meetings.

Multitasking, Collaboration, and Tools Arrangement

The results of our study show that all twelve participants perform at least one other task besides their virtual meeting. Note-taking (12/12), online searching and browsing (11/12), checking calendar and tasks (9/12), checking documents (9/12) and checking emails (4/12) are among the top tasks people perform during their online meetings. In the following few sections, we will focus on the most commonly used tasks reported by our participants.

T1. Note Taking: Of all 12 participants, nine actively take digital notes in almost every meeting. Three of twelve participants mentioned using a physical form of note taking (e.g., on paper or using sticky notes). These participants argue that taking notes on paper can be quicker, more convenient, and more free-formed, as both writing and sketching are available easily. For instance, P8 mentioned, I'm really comfortable with the old-fashioned pen and paper, mostly, as it is faster than opening a word document and start writing and P2 stated, I really like the feeling of a physical sticky note to write things down. However, organizing and using these physical notes could be challenging, making physical notes less efficient than flexible digital note-taking tools. P8 addressed this by saying, I'm always struggling to organize my [physical] notes and structure them properly. Some participants pointed out pen and paper might not be available during a meeting, so they use digital note-taking tools on their computer, P7 addressed this by I typically don't have like a pen and paper nearby so I use google doc which is available at any time.

T2. Online Searching and Browsing: Using a browser for online searching was a second common task our participants reported. They perform the searching task mostly when they need further information about something new discussed in the meeting (e.g., the definition of new terminology and finding a new research paper). We realized that our participants tend to search if they know they can get the information in a short period of time, in order to be connected with the meeting and not be distracted. Hence, to minimize getting distracted from the meeting, it is crucial to perform this task as quickly as possible.

T3. Calendar Checking: Of all participants, 75% of them actively use their calendars in the meetings. The main reason to check a calendar is to set up deadlines for the assigned tasks and their availability for follow-up meetings. The Calendar is commonly used at the end of the meeting and mostly in conjunction with a task manager or email client. Helping with either of the tasks reduces one extra focus switch from the users' side. Thus, the system should remove unnecessary switching to Calendar or its side applications by showing hints or running the corresponding task.

T4. Collaboration: Most participants (11/12) pointed out that the majority of their meetings involve a great amount of collaboration with other meeting attendees. The collaboration involved working on shared documents (e.g., Google Docs and Overleaf documents), brainstorming, designing a product, debugging and deploying software, and working on presentations (e.g., google slides).

T5. Tool Arrangement: Our analysis showed that 11 out of 12 participants used a second monitor to arrange their tools (e.g., a laptop and an external monitor). They mostly use the primary monitor (the monitor in front of the user) for the meeting platform, as it shows their attention to the meeting, and the external monitor for multitasking (e.g., taking notes and browsing).

For further investigation and to identify how participants arrange their tools, we asked our participants to assume there is only a monitor and they need to place four different applications. These applications are the meeting application (assuming no screen sharing), their note-taking software, an application to draw a diagram, and a browser. (Figure 4.2) shows a summary of the result.

The minimum space was allocated to the meeting platform among all four windows. Participants' comments indicate that as long as there is no screen sharing, which requires more attention, they did want to minimize the allocated space to the meeting window. More space was allocated to the drawing component, making the drawing task easier. Drawing on smaller spaces using a computer mouse and trackpad, could be a challenging task for the users. Participants preferred to allocate narrow vertical spaces for both the searching and text components. This shows that participants follow a narrow linear top-to-bottom flow to take notes in text-based tools.

Smart Assistant

In this section, we will discuss about the potential applications and functionalities of an intelligent assistant in an online meeting platform suggested by our participants. We will focus on how such an intelligent assistant may improve attendees' efficiency and performance.

Suggested Features: We realized that the main focus of our participants was to take advantage of the intelligent assistant to automate the procedures that could be done automatically with no or least interaction with them to minimize the distraction from the meeting. The following requirements were discussed.

F1. Smart Task Integration: One of the most commonly suggested features was task integration with the assistant, such as calendars. Many of our participants (10/12) suggested having an assistant that can extract date-related information from the conversation in a meeting and check if the user is available or not. If the user is available, the assistant could add such an event to their calendar with a confirmation from the user. One participant took a step further and suggested an integration of all users' calendars with the system so the system could offer the best date and time that works for all attendees (similar to a voting system). In this regard, P4 suggested that Intelligent assistant can be used to pick the best date and time [for the next meeting], as it knows all the availabilities of people in the meeting.

F2. Intelligent Information Seeking: Another commonly requested feature was triggering a space-efficient intelligent search tool, integrated with the virtual meeting platform, to detect the keywords from attendees' conversations that the users might want to search further and get more information. This minimizes the distraction as the search process can be done automatically, and the result is integrated with the meeting platform preventing the user from entirely switching to their browser to perform the search task.

F3. Adaptive Conversation: Selecting, copying, and pasting live transcript and using it in their notes, and saving the transcript of the entire meeting were two features our participants suggested in the system. With such a tool, important content can be easily extracted from speech and used in users' notes, making the note-taking process a more efficient procedure with less distraction. P8 stated, It saves my time if I use the transcript while taking notes.

Implementation and Privacy: Our analysis showed that our participants had a number of critical concerns about the system implementation. The first concern was about triggering the smart assistant. If the assistant keeps listening to the discussion between the attendees, it may activate a command that was not the user's intention. For instance, if the user is talking about a date and time, that does not necessarily mean they want to set up a deadline or next meeting.

Privacy was another issue raised by the participants. Although seven participants had no problem with the assistant listening to their conversation, the rest had concerns about the assistant proactively listening to their conversation. This is why some of the participants proposed using a wake word to activate the assistant and then giving a direct command to the assistant, similar to Google Assistant and Amazon Alexa. However, this method to activate the assistant could be an awkward situation and is not a practical implementation in the online meeting context.

4.2 IVA System

Informed by the literature [14, 67, 91] and our formative study, we designed and developed IVA. Due to the current limitation of AI technologies, it is not possible to develop a reliable back-end that can accurately detect the timing and assist with the tasks we identified. Therefore, we chose to implement our system via WOZ so that users can receive a nearly-perfect experience. This also facilitates us with empirically investigating users' opinions towards AI assistants during online meetings.

4.2.1 System Overview

As shown in (Figure 4.1), the system consists of three modules: (a) a back-end and data storage platform, (b) a front-end visual interface packed with valuable features and interactions, and (c) a WOZ administrative interface to facilitate the simulation of an AI assistant. Based on the formative study, we encapsulate the tools and services for online meetings into a set of interactive, modularized *components* in IVA, which is also extensible in the future.

The front-end consists of three main views: an Infinite Canvas, an Assistant Interface, and a Component Toolbar (Figure 4.3). The Infinite Canvas is designed and implemented to maximize the screen's real estate, no matter the monitor size [67, 71, 111]. It allows users to freely move

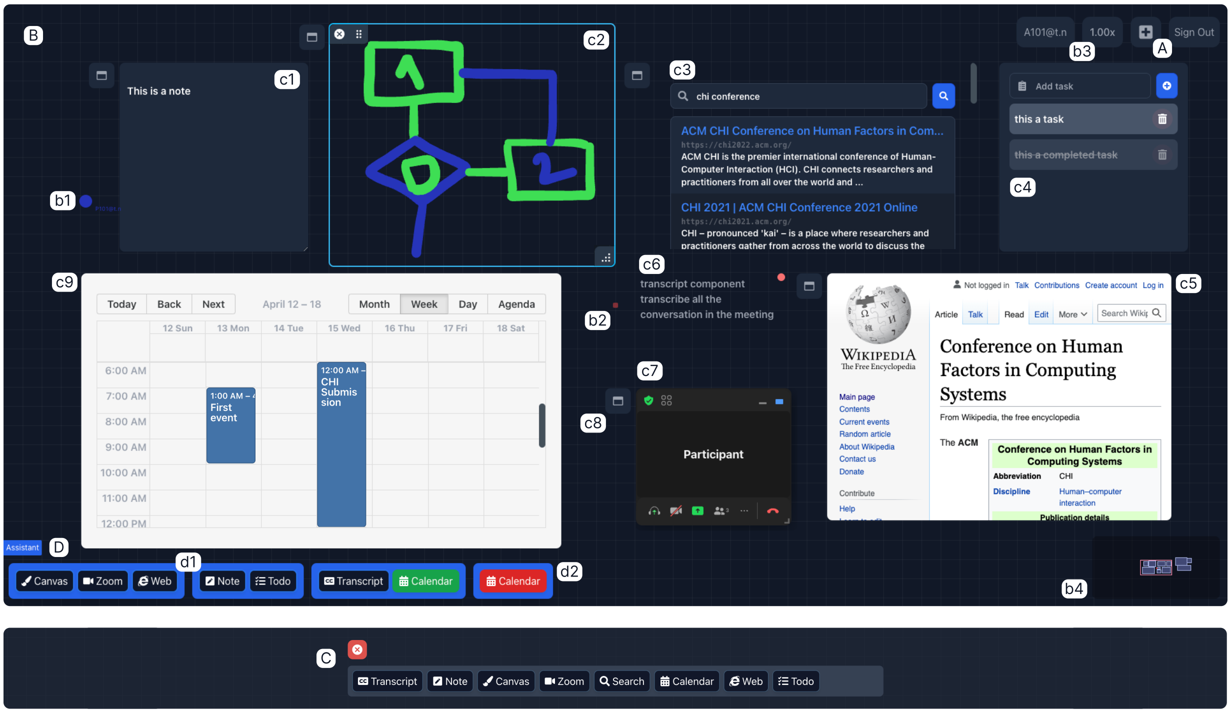


Figure 4.3: The IVA user interface shows all the different modules of the system: (A) the component toolbar button, which opens the toolbar shown with the label C, (B) the Infinite Canvas supports zooming and panning interactions for better multitasking and collaboration, (C) the Component Toolbar has a set of tools mainly used for multitasking, and (D) the Assistant IVA, showing temporarily the shipped Suggested Islands and groups of components in them.

any components in order to focus on their work or collaborate with others in the same area. The Assistant Interface hosts suggested components from the WOZ admin panel, placed in a less distracting part of the screen to help users manage the meetings without causing much attention drift [69, 73]. Finally, the Component Toolbar hosts various developed components to allow users to easily add and interact with them on the Infinite Canvas.

The WOZ administrative interface contains three modules: a JSON Editor, a Component Toolbar, and a Quick Suggestion List (Figure 4.4). The Component Toolbar is developed with the same look as the one on the client side, removing any extra learning steps for the administrator. The JSON Editor gives the administrator complete control of the data passed to each component, allowing for prefilled component suggestions. Lastly, the Quick Suggestion List allows the administrator to add a list of components that can be suggested as quickly as one click.

The front-end interface and the WOZ interface communicate through the back-end, which stores all the information regarding users, the infinite canvas, and recommendations in the PostgreSQL database. For a seamless collaboration on the users' side, the back-end also provides a web socket connection for both interfaces, which allows every system module to get all updates in real-time.

4.2.2 Implemented Components for Tools and Services

In this section, we will go through all the components used in the IVA. Based on the formative study and the previous research, IVA includes a set of best-suited tools that users regularly reported using them, both on the formative study and the previous work, in conjunction with online meeting platforms [14, 93, 123]. Each component can be added to the canvas directly from the Component Toolbar (Figure 4.3-C), and it will be located in the middle of the viewport. There is no limitation on the number of components a user can add to the canvas. The data in all components is synchronized with all users of the system. Therefore, any change on a component will result in immediate change for all users, facilitating collaboration. Selecting a component will indicate an active state using a blue border (Figure 4.3-c2). The active component will automatically come to the top, letting users combine and overlap components to fulfill different purposes. For example, drawing circles for important information on a component can happen using a drawing canvas component on top of any other component.

Notepad. The notepad component helps users to write down key points during their discussions in the meetings (Figure 4.3-c1) (**T1**). To keep note-taking simple, we use the `<textarea>` tag for multi-line input. The component mimics the same interaction as a notepad. However, note-taking components do not offer any rich text editor functionalities. Users can either use the editor to take notes linearly, similar to all other tools or use multiple note-taking components to make a non-linear note.

Drawing Canvas. The drawing canvas component helps with both sketching and annotating tasks (Figure 4.3-c2) (**T1**). The component is built using `<svg>` and `<path>` tags which results in sharp output. The canvas is transparent so that users can easily overlap it on top of other components to annotate and highlight information. The pen colour for each user is generated based on their email and is unique to them, so that each user's contribution is obvious on the canvas (**T4**). Right now in our implementation, users cannot choose the colour or erase what they draw on the canvas.

Calendar. The calendar component supports any scheduling tasks in general, such as adding, removing, and updating events (Figure 4.3-c9). The `react-big-calendar` plugin [87] is the backbone of this component which looks similar to generic calendar applications such as Google

Calendar to avoid any learning curve, considering that users mentioned using calendar actively in their meetings (**T3, F1)**.

Todo List. IVA offers a to-do list component in conjunction with calendar component to keep track of actionable (**T3**) in each meeting (Figure 4.3-c4). The component allows users to add, remove, or toggle task items. Completed tasks will be faded, and the text is crossed over to indicate the status change.

Transcripts. This component provides transcripts of the meeting session in real-time to let users keep informed about what is happening (Figure 4.3-c7) (**F3**). The `react-speech-recognition` plugin [9] provides real-time meeting transcripts using the Web Speech API. The process happens on the front-end side and synchronizes on all devices. Users can easily copy and paste text from the transcripts to their notepad components.

Video Conferencing. IVA currently supports the most used online meeting platform, Zoom, integrated as a component on the Infinite Canvas (Figure 4.3-c7). This video conferencing component uses the official Zoom Meeting Web SDK. It looks and behaves the same as the Zoom client, which allows users to interact naturally with the component. Based on the results of the formative study (Figure 4.2), the Zoom component is minimized so that users can use the remaining space for other components.

Search and Web. Many of the participants in the formative study mentioned searching and surfing while they were in meetings (**T2, F2**), and thus IVA offers search (Figure 4.3-c3) and web (Figure 4.3-c5) components. However, they prevent any focus change, like changing tabs or opening other windows. The search component uses the Google Search API, and results directly come from Google. When users click on a result shown in the list, the website will open in a web component. IVA uses the `<object>` tag to embed external websites into the application so that users interact with them directly inside of it.

Component Capture. To enable users to save their work or keep information accessible and easy to recall (**T2**), IVA allows users to take snapshots of any component (Figure 4.3-c8). On the left side of some of the components: Notepad, Drawing Canvas, Video Conferencing, Search, and Web , there is a camera button to take a capture from that component. Then, the snapshot will be added to the canvas as an image, which is another component.

Containers. All the components in IVA are wrapped in containers upon adding them to the canvas. The container handles all the interactions regarding moving, removing, and resizing the component inside (Figure 4.3-c2). It also hosts helper items such as a snapshot button for each component. The container uses the React `dnd-kit` plugin [21], which provides an API for dragging & dropping components in the application. IVA uses browser events for container resize handling.

4.2.3 Infinite Canvas for Collaboration and Multitasking

The most prominent building block of the IVA is the infinite canvas, which hosts all the components and enables users to expand their screen real-estate freely (Figure 4.3-B). The canvas supports numerous of interactions such as zoom, pan, and pinch, letting users navigate their workspace easily. It uses the `react-zoom-pan-pinch` plugin [86] to provide all these interactions in a cross-platform manner. All the newly-added components will be at the center of the user's viewport. The components initially get arranged in a sense to prevent overlapping (**T5**), but users can overlay them. Also, the components have an initial size based on their use cases. For example, the drawing canvas is relatively larger to allow users to collaborate easily (**T5**). The canvas persists all the information on it, so users can come back to what they did afterwards.

To promote collaboration and awareness (**T4**), users can see the mouse cursors for all active users in a session, along with their emails by the cursors (Figure 4.3-b1). The information gets synchronized through the web socket. The colour is unique to each user and is also used on the drawing canvas to facilitate collaboration (**T4**).

The infinite canvas provides infinite space for the users, but they may get lost in the workspace. There are three mechanisms implemented in IVA to help mitigate this problem. First, the reset transform button brings the user back to the center of the screen (Figure 4.3-b3). Second, there is a tiny dot in the center of the canvas, which is a simple visual cue for the users to let them know how far they are from the center (Figure 4.3-b2). Finally, a minimap at the corner of the screen shows all the components within and beyond the viewport, indicating how components are positioned on the infinite canvas (Figure 4.3-b4). The same small red dot is also shown to help them locate themselves on the canvas (Figure 4.3-b2).

4.2.4 Smart Assistant for Component Recommendations

Within our system, we designed a smart assistant using WOZ to study how an AI assistant can work with users to facilitate their collaboration and multitasking during online meetings. The main function of the assistant for the users is to suggest tasks based on their conversation with unobtrusive recommendations (**F1**). It is always the user's choice whether to use the recommended components (tools and services) or not.

Assistant "Iva"

The assistant, with a person's name *Iva*, is placed at the bottom left corner of the screen, to avoid big distractions (Figure 4.3-D). Initially, a badge is displayed showing where recommendations can be expected from the assistant (Figure 4.3-D). During the conversation, Suggestion Islands will appear and disappear, indicating groups of recommendations (Figure 4.3-d1). The assistant can suggest more than one component at a time. For example, during a task of scheduling the next meeting, typically a user opens the calendar, email, and todo list applications, in order to add the next meeting to the calendar, inform everyone, and write down all actionable for the next meeting.

The Suggestion Islands have a blue background to grab attention, and all the suggested components look the same as how they look in the Component Toolbar for consistency (Figure 4.3-d1). The Suggestion Islands are temporary and only on the screen for ten seconds. This ten-second window helps prevent having a long list of suggestions from distracting users. The blue background can change to red or green for some of the suggested components indicating the availability or possibility of acting using that component (Figure 4.3-d2). For example, in case of a scheduling conflict, the suggested calendar component will have a red background indicating that it is impossible to schedule a meeting on the discussed date. Otherwise, it will be indicated as a green background. All in all, the assistant is designed to play a moderator's role in meetings, reminding people of what they have and what to do.

WOZ Administration

As mentioned earlier, the smart assistant in our system is implemented using WOZ. To provide the "best" experience with the assistant, we consider three criteria during our design of the WOZ interface, including: the accuracy of recommendation, immediateness of action, and easiness of operation. To fulfill these criteria, we implemented three modules: a JSON Editor, a Component Toolbar, and a Quick Suggestion List.

The JSON Editor helps with prefilling data (**F3**) into each component, which is represented as a JSON object (Figure 4.4-B). The administrator can interact with the suggestion object directly by attaching component data. The component data translates to what the component will show when the user selects it. The data can vary from a simple string to complex objects to pre-load a drawing on the Drawing Canvas.

The Component Toolbar, similar to that of the visual interface, is provided, allowing the administrator to have a better feeling and understanding of what is being recommended to the user (Figure 4.4-C). The difference is that the buttons have three states: inactive, active, and selected, and the administrator can easily switch between all these three states. At a time, only one component is active and can be edited on the JSON Editor. But multiple components can be selected as a suggestion group. After the preparation, clicking on the send button ships all the recommendations grouped as a Suggestion Island to the users in the meeting (**F1**, **F2**). This will

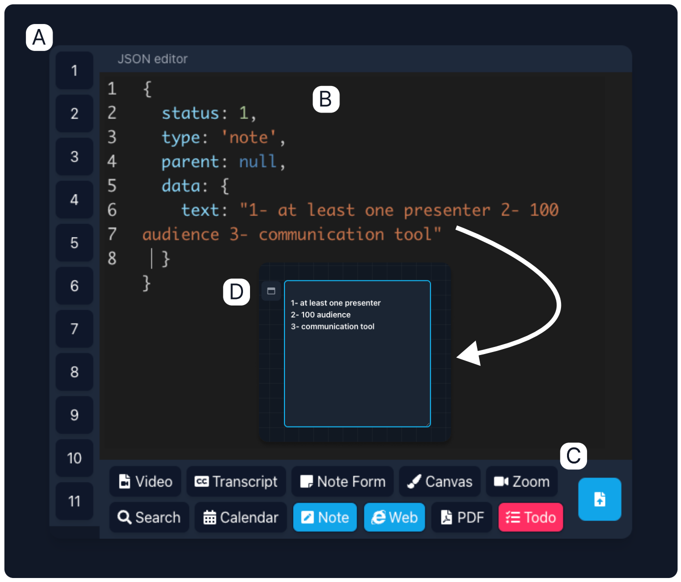


Figure 4.4: The WOZ administrative panel user interface includes: (A) the Quick Suggestion List to help administrator provide predefined recommendations as quick as a computer would, (B) the JSON Editor, which lets administrators customize their suggestions and ship them with prefilled information, (C) the Component Toolbar hosts all the components shown as buttons with three different states: inactive, active, and selected, easily switchable, and (D) the component shown to user after clicking on the recommendation.

also reset the states of all the buttons to inactive, so the administrator can quickly start preparing the next batch of recommendations.

The Quick Suggestion List is designed to help quickly ship a set of predefined recommendations to users (Figure 4.4-A). This is useful for supporting studies that have certain procedures. In execution, The administration can prepare all possible suggestions for the study and load the recommendations beforehand to the system. The suggestions will be shipped as soon as clicking on any of the list buttons. This feature reduces the recommendation delay to almost zero, which makes the WOZ assistant feel the same as being driven by a computer.

4.3 Study II: Understanding the Effectiveness and Values of IVA

This *evaluation study* aims to assess how end users interact with our virtual meeting platform, IVA. The findings of this study reveal how practical and valuable an AI-based meeting assistant can be to meeting attendees and whether it impacts the users' efficiency and performance during a remote meeting. It will also lead us to the design considerations we need to take into account in developing virtual meeting platforms to support interactive collaboration and multitasking.

4.3.1 Experimental Design

To evaluate IVA, we designed a mockup one-to-one online meeting with our participants, using IVA as our meeting platform, which was implemented using the Wizard-of-Oz prototyping, as a proof of concept. To achieve our goal, we developed a virtual meeting scenario where we could prompt all IVA's integrated tools and services to our participants, based on the discussed topics in different parts of the scenario. The mockup meeting involves one researcher and one participant having a brainstorming session to solve a well-defined problem. During the meeting, one other researcher would act as the AI assistant by listening to the conversation between two attendees, in the background, based on our WOZ implementation. Participants were not informed of the third researcher (the wizard).

The topic we picked for our brainstorming session was "design a platform to support remote presentations" (online tutoring or conference presentations). We chose this topic as it does not require additional background knowledge, and is simple and easy to understand. We also decided to conduct a brainstorming session as our participants reported brainstorming as one of their common tasks during their meetings, involving interaction, collaboration and multitasking. As summarized in the following, the scenario consists of four main phases, each consisting of a theme guided by several prompt questions or discussion points.

Warm Up. In the beginning of the brainstorming session with our participants, we explained the general objective of this brainstorming session and what are the issues we wanted to solve. We let them know that we would list the basic general features of the system [IVA: suggesting an empty note-taking component, in case they wanted to take note of the features the researcher, was about to list]. For all the suggestions, our participants could click on the assistant icon to fully open the suggested component and start interacting with it, or they could ignore it for 10 seconds, and the suggestion would disappear.

Feature Proposal. We then proposed and explained the three primary features necessary in an online presentation system [IVA: suggesting a note-taking component containing a list of features explained to the participant in real time]. These three criteria were 1) there are one to three presenters, and they should be able to present at the same time (e.g., co-presentation at a conference), 2) There is a group of audience (maximum of 50 individuals), and 3) there should be a communication channel between presenters and the audience. We picked these criteria as we wanted them to be simple and high-level so our participants could propose their questions and ideas through interactive communication with the researcher. In the next step, our participants were asked to list five essential features they wanted to see in such a system [IVE: proposing an empty note-taking component so they can list the features and share their ideas]. Throughout the experiment, we made sure that there was an interaction between the researcher and the participant discussing participants' suggestions.

Interface Design. As a follow-up, we asked our participants to describe the system's user interface, considering the proposed features discussed in the previous step [IVA: prompting an empty drawing component and the note-taking component containing the list of features discussed in the previous step as a reminder]. For more inspiration and to see how the users use the search component, we named one other similar existing platform, Axure, so our participants could use the search component to check the features and UI [IVA: suggesting a search component. By clicking on the suggestion, participants could see the result of a google search]

Wrap up. To see how our participants use IVA's calendar component, we informed our participants that we needed to finalize the design on another specific day, in the following week, at a particular time [IVA: shows that the date and time are unavailable, it also suggests the calendar component for further details if needed]. To check their availability, our participants could check IVA's prompt at the bottom of the screen or directly open and check the integrated calendar component. We also assigned them five tasks they needed to finish before the next meeting [IVA: prompting a task component with all the assigned tasks in real time]

Although the scenario was made to ensure we could suggest all the designed components of IVA to our participants, to explore how the users would utilize these tools, we let them know that it was fine if they wanted to skip using the suggested prompts from IVA (e.g., memorizing items and describing the ideas orally as opposed to using the note-taking and drawing tools). To see how false activation may affect the users and make the experiment more realistic, we randomly prompted one of the tools every 8 minutes.

4.3.2 Participants

For this experiment, twelve new participants were recruited, from a local university campus from various fields (3 females and 9 males; average age 25.1, 2.3). Similar to the first interview, we ensured our participants were familiar with virtual meetings (at least three weekly meetings for

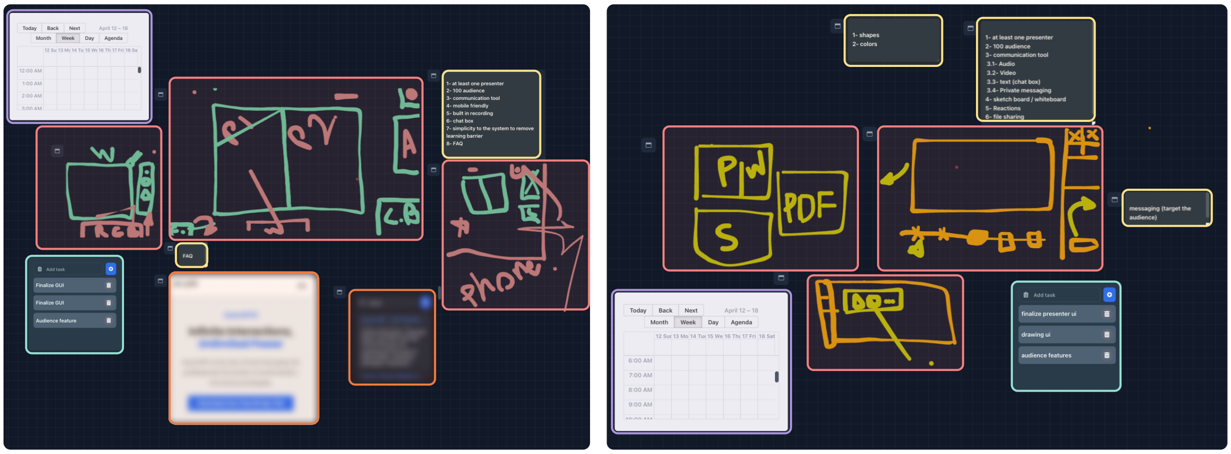


Figure 4.5: The final canvases after two of the studies. Each color represents one type of component used on the canvas, and some components are used several times.

the past four weeks). The entire experiment took 45 minutes, and participants were compensated with a $15 gift card for their time.

4.3.3 Procedure

To explain the experiment's objectives and instructions, we first set up a virtual meeting with our participants using the Zoom platform. After reading and signing the consent form by our participant, a link to our virtual meeting platform was provided to the participants. One of the researchers spent approximately ten minutes introducing our virtual meeting platform and showing how the intelligent assistant can automatically provide the tools and services designed in our system based on the discussion between the two sides of the meeting. When we ensured our participants knew how to use the system, we started the brainstorming session following the scenario explained earlier. At the end of the experiment, we asked our participants to fill out a questionnaire to assess the effectiveness and usefulness and conducted a semi-structured interview to explore how participants used the tools and services of the system. (Table 4.2) shows the details of these questions.

4.3.4 Results

In this section, we report the results of the second experiment broken down into the following themes. Overall, the results of our study show that IVA could be a suitable alternative to current

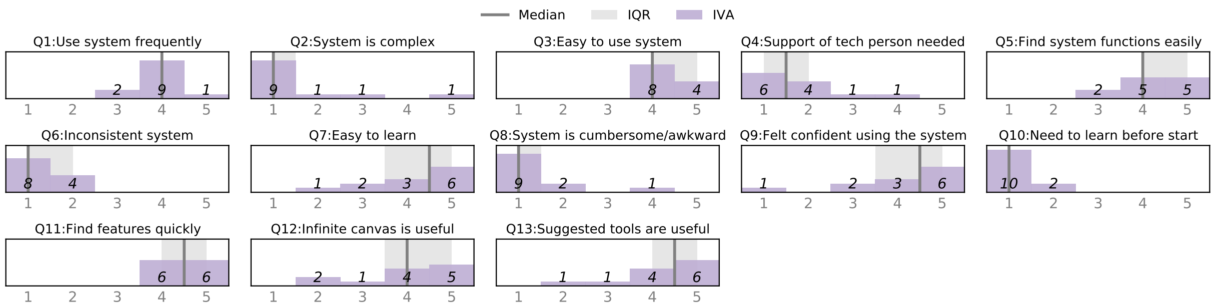


Figure 4.6: Participants' questionnaire ratings on a 5-point Likert scale (1 = "strongly disagree" and 5 = "strongly agree").

virtual meeting platforms by automating some of the common tasks people perform during their meetings. As shown in (Figure 4.6), for Q1 regarding the willingness to use the system, the median was 4, with only two participants with a minimum score of 3 (neutral), indicating our participants see themselves using IVA as their daily virtual meeting platform. In general, the answers to the usability-related questions show that our platform was not complex (Q2, 1), was easy to use (Q3, 4) and easy to learn (Q7, 4). In the following sections, we will report the results focusing on specific features of IVA.

Collaboration and Multitasking

Participants' responses show that the infinite canvas can support collaborative tasks and multitasking (Q12, 4). The floating adjustable windows for components played a significant role in facilitating multitasking and collaboration. Many participants commented on the dynamic location of tools on the platform as an advantage, as all the users' required tools are located beside each other, integrated into the infinite canvas, which makes the transition between tools faster and easier with less distraction. For example, P8 mentioned, I could easily search inside the meeting which I preferred rather than searching on my browser changing my focus. P6 and P11 also mentioned Current tools including Zoom would not allow for a smooth transition between applications during a meeting and It would reduce the time of meeting spent outside of the virtual meeting for searching and collaborative editing. These comments show how presenting and arranging multiple tools can improve the efficiency and performance of the users. P5 and P12 confirmed this by saying, having all the components being accessible at once is very nice, and there is no need to jump through several windows to search for something during meetings!

Although we designed a simple algorithm to arrange the floating windows, to ensure there is no overlap between them, one of the participants re-arranged the suggested layout as they needed more space for the search component. However, this shows that our system is a customizable tool which can be personalized based on the user's preference. In addition, P5 mentioned that a fixed location for each system component would be more accessible for the user to locate these components the arrangement of components gets challenging as their number increases.

The screen-sharing components of existing virtual meeting platforms are not designed to support collaborative tasks. In many of these commercial platforms (e.g., Zoom or MS Teams), it is very common that one person is responsible for sharing the screen and working on the document while discussing with other attendees. P7 addressed by the problem with screen-share is that only one person can make changes, and sharing docs is distracting as you need to work on two separate platforms. However, there are cases where screen sharing is necessary. For instance, if the user wants to show how an application or code works on their computer, for instance, P10 mentioned sharing screen might still be a good idea if you want to show a picture or system behaviour.

Integrated Tools and Services

All participants found IVA's suggested tools and services helpful and related to what they needed during the virtual meeting (Q13, 4.5). The integrated note-taking component was the most used in the meetings. Out of 48 prompts for the note-taking component, by the meeting assistant, across all participants, 43 of them were directly used by our participants. Among all the note-taking suggestions, 24 had some pre-defined text suggested to the users (items and features discussed with the participants explained in the scenario section). Our participants found the proposed text valuable and relevant, making the note-taking process easier and faster. In this regard, P8 mentioned, It was faster to take notes as the assistant suggested relevant notes several times.

The integrated search feature was the second most helpful component of the infinite canvas. Our participants liked the idea of suggesting a google search and showing the result when there is a discussion they need further information to engage with other attendees.

Although we dedicated a tiny portion of the display to the assistant, to minimize the distraction with the users during the meeting, some participants suggested more noticeable prompts from the assistant. For instance, P1 suggested making a bigger popup, make the assistant button bigger to increase its visibility indicating that the assistant prompt should be noticeable and glanceable. However, only one of the participants reported that they were distracted a few times during the meeting session, especially when they were typing to take notes I was surprised by how good it works, but it was distracting for me while I was taking notes.

One of the features not addressed by participants in the formative study and found useful in this study was the save option, the capability to screenshot all the components on canvas (including note-taking, drawing, calendar, and tasks components). P5 mentioned, the fact that I could have various canvases and take screenshots to save them was cool. This feature could be utilized as the summary of the meeting for the next follow-up meetings or as a report of what was discussed in the meeting. P1 addressed this by saying, It [Saving option] can keep a record of everything which sometimes I personally forget to write down.

The transcript feature of the platform was not used during our sessions with the participants; however, the users suggested potential use cases for the transcript component. Some participants suggested that the transcript can be used for people who joined the meeting later to keep track of what happened in the meeting before they joined the meeting; P2 addressed The transcript option helps participants keep track of the session. In addition, the users can take advantage of the co-existing live transcript along with the note-taking feature to take note of critical points of the conversions in the meeting. For example, exporting part of the transcript into the notes.

Intelligent Assistant

The assistant was the most used component of the entire system. All participants found it a valuable and practical tool to mitigate the existing concerns with commercial meeting platforms. P3 addressed this by It's very useful! When you talk about a calendar, a web search, a drawing, etc., it suggests you the required component automatically.

The intelligent assistant was also helpful in simplifying many of the functionalities of virtual meeting platforms. Five participants addressed that they felt the assistant simplified many of the tasks for them and reduced the complexity of the platform by performing the tasks and services automatically, causing them to pay attention to the meeting rather than configuring tools and configurations during the meeting. P4 commented They [existing commercial virtual meeting platforms] are so complicated with too many features that make me confused. This one is light and more user-friendly, and P12 mentioned, My current virtual platform is more complicated to use!

Two participants provided suggestions for further potential improvement in IVA. In general, they asked for more advanced automated services that could facilitate performing various tasks without direct interaction with the users (e.g., the AI shows if the user is available on a specific day and time without opening the calendar) P5 commented It was nice, but it was just a shortcut (except for the calendar). So, maybe if it was a bit more advanced I would like it more. However, our participants addressed that IVA was beneficial just to remind them of the tools and services of the system that could be potentially useful for them. I forgot that there is a feature for a matter, but the assistant suggested it anyway, was explained by P10. A similar point was also addressed by our participants in the formative study. Many participants pointed out that the assistant could be used as a reminder about the features they can use during the meeting (e.g., recording the meeting).

Privacy

Six of our participants did not have any privacy concerns with IVA. However, some participants had concerns about sharing their calendars and private events with others using the calendar component. P5 mentioned, I might prefer it [calendar] not be visible to everyone in the meeting as there might be some private events on my calendar. One of the participants, p8, proposed adding a new feature to IVA so that the user could configure their tools (e.g., notes, drawings, and/or calendars) in a private space, without fully or partially sharing them with others. In this regard, P4 mentioned, Makes some privacy for the users to write down their own notes without revealing to other people. In addition, P2 commented, I always google things during meetings but if other people aren't able to see that I would be happier.

Confirming the results we collected from the interview session, many participants were worried about the location of collected data, whether it was collected and processed on the local machine or a cloud server. Almost all participants were fine with the data collected and processed on their local computers.

Comparing with Existing Virtual Meeting Platforms

Eleven out of twelve participants preferred our virtual meeting platform over current tools. A significant portion of the participants liked the idea of integrating tools such as note-taking, drawing, search and calendar with virtual meeting platforms. P2 commented, The current system is by far more useful platform that I'm using these days due to having integrated features such as integrated writing, task management, and search feature.

Eight participants found more interactivity in IVA, compared to their current tools, the system is definitely more interactive, as mentioned by P9. One of the main reasons IVA supports collaborative interaction between users is the integration of sharing and using various tools across attendees, working together to achieve their goals. They also mentioned that there is no need to switch their entire attention from their meeting to other tools for collaboration or multitasking (e.g., using multiple monitors or minimizing/maximizing applications during meetings). All the tools could be available and present simultaneously beside each other, resulting in minimum distraction. In this regard, P4 mentioned, [In current platforms] you have to swipe between apps and click on them to choose, but in your platform everything is ready to use by dragging.

4.4 Discussion

The feedback collected from our participants shows that IVA, our virtual meeting platform with an embedded intelligent assistant, has the potential to become an alternative to current commercial remote meeting platforms, supporting interactive collaboration and multitasking during an online meeting.

4.4.1 Overall Findings

Part of the participants' feedback was about the low-level details of the implementation of IVA's components. For instance, although we provided some basic editing options in the note-taking component, three participants asked to see more advanced text customization options. In a similar comment, two participants suggested predefined basic shapes in the drawing component. In addition, we received a few comments about the actual integration of our system with their personal calendar. Although these minor improvements that could make our prototype as close as possible to the final product and more realistic, we did not add these details to our system in the first place, to make the system easy to use, understand and practice for the experiment. However, we implemented all essential tools and services for each of the components of IVA for the sake of the experiment's simplicity.

Although some of our participants did not have any privacy concerns with IVA, it was the main concern of many participants in this experiment. However almost all of the participants were fine if they knew the data was collected and processed on their local computer. In general, from participants' comments we can come to the conclusion that there are potential benefits and advantages of using IVA, as a virtual meeting platform, to perform collaborative, and interactive tasks. As multi tasking is common during virtual meetings [14, 66], confirmed by the result of the formative study, our platform were designed to support multitasking during an online meeting. The infinite canvas feature was designed in our platform to provide a dynamic space to our users so they could use multiple system integrated tools at the same time without being disconnected from the meeting. Our evaluation showed that participants used the infinite canvas to work with multiple components of the system (e.g., task management, and calendar components).

The first interview revealed that note-taking, searching, drawing, task-management, and calendar are among the common tools users need during their meetings. The collected data from this interview also showed many use cases of an AI-based assistant that can help the users using these tools in a more productive and efficient way. Therefore we integrated these tools into IVA, based on participants' comments from the interview session. Tested by our participants, the integration of tools with the infinite canvas was the most helpful component of our system to support multitasking and simultaneous collaboration.

Privacy was a major concern of our participants in both studies. Many of participants mentioned that the privacy could be an issues depending on the meeting topic as well as other attendees. Some of the participants did not want to use assistant during their personal meetings. However, almost all of the participants in both experiments were fine with having the intelligent assistant listening to their conversions beyond personal meetings (e.g., work- or education-related meetings).

4.4.2 Take Away Lessons

Based on the result of our studies, we propose the following key take aways:

- Multitasking and collaboration are very common in online meetings. Switching the attention between the meeting platform and other applications results in users' distraction from the meeting. It would be advantageous to design virtual meeting platforms to support collaboration and multitasking, with minimal user distraction.
- Many of the tasks users perform during online meetings can be done automatically without requiring users' attention, meaning less distraction.
- An AI-based assistant integrated with the virtual meeting platform can be designed to automate users' tasks, enhancing users' performance and efficiency.
- Privacy is a concern, especially in personal meetings, but people will accept the trade-off if a system provides adequate value and control. Besides, running all the process that deals with sensitive data locally can help address privacy concerns.

4.4.3 Limitations and Future Work

We implemented all the essential tools and services addressed by our participants in the formative study. However, a number of subtle tools were not added to the system to prevent adding unnecessary complexity and making it hard and time consuming process to learn for the evaluation experiment. For instance, adding more editing tools in the text editor component or predefined shapes in the drawing component were asked from our participants in the second experiment. However, as the main focus of our experiment was to see if such system, as whole, can be beneficial in users workflow using virtual meetings, such detailed customization was not added to IVA.

The other limitation of our experiment was using the Wizard-of-Oz methodology to test and evaluate our system. Although implementing the whole system using different computer science techniques, for instance natural language processing and text analysing, could be closer to our primary goal, which is fully functional autonomous intelligent assistant, this was not necessary for two reasons. The first reason is that our participants did not know we were using WOZ implementation in our virtual meeting platform. So they though the suggested suggestions by the assistant were proposed by an actual AI. Implementing IVA with real AI and more customization capabilities will be our future work.

Although, in the formative study, most of our participants reported that they have meetings with more than two persons in their virtual meetings, we decided to evaluate our system with only two persons in our mock-up meeting. Having more than one participant means less control over the experimental design and scenario, which can be translated to more complexity. Hence, we decided to pick only one participant per meeting to evaluate our system. Also, To ensure a smooth and controlled meeting, we needed a researcher present, especially since the study and session were conducted online. Having the researcher join the online meeting was the best solution considering that participants could not join two meetings simultaneously. However, the scenario for the evaluation study was a highly collaborative interactive scenario between the participant and the researcher, which introduces biases for sure.

Table 4.1: Formative study interview questions.

Part 1
Q1: How do your online meetings look like?
Q2: What are the main objectives of your meetings? (e.g., manager/employee talking about a project, student/supervisor talking about a research paper, etc.)
Q3: How many meetings have you had in the previous week?
Q4: How many of these meetings are scheduled meetings?
Q5: What proportion of your meetings is one-to-one, and how many attendees attend the group meetings?
Part 2
Q6: Do you take notes or draw (of any form - physical or digital) during virtual meetings?
Q6.1: What are the tools you use to take notes (and/or draw)?
Q6.2: What triggers you to take notes (and/or draw)?
Q6.3: Could you explain at least two scenarios?
Q7: What other tools are you using during online meetings? (e.g., browser, calendar, task management application, presentation, PDF, etc.)
Q7.1: Could you explain some examples for each tool?
Q7.2 What triggers you to take notes (and/or What triggers you to use these tools?
Q8: Do you multitask in your meetings? (if yes, at least two scenarios)
Q8.1: What are the tools and how you arrange these tools while multitasking?
Q8.2: Assume you have the online meeting, note taking, drawing applications and browser. How would you arrange these windows on one display? Why?
Q9: What proportion of your meetings are collaborative meetings?
Q9.1: How do you communicate with your collaborators in your online meetings?
Q9.2: What are the tools you are using for collaboration.
Part 3
Q10: What are your current problems and concerns with existing online meeting platforms?
Q11: How a smart assistant can help you resolve your concerns during virtual meetings? (at least five different examples)
Q12: What are the tasks and services a smart assistant can do automatically? (at least five examples)
Q13: How would you want to trigger this assistant? (at least five examples)
Q13.1: Actively listening to the content of the meeting and proposing suggestions automatically?
Q13.2: Using button or wake word to trigger the assistant?

Table 4.2: Evaluation study questionnaire and interview questions.

Likert Scale Questions
Q1: I think that I would like to use this system frequently.
Q2: I found the system unnecessarily complex.
Q3: I thought the system was easy to use.
Q4: I think that I would need the support of a technical person to be able to use this system.
Q5: I found the various functions in this system were well integrated.
Q6: I thought there was too much inconsistency in this system.
Q7: I would imagine that most people would learn to use this system very quickly.
Q8: I found the system very cumbersome/awkward to use.
Q9: I felt very confident using the system.
Q10: I needed to learn a lot of things before I could get going with this system.
Q11: I could find the features that I was looking for quickly in the interface.
Q12: I found the infinite canvas useful during the meeting.
Q13: I found suggested tools useful during the meeting.

Interview Questions
Q14: What are the pros and cons of IVA?
Q15: How would you improve IVA?
Q16: How IVA can help you in your virtual meetings?
Q17: Explain if you have privacy concerns.
Q18: Which one of the features of the system do you think would be the most useful feature for you? why?
Q19: How would you compare our System with your current virtual meeting platform?
Q20: Did you use the suggested tools from the assistant? How would you describe your experience with the assistant?
Q21: Did you find the suggestions from the assistant disturbing? If yes, how would you improve it? If no, please elaborate on the experience.
Q22: How would you compare our system with your current virtual meeting platform in terms of multi-tasking and collaboration?
Q23: How would you compare your performance and efficiency compared to existing online meeting platforms?
Q24: How did you find our system distracting compared to doing

Chapter 5

Conclusion

Informed by the literature and a formative study with UX experts, we designed a visual analytics tool, CoUX, to support UX practitioners in both independently analyzing a usability test video and collaborating. CoUX extracts acoustic, textual, and visual features from think-aloud videos using machine learning and includes a chatbox-like interface for problem annotation and discussion, among others. We conducted an exploratory user study with six pairs of UX practitioners in collaborative video analysis tasks. The results show that CoUX helped them improve the completeness and reliability of their analyses. The results also show that different features allowed them to spot problems they might otherwise have neglected and to have focused conversations to seek clarification from and respond to their partners. In sum, our work has taken the first step to create an integrated environment to support the analysis and collaboration of usability test videos among UX practitioners and highlighted further research directions.

For the second system to better support collaboration in meetings, we presented IVA, an intelligent assistant designed for virtual meetings that can automatically suggest tools and services during an online meeting, to enhance the efficiency and performance of the meeting attendees. We first conducted an exploratory formative study to learn users' practices and challenges in online meetings. The results reveal that users struggle with multitasking and collaboration while they are in the meeting switching their focus between different independent applications. The results also allow us to identify users' tasks, which could be done automatically (fully or partially) by an intelligent assistant during remote meetings. We thus considered how an AI-based assistant could improve users' performance and efficiency and designed IVA. Moreover, our second study evaluated IVA using a WOZ implementation. The results show that our participants found the suggested tools and services relevant, useful, easy-to-use, and time-saving by preventing the users from spending too much time outside the meeting and performing various tasks using different tools.

References

[1] Video conferencing statistics 2022, Jul 2022.

[2] Anand Agarawala and Ravin Balakrishnan. Keepin' it real: Pushing the desktop metaphor with physics, piles and the pen. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, page 1283–1292, New York, NY, USA, 2006. Association for Computing Machinery.

[3] B.B. Allison, Marissa Shuffler, and A.M. Wallace. *The successful facilitation of virtual team meetings*, pages 680–705. 01 2015.

[4] Terence Andre, H Hartson, and Robert Williges. Determining the Effectiveness of the Usability Problem Inspector: A Theory-Based Model and Tool for Finding Usability Problems. *Human factors*, 45:455–82, February 2003.

[5] Maksmi Belitski and Marcello Mariani. The effect of knowledge collaboration on business model reconfiguration. *European Management Journal*, 2022.

[6] Raquel Benbunan-Fich. Using protocol analysis to evaluate the usability of a commercial web site. *Information & management*, 39(2):151–163, 2001.

[7] Xiaojun Bi, Seok-Hyung Bae, and Ravin Balakrishnan. Walltop: Managing overflowing windows on a large display. *Human-Computer Interaction*, 29:153–203, 03 2014.

[8] T. Blascheck, M. John, K. Kurzhals, S. Koch, and T. Ertl. VA2: A Visual Analytics Approach for Evaluating Visual Analytics Applications. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):61–70, January 2016.

[9] James Brill. React-speech-recognition, 2022.

[10] Frederik Brudy, David Ledo, Michel Pahud, Nathalie Henry Riche, Christian Holz, Anand Waghmare, Hemant Bhaskar Surale, Marcus Peinado, Xiaokuan Zhang, Shannon Joyner, Badrish Chandramouli, Umar Farooq Minhas, Jonathan Goldstein, William Buxton, and Ken Hinckley. Surfacefleet: Exploring distributed interactions unbounded from device, application, user, and time. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, UIST '20, page 7–21, New York, NY, USA, 2020. Association for Computing Machinery.

[11] Angelika Bullinger-Hoffmann, Michael Koch, Kathrin Möslein, and Alexander Richter. Computer-supported cooperative work – revisited. *i-com*, 20(3):215–228, 2021.

[12] Carrie J. Cai, Samantha Winter, David Steiner, Lauren Wilcox, and Michael Terry. "hello ai": Uncovering the onboarding needs of medical practitioners for human-ai collaborative decision-making. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW), nov 2019.

[13] Guendalina Caldarini, Sardar Jaf, and Kenneth McGarry. A literature survey of recent advances in chatbots. *Information*, 13(1), 2022.

[14] Hancheng Cao, Chia-Jung Lee, Shamsi Iqbal, Mary Czerwinski, Priscilla N Y Wong, Sean Rintel, Brent Hecht, Jaime Teevan, and Longqi Yang. Large scale analysis of multitasking behavior during remote meetings. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery.

[15] Brandon Castellano. Scenedetect: A cross-platform, OpenCV-based video scene detection program and Python library.

[16] Parmit K. Chilana, Jacob O. Wobbrock, and Andrew J. Ko. Understanding usability practices in complex domains. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems - CHI '10*, pages 2337–2346, Atlanta, Georgia, USA, 2010. ACM Press.

[17] L. Cooke. Assessing Concurrent Think-Aloud Protocol as a Usability Test Method: A Technical Communication Approach. *IEEE Transactions on Professional Communication*, 53(3):202–215, September 2010.

[18] Ana-Paula Correia, Chenxi Liu, and Fan Xu. Evaluating videoconferencing systems for the quality of the educational experience evaluating videoconferencing systems for the quality of the educational experience. *Distance Education*, 41, 09 2020.

[19] Ana-Paula Correia, Chenxi Liu, and Fan Xu. Evaluating videoconferencing systems for the quality of the educational experience evaluating videoconferencing systems for the quality of the educational experience. *Distance Education*, 41, 09 2020.

[20] Mary Czerwinski, Eric Horvitz, and Susan Wilhite. A diary study of task switching and interruptions. *Conference on Human Factors in Computing Systems - Proceedings*, 6, 02 2004.

[21] Claudéric Demers. Dnd kit – a modern drag and drop toolkit for react, 2022.

[22] Wenwen Dou, Dong Hyun Jeong, Felesia Stukes, William Ribarsky, Heather Richter Lipford, and Remco Chang. Recovering Reasoning Processes from User Interactions. *IEEE computer graphics and applications*, 29(3):52–61, May 2009.

[23] Bob Evans. The zoom revolution: 10 eye-popping stats from tech's new superstar, Jun 2020.

[24] Mingming Fan, Yue Li, and Khai N. Truong. Automatic Detection of Usability Problem Encounters in Think-aloud Sessions. *ACM Transactions on Interactive Intelligent Systems*, 10(2):1–24, June 2020.

[25] Mingming Fan, Jinglan Lin, Christina Chung, and Khai N. Truong. Concurrent Think-Aloud Verbalizations and Usability Problems. *ACM Transactions on Computer-Human Interaction*, 26(5):1–35, September 2019.

[26] Mingming Fan, Serina Shi, and Khai N Truong. Practices and Challenges of Using Think-Aloud Protocols in Industry: An International Survey. *Journal of Usability Studies*, 15(2):85–102, 2020.

[27] Mingming Fan, Ke Wu, Jian Zhao, Yue Li, Winter Wei, and Khai N. Truong. VisTA: Integrating Machine Intelligence with Visualization to Support the Investigation of Think-Aloud Sessions. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):343–352, January 2020.

[28] Mingming Fan, Xianyou Yang, TszTung Yu, Q. Vera Liao, and Jian Zhao. Human-ai collaboration for ux evaluation: Effects of explanation and synchronization. *Proc. ACM Hum.-Comput. Interact.*, 6(CSCW1), apr 2022.

[29] Mingming Fan, Qiwen Zhao, and Vinita Tibdewal. Older adults' think-aloud verbalizations and speech features for identifying user experience problems. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery.

[30] Asbjørn Følstad, Effie Lai-Chong Law, and Kasper Hornb. Analysis in practical usability evaluation: A survey study. In *Proceedings of the 30th SIGCHI Conference on Human Factors in Computing Systems - CHI '12*, pages 2127–2136, Austin, Texas, 2012. ACM Press.

[31] Asbjørn Følstad, Effie Lai-Chong Law, and Kasper Hornbæk. Analysis in usability evaluations: An exploratory study. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, NordiCHI '10, pages 647–650, New York, NY, USA, October 2010. Association for Computing Machinery.

[32] FullStory. FullStory — Robust Analytics, Session Replay, Heatmaps, Dev Tools, and more. https://www.fullstory.com/platform.

[33] Frieda Goldman-Eisler. Psycholinguistics: Experiments in spontaneous speech. 1968.

[34] Gong.Io. Gong.io. https://www.gong.io/, 2021.

[35] Óscar González-Benito, Pablo Muñoz-Gallego, and Evelyn García-Zamora. Role of collaboration in innovation success: differences for large and small businesses. *Journal of Business Economics and Management*, 17:645–662, 07 2016.

[36] Google. Google Sheets: Free Online Spreadsheets for Personal Use. https://www.google.ca/sheets/about/, 2021.

[37] Bart Green and Claire Johnson. Interprofessional collaboration in research, education, and clinical practice: Working together for a better future. *The Journal of chiropractic education*, 29, 01 2015.

[38] Julián Grigera, Alejandra Garrido, José Matías Rivero, and Gustavo Rossi. Automatic detection of usability smells in web applications. *International Journal of Human-Computer Studies*, 97:129–148, 2017.

[39] Gabriel Haas, Michael Rietzler, Matt Jones, and Enrico Rukzio. Keep it short: A comparison of voice assistants' response behavior. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA, 2022. Association for Computing Machinery.

[40] Janine Hacker, Jan vom Brocke, Joshua Handali, Markus Otto, and Johannes Schneider. Virtually in this together – how web-conferencing systems enabled a new virtual togetherness during the covid-19 crisis. *European Journal of Information Systems*, 29(5):563–584, 2020.

[41] Said Hadjerrouit. A collaborative writing approach to wikis: Design, implementation, and evaluation. *Issues in Informing Science and Information Technology Volume*, 8, 01 2011.

[42] Patrick Harms. Automated usability evaluation of virtual reality applications. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 26(3):1–36, 2019.

[43] Martha A. Harsanyi. Multiple authors, multiple problems: bibliometrics and the study of scholarly collaboration: a literature review. *Library & Information Science Research*, 15:325–354, 1993.

[44] Jeffrey Heer, Fernanda B. Viégas, and Martin Wattenberg. Voyagers and voyeurs: Supporting asynchronous collaborative visualization. *Communications of the ACM*, 52(1):87–97, January 2009.

[45] Morten Hertzum and Niels Ebbe Jacobsen. The Evaluator Effect: A Chilling Fact About Usability Evaluation Methods. *International Journal of Human-Computer Interaction*, 15(1):183–204, 2001.

[46] Morten Hertzum, Rolf Molich, and Niels Ebbe Jacobsen. What you get is what you see: Revisiting the evaluator effect in usability tests. *Behaviour & Information Technology*, 33(2):144–162, April 2013.

[47] Kevin Hogan, Annabelle Baer, and James Purtilo. Diplomat: A conversational agent framework for goal-oriented group discussion, 2021.

[48] Jess Hohenstein and Malte Jung. Ai-supported messaging: An investigation of human-human text conversation with ai support. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI EA '18, page 1–6, New York, NY, USA, 2018. Association for Computing Machinery.

[49] C J Hutto and Eric Gilbert. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. In *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media (ICWSM-14)*, pages 216–225, Ann Arbor, MI, 2014.

[50] Shamsi T. Iqbal, Jonathan Grudin, and Eric Horvitz. Peripheral computing during presentations: Perspectives on costs and preferences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, page 891–894, New York, NY, USA, 2011. Association for Computing Machinery.

[51] Katherine Isbister, Hideyuki Nakanishi, Toru Ishida, and Cliff Nass. Helper agent: Designing an assistant for human-human interaction in a virtual meeting space. In *Proceedings of*

the SIGCHI Conference on Human Factors in Computing Systems, CHI '00, page 57–64, New York, NY, USA, 2000. Association for Computing Machinery.

[52] Petra Isenberg and Danyel Fisher. Collaborative brushing and linking for co-located visual analytics of document collections. In *Proceedings of the 11th Eurographics / IEEE - VGTC Conference on Visualization*, EuroVis'09, pages 1031–1038, Chichester, GBR, June 2009. The Eurographs Association & John Wiley & Sons, Ltd.

[53] Yannick Jadoul. Praat-parselmouth: Praat in Python, the Pythonic way. https://parselmouth.readthedocs.io/en/stable/, 2019.

[54] JongWook Jeong, NeungHoe Kim, and Hoh Peter In. Detecting usability problems in mobile applications on the basis of dissimilarity in user behavior. *International Journal of Human-Computer Studies*, 139:102364, 2020.

[55] Eunseo Kim, Jeongmin Hong, Hyuna Lee, and Minsam Ko. Colorbo: Envisioned mandala coloringthrough human-ai collaboration. In *27th International Conference on Intelligent User Interfaces*, IUI '22, page 15–26, New York, NY, USA, 2022. Association for Computing Machinery.

[56] Ned Kock. What is e-collaboration? 1, 01 2005.

[57] Shota KUSAJIMA and Yasuyuki Sumi. Activating group discussion by topic providing bots. *IEICE Transactions on Information and Systems*, E101.D:856–864, 04 2018.

[58] Anastasia Kuzminykh and Sean Rintel. Classification of functional attention in video meetings. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–13, New York, NY, USA, 2020. Association for Computing Machinery.

[59] Anastasia Kuzminykh and Sean Rintel. Low engagement as a deliberate practice of remote participants in video meetings. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, page 1–9, New York, NY, USA, 2020. Association for Computing Machinery.

[60] Marjan Laal and Seyed Mohammad Ghodsi. Benefits of collaborative learning. *Procedia - Social and Behavioral Sciences*, 31:486–490, 2012. World Conference on Learning, Teaching & Administration - 2011.

[61] Clayton Lewis. *Using the "Thinking Aloud" Method in Cognitive Interface Design*. IBM T.J. Watson Research Center, 1982.

[62] Jane Li, Christian Muller-Tomfelde, and Toni Robertson. Designing for distributed scientific collaboration: A case study in an animal health laboratory. In *2012 45th Hawaii International Conference on System Sciences*, pages 373–381, 2012.

[63] Zhuying Li, Yan Wang, Wei Wang, Stefan Greuter, and Florian 'Floyd' Mueller. Empowering a creative city: Engage citizens in creating street art through human-ai collaboration. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, page 1–8, New York, NY, USA, 2020. Association for Computing Machinery.

[64] Tatiana Losev, Sarah Storteboom, Sheelagh Carpendale, and Søren Knudsen. Distributed synchronous visualization design: Challenges and strategies. In *2020 IEEE Workshop on Evaluation and Beyond - Methodological Approaches to Visualization (BELIV)*, pages 1–10, 2020.

[65] N. Mahyar and M. Tory. Supporting communication and coordination in collaborative sensemaking. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1633–1642, 2014.

[66] Jennifer Marlow, Eveline van Everdingen, and Daniel Avrahami. Taking notes or playing games? understanding multitasking in video communication. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, CSCW '16, page 1726–1737, New York, NY, USA, 2016. Association for Computing Machinery.

[67] Thomas Marrinan, Jillian Aurisano, Arthur Nishimoto, Krishna Bharadwaj, Victor Mateevitsi, Luc Renambot, Lance Long, Andrew Johnson, and Jason Leigh. Sage2: A new approach for data intensive collaboration using scalable resolution shared displays. In *10th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing*, pages 177–186, 2014.

[68] Sharon McDonald, Helen M. Edwards, and Tingting Zhao. Exploring Think-Alouds in Usability Testing: An International Survey. *IEEE Transactions on Professional Communication*, 55(1):2–19, March 2012.

[69] Moira McGregor and John C. Tang. More to meetings: Challenges in using speech-based technology to support meetings. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17, page 2208–2220, New York, NY, USA, 2017. Association for Computing Machinery.

[70] O. Miksik, I. Munasinghe, J. Asensio-Cubero, S. Reddy Bethi, S-T. Huang, S. Zylfo, X. Liu, T. Nica, A. Mitrocsak, S. Mezza, R. Beard, R. Shi, R. Ng, P. Mediano, Z. Fountas, S-H.

Lee, J. Medvesek, H. Zhuang, Y. Rogers, and P. Swietojanski. Building proactive voice assistants: When and how (not) to interact, 2020.

[71] Miro. The visual collaboration platform for every team: Miro, 2022.

[72] Kate Moran and Kara Pernice. Remote Moderated Usability Tests: Why to Do Them. https://www.nngroup.com/articles/moderated-remote-usability-test-why/, April 2020.

[73] Omar Mubin, Thomas D'Arcy, Ghulam Murtaza, Simeon Simoff, Chris Stanton, and Catherine Stevens. Active or passive?: Investigating the impact of robot role in meetings. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, pages 580–585, 2014.

[74] Jakob Nielsen. 10 Usability Heuristics for User Interface Design. https://www.nngroup.com/articles/ten-usability-heuristics/, 1994.

[75] Jakob Nielsen. Severity Ratings for Usability Problems. https://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/, 1994.

[76] Jakob Nielsen. Thinking Aloud: The #1 Usability Tool. https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/, 2012.

[77] Noldus. Record & annotate - Recording options and easy annotation. https://www.noldus.com/viso/record-annotate, 2020.

[78] Mie Nørgaard and Kasper Hornbæk. What do usability evaluators do in practice? an explorative study of think-aloud testing. In *Proceedings of the 6th Conference on Designing Interactive Systems*, DIS '06, pages 209–218, New York, NY, USA, June 2006. Association for Computing Machinery.

[79] Don Norman. *The design of everyday things: Revised and expanded edition*. Basic books, 2013.

[80] Minna Nykopp, Miika Marttunen, and Gijsbert Erkens. Coordinating collaborative writing in an online environment. *Journal of Computing in Higher Education*, 31(3):536–556, 2018.

[81] OpenCV. Optical Flow function. https://docs.opencv.org/3.4.13/d4/dee/tutorial_optical_flow.html, 2020.

[82] Asil Oztekin, Dursun Delen, Ali Turkyilmaz, and Selim Zaim. A machine learning-based usability evaluation method for elearning systems. *Decision Support Systems*, 56:63–73, 2013.

[83] Sun Young Park, Pei-Yi Kuo, Andrea Barbarin, Elizabeth Kaziunas, Astrid Chow, Karandeep Singh, Lauren Wilcox, and Walter S. Lasecki. Identifying challenges and opportunities in human-ai collaboration in healthcare. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing*, CSCW '19, page 506–510, New York, NY, USA, 2019. Association for Computing Machinery.

[84] Fabio Paternò, Antonio Giovanni Schiavone, and Antonio Conti. Customizable automatic detection of bad usability smells in mobile accessed web applications. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 1–11, 2017.

[85] Walter Powell, Kenneth Koput, and Laurel Smith-Doerr. Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. *Administrative Science Quarterly*, 41:116–145, 03 1996.

[86] Maciej Pyrc. React-zoom-pan-pinch, 2022.

[87] Jason Quense. React-big-calendar, 2022.

[88] Jeba Rezwana and Mary Lou Maher. Designing creative ai partners with cofi: A framework for modeling interaction in human-ai co-creative systems. *ACM Trans. Comput.-Hum. Interact.*, feb 2022. Just Accepted.

[89] Heather Richter Lipford, Felesia Stukes, Wenwen Dou, Matthew Hawkins, and Remco Chang. Helping Users Recall Their Reasoning Process. In *Proceedings of the IEEE Conference on Visual Analytics Science and Technology*, pages 187–194, Salt Lake City, Utah, USA, October 2010.

[90] Patrick Riehmann, Manfred Hanfler, and Bernd Froehlich. Interactive sankey diagrams. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, pages 233–240. IEEE, 2005.

[91] Rutger Rienks, Anton Nijholt, and Paulo Barthelmess. Pro-active meeting assistants: Attention please! *AI Soc.*, 23(2):213–231, aug 2008.

[92] Anthony C. Robinson. Collaborative synthesis of visual analytic results. In *VAST'08 - IEEE Symposium on Visual Analytics Science and Technology, Proceedings*, pages 67–74, December 2008.

[93] Theerasak Rojanarata. How online whiteboard promotes students' collaborative skills in laboratory learning. In *Proceedings of the 2020 8th International Conference on Information and Education Technology*, ICIET 2020, page 68–72, New York, NY, USA, 2020. Association for Computing Machinery.

[94] Paolo Romano, Rosalba Giugno, and Alfredo Pulvirenti. Tools and collaborative environments for bioinformatics research. *Briefings in bioinformatics*, 12:549–61, 11 2011.

[95] Luc Rubinger, Aaron Gazendam, Seper Ekhtiari, Nicholas Nucci, Abbey Payne, Herman Johal, Vikas Khanduja, and Mohit Bhandari. Maximizing virtual meetings and conferences: a review of best practices. *International Orthopaedics*, 44, 05 2020.

[96] Joshua Rubinstein, David Meyer, and Jeffrey Evans. Executive control of cognitive processes in task switching. *Journal of experimental psychology. Human perception and performance*, 27:763–97, 09 2001.

[97] Maria Saiz-Manzanares, Raúl Sánchez, and Javier Ochoa-Orihuel. Effectiveness of using voice assistants in learning: A study at the time of covid-19. *International Journal of Environmental Research and Public Health*, 17:5618, 08 2020.

[98] Ali Sarvghad and Melanie Tory. Exploiting analysis history to support collaborative data analysis. In *Proceedings of the 41st Graphics Interface Conference*, GI '15, pages 123–130, CAN, June 2015.

[99] Ali Sarvghad, Melanie Tory, and Narges Mahyar. Visualizing Dimension Coverage to Support Exploratory Analysis. *IEEE transactions on visualization and computer graphics*, 23(1):21–30, January 2017.

[100] Andrew Sears. Heuristic Walkthroughs: Finding the Problems Without the Noise. *International Journal of Human-Computer Interaction*, 9(3):213–234, September 1997.

[101] Amine Sehili. Auditok: A module for Audio/Acoustic Activity Detection.

[102] Silverback. Silverback 3. https://silverbackapp.com/, 2019.

[103] Mikael B. Skov and Jan Stage. Supporting problem identification in usability evaluations. In *Proceedings of the 17th Australia Conference on Computer-Human Interaction: Citizens Online: Considerations for Today and the Future*, OZCHI '05, pages 1–9, Narrabundah, AUS, November 2005. Computer-Human Interaction Special Interest Group (CHISIG) of Australia.

[104] Slack. Slack: Where work happens. https://slack.com/.

[105] Ehsan Jahangirzadeh Soure, Emily Kuang, Mingming Fan, and Jian Zhao. Coux: Collaborative visual analysis of think-aloud usability test videos for digital interfaces. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):643–653, 2022.

[106] Siegfried Ludwig Sporer and Barbara Schwandt. Paraverbal indicators of deception: A meta-analytic synthesis. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 20(4):421–446, 2006.

[107] Willem Standaert, Steve Muylle, and Amit Basu. An empirical study of the effectiveness of telepresence as a business meeting mode. *Inf. Technol. and Management*, 17(4):323–339, dec 2016.

[108] Jill Steel, Joanne M. Williams, and Sarah McGeown. Teacher-researcher collaboration in animal-assisted education: Co-designing a reading to dogs intervention. *Educational Research*, 64(1):113–131, 2022.

[109] Mark Stefik, Gregg Foster, Daniel G. Bobrow, Kenneth Kahn, Stan Lanning, and Lucy Suchman. Beyond the chalkboard: Computer support for collaboration and problem solving in meetings. *Commun. ACM*, 30(1):32–47, jan 1987.

[110] Minhyang (Mia) Suh, Frank Bentley, and Danielle Lottridge. ”it’s kind of boring looking at just the face”: How teens multitask during mobile videochat. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW), nov 2018.

[111] Desney S. Tan, Darren Gergle, Peter Scupelli, and Randy Pausch. With similar visual angles, larger displays improve spatial performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’03, page 217–224, New York, NY, USA, 2003. Association for Computing Machinery.

[112] Yla R. Tausczik and James W. Pennebaker. Improving teamwork using real-time language feedback. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’13, page 459–468, New York, NY, USA, 2013. Association for Computing Machinery.

[113] TechSmith. Morae 3 Tutorials. https://www.techsmith.com/tutorial-morae-current.html.

[114] George Terzopoulos and Maya Satratzemi. Voice assistants and smart speakers in everyday life and in education. *Informatics in Education*, 19:473–490, 09 2020.

[115] Philipp Tschandl, Christoph Rinner, Zoe Apalla, Giuseppe Argenziano, Noel Codella, Allan Halpern, Monika Janda, Aimilios Lallas, Caterina Longo, Josep Malvehy, John Paoli, Susana Puig, Cliff Rosendahl, H. Peter Soyer, Iris Zalaudek, and Harald Kittler. Human–computer collaboration for skin cancer recognition. *Nature Medicine*, 26(8):1229–1234, Aug 2020.

[116] Gokhan Tur, Andreas Stolcke, Lynn Voss, Stanley Peters, Dilek Hakkani-Tur, John Dowding, Benoit Favre, Raquel Fernandez, Matthew Frampton, Mike Frandsen, Clint Frederickson, Martin Graciarena, Donald Kintzing, Kyle Leveque, Shane Mason, John Niekrasz, Matthew Purver, Korbinian Riedhammer, Elizabeth Shriberg, Jing Tien, Dimitra Vergyri, and Fan Yang. The calo meeting assistant system. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(6):1601–1611, 2010.

[117] UserTesting. Usertesting: The human insight platform. https://www.usertesting.com/.

[118] Siv Vangen and Chris Huxham. Nurturing collaborative relations: Building trust in interorganizational collaboration. *The Journal of Applied Behavioral Science*, 39:5–31, 03 2003.

[119] F. B. Viegas, M. Wattenberg, F. van Ham, J. Kriss, and M. McKeon. ManyEyes: A Site for Visualization at Internet Scale. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1121–1128, November 2007.

[120] Dakuo Wang, Justin D. Weisz, Michael Muller, Parikshit Ram, Werner Geyer, Casey Dugan, Yla Tausczik, Horst Samulowitz, and Alexander Gray. Human-ai collaboration in data science: Exploring data scientists' perceptions of automated ai. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW), nov 2019.

[121] M. Wattenberg and J. Kriss. Designing for social data analysis. *IEEE Transactions on Visualization and Computer Graphics*, 12(4):549–557, July 2006.

[122] Christopher Wickens. Multiple resources and mental workload. *Human factors*, 50:449–55, 07 2008.

[123] Brooke Wilkerson, Anaely Aguiar, Christina Gkini, Igor Oliveira, Lars-Kristian Trellevik, and Birgit Kopainsky. Reflections on adapting group model building scripts into online workshops. *System Dynamics Review*, 36, 09 2020.

[124] Wesley Willett, Jeffrey Heer, Joseph Hellerstein, and Maneesh Agrawala. CommentSpace: Structured support for collaborative visual analysis. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, pages 3131–3140. ACM, 2011.

[125] Amy X. Zhang, Michael Muller, and Dakuo Wang. How do data science workers collaborate? roles, workflows, and tools. *Proc. ACM Hum.-Comput. Interact.*, 4(CSCW1), may 2020.

[126] Anthony Zhang (Uberi). SpeechRecognition: Library for performing speech recognition, with support for several engines and APIs, online and offline.

[127] Jian Zhao, Michael Glueck, Simon Breslav, Fanny Chevalier, and Azam Khan. Annotation graphs: A graph-based visualization for meta-analysis of data based on user-authored annotations. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):261–270, Jan 2017.

[128] Jian Zhao, Michael Glueck, Petra Isenberg, Fanny Chevalier, and Azam Khan. Supporting handoff in asynchronous collaborative sensemaking using knowledge-transfer graphs. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):340–350, 2017.

[129] Zoom. Zoom - Video Conferencing, Web Conferencing, Webinars, Screen Sharing. https://zoom.us/, 2021.

[130] Dilawar Shah Zwakman, Debajyoti Pal, and Chonlameth Arpnikanondt. Usability evaluation of artificial intelligence-based voice assistants: The case of amazon alexa. *SN Computer Science*, 2, 02 2021.

[131] Önder Nomaler, Koen Frenken, and Gaston Heimeriks. Do more distant collaborations have more citation impact? *Journal of Informetrics*, 7(4):966–971, 2013.