Looking at Yourself on Zoom?

Karolína Balogová
HCI-E MSc Final Project Report 2021
UCL Interaction Centre, University College London
Supervisor: Duncan Brumby

ABSTRACT
Since the COVID-19 pandemic we are all spending prolonged hours on video calls. However, it is unclear how people setup their videoconferencing tools now that they are highly accustomed to this communication medium. What are users' opinions about the default self-view window? How does it affect where they look during a conversation? To address these questions, two studies were conducted. First, an online questionnaire explored 115 users' videoconferencing setup preferences in their current day-to-day. Second, the effect of the self-view window and the role played in conversation (speaking, listening) on gaze allocation was investigated in a small-scale eye-tracking experiment with 3 participants, in a natural educational setting. Results of the survey found that even though most people prefer to leave the self-view window on to ensure there is nothing wrong with the way they are presented, many reported having it there due to not being aware of the possibility to turn it off. Results of the eye-tracking study found that when the self-view is hidden, people do not dwell on other meeting attendees more, but instead spend more time looking away from the screen and engaging in multitasking activities. Self-observing behaviour can be linked to one's active role in the conversation, in that people look at themselves more when they are speaking, which is likely to impact gaze behaviour important for effective turn-taking. These findings imply that videoconferencing tools should make the option to hide the self-view window more readily available.

Author Keywords
gaze allocation; video-mediated communication; self-view window; eye-tracking; turn-taking

ACM Classification Keywords
Information systems---Internet communication tools---Web conferencing

MSc Contribution Type
Empirical

MSC HCI-E FINAL PROJECT REPORT
Project report submitted in part fulfilment of the requirements for the degree of Master of Science (Human-Computer Interaction with Ergonomics) in the Faculty of Brain Sciences, University College London, [YEAR].

1. INTRODUCTION

Due to the COVID-19 pandemic, people are spending more time on video calls than ever; Zoom, one of the most popular videoconferencing tools, has seen an increase in daily meeting participants from 10 million in 2019 to 350 million in 2020 [16]. Video calls have replaced in-person work meetings, physical classrooms and allow us to keep in touch with our loved ones. However, it is speculated that spending prolonged hours on online meetings causes exhaustion, often referred to as "Zoom fatigue" [3,23], which makes it increasingly evident that video-mediated communication (VMC) is far from being equivalent to that of face-to-face (FTF). To tighten the gap between these two communication mediums, it is important to understand the use of VMC in the current context and explore the differences between the two.

Whilst most studies regarding VMC has been conducted at times when videoconferencing was a rarity, there is emerging research focusing on understanding users' behaviour on remote work meetings during the pandemic [23]. A large-scale study on multitasking during remote work meetings found that people spend a considerable amount of time attending to other activities, be it work or non-work related [6]. When it comes to the camera set-up, people were turning their video off more often as working from home became the new norm [5]. The main reasons include connection issues, the effort of having to be constantly attentive and presentable, other attendees turning their video off, feeling self-conscious, and the desire to multitask.

However, it is still not very well understood how people setup other aspects of their video communication tools even outside the context of work, and what the reasons for their preferences are. Do they prefer seeing all the participants aligned on a grid or being focused on the speaker? Most importantly, what are users' opinions about the presence of the self-view window? Do they leave it on, or do they turn it off? How often and when does one look at their own image?

All today's popular video conferencing tools provide the self-view feature by default as most people prefer having a visual feedback available to check on their appearance, especially at the beginning of the call [24]. Several studies investigated the effects of this interface design choice. For example, it was found that it increases self-awareness [25], causes vanity, discomfort and distraction [24], intensifies negative emotions [26], and impairs task performance [10],

but also encourages pro-social behaviour [18] and in certain situations reduces anxiety for high socially anxious individuals [19].

Nonetheless, only a few of these experiments have tracked the eye gaze of participants and it is therefore unclear how the presence of visual feedback directly affects gazing behaviour. With work meetings and educational workshops moving online, video meetings between established teams have become more and more common, however, previous research observed only one-to-one interactions between strangers. Furthermore, given the high cost of eye-trackers all the studies concerning gaze allocation within VMC have been conducted in a laboratory environment using an artificial task scenario, and are thus lacking ecological validity.

To fill the identified research gaps, two studies were conducted. First, an online questionnaire exploring users' videoconferencing setup preferences in their current day-to-day was distributed to the wider audience. We were particularly interested in users' opinions regarding the presence of the self-view window and its subjective impact on their gazing behaviour. The effect of the self-view window on gaze allocation was then empirically validated in a small-scale experiment. This was conducted during a set of regular meetings within a natural educational setting, using remote eye-trackers that were set up on participants' personal computers. The second study was aiming to address the following questions: (1) How does gaze behaviour differ when self-view is present and when it is not? (2) How does the role played in the conversation (speaking, listening) affect gaze allocation in VMC? Given the importance of eye gaze in social interactions, understanding why and to what extent gaze allocation differs in VMC compared to that of FTF could inform the design of videoconferencing tools, and help overcome one of the most frequently reported issue in VMC – the difficulty to take turns [5].

The structure of the dissertation is as follows; First, prior literature on related topics is reviewed, including an overview on video-mediated communication during the pandemic, self-focused attention and turn-taking in VMC, and gaze allocation in FTF social interactions. After describing the methodologies, reporting the findings, and discussing the main takeaways from both studies, a general discussion is held on the limitations and implications of this work.

2. LITERATURE REVIEW

2.1. VMC During the Pandemic

Most studies regarding VMC has been conducted at times when this communication medium was not common. For example, in De Vasconcelos Filho et al.'s (2009) study that was run in 2009, 79% of participants reported using videoconferencing rarely or never. Even pre-pandemic the proficiency and usage patterns of videoconferencing apps were incomparable to those today [28]. Microsoft, one of the research leaders in the sphere of remote and hybrid work meetings has run a comprehensive set of studies investigating the impact of the pandemic on work practices [23]. To learn about employees' experiences specifically in regards to remote meetings, a longitudinal diary study was conducted that collected data from 715 Microsoft employees over the course of 4 months [27].

Multiple of these studies report an increased meeting load during the pandemic [21,23]. Not only there are more of them, but they are also longer than they used to be in-person [5]. As a result of spending prolonged hours on video calls, many employees are feeling exhausted. The main reasons of this "Zoom fatigue" include the lack of non-verbal cues and the inability to read people on video calls, the need to be constantly attentive and presentable, low media quality, the need for cognitive multi-tasking in large group meetings and the lack of break between them [5].

People also reported that in-meeting multitasking became more frequent as work from home became the new normal [6]. Because the extended amount of time spent on meetings, there is less time to do actual work, and office workers are keen to adjust their meeting habits to avoid long working hours. This multitasking behaviour was explored in a large-scale study [6] that collected 4 week-long snapshots thorough February to May 2020 of U.S. Microsoft employees containing metadata on remote meetings including duration, meeting type (ad hoc, scheduled, recurring, broadcast) and size as well as email usage and file edits. It was found that office workers tend to multitask more often in the mornings, during longer, larger and ad-hoc virtual meetings. Extrinsic factors influencing multitasking include catching up on other work, external distractions, and anxiety relief. People engage in both work (checking emails, running scripts, taking notes) and non-work related (eating, exercise, chores) tasks.

This shift in length, nature and behaviour during work meetings also led to people gradually adjusting their videoconferencing setup preferences to better match their needs. For example, it was found that people started turning off their video more and more often as pandemic progressed [5]. The main reasons included connection issues, the effort of having to be constantly attentive and presentable, other attendees turning theirs off, feeling self-conscious, or the desire to multitask [5]. The ability to see others was also perceived as distracting by some, especially during large group meetings and presentations [5]. Video was seen as adding social and engagement value rather than improving work efficiency or the quality of conversation. This is consistent with previous research, according to which VMC is no different to that of audio-only when it comes to turn-taking or conversation structure [22]. This will be further discussed in section 2.3.

Microsoft's comprehensive research provides us with insight into users' behaviour during remote work meetings and their videoconferencing setup preferences in the current context. However, the focus of their efforts has been exclusively on professional work meetings. Nowadays, video calls are also extensively used for personal, educational and wellbeing purposes. Furthermore, there are a few gaps that have not been investigated in much detail. Particularly, it is unclear what users' opinions are about the presence of the self-view window, which is prominently displayed by default in all popular videoconferencing apps and therefore acts as an all-day mirror.

2.2. Self-focused Attention in VMC

The real-time visual feedback of one's appearance is probably the most notable difference of VMC when compared to FTF. De Vasconcelos Filho et al. (2009) were the first ones to measure how often participants look at themselves and examine how comfortable they are with their self-image. Twelve pairs were tasked to complete two 5-minute brainstorming exercises in which they had to plan a social event and a training course. The study was conducted in a lab, with each participant placed in a private room. Eye-tracking data was recorded thorough the session. The participants filled out a post-session questionnaire about their experience with different conditions.

The results of De Vasconcelos Filho et al.'s (2009) study show that 71% of participants wanted to see their feedback video with the main reasons being the desire to know what the other person's view of them was and to ensure they looked alright. This finding was supported by the observation that 55% of glances occurred during the first minute of the task. However, several participants also referred to the video feedback being distracting and making them feel self-conscious. Some participants glanced at themselves more often than others (e.g., one participant looked at themself 133 times, compared to a mean of 29).

Even though eye-tracking data was collected, the results focused mainly on the qualitative insights. There was no statistical analysis nor any comparison of the two experimental conditions (with and without video feedback). Furthermore, as previously already mentioned, compared to today's pandemic situation, video conferencing in 2009 was rarity. There is a need to investigate the behaviour and subjective experience of users highly accustomed to this communication medium.

In a more recent study, Azriel et al. (2020) used eye-tracking within VMC with the aim to understand how socially anxious individuals behave during online video meetings. Visual allocation patterns of low and high socially anxious individuals were investigated in two different contexts: during an interview with a confederate and when delivering a presentation. Both tasks lasted approximately 4 minutes. The study was conducted in a lab. Gaze allocation was attributed within three main areas of the virtual set-up: self-image of the participant, confederate's image, and any other non-face areas. Relative total dwell time and number of fixations per minute were calculated for each.

Azriel et al. (2020) found that participants looked at themselves around 9% of time, which is significantly less than on the confederate (47%) or other non-face areas (44%). Socially anxious individuals reported higher levels of anxiety and demonstrated different gazing behaviour within both tasks. However, they did not look at their self-image more often than the low socially anxious individuals. The potential reasoning for this is that heightened self-focus might be triggered only under certain circumstances (e.g., when being critiqued). The results suggest that gazing behaviour depends not only on individual differences but also context.

However, it is unclear how the presence of the self-view window influences users' gaze behaviour as this variable was not manipulated in Azriel et al.'s (2020) study. It was also focusing on purely one-to-one interactions and is lacking ecological validity, being limited to laboratory environment and short staged tasks. Azriel et al. (2020) hypothesise that the observed gaze allocation patterns would be similar in a FTF setting because participants do not look at themselves too often. However, we have a reason to believe that this is not true. The basis of this hypothesis will be discussed in the next section.

2.3. Turn-taking in VMC

The development of VMC held the promise of enhancing audio-only communication with the many benefits that FTF communication has, including non-verbal cues such as gaze, posture, and facial expressions [22]. However, it has soon became evident and has been empirically validated time and time again that despite the ability to see the head and shoulders of the conversing partner, in many aspects VMC is has a long way to go to be comparable to FTF communication [7,8,22].

One of the early studies investigated the differences of face-to-face, audio-only and video-mediated communication, by comparing the structure of a conversation within a collaborative task [22]. Sellen (1995) found that there is less interactivity in both audio-only and VMC compared to that of FTF, with the interactions having less overlapping speech and being more formal. Having the video present does not add many benefits to the conversation. This is also evident even from users' subjective experiences. According to Microsoft's recent study in which participants were asked their opinions about remote work meetings, turn-taking is the most common interactivity issue in VMC and users are constantly struggling to find strategies to cope with overlapping talk [5].

A reasonable explanation as to why turn-taking suffers in VMC is that most of the commercially available

videoconferencing systems do not support eye contact. This is due to the disparity of the camera position and the location of the meeting attendees' video stream on the screen. In fact, it is completely impossible to infer where others are looking at during an online video conversation, because of the lack of shared spatial information. Given the importance of eye gaze in facilitating turn-taking [1], it has been often assumed that the lack of it in VMC causes friction in the conversation manifesting itself in awkward silences and talking over each other [4].

To validate whether problems with turn-taking in VMC are associated with insufficiently conveyed eye contact information, Mukawa et al.'s (2005) study aimed to compare two systems, with/without eye contact, from psychological and engineering perspective. Mukawa et al. (2005) used an ethnographic methodology and analysed users' verbal and non-verbal cues. Participants' behaviour was compared within the three-step conversation opening protocol (audio-visual link, person identification, conversation acceptance) which has been used in previous studies to compare telephone and face-to-face communication channels.

Mukawa et al. (2005) found that in VMC systems with eye gaze correction that allow for mutual eye contact to take place, the behaviour closely resembles that of face-to-face. Participants have confirmed the audio-visual link by repeatedly establishing and averting eye contact accompanied by smiling. This is consistent with the intimate equilibrium theory according to which prolonged gaze in close proximity is considered rude. On the other hand, in the non-eye-contact system the participants looked at their conversing partner for longer, averted their gaze less often and used exaggerated gestures such as waving and saying "Hi" to draw the attention of their conversing partner.

We now know that both the presence of the self-view window and the inability to make or infer eye-contact during conversation are expected to have an impact on one's natural eye gaze behaviour which plays an important role in FTF social interactions [1]. One of the main limitations of Mukawa et al.'s (2005) study is that the participants' gaze was observed by the researchers and was not tracked via an eye-tracking system. To investigate how both these aspects affect gaze allocation in VMC that would allow us to compare it to FTF communication, a more robust methodology, that is based on the understanding of gaze allocation in FTF social interactions, is needed.

2.4. Gaze Allocation in FTF Social Interactions

Gaze plays a crucial role in social interactions; it provides feedback on how to conversation is going, communicates emotions and personal closeness, and facilitates turn-taking [1].

Existing eye-tracking studies found that gaze allocation in face-to-face spoken communication depends on the task at hand (i.e., speaking, listening), relationship of the conversing partners and broader social context (e.g., role in a co-operative task). Listeners look at their conversing partner's face more often than speakers [13,14]. The person talking ends their turn by looking at their partner whilst the person listening starts their turn initially looking away [14]. In a multi-person interaction, unaddressed participants can anticipate turn-taking and often shift their gaze towards the next speaker before the current stops talking [15].

Hessels et al. (2019) made use of these empirical findings to compare gaze behaviour in a real-time FTF conversation and watching pre-recorded videos. In their first experiment, participants were assigned to a confederate and asked to talk about their work and holiday experience. The conversing partners were in the same room but could see each other through a screen using a state-of-the-art video-mediated set-up that allowed for mutual eye-contact to take place. SMI RED eye trackers were used within the dual eye tracking system. The task was designed to investigate the relationship between gaze allocation to facial features and task structure (speaking, listening) and to examine how the gaze behaviour changes in relation to the conversing partners. The second experiment differed in that participants watched pre-recorded videos instead of engaged in a face-to-face interaction. It was found that the gaze behaviour of a conversational partner influences the other only in a real-time face-to-face conversation.

It is unclear what gazing patterns are present in conversations conducted via today's video conferencing tools that do not allow for mutual eye contact to take place but are conducted in real-time. Hessels et al. (2019) also suggest that longer conversational settings might be more suitable to investigate this turn-taking gazing behaviour. However, taking a similar methodological approach to Hessels et al.'s (2019) study would allow us to investigate how a particular sub-task (speaking or listening) affects gaze allocation in VMC when self-view is present and when it is not, and how this differs to the gaze behaviour observed in FTF communication.

3. CURRENT STUDIES

Previous research suggests that people are constantly adjusting their videoconferencing setup preferences to better match their needs, which have drastically changed during the COVID-19 pandemic [6,21,23]. However, it is unclear how users feel about the presence of the self-view window, which has shown to attract around 9% of one's attention and is expected to have an impact on natural eye gaze behaviour [2]. Gaze plays an important role in turn-taking [1], one of the biggest struggles in VMC [5]. Whilst previous literature has tracked eye gaze of conversing partners during FTF communication to understand the role gaze plays in a conversation [13,14,20], this has not yet

been investigated in VMC with the presence of the self-view window in the equation.

To fill in the above gaps, we now present two studies. First, an online questionnaire is going to explore 115 users' videoconferencing setup preferences in their current day-to-day now that they are highly accustomed to VMC and even outside the context of work. Second, the effect of the self-view window and the role played in conversation (speaking, listening) on gaze allocation is going to be investigated in a small-scale eye-tracking experiment with 3 participants, in a natural educational setting.

4. STUDY 1: VIDEOCONFERENCING TOOL SETUP PREFERENCES

The aim of the first study is to investigate users' videoconferencing setup preferences, with a particular focus on the presence of the self-view window. This is going to be achieved by distributing a short 5-minute online questionnaire to the wider audience.

We are going to address the research questions listed below and based on previous research expect to find the following:

(1) *How do people setup their videoconferencing tool layout, camera and self-view window?*

As pandemic progressed, people turned off their video more often, mainly due to bandwidth issues, having to always look presentable, multitasking and feeling self-conscious [5]. We therefore expect a significant number of users to have their camera turned off as their preferred option for similar reasons that were found in Baym et al.'s (2021) study. When it comes to the self-view window, most people prefer having it available to check on their appearance, especially at the beginning of the call [24]. Given its default presence we assume users got used to this feature and leave it on most of the time. Furthermore, because that the grid view layout leads to increased levels of stress and concentration [5], we would assume the speaker view to be the preferred layout option.

(2) *How distracting, comfortable, and important do users find the presence of the self-view window?*

In a previous study, 71% (17) of participants found the presence of the self-view window important, with the main reason being the desire to know what the conversing partner's view of them was [24]. However, several participants also found the visual feedback distracting [24]. Given that people grew accustomed to seeing themselves as this communication medium became more popular, especially during the pandemic, we would expect fewer participants being distracted by their own self-image.

(3) *How often do people think they engage in activities that distracts their attention from the video call (looking at themselves and multi-tasking)?*

Previous studies suggest that around 30% of remote meetings involve email multitasking [6] and that people tend to look at themselves around 9% of the time [2]. However, none of this research measured the multitasking and self-observing behaviour subjectively. We are assuming that participants would be less willing to admit to both multitasking and looking at themselves during video calls, and therefore expect the self-reported metrics to be lower.

4.1. Method

4.1.1. Participants

115 participants participated in the study (50 females, 62 males, 3 undisclosed). The vast majority (94.8%) were between 18 to 54 years of age, with 25 to 34 years being the largest age category (48.7%). All but one participant reported being a daily user of videoconferencing tools.

4.1.2. Materials

The survey was built using Microsoft Forms. It consisted of seven main sections: (1) About yourself, in which we collected information about users' age and gender; (2) Video communication tools usage, where we asked about the tools our participants use, for what purpose and how much time they spend on for video calls; (3) Video layout setup, in which we gave users to choose their typical video conferencing layout (speaker / gallery view / other) and describe the main reasons behind their preference; (4) Camera setup, where we assessed on a 5-point Likert scale how often they turn their camera and self-view during video calls and provided a freeform field for them to explain the main reasons for doing so; (5) Self-view window setup, which in addition to the type of questions in the previous section also assessed the distraction, comfort and importance of the self-view window on 5-point Likert scales; (6) Looking at yourself and (7) Multitasking where we asked about the frequency users think they look at themselves and multitask during video calls on a 6-point Likert scale and also contained open-ended questions asking about the main reasons for looking at self and the most common multitasking activities.

4.1.3. Procedure

The participants were recruited via convenience and snowball sampling. The survey was advertised on social media sites such as LinkedIn or Twitter. Some participants were found on SurveyCircle, a community for mutual support in online research, and some were directly approached via a personal message.

4.1.4. Data Analysis

The survey results were imported to Excel. Validity of the collected data was checked. None of the responses was excluded from analysis. Basic descriptive statistics was used to analyse the data. For the closed-ended questions, frequency distribution was calculated, and the results were visualised using bar charts. For the open-ended questions, bottom-up thematic analysis was used to identify key codes.

4.1.5. Results

Our results show that people spend a considerable amount of time on video calls in their day-to-day. Almost half of the participants (47%, 54) spend 1 - 3 hours a day on virtual calls on average. 38% (44) reported spending even more time than that (more than 3 hours per day). 14% (16) use videoconferencing tools for less than an hour a day. Only 1 participant is not a daily user.

Video calls are mostly used for professional (90%, 104) and social purposes (88%, 101). 37% (43) of the participants also use them for educational and 17% (20) for health & wellbeing reasons.

According to the survey results, Zoom is by far the most popular videoconferencing tool. Almost all the participants (96%, 110) report using this software for video calls. Its usage is almost twice of Microsoft Teams' (54%, 62), which is closely followed by WhatsApp (53%, 58) and FaceTime (36%, 41). Only 20% (23) of the participants use Skype for video calls. Other tools that were mentioned include Google Meet, Facebook Messenger, Discord, Cisco WebEx, Slack, Hangouts, Jitsi and a few others.

Video Layout Preferences

Participants prefer gallery view (i.e., all meeting attendees are of equal size, aligned on a grid) over speaker view (i.e., the person speaking is the largest). 63% (73) chose gallery view as their typical video layout setup as opposed to 30% (34) opting for the speaker view. Alternative setups mentioned include screensharing view as well as having gallery and speaker views in two separate windows.

However, multiple participants explicitly stated that their preference hugely varies depending on the context of the meeting. Speaker view is preferable for large meetings or when *"there is only one or two people talking/presenting"* (P24), such as online lectures. Gallery view, on the other hand, is preferred for social calls and smaller meetings requiring collaboration. To give an idea, one participant commented the following in the provided freeform field:

> *"Speaker view is my main layout setup as I want to focus on the person speaking (for example, during an online seminar). However, when working in small groups (3-8 people), gallery view is preferred so that I can focus on all members of the group at the same time. It also allows for more cohesive collaboration between group members." (P98)*

Most participants opting for the speaker view justified their preference by stating they *"wish to focus on the person speaking"* (P14), rather *"than watching other people listening"* (P111). A few were also the opinion that this layout is less distracting because it is *"easier to focus on one persona at a time than many at once"* (P46) and because there is not a *"large box of me"* (P71). However, several dislike the *"jumpiness of the speaker view always changing"* (P52), which is one of the reasons they switch to the gallery layout.

The main reason for choosing gallery view is the desire to see all the call attendees, their reactions, expressions, and body language, which makes it easier to *"involve people who want to intervene in the conversation"* (P7), *"gauge interest level of the audience"* (P31, P51), and *"read the room better"* (P56, P59, P102). Participants referred to this setup as being more natural, realistic, and often compared it to *"being in a room of people when you're in person"* (P60). Interestingly, whilst few are opting for the speaker view because they do not want to be looking at themselves *("I don't want to see a large box of me"* – P71), others are choosing gallery view with a larger self-view window for the exact opposite reason *("I like to see everyone in the session, myself included."* – P110).

Camera Preferences

Figure 1 shows responses to the questions asking how often participants turn their camera and self-view off during video calls. It shows that half the participants (57) reported having their camera usually or always on. 29% (33) turn their video off about half the time, and 22% (25) rarely or never show their video feed to other meeting attendees.

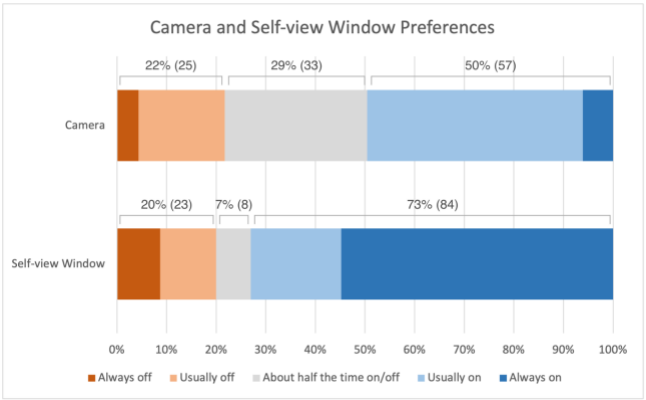

Figure 1. Camera and self-view window preferences

The main reason for turning off the camera is the desire to multitask. Eating was a commonly mentioned side activity associated with disabling video. Whilst multitasking is often voluntary, many blamed interruptions as a trigger for attending to other activities.

"…I'm doing stuff I'm not supposed to do during the call (being on my phone, cleaning my room, etc.)" (P43)

"Need to address something that is outside of work needs or bio break or snack or something quick like that." (P57)

"If eating or being disrupted during the meeting" (P63)

"When I turn off my camera, it means that I was interrupted by other things." (P12)

Some participants elaborated on this behaviour by stating that they *"don't want to distract others"* when doing *"multiple activities whilst also listening"* (P34). Others simply want to avoid looking disrespectful by not *"having 100% of my attention on the screen"* (P3).

"Multi-tasking at times so it doesn't look good if the other person thinks we are not concentrating on what they are saying." (P48)

"It's mainly when I am interrupted by an email or slack message that needs my immediate attention, but I don't want others think that I am not paying attention" (P11)

Another common reason why participants do not turn on their camera is because *"everyone else has it off"* (P38, P40, P104), either not to *"apply pressure"* (P34) or to avoid *"being the only one with the camera on"* (P67), *"dominate other peoples' screens"* (P97), and as a result end up feeling *"like a fish in the bowl"* (P34). A few shared this concern over having *"a lot of eyes on you"* (P101):

"Dislike if being constantly observed. After all, if somebody kept staring at you in real life it would be disturbing." (P41).

"I think it's creepy if people can see me staring at my computer screen, especially if they are strangers." (P106)

Several do not feel the need to be visible when they act *"as an observer"* (P44) more than a *"key participant"* (P22) and are *"just listening and not speaking in the meeting"* (P60). This is often the case during large meetings.

"I will turn my camera off if the meeting has a large number of attendees and my participation is not really required. Usually during town hall style meetings." (P75)

"If a meeting has too many attendees and my main goal is just to listen, I prefer to save energy." (P77)

Another reason for turning off the camera include the *"need to get up from the desk"* (P106). One of the participants elaborates on this behaviour as follows:

"If I have 7 hours of meetings in one day, I don't want to be locked in to one position. With video on, I can't even really shift my body position too much! I turn the camera off to get water from the kitchen, take one of my calls from the couch, walk around, etc." (P86)

Others prefer not to share their video feed when they are *"not looking presentable"* (P14, P26, P38, P43, P56, P91), are experiencing *"bandwidth (connection) issues"* (P16), or are concerned over their privacy and environment, such as being in a *"busy home environment"* (P70), *"people passing through"* (P26) or having an *"unorganised background"* (P69).

A few participants expressed a strong preference for having the video on and emphasized its importance when it comes to maintaining connections.

"In the new normal of remote working, I'd prefer if everyone had their cameras on most of the time to create that level of connection, so I make sure I keep mine on except for rare occasions (eating lunch, etc.)" (P52)

"…I generally try to keep my camera on to help improve relationships" (P86)

On the other hand, some have expressed completely opposite opinions by stressing the unimportance of the visual channel during virtual calls.

"Don't care to look at myself, and no need to visually see each other when the main reason is to hear voice communications." (P46)

Finally, a small portion of participants turn off their video feed because they find the self-view window distracting and do not want to be looking at their own reflection during the call.

"Seeing yourself is really distracting…" (P10)

"I do not enjoy seeing what I look like to other people." (P19)

"I get tired of looking at myself and being self-conscious about my expressions and body language." (P3)

"I hate seeing my own face all meeting..." (P8)

Self-view Window Preferences

Despite some of the negative comments about this feature, most participants are comfortable with the presence of their own visual feedback (56%, 64) and find it only slightly or not distracting at all (60%, 69). However, a considerable 25% (29) of participants still find it very of fairly distracting. On the other hand, only 12% (14) have strong negative opinions when it comes to their comfort levels with the self-view window. (See Figures 2 and 3)

However, the perceived importance of the self-view window shows mixed results (see Figure 4). 38% (44) find the self-view window of only slight or no importance at all.

28% (32) believe its presence is important and 34% (39) are the opinion that it is indeed fairly or very important.

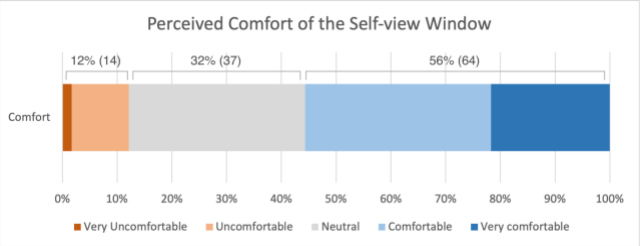

Figure 2. Perceived comfort of the self-view window

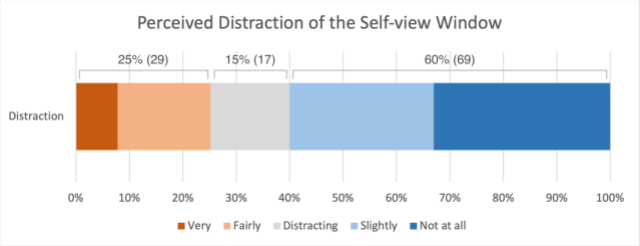

Figure 3. Perceived distraction of the self-view window

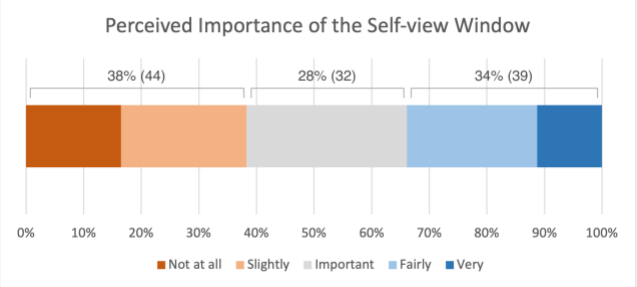

Figure 4. Perceived importance of the self-view window

Compared to the perceived discomfort, distraction and importance of the self-view window, a surprisingly large number of participants leave this feature untouched (See Figure 1). 55% (63) always leave their self-view on, with a further 18% (21) having it usually present. 7% (8) reported having their video feedback on about half the time. Only 20% (23) prefer turning it off. However, this might include cases when the self-view window is not present due to the camera being turned off, suggesting that only a very few participants make an active effort to disable their self-view window whilst still projecting their video feed to other meeting attendees.

A reasonable explanation for why the number of people leaving the self-view window on is so high could be the fact that many participants reported not knowing about or not having the possibility to turn the feature off. In fact, it is one of the most mentioned reasons for leaving the video feedback on. For example, P20 stated *"I did not know you could do that might be doing it from now on!"* Furthermore, many do not mind having the self-view window present and therefore they *"don't feel the need to bother"* (P52) to change this default setting.

Most participants, however, leave their self-view on because they find it important to know how other people see them. They use it to check up on their appearance, environment, camera angle and lighting. According to one participant, *"it's easy to look not good in video calls"* (P42) and therefore it is useful to have an *"easy self-check when you need it"* (P16).

> *"I like to check how other people on the call see me and make sure they do not see things I do not want them to see (me playing on my phone, my messy room or me wearing pyjama bottoms"* (P43)

> *"It's nice to see what exact angle my camera is facing. What if there is a light reflection/glare effecting my presence on camera? It's nice to be able to adjust based on the picture others are seeing."* (P25)

> *"I like to make sure, I can see what my attendees can see. What if I had a booger...the horror!"* (P45)

In most of these answers there is an underlying fear that there might be something wrong with the way one is presented on camera. Some explicitly expressed these feelings and attempted to explain them as follows:

> *"It makes absolutely no sense because I don't carry a mirror with me when I speak with people in-person, but I just feel uncomfortable when I'm on a call and I can't see self-view…"* (P106)

> *"Honestly I'm not too sure. As self-conscious as I am about how I look, the idea of not knowing somehow makes me more paranoid."* (P3)

> "*Helps alleviate uncertainty about my appearance, but also makes me more aware of it and it is a distraction."* (P40)

It is therefore reasonable to conclude that for most participants self-view acts as a reminder that they are on camera which prevents them from embarrassments and inappropriate behaviour.

Besides this reason, some mentioned that the self-view window helps them monitor their reactions and control impressions.

> *"I like to see how others are seeing me. During meetings I want to project confidence and competence. Being able to see myself during the meeting gives me a good gauge for what kind of persona I am projecting…"* (P59)

> *"It's interesting to see my reactions. I realize I do not have a very good poker face."* (P86)

Finally, according to few it even makes them more focused (*"I think looking at myself will help me become more focused."* – P12) and confident (*"Feel more confident speaking if I can see my video feed."* – P112)

On the other hand, those who typically turn their self-view window off justified their preference by stating that its presence prevents them from focusing on other video call attendees.

> *"I find it distracting to constantly be confronted with the sight of my own face on my screen while I am trying to focus on other people. I also feel self-conscious and uncomfortable about it whenever I do accidentally catch sight of myself, so keeping the self-view window off allows me to avoid that."* (P98)

Some said they turn the self-view window off because they prefer fitting more people on the screen. For example, P104 said *"Limited number of participants can show in the gallery view when someone is screen sharing. I'd rather not waste a spot for myself."* Some simply do not want to keep looking at themselves *("I am not interested in looking at myself when I am speaking/listening"* – P47). Finally, a few stated that they find the self-view window useful only at the beginning of the call.

> *"I will be shown how my screen looks when I join the call, and I don't need to keep viewing it after that."* (P74)

> *"... Realistically, I think I only really need self-view at the beginning of the meeting to make sure I'm clearly visible and that my background is not distracting."* (P58)

Multitasking and Looking at Self

According to their own judgement, our participants spend a considerable amount of time engaging in activities that distracts their attention from the video call (see Figure 5). 57% (66) said they engage in frequent multitasking behaviour and a further 32% (37) admitted to occasionally engage in other side-activities. Only 10% (12) stated that they rarely or never multitask. Similar results are shown for the frequency of self-observing behaviour, with 44% (51) looking at themselves very frequently or frequently, 38% (44) occasionally and 17% (20) rarely, very rarely or never.

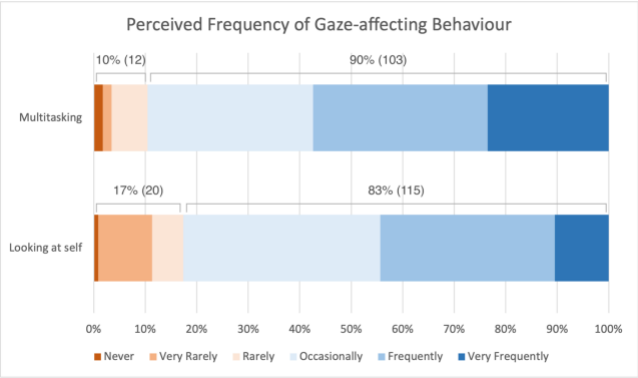

Figure 5. Perceived frequency of gaze-affecting behaviour

The most common side-activities include checking or replying to emails, sending messages to colleagues, engaging in work-related or call-related tasks, note-taking, and web-browsing. A few also mentioned using their phone, checking the news or social media, eating and doing chores.

When we asked our participants to explain the main reasons behind looking at themselves, apart from the already mentioned appearance checks and attempts to avoid embarrassments, many could not clearly explain this commonly occurring phenomenon. It was often described as a *"reflex"* (P7, P76), something that *"happens unconsciously, like looking in a mirror"* (P95). Others attributed this naturally occurring act to vanity *("I just look so good :)"* – P34) and boredom.

> *"I think this is a natural reaction, I can't really help it, or explain it. My best attempt to verbalise my reasons is that it's a somewhat self-regulatory action to make sure I am not looking silly."* (P11)

> *"Again, I'm not really sure, maybe because like most people I'm slightly self-obsessed."* (P106)

A few particularly self-conscious and attentive individuals have noticed that they tend to look at themselves mainly when speaking.

> *"Not on purpose, but my eye is drawn there, particularly when I speak. I use my hands a lot, and I guess the movement of my hands draws my attention. Not intending to, just end up doing it."* (P97)

> *"I find it easier to articulate a point if I can see myself forming the words."* (P26)

4.1.6. Discussion

Our findings confirmed that the proficiency and usage patterns of videoconferencing tools have drastically changed over the years. Compared to De Vasconcelos Filho et al.'s (2009) study in which 79% of participants reported using VMC rarely or never, only one participant of our study reported not being a daily user of videoconferencing tools. VMC currently dominates in many aspects of our lives, being used not for professional, but also social, education and health & wellbeing purposes.

The preferences of the videoconferencing layout vary depending on the context of the meeting, with the speaker view (i.e., the person speaking is the largest) being chosen for large meetings with a single key presenter, and the grid view (i.e., all meeting attendees are of equal size, aligned on a grid) being preferred for social calls and meetings requiring collaboration. However, overall grid view scored higher compared to speaker view. This is surprising given Baym et al. (2021) found the layout is causing stress and requires higher concentration due to having to focus on many personas at once. Nevertheless, our results show that

people prefer this setup because it allows them to gauge the interest levels of the audience and feels more realistic. What is more, the speaker view can be as much if not more distracting due to its constantly changing nature. It is worth noting that Teams' new layout, "Together Mode", holds the promise of overcoming some of the existing setups' shortcomings [5].

Despite previous research showing that people are becoming more and more reluctant to turn their cameras on during virtual calls [5], our results show that only a few opt for turning their video feed off in their usual day-to-day. The reasons for doing so are consistent with Baym et al.'s (2021) study and include the desire to multitask, not wanting to be seen on camera eating, other participants having it off, having to look constantly attentive and presentable, wanting to move away from desk, connection issues, concerns over the background, and not wanting to look at self. The presence of the video seems to be important mainly for maintaining connections rather than improving communication.

Our results regarding the perceived importance, comfort and distraction of the self-view window are consistent with De Vasconcelos Filho et al.'s (2009) study, despite it being conducted over 10 years ago. Most people are comfortable with the presence of their own visual feedback and find it only slightly or not distracting at all. However, the perceived importance of the self-view window shows mixed results. A few have mentioned that they find the self-view window important only at the beginning of the call. We would have expected this answer to be more prominent based on De Vasconcelos Filho et al.'s (2009) finding that 55% of glances to self occur in the first minute of the call. Given the variety of opinions about the importance of the self-view window and that a considerable number of people still find it distracting suggests that it is not simply a matter of getting used to one's visual feedback over time, but that there are indeed individual differences worth investigating.

We found that only a very few users make an active effort to disable their self-view window whilst still projecting their video feed to other meeting attendees. For most, self-view acts as an easy self-check, provides a sense of security and acts as a reminder of one's manners in front of the camera. Surprisingly however, many do not even know about the possibility to turn the feature off. Indeed, many videoconferencing tools do not even provide the ability to hide the self-view window (e.g., Teams, Skype), and some make the option rather difficult to access (e.g., Zoom). Whilst most people do not mind the presence of the self-view window, some find looking at themselves undesirable, stating it prevents them from focusing on other video call attendees. It is therefore questionable why videoconferencing tools do not make the option to hide the self-view window more readily available.

Consistent with previous research, people spend a considerable amount of time engaging in activities that distracts their attention from the video call, both multitasking and looking at themselves. The most common multitasking activities include checking or replying to emails, sending messages to colleagues, engaging in work-related or call-related tasks, note-taking, and web-browsing. Interestingly, looking at one's self-image was often described as a reflex, or something that happens unconsciously, like looking in a mirror. This suggests that the self-reported metrics might underestimate the actual self-observing behaviour. A few particularly self-conscious and attentive individuals have noticed that they tend to look at themselves mainly when speaking, but it is unclear whether this is a general pattern or not.

The current study raises a lot of unanswered questions regarding gazing behaviour in VMC, which might significantly differ to FTF communication. Not only it is impossible to establish eye contact during virtual calls, but people also spend a significant amount of time multitasking and looking at self. However, we have a reason to believe that the self-reported metrics on the frequency of these activities might not reflect the reality. How often do people actually look at themselves during a video call? Does the presence of the self-view window prevent users from focusing on other video call attendees, as suggested by some? Furthermore, because glances to self seem to occur unconsciously, it is difficult to determine any observable patterns in gazing behaviour from self-reported data. Do people look at their self-image mainly at the beginning of the call, or when speaking? We are now going to present a small-scale eye-tracking study which is aiming to find answers to these questions.

5. STUDY 2: GAZE ALLOCATION IN VMC

Our first study focused on videoconferencing setup preferences and subjective opinions about the presence of the self-view window. However, it raised some questions regarding gazing behaviour in VMC, which might significantly differ to FTF communication. The aim of the second study is to empirically investigate these gaze allocation patterns in VMC. More specifically, the effect of the self-view window and the role played in conversation (speaking, listening) on gaze allocation is going to be investigated in a small-scale eye-tracking experiment with 3 participants, in a natural educational setting.

We are going to address the research questions listed below and based on previous research expect to find the following:

(1) *How does gaze behaviour differ in VMC when self-view is present and when it is not?*

According to [24], 55% of glances to self occur during the first minute of the meeting. In a recent study, it was

found that people look at their own self-image around 9% of the time [2]. It is expected that these numbers would be similar in the present study. However, none of the previous eye-tracking research has investigated where users allocate their gaze when visual feedback is turned off. Because the presence of the self-view window increases cognitive load and impairs task performance [10], we hypothesise that participants would spend attending most of the time they would normally look at themselves on other meeting attendees. However, given users tendency to multi-task during virtual meetings [6], it is possible that they would also spend a portion of this time doing other side-activities.

(2) How does the role played in the conversation (speaking, listening) affect gaze allocation in VMC?

According to Hessels et al.'s (2019) study on gaze allocation in FTF communication, people dwell one their conversing partner's facial features for a longer total duration when listening compared to when speaking. More specifically, participants looked at others around 80% of time when they were listening compared to 40% when speaking. We hypothesize that this gaze pattern would be also present in VMC. However, Mukawa et al. (2005) observed differences in gaze behaviour when comparing videoconferencing systems with and without eye-contact. In the non-eye-contact system participants looked at their conversing partner for longer and averted their gaze less often. It would be therefore reasonable to assume that the overall percentage of looking at others would be higher in VMC. However, given there are many distractions, including the self-view window and the possibility to multi-task, we are not sure whether this will hold true. Furthermore, in the presence of the self-view window, we assume people would be looking at themselves more when speaking or when there is silence, compared to when listening. This is based under the assumption that people are generally attentive to the speaker.

5.1. Method

5.1.1. Participants

Three UCL students were recruited to participate in the study (3 females, M = 23.7, SD = 1.5). Participants belonged to the same HCI MSc final project group supervised by Professor Duncan Brumby. The group consisted of the supervisor, 2 PhD assistants and 6 MSc students and was meeting regularly over Zoom. Additional inclusion criterion was the ownership of a PC or laptop with a Windows operating system, due to the technical limitations of the eye-tracking equipment. All participants had a normal or corrected-to-normal vision and spoke English as their second language. They were rewarded with a £50 Amazon voucher. Ethical approval was obtained from the UCL Interaction Centre (UCLIC) Research Department's Ethics Committee and participants were asked to give informed consent at the beginning of the study.

5.1.2. Design

The experiment employed a 2x2 within-subject design, in which the presence of the self-view window (on, off) and meeting size (small, large) independent variables were being manipulated. The order of the two experimental conditions was counter balanced. The video layout was always set to gallery view (i.e., the video feed of all meeting attendees is of equal size, aligned on a grid) for all participants and thus acted as a controlled variable. Eye gaze data was recorded thorough the whole duration of the meetings. Fixations were defined as at least 120ms of stable fixation within an area of interest (see Figure 6). Relative total dwell time, average dwell time and number of fixations per minute were being calculated for each AOI. Ecological validity and subjective experience of each experimental condition was assessed in a short post-meeting questionnaire. Demographical data and overall preferences were collected in a survey distributed to the participants at the end of the study period.

5.1.3. Materials

Tobii 5 (2pc) and Tobii 4C (1pc) eye-trackers were used to track the eye gaze of participants. Both are costumer facing eye-trackers engineered for PC gaming. They were chosen because of their affordability and lightweight design, which allowed for easy distribution and set-up on participants' personal computers.

The biggest limitation of the devices is that they do not provide access to raw eye movement data as the high-end eye-trackers for research purposes do. However, the researcher found a unique solution to this constraint. With the use of a free extension called Tobii Ghost (https://gaming.tobii.com/software/ghost/) which *"enables streamers to give their audience the insights of seeing where they're looking in-game or in applications"* and OBS Studio (https://obsproject.com/), an open-source software for video recording and live streaming, it was possible to overlay participants' gaze trace on top of a screen recording. Participants did not see the trace when on the Zoom call and therefore it did not interfere with the ecological validity of the study, but it was visible in the final recording. With the help of a custom Python script written by the researcher, which will be introduced in more detail in section 5.5., we were able to extract the raw eye-movement data from the recordings. This unique approach allowed us to conduct a, to our knowledge first of its kind research study with low-cost eye-trackers.

Both questionnaires were built using Microsoft Forms. in the post-meeting survey, we first asked participants which experimental condition they just experienced. For the self-view off condition, we then asked what they liked and disliked about not being able to see their own reflection during a video call. For the self-view on condition, we also assessed our participants' subjective frequency of self-observing behaviour on a 6-point Likert scale and asked them to describe a concrete occasion when they looked at themselves. After each meeting, regardless of the experimental set-up, subjective frequency of multitasking behaviour was assessed on a 6-point Likert scale and a free-form field was provided for describing the context and reasons for engaging in a specific side-activity. The final section of the questionnaire focused on evaluating the ecological validity of the study, which examined whether (and if yes, how) the typical Zoom set-up of our participants differs from the one they just experienced for the study's purposes. Participants had to rate on a 5-point Likert scale how conscious they were of their gaze being recorded and describe anything that prevented them from behaving naturally, as they would under normal circumstances under this type of meeting.

The post-study questionnaire consisted of only eight questions. Participants were asked about their (1) age, (2, 3) self-view condition preferences for the individual and group meetings, (4) the perceived differences in behaviour or attitudes when self-view was off compared to when it was on, (5) the urge of turning the self-view window back on, (6) activities engaged in when looking off the screen, (7) the frequency at which they disabled their self-view window prior to participating in this study, (8) future intentions of their self-view window set-up after experiencing both conditions as part of the research study.

5.1.4. Procedure

The eye-trackers were delivered to participants in person. The set-up and installation of the eye-trackers on their personal computers took approximately 1 hour with the researcher. This included installing 3 software (Tobii Experience, Tobii Ghost, OBS Studio) necessary for the calibration of the eye-tracker and for burning the gaze trace into a screen recording. Participants were given written instructions on how to calibrate their equipment and how to start and stop the recording.

The data collection was preceded by a week-long pilot study. Its main aim was to ensure that participants are accustomed to and comfortable with setting up the equipment, and that things are running smoothly from a technical point of view. During the data collection period, participants were asked to record their screen together with their eye gaze during their weekly group meetings and one-to-one meetings with their supervisor. The 1-hour weekly group meetings were informal catchups aimed primarily as a social/community activity. Their purpose was to discuss individual progress within a small and supportive group and discuss shared problems. Although there was no formal agenda, the students were encouraged to actively participate. Each student also had a weekly 30-minute one-to-one meeting with their supervisor. For one student these individual meetings were joined by an industry professional due to the nature of their project. These catchups were aimed for students to get feedback on desired aspects of the project. The regular online meetings provided a natural educational setting and were therefore fit for the purposes of the study.

The official study run over 2 weeks. 3x2 group meetings and 3x2 one-to-one meetings were recorded, resulting in approximately 9 hours of video material in total. Participants were asked to use gallery view and turn their self-view window on during one week of the study and to turn it off during the other. They were encouraged to act as naturally as possible during the meetings, which included the freedom to share their screen and multi-task, as long as they did not move the Zoom window out of the recorded screen. The meetings were also recorded on Zoom so that the official transcripts could be obtained via the Cloud service. Participants had the opportunity to review all the recorded material before sharing it with the researcher. They filled in a short questionnaire about their subjective experience after each meeting. Overall preferences and demographical data were collected in a questionnaire distributed to participants at the end of the study period. Both these questionnaires took approximately 5 minutes to complete.

5.1.5. Data Preparation

The screen recordings obtained via the OBS Studio were in *mkv* format, with a resolution below 900px (the exact measures depended on the participants' screen ratio) and a frame rate of 30 frames per second (fps). A Python script was written by the researcher to extract essential metadata for each frame, including the gaze coordinates, gaze target area, and the person currently speaking.

First, the exact position of the eye gaze (represented by a red circle) was detected for each frame using OpenCV real-time computer vision library. The programme then determined which area of interest (AOI) do the gaze coordinates belong to (see Figure 6): (a) Self AOI, (b) Speaker AOI or (c) Other AOI (distinguished by unique IDs), (d) Non-face AOI, or (e) Off-screen AOI. To make this possible, the AOI coordinates were manually measured in Adobe Photoshop and subsequently fed into the Python script. Occasionally, the grid of meeting attendees would shift on the participants' screen due to them resizing the window, multi-tasking, an attendee turning their camera on/off or joining/leaving the call. Because the script could not handle this transition automatically, the screen recording was split using Adobe Premiere Pro every time a change in layout occurred and each of the video sections

was processed by the script separately, before being merged again. Eye gaze data on occasions when the participant switched windows or shared their screen was treated as off the screen. Dwells were defined as a minimum of four consecutive video frames (120 ms) in which the gaze position was on the same AOI [13]. Shorter fixations were excluded from analysis. Dwells of one video frame (33.3 ms) surrounded by dwells on the same AOI were treated as noise and transformed into the AOI of the surrounding dwells [12].

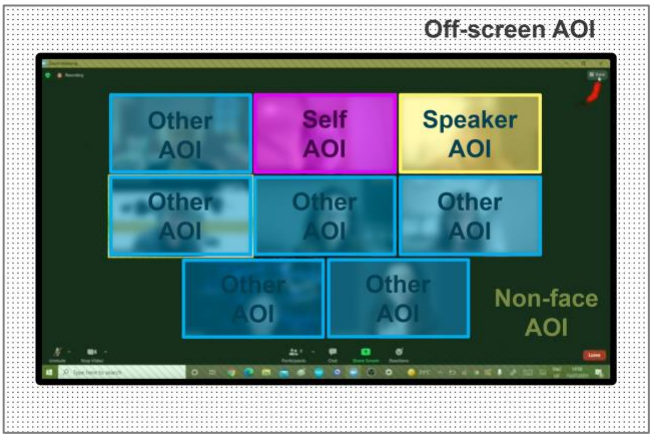

Figure 6. 5 areas of Interest (AOIs) within the Zoom setup – (1) Self AOI, (2) Speaker AOI, (3) Other AOI, (4) Non-face AOI and (5) Off-screen AOI

To determine participant's active role played in the conversation for each frame and differentiate between the Speaker AOI and Other AOI gaze target areas, the *vtt* transcript file was transformed into a list representing a timeline of speakers (distinguished by the same unique IDs as for the gaze metadata), with frames as the main unit, and merged with the rest of the eye gaze metadata. By comparing the gaze target area ID with the speaker ID, we could determine whether the participant was looking at the Speaker AOI or Other AOI. To ensure the data from both sources would be in sync, the transcript and the video containing the participant's gaze trace were first synchronised on a single timeline in Adobe Premiere Pro.

5.1.6. Results

Given the small number of participants we cannot perform inferential statistics on our dataset. We are therefore going to make sense of the gaze data taking a more descriptive approach by simply reporting the means and standard deviations across all the conditions. Our primary measure of interest is relative dwell time (RDT) which was calculated by determining the proportion of time looking at an AOI out of the total meeting time.

Table 1 shows the means and standard deviations of RDT expressed in percentages on each AOI across all four experimental conditions. On average, when the self-view was on, participants looked at themselves 15% (SD = 11.5%, range = 1.1% – 32.7%) of time. They dwelled on the Self AOI less during individual meetings (M = 18.5%, SD = 16%) compared to group meetings (M = 11.6%, SD = 6.4%). This could be expected given the self-view window during a one-to-one meeting is much larger and takes up more screen space. In the self-view off condition the Self AOI was non-existent.

To answer whether the self-view window prevents users from looking at other meeting attendees, we compared the mean RDT on Speaker AOI and Other AOI when self-view was on and when self-view was off across both meeting type conditions. Contrary to expectations, participants attended to the speaker more in the self-view on condition (M = 37.2%, SD = 10.2%) as opposed to the self-view off condition (M = 29.6%, SD = 7.5%). This is shown in Figure 7. Similar pattern is observable for the Other AOI. Participants looked at other non-speaking attendees 14.2% (SD = 8.3%) of the total meeting time when self-view was on as opposed to 13.4% (SD = 8.7%) when self-view was off.

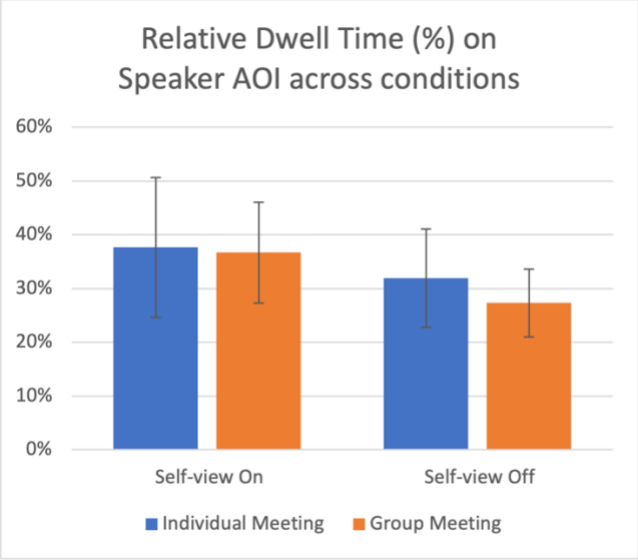

Figure 7. Mean Relative Dwell Time (%) on Speaker AOI across conditions. Error bars depict standard deviations from the mean.

There do not appear to be any notable differences in the mean RDT on Speaker AOI between individual and group meetings. For the Other AOI, participants dwelled on other attendees less during individual meetings (M = 7.6%, SD = 3%) as opposed to group meetings (M = 20%, SD = 6.6%). This could be explained by the fact that in one-to-one meetings dwells to Other AOI could occur only when the participant looked at their conversing partner when there was silence or when the participant was speaking.

Where did participants focus their gaze attention towards when the self-view was off (instead of looking at themselves), if not on other attendees? We observed a large difference in the mean RDT on Off-screen AOI between the

Measure	Meeting Type	Self-view On	Self-view Off
RDT on Self AOI (%)	Individual	18.5% (16%)	-
	Group	11.6% (6.4%)	-
	Total	15.0% (11.5%)	-
RDT on Speaker AOI (%)	Individual	37.7% (13%)	31.9% (9.1%)
	Group	36.7% (9.3%)	27.3% (6.3%)
	Total	37.2% (10.2%)	29.6% (7.5%)
RDT on Other AOI (%)	Individual	8% (4.2%)	7.1% (2.2%)
	Group	20.3% (6.5%)	19.7% (8.1%)
	Total	14.2% (8.3%)	13.4% (8.7%)
RDT on Non-face AOI (%)	Individual	8.2% (2.9%)	9.1% (0.8%)
	Group	8.8% (3.9%)	9% (5%)
	Total	8.5% (3.1%)	9.1% (3.2%)
RDT on Off-screen AOI (%)	Individual	25.8% (21.6%)	46.1% (12.1%)
	Group	19.4% (9.3%)	41.2% (18.9%)
	Total	22.6% (15.3%)	43.6% (14.4%)

Table 1. Means and standard deviations of relative dwell time (RDT) expressed in percentages on each AOI across all four experimental conditions.

two self-view conditions (see Figure 8). Participants spent more time looking away from the screen or attended to a different window when self-view was off (M = 43.6%, SD = 14.4%) compared to when self-view was on (M = 22.6%, SD = 15.3%).

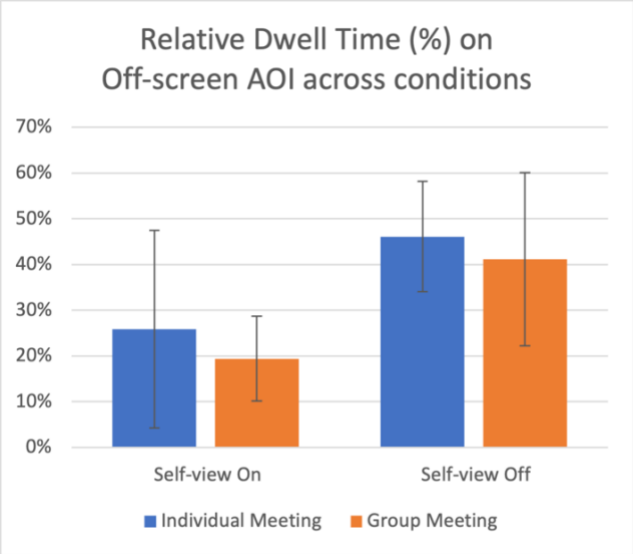

Figure 8. Mean Relative Dwell Time (%) on Off-screen AOI across conditions. Error bars depict standard deviations from the mean.

There are also small but observable differences in the mean RDT on Off-screen AOI between the two meeting types, with an average of 36% (SD = 19.3%) during individual meetings and 30.3% (SD = 17.9%) during group meetings. This is surprising given that people tend to multi-task more during large remote meetings [6].

There do not appear to be any notable differences in the mean RDT on Non-face AOI across any of the conditions. Participants looked at non-face areas of the screen on average 8.5% (SD = 3.1%) of time when self-view was on and 9.1% (SD = 3.2%) of time when self-view was off.

To explore how self-observing behaviour changes over the duration of a meeting, we visualised the proportion of time gaze was allocated to each AOI within every minute for each participant. To then further investigate how gaze allocation differs depending on the role played in a conversation, the graph was complemented with a timeline of the three possible episodes (speaking, listening, silence). Figure 5 depicts P1's gaze allocation to each AOI and their role played in the conversation over time for both individual and group meetings in the self-view on condition.

Contrary to expectations, there is no clear sign that participants attend to their self-image more often at the beginning of the call. Rather, self-observing behaviour can be linked to one's active role in the conversation, in that participants looked at themselves more when they were

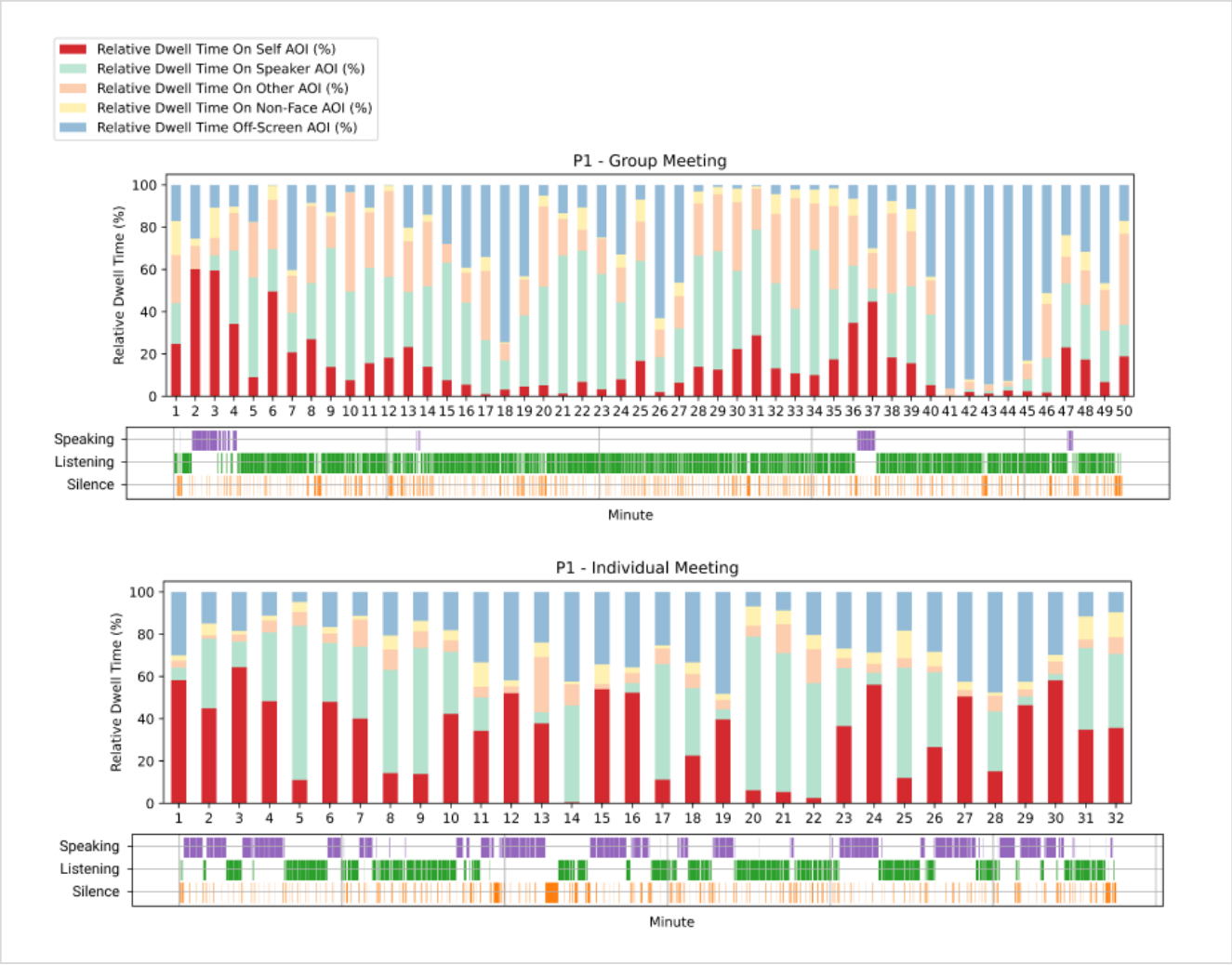

Figure 9. P1's gaze allocation to each AOI and their role played in conversation over time for both individual and group meetings in the self-view on condition.

speaking. This behaviour can be observed on Figure 9. By calculating mean RDT per AOI for each episode (speaking, listening, silence), we have found notable differences between RDT on the Self AOI when speaking (M = 41%, SD = 33.1%) compared to when listening (M = 7.5%, SD = 4.2%) and when there was silence (M = 12.6%, SD = 9.3%). See Table 2 for more details.

Participants dwelled on other meeting attendees (be it the speaker or other listeners) more when listening, compared to when speaking. The presence of the self-view window does not seem to influence this behaviour. When participants could see themselves, they attended to the speaker 56.2% (SD = 19.3%) and to others 8.3% (SD = 9.3%) of their total listening time. In contrast, they attended to others only 18.9% (SD = 16.5%) of their total speaking time. In the self-view off condition (see Table 3), the speaker got 46.6% (SD = 15.7%), and other listeners got 7.1% (SD = 9%) of participants' attention. However, when speaking, they looked at other meeting attendees only 28.1% (SD = 17.5%) of their total speaking time.

Participants dwelled on Off-screen AOI the most when speaking and the least when they were listening. This holds true across both self-view conditions. With self-view on, participants dwelled on Off-screen AOI 29.8% (SD = 24.1%) of their total speaking time, 25.4% (SD = 16.8%) of the total silence time and 17.6% (SD = 12%) of their total listening time. With self-view off, participants dwelled on Off-screen AOI 55.8% (SD = 15.8%) of their total speaking time, 43.6% (SD = 16.4%) of the total silence time and 34.5% (SD = 15.8%) of their total listening time.

There do not appear to be any notable differences in mean RDT on Non-face AOI across episodes.

Looking at the gaze distribution between AOIs during each episode, we have further observed the following; When participants spoke with the self-view on, they dwelled the most on the Self AOI (M = 41%, SD = 33.1%) compared to

other AOIs. However, when the self-view was off, participants spent most of their time speaking looking away from the screen (M = 55.8%, SD = 15.8%). When listening, participants attended to the speaker the most regardless of whether self-view was present (M = 56.2%, SD =19.3%) or not (M = 46.6%, SD = 15.7%). Finally, during silence participant spent most of the time looking at others with self-view on (M = 50.6%, SD = 14.6%). Interestingly, when self-view was off, they dwelled on the Off-screen AOI (M = 43.6%, SD = 16.4%) more.

Measure	Speaking	Listening	Silence
RDT on Self AOI (%)	41% (33.1%)	7.5% (4.2%)	12.6% (9.3%)
RDT on Speaker AOI (%)	-	56.2% (19.3%)	-
RDT on Other AOI (%)	18.9% (16.5%)	8.3% (9.3%)	50.6% (14.6%)
RDT on Non-face AOI (%)	7.3% (4.5%)	8.1% (4%)	8.8% (2.9%)
RDT on Off-screen AOI (%)	29.8% (24.1%)	17.6% (12%)	25.4% (16.8%)

Table 2. Means and standard deviations of Relative Dwell Time (RTD) in percentages per AOI for each episode (speaking, listening, silence) in the self-view on condition.

Measure	Speaking	Listening	Silence
RDT on Self AOI (%)	-	-	-
RDT on Speaker AOI (%)	-	46.6% (15.7%)	-
RDT on Other AOI (%)	28.1% (17.5%)	7.1% (9%)	36% (14.6%)
RDT on Non-face AOI (%)	9.9% (4.3%)	8.5% (3.3%)	11.1% (6.8%)
RDT on Off-screen AOI (%)	55.8% (15.8%)	34.5% (15.8%)	43.6% (16.4%)

Table 3. Means and standard deviations of Relative Dwell Time (RTD) in percentages per AOI for each episode (speaking, listening, silence) in the self-view off condition.

5.1.7. Discussion

Our results show that people look at themselves during video calls around 15% of time, which is slightly higher than previous studies suggest [2]. This discrepancy could be attributed to our small sample size. Furthermore, we found that people dwell on their self-image less during individual meetings compared to group meetings, which is expected given the self-view window in the grid view layout during a one-to-one meeting is larger and takes up more screen space.

In contrast to De Vasconcelos Filho et al.'s (2009) study, our results show no clear sign of people attending to their self-image more often at the beginning of the call. Rather, self-observing behaviour can be linked to one's active role in the conversation, in that people look at themselves more when they are speaking, compared to when listening or when there is silence. What is more, we found that when speaking in the presence of the self-view window people dwell on it almost half the time, which is more than on anything else.

Most importantly, according to our findings people do not spend attending the time they would normally look at themselves on other meeting attendees, as we initially expected. Rather, they spend it looking away from the screen. This gaze allocation pattern is most observable during speaking. When the self-view window is hidden, people spend over half of their time speaking looking away from the screen, which is more than anywhere else. This resembles gaze allocation within FTF communication, in which the speaker's averted gaze indicates to the listener their intention to hold their turn [14]. The fact that this behaviour was observed even within a videoconferencing system that does not support eye contact, and therefore it is not likely to serve its purpose, suggests that this gaze behaviour may be an overlearned response. This tendency to look away when speaking and the fact that our participants spoke more during one-to-one meetings could also explain why there were longer gazes off the screen during individual compared to group meetings.

However, people spend more time looking away from the screen when the self-view is hidden also when listening and during silence. We hypothesise that meeting attendees spend more time multitasking when they do not have their visual feedback present which would help them keep their manners constantly in check and encourage pro-social behaviour [18]. This could either suggest that the self-view window helps users to be more attentive to the virtual conversation by preventing them from multitasking, or that it detracts their attention from the meeting by causing them to worry about their projected image. Based on the mixed opinions about the presence of the self-view window in Study 1 we have a reason to believe that this differs on an individual basis. Given that during silence with self-view on our participants spent most of the time looking at other meeting attendees, possibly trying to infer from their bodily cues who is trying to take the next turn speaking, but with a hidden self-view they were spending most of the time looking off the screen, we are assuming that within our sample the former of the possible effects occurred.

Finally, we found that people dwell on other meeting attendees more when listening compared to when speaking, which is a commonly observed gaze pattern in FTF communication. This behaviour exists regardless of the

presence of the self-view window. However, comparing Hessels et al.'s (2019) results of gaze allocation in FTF communication with ours suggest that the overall percentage of looking at others is lower in VMC. This contradicts the findings of Mukawa et al. (2005)'s lab study, in which participants looked at their conversing partner for longer during a video call in which it was impossible to make or infer eye-contact. We explain this discrepancy by users' tendency to multitask during virtual meetings in their natural environment [6].

6. GENERAL DISCUSSION

Results of Study 1, in which we asked 115 participants about their videoconferencing tool preferences, found that most people prefer to leave the self-view window on to ensure there is nothing wrong with the way they are presented. In fact, the absence of the visual self-check can increase the feelings of discomfort and worry. However, many reported having it there due to not being aware of the possibility to turn it off. Given some still find the self-view window distracting, even after spending extensive amount of time on video calls during the pandemic, suggests that it is not simply a matter of getting used to one's visual feedback over time, but that there are indeed individual differences worth investigating. The fact that looking at one's self-image was often described as something that happens unconsciously, suggests that the self-reported metrics might underestimate the actual self-observing behaviour, and its investigation requires an empirical approach.

Our findings from Study 2, in which we tracked the gaze of 3 participants during a set of one-to-one and group meetings in an educational context, suggest that the presence of the self-view window has multiple effects on gaze allocation within VMC; First, it impacts gaze behaviour important for effective turn-taking by drawing users' attention to their own self-image when speaking. Whilst it is unlikely that this would have any negative effects in videoconferencing systems that do not support eye-contact in the first place, it is likely to impair turn-taking in modern VMC solutions within which it is possible to infer others' eye gaze. Second, the absence of the self-view window causes users to avert their gaze attention away from the screen, likely to engage in multitasking activities.

The findings of this research could inform the design of videoconferencing tools and their self-view window feature. First, given some people find their visual feedback distracting, videoconferencing tools should make the option to hide the self-view window readily available. A more advanced enhancement could include automated detection of movement, position, and environment changes, and only display the self-view window at times when something does not look right (e.g., when the user moves out of the frame) to remind them that they are on a video call. Knowing that one would get informed about undesired events when they occur could decrease the feelings of discomfort and worry that are often present in the absence of the self-view window. At the same time, having the self-view window hidden by default would prevent users from getting distracted by their own image, and avoid the undesired effects that have been previously associated with its presence [10,24–26].

Second, whilst in the current videoconferencing tools turn-taking is already impaired by the inability to make eye-contact, for modern solutions aiming to introduce eye-contact in VMC to be effective in enhancing communication [11,17], one really needs to minimise the presence of the self-view window, especially when speaking. Whilst some find it uncomfortable not being able to observe themselves during video calls when presenting, we hypothesise that this might be a matter of habit more than anything else, as we could not find a reasonable explanation to why looking at oneself during speaking might be important or beneficial. Therefore, we believe that enhancing communication should be prioritised over the initial comfort of the users.

Another implication of this study is around its research methodology. We conducted a, to our knowledge first of its kind research study with low-cost eye-trackers, by implementing a unique programmatic approach to access the raw gaze data customer-facing eye-trackers do have readily available. We are hoping this would inform future work in the area and will make eye-tracking studies more accessible to the research community.

Our study had various limitations. Whilst we complemented the small-scale eye tracking study with survey data that collected responses from the wider audience to increase external validity, the small number of participants of the second study remains the biggest limitation of this research. An obvious way to solve this problem would be to purchase more low-cost eye-trackers that could be distributed to participants' homes. However, these devices are not designed for research purposes and come with many constraints, the lack of compatibility with the MacOS operating system being only one of them. To further emphasize the difficulty to overcome this limitation, it is worth pointing out that the average cost of an eye-tracker is around $17,500 with the low-cost eye-trackers ranging from $100 to $1000 which is still not a small investment [9].

When it comes to the procedure, given the eye-trackers were set-up in the participants home environments, their calibration prior to each meeting was not supervised by the researcher. Despite our participants were given clear instructions on how to calibrate the equipment and record their screen, the gaze data could have been easily compromised due to incorrect calibration, positioning the eye-tracker under the wrong angle, or forgetting to run all the required software.

In terms of methodology, our collected data did not differentiate between off-screen gazes and multitasking behaviour. What is more, eye gaze data on occasions when the participants shared their screen was treated as off the screen as well, although this happened very rarely and is unlikely to compromise our results. However, by treating window-switching, looking away from the screen and screen-sharing as separate events we could have gained a more granular insight into users' gaze allocation patterns in VMC.

The lack of external validity of the eye-tracking study is compensated by its high ecological validity. The length of the meetings allowed participants to get comfortable during the experiment and forget about their gaze being recorded. This was supported by the fact that participants reported being only slightly aware of their gaze being tracked in the post-meeting questionnaires. This research has given us a general idea of what people are doing during virtual meetings in their natural environment.

Future studies could validate these findings with a bigger sample size and address more specific research questions in a controlled laboratory environment. For example, one could investigate in more detail the relationship between the presence of the self-view window and multitasking, measure the emotional state of participants in both visual feedback conditions, or investigate the individual differences of gaze allocation in VMC and how they correlate with users' preferences and attitudes. Another interesting research direction one could take is to investigate the effects of the self-view window over extended amount of time in a longitudinal study. Does it have an impact on our emotional well-being and how do users' attitudes about its presence change over time?

7. CONCLUSION

Two studies were conducted as part of this dissertation. First, an online questionnaire explored 115 users' videoconferencing setup preferences in their current day-to-day. Second, the effect of the self-view window and the role played in conversation (speaking, listening) on gaze allocation was investigated in a small-scale eye-tracking experiment with 3 participants, in a natural educational setting. Results of the survey found that even though most people prefer to leave the self-view window on to ensure there is nothing wrong with the way they are presented, many reported having it there due to not being aware of the possibility to turn it off. Given some still find the self-view window distracting suggests that it is not simply a matter of getting used to one's visual feedback over time, but that there are individual differences worth further investigation. Results of the eye-tracking study found that self-observing behaviour can be linked to one's active role in the conversation, in that people look at themselves more when they are speaking, compared to when listening or when there is silence. This is likely to impact gaze behaviour important for effective turn-taking. When the self-view is off, people do not dwell on other meeting attendees more but instead spend more time looking away from the screen, likely to engage in multitasking activities. Further research is needed to explore the relationship between the presence of the self-view window and multitasking behaviour in more detail. The findings of this research could inform the design of videoconferencing tools, especially their self-view window feature. We are also hoping that the novel methodology of this study will inform future work in the area and make eye-tracking studies more accessible to the research community.

ACKNOWLEDGMENTS

I would like to thank my supervisor, Prof. Duncan Brumby for his continuous support, guidance and feedback over the past few months and our little student research group that made the whole experience enjoyable; my colleagues from Decibel at Medallia who not only supported me in my university studies thorough the whole year but also keenly participated in my study; the participants of the eye-tracking study who went above and beyond to ensure that the collected data is of the best quality; my brother, who is always willing to help, for his expert input and patience, and last but not least my parents without whom I would not be able to achieve any of this.

REFERENCES

[1] Michael Argyle, Mansur Lalljee, and Mark Cook. 2017. The Effects of Visibility on Interaction in a Dyad. *Philosophy*, 67–73.

[2] Omer Azriel, Amit Lazarov, Adva Segal, and Yair Bar-Haim. 2020. Visual attention patterns during online video-mediated interaction in socially anxious individuals. *J. Behav. Ther. Exp. Psychiatry* 69, December 2019 (2020), 101595. DOI:https://doi.org/10.1016/j.jbtep.2020.101595

[3] Jeremy N. Bailenson. 2021. Nonverbal overload: A theoretical argument for the causes of Zoom fatigue. *Technol. Mind, Behav.* 2, 1 (2021). DOI:https://doi.org/10.1037/tmb0000030

[4] By Nancy Baym, Rachel Bergmann, Adam Coleman, Ricardo Reyna Fernandez, Sean Rintel, Abigail Sellen, Tiffany Smith, Work Practices, Jaime Teevan, Brent Hecht, and Sonia Jaffe. 2021. Collaboration and Meetings. (2021).

[5] Nancy Baym, Rachel Bergmann, Adam Coleman, Ricardo Reyna Fernandez, Sean Rintel, Abigail Sellen, and Tiffany Smith. 2021. *Collaboration and Meetings – Chapter 1 of the 2021 New Future of Work report*.

[6] Hancheng Cao, Chia-Jung Lee, Shamsi Iqbal, Mary Czerwinski, Priscilla Wong, Sean Rintel, Brent Hecht, Jaime Teevan, and Longqi Yang. 2021.

Large Scale Analysis of Multitasking Behavior During Remote Meetings. In *CHI 2021*, ACM. Retrieved from https://www.microsoft.com/en-us/research/publication/large-scale-analysis-of-multitasking-behavior-during-remote-meetings/

[7] Derek S. Chapman, Krista L. Uggerslev, and Jane Webster. 2003. Applicant Reactions to Face-to-Face and Technology-Mediated Interviews: A Field Investigation. *J. Appl. Psychol.* 88, 5 (2003), 944–953. DOI:https://doi.org/10.1037/0021-9010.88.5.944

[8] Gwyneth Doherty-Sneddon, Claire O'Malley, Simon Garrod, Anne Anderson, Steve Langton, and Vicki Bruce. 1997. Face-to-Face and Video-Mediated Communication: A Comparison of Dialogue Structure and Task Performance. *J. Exp. Psychol. Appl.* 3, 2 (1997), 105–125. DOI:https://doi.org/10.1037/1076-898X.3.2.105

[9] Bryn Farnsworth. 2019. Eye Tracker Prices – An Overview of 20+ Eye Trackers. *iMotions*. Retrieved from https://imotions.com/blog/eye-tracker-prices/

[10] Martin Hassell and John Cotton. 2017. Some things are better left unseen: Toward more effective communication and team performance in video-mediated interactions. *Comput. Human Behav.* 73, (2017), 200–208. DOI:https://doi.org/10.1016/j.chb.2017.03.039

[11] Muchen He. *Are You Looking at Me ? Eye Gazing in Web Video Conferences*. Association for Computing Machinery.

[12] Roy S. Hessels, Tim H.W. Cornelissen, Ignace T.C. Hooge, and Chantal Kemner. 2017. Gaze Behavior to Faces during Dyadic Interaction. *Can. J. Exp. Psychol.* 71, 3 (2017), 226–242. DOI:https://doi.org/10.1037/cep0000113

[13] Roy S. Hessels, Gijs A. Holleman, Alan Kingstone, Ignace T.C. Hooge, and Chantal Kemner. 2019. Gaze allocation in face-to-face communication is affected primarily by task structure and social context, not stimulus-driven factors. *Cognition* 184, December 2018 (2019), 28–43. DOI:https://doi.org/10.1016/j.cognition.2018.12.005

[14] Simon Ho, Tom Foulsham, and Alan Kingstone. 2015. Speaking and listening with the eyes: Gaze signaling during dyadic interactions. *PLoS One* 10, 8 (2015). DOI:https://doi.org/10.1371/journal.pone.0136905

[15] Judith Holler and Kobin H. Kendrick. 2015. Unaddressed participants' gaze in multi-person interaction: Optimizing recipiency. *Front. Psychol.* 6, FEB (2015), 1–14. DOI:https://doi.org/10.3389/fpsyg.2015.00098

[16] Mansoor Iqbal. 2021. Zoom Revenue and Usage Statistics (2021). *Business of Apps*.

[17] Aleš Jaklič, Franc Solina, and Luka Šajn. 2017. User interface for a better eye contact in videoconferencing. *Displays* 46, (2017), 25–36. DOI:https://doi.org/10.1016/j.displa.2016.12.002

[18] Matthew K. Miller, Regan L. Mandryk, Max V. Birk, Ansgar E. Depping, and Tushita Patel. 2017. Through the looking glass: Effects of feedback on self-awareness and conversation during video chat. *Conf. Hum. Factors Comput. Syst. - Proc.* 2017-May, (2017), 5271–5283. DOI:https://doi.org/10.1145/3025453.3025548

[19] Matthew K Miller, Martin Johannes Dechant, and Regan L Mandryk. 2021. Meeting You, Seeing Me: The Role of Social Anxiety, Feedback, and Interface Layout in a Get-to-Know-You Task via Video Chat. *Proc. 2021 CHI Conf. Hum. Factors Comput. Syst.* (2021).

[20] Naoki Mukawa, Tsugumi Oka, Kumiko Arai, and Masahide Yuasa. 2005. What is connected by mutual gaze?-User's behavior in video-mediated communication. *Conf. Hum. Factors Comput. Syst. - Proc.* (2005), 1677–1680. DOI:https://doi.org/10.1145/1056808.1056995

[21] Mark Powers and Jasminder Thind. 2020. Lessons from China: A return to the office, but not to the old way of working.

[22] Abigail J. Sellen. 1995. Remote Conversations: The Effects of Mediating Talk With Technology. *Human-Computer Interact.* 10, 4 (1995), 401–444. DOI:https://doi.org/10.1207/s15327051hci1004_2

[23] Jaime Teevan. 2011. The New Future of Work. *Policy* 27, 2 (2011), 33–39.

[24] Jose Eurico De Vasconcelos Filho, Kori M. Inkpen, and Mary Czerwinski. 2009. Image, appearance and vanity in the use of media spaces and videoconference systems. *GROUP'09 - Proc. 2009 ACM SIGCHI Int. Conf. Support. Gr. Work* (2009), 253–261. DOI:https://doi.org/10.1145/1531674.1531712

[25] Noortje Vriends, Yasemin Meral, Javier A. Bargas-Avila, Christina Stadler, and Susan M. Bögels. 2017. How do I look? Self-focused attention during a video chat of women with social anxiety (disorder). *Behav. Res. Ther.* 92, (2017), 77–86. DOI:https://doi.org/10.1016/j.brat.2017.02.008

[26] Jürgen Wegge. 2006. Communication via videoconference: Emotional and cognitive consequences of affective personality dispositions,

seeing one's own picture, and disturbing events. *Human-Computer Interact.* 21, 3 (2006), 273–318. DOI:https://doi.org/10.1207/s15327051hci2103_1

[27] Priscilla Wong, Advait Sarkar, Abigail Sellen, and Sean Rintel. 2020. Methodology and Participation for 2020 Diary Study of Microsoft Employees Experiences in Remote Meetings During COVID-19. (2020), 1–46.

[28] 2021. Growth in downloads of select video conferencing apps as of March 2020 vs. weekly average for Q4 2019, by country. *Statista Research Department*. Retrieved from https://www.statista.com/statistics/1109875/download-growth-video-conferencing-apps/