

The Shortcomings of Video Conferencing Technology, Methods for Revealing Them, and Emerging XR Solutions

Abstract

Video conferencing has become a central part of our daily lives, thanks to the COVID-19 pandemic. Unfortunately, so have its many limitations, resulting in poor support for communicative and social behavior and ultimately, “zoom fatigue.” New technologies will be required to address these limitations, including many drawn from mixed reality (XR). In this paper, our goals are to equip and encourage future researchers to develop and test such technologies. Toward this end, we first survey research on the shortcomings of video conferencing systems, as defined before and after the pandemic. We then consider the methods that research uses to evaluate support for communicative behavior, and argue that those same methods should be employed in identifying, improving, and validating promising video conferencing technologies. Next, we survey emerging XR solutions to video conferencing’s limitations, most of which do not employ head-mounted displays. We conclude by identifying several opportunities for video conferencing research in a post-pandemic, hybrid working environment.

I Introduction

Over the recent years of the COVID-19 pandemic, a mass move toward remote work and communication has forced many to reckon with the long-term effects of video conferencing as a primary communication method. In particular, Zoom fatigue, or video conferencing fatigue, became particularly prominent during the pandemic (Fauville et al., 2021a). Zoom fatigue is defined as physical and cognitive exhaustion resulting from intensive use of video conferencing tools (Riedl, 2021). Post-pandemic, most expect remote video conferencing to remain much more widely used than it was before COVID (Rommel, 2021), serving as both a safety precaution and a crucial enabler of a burgeoning hybrid home/office work environment as public health precautions end. Given this, understanding the challenges and opportunities of video conferencing is particularly important, both to prevent negative consequences, and to realize benefits in the long term. This is especially pertinent as use of previously unconventional meeting environments, such as virtual reality, grows.

In seeking this understanding, our guiding questions are: (a) what shortcomings limited conferencing effectiveness prior to the pandemic, and how do they

contribute to Zoom fatigue, (b) what solutions have addressed these shortcomings and might ease Zoom fatigue in the post-pandemic hybrid working environment, and (c) what potential do nontraditional conferencing interfaces, such as XR, have to address these same shortcomings?

Toward this end, we survey research and opinion on video conferencing across several disciplines, both technical and social, and across several technologies, including traditional interfaces as well as mixed and virtual reality. We focus on video conferencing's shortcomings and evaluative methods for finding them, as defined both pre- and post-pandemic. We give particular attention to the cognitive and technological factors that contribute to Zoom fatigue, which emerged during the pandemic. Finally, we survey the emerging solutions that address known shortcomings, including several in VR and XR.

With this survey, we aim to provide future researchers with a foundation enabling video conferencing improvements reducing Zoom fatigue, especially in the post-pandemic, hybrid working environment.

2 Survey Methods

Our survey fits within the narrative review framework, defined as a qualitative method seeking to describe current literature, without quantitative synthesis (Shadish et al., 2001). Given the limited literature on video conferencing over the past decades and the recent drastic increase in its use, video conferencing research is still nascent. For this reason, we turned to this narrative review method, which “focuses on formulating general relations among a number of variables of interest.” Two examples of this sort of survey include Lam et al. (2011) and Perer and Schneiderman (2009).

We sought to survey references describing video conferencing systems and challenges, as well as the human behavior they sought to support and methods for evaluating that support. To collect references, we therefore first searched Google Scholar for material using the keywords “zoom fatigue,” “fatigue,” “video conferencing,” or “gaze awareness.” For references older than 2015, we

set a threshold of 10 citations for acceptability, to restrict our survey to work that had had at least minimal impact, when there had been time for the research community to respond. The initial search returned approximately 180 papers. Both authors then examined each paper, discarding any that they agreed did not discuss video conferencing, communicative behavior, or methods for measuring and evaluating that behavior. We then studied the bibliographies of each remaining paper, examining any cited paper with a title that contained our search keywords, and discarding any that we agreed did not meet our inclusion criteria. We were left with 65 papers.

We next categorized these references using an iterative open coding (Creswell, 2014) methodology that began with three labels reflecting our concerns as we began our survey work: *Measures*, or methods for evaluating video conferencing success; *Shortcomings*, or weaknesses of video conferencing solutions; and *Fatigue*, or the general feeling of exhaustion that many report feeling after video conferencing. Within the *Measures* label, our sub-labels were *technical*, objective methods of measurement; *behavioral* observational and experimental methods capturing human use; and *subjective* methods that asked users to offer their judgements of video conferencing systems. Within the *Shortcomings* label, our sub-labels were *delay*, the lag users encounter between the moment they act and the moment that act is displayed to other conference participants; and *gaze*, the degree to which a system communicates where users are looking. Finally with the *Fatigue* label, we used a *zoom* sub-label to mark discussion of fatigue in the context of video conferencing, and a *general* sub-label to indicate discussion of fatigue more generally.

As we iterated through the papers we found, we proposed additional labels and sub-labels, and adopted them if both authors approved. To the *Shortcomings* label, we added an *objects of discussion* sub-label referring to video conferencing support for referencing items of participant interest, and a *nonverbal cues* sub-label marking support for nonverbal communicative signals beside gaze and discussed objects. We also added a *Focus* label, referencing the type of knowledge a paper contains, with the sub-labels *solution* for an engineering improvement to video conferencing systems, *new measures* for

methods of measuring communicative behavior; *review* for a survey of video conferencing research or communicative behavior, and *study* for an experiment examining communication in conferencing systems.

Table 1 shows how we categorized the papers in this survey with our labels. We use these categories to structure the following review.

3 (Zoom) Fatigue

The pandemic has drastically increased use of video conferencing, resulting in the widespread experience of what has come to be called “Zoom fatigue.” In this section, we first review recent popular literature addressing Zoom fatigue. Next, we investigate research literature on the general phenomenon of fatigue (Table 1, Fatigue, General Fatigue column), and Zoom fatigue itself (Table 1, Fatigue, Zoom Fatigue column): specifically, its definition, measurement and analyses of its components.

3.1 Recent Expert and Popular Opinion

The massive increase in use of video conferencing during the pandemic created a rush of opinion about video conferencing’s shortcomings, many focusing on apparent long-term consequences. The phrase “Zoom fatigue” was quickly coined (Sklar, 2020; Fosslien & Duffy, 2020; Degges-White, 2020; deHahn, 2020; Rosenberg, 2020; Robert, 2020), referring to a lasting fatigue born from the unique stresses of remote work using video conferencing (this phrase has become a standard, despite the existence of many video conferencing alternatives). A number of possible causes have been suggested. These include *physical issues*, related to the bad ergonomics of turning a freely moving, immersive meeting into a constrained event passing through a small rectangle (Degges-White, 2020); and *emotional issues*, like the turmoil and isolation of a pandemic (Degges-White, 2020). But most thinking has dwelled on *nonverbal and social cues*, including poor eye contact and seeming inattentiveness (Degges-White, 2020), “constant gaze” with a gallery of other faces creating the percep-

tion of being watched (Rosenberg, 2020; Jiang, 2020) (Bailenson, 2020), deficiency of gesture that requires a draining hyper-focus to pick up the little body language that remains (Hickman, 2020; deHahn, 2020), “big face,” in which faces appear larger than they would at a comfortable interpersonal distance (Bailenson, 2021), and poor backchanneling, which makes it harder to recover from misunderstandings and attentional lapses (Fosslien & Duffy, 2020; Sklar, 2020) through asides with other participants. Appropriately, much of this popular conjecture has been investigated through formal research.

3.2 Defining and Measuring Zoom Fatigue

Fatigue is “an unpleasant physical, cognitive, and emotional symptom described as a tiredness not relieved by common strategies that restore energy. Fatigue varies in duration and intensity, and it reduces, to different degrees, the ability to perform the usual daily activities” (Aaronson et al., 1999). When measuring subjective fatigue, “(1) There was found to be a high correlation between the frequency of complaints of fatigue and the feeling of fatigue; (2) The amount of feeling of fatigue is different for the type of symptom” (Yoshitake, 1971). These considerations should apply to Zoom fatigue, which is currently measured subjectively (Nesher Shoshan & Wehrt, 2021).

Riedl et al. (2021) define “Zoom fatigue” as “somatic and cognitive exhaustion that is caused by the intensive and/or inappropriate use of video conferencing tools, frequently accompanied by related symptoms such as tiredness, worry, anxiety, burnout, discomfort, and stress, as well as other bodily symptoms such as headaches.” Riedl et al. also posit a conceptual framework for Zoom fatigue (see Figure 1), in which both a lack of information (asynchronicity of communication; coordination difficulty; lack of body language, eye contact, and shared attention) and an information overload (self-awareness or constant gaze, interruption of automaticity, interaction with multiple faces) contribute to cognitive effort and stress. Recently, Fauville et al. (2021a) established a survey measure for Zoom

Table 1. A Table Categorizing the Research Reviewed in This Paper, by Type of Evaluation Measure, Video Conferencing Shortcoming, Type of Fatigue, and Whether the Paper Describes New Solutions, New Evaluative Measures, or Is a Survey of Other Work

Authors	Measures			Shortcomings of Video Conferencing			
	Technical	Behavioral	Subjective	Delay	Gaze	Objects of Discussion	Nonverbal Cues
Aaronson et al. (1999)							
Angelopoulos et al. (2021)							
Bailenson (2020)							
Bailenson (2021)							
Becher et al. (2020)							
Bennett et al. (2021)							
Berndtsson et al. (2012)							
Bohannon et al. (2013)							
Bos et al. (2002)							
Boyaci et al. (2009)							
D'Angelo and Begel (2017)							
D'Angelo and Gergle (2018)							
deHahn (2020)							
Driskell and Radtke (2003)							
Driskell et al. (2003)							
Edelmann et al. (2013)							
Fauville et al. (2021a)							
Fauville et al. (2021b)							
Feick et al. (2018)							
Friston and Steed (2014)							
Gan et al. (2020)							
Gasteratos et al. (2021)							
Gergle et al. (2012)							
Grayson and Monk (2003)							
Gunkel et al. (2015)							
He et al. (2021)							
He et al. (2020)							
Homaecian et al. (2021)							
Hopkins and Benford (1998)							
Ishibashi et al. (2006)							
ITU (2000)							
ITU (2007)							
Jansen and Bulterman (2013)							

Table 1. *Continued.*

Authors	Measures			Shortcomings of Video Conferencing			
	Technical	Behavioral	Subjective	Delay	Gaze	Objects of Discussion	Nonverbal Cues
Jansen et al. (2011)							
Kobayashi et al. (2021)							
Kraut et al. (1998)							
Kuster et al. (2012)							
Lawrence et al. (2021)							
Li et al. (2020)							
Lombard et al. (2009)							
Monk and Gale (2002)							
Mota and Pimenta (2006)							
Nardi and Whittaker (2001)							
Nesher Shoshan and Wehrt (2021)							
Nguyen and Canny (2005)							
Nguyen and Canny (2007)							
O'Malley et al. (1996)							
Oducado et al. (2021)							
Okada et al. (1994)							
Olbertz-Siitonen (2015)							
Orlosky et al. (2021)							
Pachnowski (2002)							
Ratan et al. (2021)							
Riedl (2021)							
Roberts et al. (2009)							
Schmitt et al. (2014)							
Schoenenberg (2016)							
Schoenenberg et al. (2014)							
Stotts et al. (2010)							
Su et al. (2014)							
Tam et al. (2012)							
Vertegaal et al. (2003)							
Whittaker (2003)							
Wu et al. (2021)							
Yan et al. (2020)							
Yoshimura and Borst (2021)							
Yu et al. (2021)							

Table 1. *Continued.*

Authors	Fatigue		Focus			
	Zoom Fatigue	General Fatigue	Solution	New Measures	Review	Study
Aaronson et al. (1999)						
Angelopoulos et al. (2021)						
Bailenson (2020)						
Bailenson (2021)						
Becher et al. (2020)						
Bennett et al. (2021)						
Berndtsson et al. (2012)						
Bohannon et al. (2013)						
Bos et al. (2002)						
Boyaci et al. (2009)						
D'Angelo and Begel (2017)						
D'Angelo and Gergle (2018)						
deHahn (2020)						
Driskell and Radtke (2003)						
Driskell et al. (2003)						
Edelmann et al. (2013)						
Fauville et al. (2021a)						
Fauville et al. (2021b)						
Feick et al. (2018)						
Friston and Steed (2014)						
Gan et al. (2020)						
Gasteratos et al. (2021)						
Gergle et al. (2012)						
Grayson and Monk (2003)						
Gunkel et al. (2015)						
He et al. (2021)						
He et al. (2020)						
Homacian et al. (2021)						
Hopkins and Benford (1998)						
Ishibashi et al. (2006)						
ITU (2000)						
ITU (2007)						
Jansen and Bulterman (2013)						
Jansen et al. (2011)						
Kobayashi et al. (2021)						
Kraut et al. (2002)						
Kuster et al. (2012)						
Lawrence et al. (2021)						

Table 1. *Continued.*

Authors	Fatigue		Focus			
	Zoom Fatigue	General Fatigue	Solution	New Measures	Review	Study
Li. et al. (2020)						
Lombard et al. (2009)						
Monk and Gale (2002)						
Mota and Pimenta (2006)						
Nardi and Whittaker (2001)						
Nesher Shoshan and Wehrt (2021)						
Nguyen and Canny (2005)						
Nguyen and Canny (2007)						
O'Malley et al. (1996)						
Oducado et al. (2021)						
Okada et al. (1994)						
Olbertz-Siitonen (2015)						
Orlosky et al. (2021)						
Pachnowski (2002)						
Ratan et al. (2021)						
Riedl (2021)						
Roberts et al. (2009)						
Schmitt et al. (2014)						
Schoenenberg (2016)						
Schoenenberg et al. (2014)						
Stotts et al. (2010)						
Su et al. (2014)						
Tam et al. (2012)						
Vertegaal et al. (2003)						
Whittaker (2003)						
Wu et al. (2021)						
Yan et al. (2020)						
Yoshimura and Borst (2021)						
Yu et al. (2021)						

exhaustion, the Zoom Exhaustion and Fatigue (ZEF) Scale, which consists of 15 items across five dimensions of fatigue: general, social, emotional, visual, and motivational.

3.3 The Components of Zoom Fatigue

Video conferencing delivers several types of information not present in face-to-face meetings, creating a

stressful information overload. Much of this information is delivered nonverbally, as indicated by Figure 1. One study suggests that nostalgia for a time before the pandemic may contribute to Zoom fatigue (Nesher Shoshan and Wehrt, 2021), but most thought points toward causes present prior to the COVID-19 crisis. The prominence of self-video in video conferencing has led to a rise in facial dissatisfaction, or mirror anxiety (Fauville et al., 2021b), with some cases extending

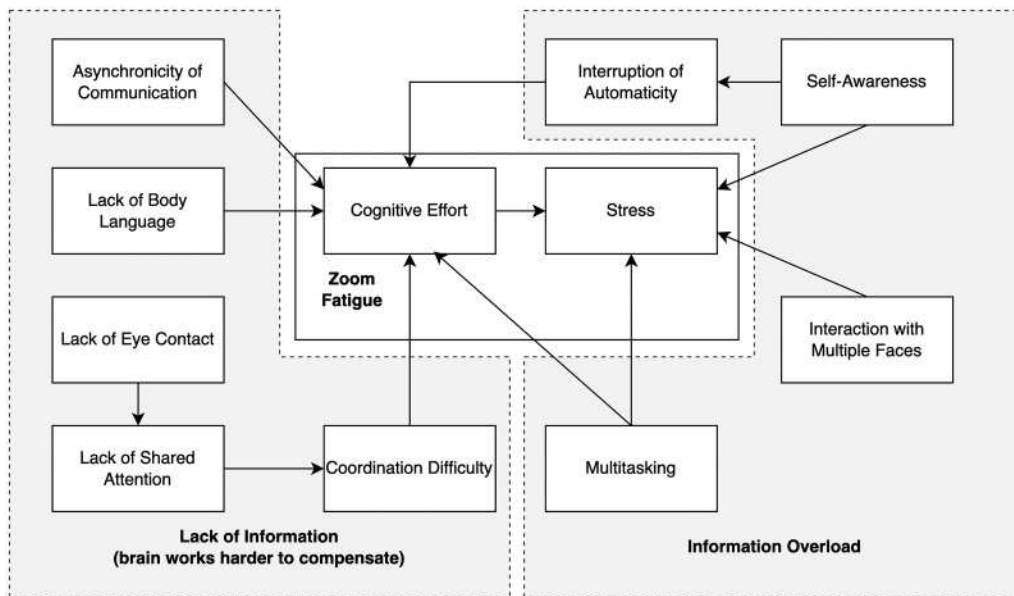


Figure 1. A diagram depicting Riedl's hypothesis for a conceptual model of what factors contribute to Zoom Fatigue, adapted from Riedl's study (2021).

into “Zoom dysmorphia,” driving an increase in plastic surgery (Gasteratos et al., 2021), particularly among women (Ratan et al., 2021). Hyper-gaze from a grid of staring faces is yet another informational challenge (Fauville et al., 2021b). On the other hand, much of the information normally present in-person is missing in video conferencing: the combination of “being physically trapped” in front of the screen and “the cognitive load from producing and interpreting nonverbal cues” (Fauville et al., 2021b) makes referencing a common context and creating shared attention and connection difficult. Academic classrooms and workplaces, marked by a high frequency and intensity of video conferencing, were shown to exacerbate Zoom fatigue, as did factors such as lower economic status, poor academic performance, and unstable internet connections (Oducado et al., 2021).

Yet video conferencing users need not wait for technological upgrades to reduce their fatigue. Bennett et al. (2021) offer a number of recommendations for reducing Zoom fatigue, as illustrated in Table 2. Concrete recommendations include better meeting times, improved group belongingness, and muting microphones when not speaking; less certain recommendations in-

clude turning off webcams, using “hide self” view, taking breaks, and establishing group norms.

Zoom fatigue is a constellation of many different communicative problems with current video conferencing. While the recommendations in Table 2 are a solid step in the right direction, they do not address the conferencing technology itself, a necessary step to begin reducing the need for users to compensate for the technology's shortcomings. Video conferencing research long predates Zoom and has also identified shortcomings, devised and applied methods for measuring communicative success, and proposed potential solutions. Below, we review these shortcomings, measures and solutions.

4 Shortcomings of Video Conferencing

Video conferencing has been with us for nearly a century (Peters, 1938), and research on its limitations predates the pandemic by decades. While research on video conferencing's long-term effects is sparse, researchers did investigate many of the same shortcomings studied by Zoom fatigue investigators such as Riedl et al. (2021). We group the most relevant work into projects

Table 2. *A Table Explaining Recommendations for Reducing Video Conference Fatigue, Adapted from Figure 6 of Bennett et al.'s (2021) Study on Video Conference Fatigue*

Recommendations supported by our quantitative study	Potential explanation for fatigue reduction
1. Hold meetings at a time that is least fatiguing for as many participants as possible based on work schedule, which may be earlier in the work period	Meetings are affect-generating events that may influence fatigue trajectory over the course of a day.
2. Enhance perceptions of group belongingness	Enhanced perception of belongingness is expected to encourage interest in participation, reducing effortful attention and fatigue.
3. Unless you are speaking, mute your microphone	Muting reduces both the potential for distracting background noise and the amount of active attention to stay quiet on the user's part.
Recommendations with inconclusive evidence from our quantitative study	Potential explanation for fatigue reduction
4. Decrease/increase webcam usage	Increased webcam usage may increase group belongingness (and reduce fatigue), while decreased usage decreases stimuli and allows detaching, also possibly reducing fatigue.
5. Consider using "hide self" view	Hiding the self camera potentially reduces stimuli and how much users worry about their appearance/background, improving belongingness.
Recommendations based on qualitative comments	Potential explanation for fatigue reduction
6. Take breaks during videoconferences and between videoconferences	Breaks between and/or during meetings allow users to detach, a key method of reducing fatigue.
7. Establish group norms (e.g., usage of mute and webcam, acceptability of multitasking, when/how to speak up)	Strong norms reduce ambiguity about acceptable behavior, and reduces active worry that contributes to fatigue. Additionally, they increase group belongingness.

addressing problems with delay, gaze, objects of discussion, and a variety of nonverbal conversational cues.

4.1 Delay

Delay (lag) remains one of the most widely researched of video conferencing's technical shortcomings

(Table 1, Shortcomings, Delay column), and represented in 12 out of 21 papers within the Measures, Technical column in Table 1. Delay for video conferencing is defined as the time elapsed between the moment of input (e.g., a joke or a smile) and the resulting response (e.g., a retort or a laugh): "glass-to-glass." As a factor often outside the control of users, the majority of

research compares delay's quantitative severity against its qualitative effects. VideoLat is one system for measuring glass-to-glass video conferencing delays (Jansen & Bulterman, 2013). Users display a QR code to the camera, and compare the times that it is detected by the camera and displayed on an output monitor. We discuss more tools for measuring delay in Section 5.1.

A number of studies defined "significant" delay as 500–650 ms (Schmitt et al., 2014; Whittaker, 2003; Tam et al., 2012). At such levels, delay causes "prolonged overlap, gap, and sequential disarray and missed attempts at turn-taking" (see Table 3) (Olbertz-Siitonen, 2015), in addition to increased interruptions in video settings (O'Malley et al., 1996), all of which contribute to lower conversational quality. Schoenberg (2016) defined a Quality of Mediated Conversation measure, composed of "the Conversational Quality, the Mediated Interaction, the Experiencer, the Interaction Partners, and the Circumstances." Even one active user with significant delay can negatively impact the entire group's Quality of Experience (QoE), with the QoE decreasing as the delay becomes more symmetrical (equally distributed across participants) (Schmitt et al., 2014). Additionally, Becher et al. (2020) found that while communicating collaboratively in an immersive virtual reality environment, increasing added delay from 300 ms to 450 ms introduced a noticeable decrease in mutual understanding, alongside a consistent decrease in task performance. Su et al. (2014) attempted to mask delay with prerecorded video or predicted motion, and found that masking was effective up until 800 ms, but frequent masking was still required at 200 ms.

In addition to disrupting conversation and its general quality, video conferencing delay has emotional effects. Delays over 100 ms have impacted user feelings of "fairness" in competitive events hosted on video conferencing, like a quiz game scenario (Ishibashi et al., 2006), with perceived fairness degrading as delay increases. Symmetrical and asymmetrical delay of 1200 ms has caused users to mistakenly attribute technical issues to personal shortcomings (Schoenberg et al., 2014; Schoenberg, 2016), where users were likely to feel that delayed conversational partners were inattentive, undisciplined, or less friendly.

Table 3. A Table Illustrating Turn-Taking in a Collaborative Puzzle Task, Adapted from Gergle et al.'s study (2012)

Turn	User A/ User B	Dialogue/[Action]
1	A	alright, um take the main black one
2	A	and stick it in the middle
3	B	[moves and places correct piece]
4	A	take the one-stripe yellow
5	A	and put it on the left side
6	B	[moves and places correct piece]
7	A	uh yeah, that's good
8	A	take the um two stripe one
9	A	and put it on top of the black one
10	B	[moves and places correct piece]
11	A	and take the half shaded one
12	A	and put it diagonally above the one that you just moved to the right
13	B	[moves and places correct piece]
14	A	yup, done.

Because it is largely a technical problem, delay's effects are difficult for users to solve themselves. For example, reducing video quality may reduce delay, but it also reduces visual communicative cues. These studies confirm that even a fraction of a second (100–500 ms) of delay can create conversational challenges. However, most of these studies examined only single video conferences. We suspect that, over several conferences (i.e., a fairly typical day of post-pandemic hybrid or remote work), even less delay (<100 ms) may suffice to increase

the cognitive effort Riedl et al. (2021) include in their model, and create Zoom fatigue.

4.2 Gaze

Gaze awareness is the ability to identify what—or importantly, who—a person is looking at. The majority of papers offering video conferencing solutions we reviewed (see Table 1, Focus/Solution column) addressed gaze. These investigations are especially important, given few of today’s common video conferencing systems can effectively depict gaze. In the real world, our view of a conversational partner and their view of us correspond. However, in video conferencing systems, because camera and display are rarely colocated, this correspondence is broken. Even minor offsets in camera and display can notably affect our ability to recognize whether we are being looked at (Grayson & Monk, 2003). Gaze depiction becomes even more problematic as the number of conference participants grows: one camera cannot colocate with many participants. Yet when gaze can be effectively communicated by video conferencing, it has notable effects, especially in light of the importance of eye contact in mediated communication (Bohannon et al., 2013).

Among the papers addressing the challenges video conferencing systems have in communicating gaze (Table 1, Shortcomings, Gaze column), GA Display (see Figure 2) was among the earliest. An experimental video conferencing system supporting gaze awareness (Monk & Gale, 2002) for two users, GA Display used translucent screens and half-silvered mirrors. Its users were much more efficient in a conversational game (55% less turns and 949 less words) than users of an audio-only system. Multiview (Nguyen & Canny, 2005, 2007) supports two groups communicating with correct gaze through a shared virtual window. Multiview users formed trust relationships more quickly than users of traditional video conferencing systems. Even when collaborative tasks are not primarily communicative and video is not used, shared gaze awareness can be helpful. For example, gaze visualization has been shown to positively affect performance (D’Angelo & Gergle, 2018). When pair programmers are refactoring code, shared

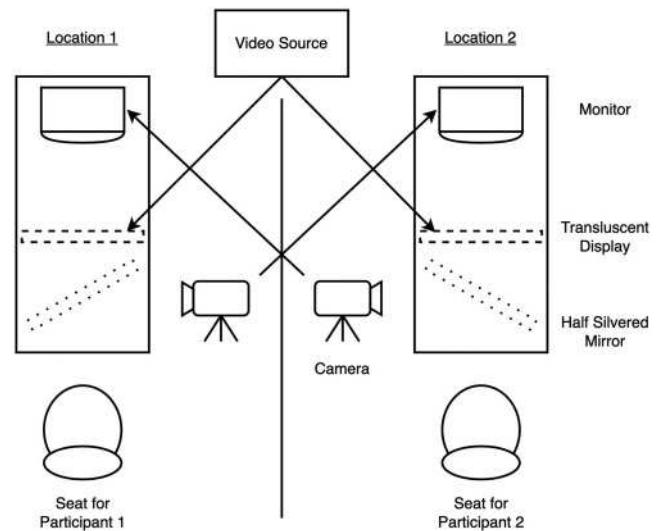


Figure 2. An illustration of how GA Display creates gaze awareness in its video conferencing solution, adapted from Monk et al.’s study (2002).

gaze awareness achieved with eye trackers and highlights showing gaze on shared in code helped increase task speed (D’Angelo & Begel, 2017).

The importance of gaze has resulted in a number of solutions in addition to GA Display and Multiview, which focus more on engineering the solutions themselves than on understanding the importance of gaze (Edelmann et al., 2013; He et al., 2021; Lawrence et al., 2021). We discuss these in more detail in Section 6. Gan et al. (2020) use ethnographic methods to study an underserved use of video conferencing, in which maintaining gaze is central: three-party video calls wherein one participant (e.g., a child) is less able to manage the technology, so that another (e.g., a grandparent) must help them speak with the third participant (e.g., a parent). They identify a range of needs for future video conferencing solutions to address.

Although few existing video conferencing solutions rely on it (e.g., D’Angelo & Begel, 2017), gaze tracking may play an important role in maintaining gaze awareness in the future. Fortunately, gaze tracking technology is already quite effective and quickly becoming more so: recent systems have achieved a refresh rate of 10,000 Hz using less than 12 Mbits of bandwidth (Angelopoulos et al., 2021), or even power draws as

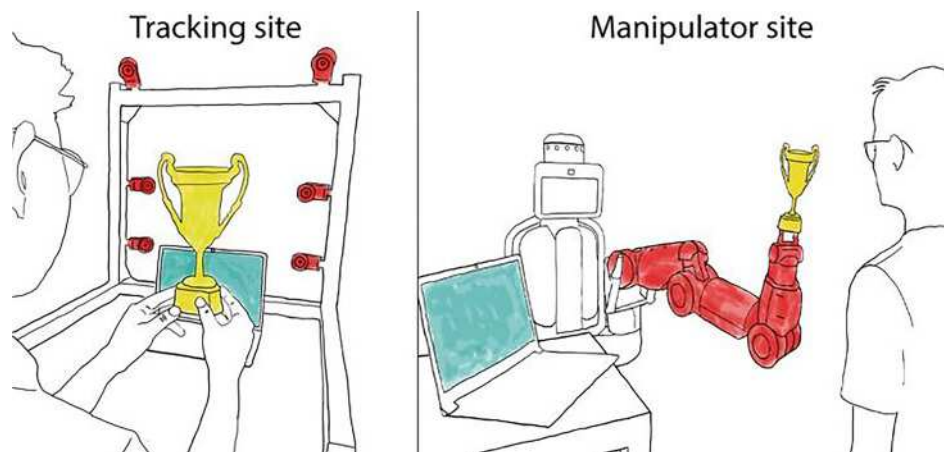


Figure 3. An illustration of how Remote Manipulator operates, provided courtesy of Feick et al. (2018).

low as 16 mW that are still accurate to within 2.67° while maintaining 400-Hz refresh rates (Li et al., 2020). Power and refresh rate concerns are especially important for XR headsets, in which power and latency can hinder not only eye-tracking effectiveness, but general comfort. Headset-less solutions to video conferencing will likely mandate sophisticated gaze tracking on top of other tracking technologies.

Like delay, lack of video conferencing support for gaze likely contributes to Zoom fatigue. As seen in Figure 1, lack of eye contact likely engenders a lack of shared attention, which in turn increases the cognitive load of video conferencing. Given the growing availability of inexpensive cameras and gaze tracking solutions, support for gaze may offer a widely applicable salve to long-term fatigue.

4.3 Objects of Discussion

Conversation often centers around a shared object, such as a whiteboard diagram, a presentation, or a working document. In such cases, discussion is filled with shorthand “deictic” references that make use of the object’s context like “that one,” “to the left of” and with pointing, “over there.” Facilitating such conversational grounding has been particularly important for improving performance on shared tasks during video conferencing. Objects of discussion are an important part of shared attention and the papers we review

(Table 1, Shortcomings, Object column), and a large portion of the papers that offer novel video conferencing solutions. With their ability to overlay virtual objects onto real world views, XR and AR are uniquely equipped to address this issue.

In two studies of communicative behavior, Gergle et al. (2012) and Kraut et al. (2002) found that clear, synchronized, and low delay shared visual information provides important feedback for successful communication. In more applied work, a novel system visualizing gaze onto code shared by pair programmers resulted in a notable improvement in performance of refactoring tasks across a range of metrics, including faster reference acknowledgement, more time with overlapping gaze and faster task completion (D’Angelo & Begel, 2017). ReMa (Feick et al., 2018), illustrated in Figure 3, tracks a user’s manipulation of an object, and maps it to a remote robotic arm manipulating a proxy object. Initial evaluations showed that users preferred ReMa when combined with video conferencing, because it allowed a more intuitive understanding of shared artifacts. Initial studies in CollaboVR (He et al., 2020), a sketch based framework for collaboration in XR, found that among the projected, side-by-side and face-to-face (mirrored) layouts, users preferred face-to-face, citing that it allowed them to focus on both the shared artifact and their collaborator at the same time.

With their ability to overlay virtual objects onto real world views, XR and AR may be uniquely equipped to

provide the grounding conversational context of objects of discussion. We could not find research examining support for objects of discussion in video conferencing over the long term, but expect that like reductions in delay and support for gaze, it may improve the quality of communication and lower Zoom fatigue.

4.4 Additional Nonverbal Cues

In addition to gaze and objects of discussion, a variety of other nonverbal cues play an important role in conversation and are not commonly well-supported by existing conferencing systems, and include spatial location, gesture, facial expression, and body language (Table 1, Shortcomings, Other Cues column). While none of these individual shortcomings is a dominant theme, collectively these shortcomings form a significant portion of the literature we review.

Gesture and expression are particularly important in speech formation and clarity, and have aided understanding in collaborative tasks (Driskell & Radtke, 2003; Driskell et al., 2003). Early studies confirmed the value of video in delivering such nonverbal cues, with video conferencing supporting more natural conversations as measured by improved turn taking, distinguishing among speakers, and better ability to interrupt/interject (Whittaker, 2003). When compared with audio-only solutions, video conferencing not only increased perceived naturalness, but also mitigated the impact of delays up to 500 ms (Tam et al., 2012). Emphasizing the importance of visuals, Berndtsson et al. (2012) found that it was more important to synchronize audio and video than to reduce audio delays. This was true even at delays lower than 600 ms (Berndtsson et al., 2012), with users preferring video conferencing over audio-only communication. In a virtual reality system without video, the introduction of more facially expressive avatars not only increased presence and social attraction, but also increased task performance (Wu et al., 2021).

Not all nonverbal conversational cues are visual. When improving the domestic video conferencing experience, Jansen et al. (2011) found that spatial audio coupled with spatial audiovisual layout were necessary additions.

With these features, communicating groups could more easily attend to central conversation, and ignore distractions in busy family environments.

Together with gaze and objects of discussion, these nonverbal cues form a suite of social markers that users cannot rely upon in common video conferencing solutions, increasing the cognitive effort and stress that form the backbone of long term fatigue. The work showing that such nonverbal cues can compensate for delays makes them a particularly promising avenue for future video conferencing solutions. XR technology should be particularly helpful in communicating location and gesture.

5 Measuring Video Conferencing Effectiveness

Any attempt at addressing video conferencing's shortcomings should be evaluated, to determine how well it supports human communication. While human-computer interface researchers know evaluative methods well, most are likely unfamiliar with evaluating communication itself. In this section, we review the methods previously used to evaluate video conferencing systems, with particular attention to those assessing communication. These fall into three categories: technical, behavioral, and subjective (Table 1, Measures columns).

5.1 Technical Measures

Technical measures are those concerned with the performance of a system, and do not typically depend on human users. Such measures are most often used by researchers working with a systems focus, creating novel or improved solutions for video conferencing. The most commonly used technical measure is delay (Edelmann et al., 2013; Su et al., 2014; Gunkel et al., 2015; Schmitt et al., 2014), and related measures such as loss rates, audio and video quality (Edelmann et al., 2013; Jansen et al., 2011). Researchers often use delay in concert with other behavioral and subjective measures (O'Malley et al., 1996; Gunkel et al., 2015; Homacian et al., 2021; Berndtsson et al., 2012).

Video conferencing researchers use many tools for measuring delay, including videoLat (Jansen & Bulterman, 2013), which measures the time elapsed between appearance on camera and on remote display (“glass-to-glass”); and vDelay (Boyaci et al., 2009), which measures delay similarly. Virtual reality researchers are also very concerned with delay, and often create communicative applications. Friston and Steed (2014) review methods of measuring latency, and describe a simple method for measuring delay, Automated Frame Counting, that makes use of a high-frame rate video camera. These tools are particularly useful for creating independent, consistent measurements of delay across varied, complex, and sometimes closed video conferencing systems. A common element of these delay measuring methods is using camera footage or visual information (Jansen & Bulterman, 2013; Friston & Steed, 2014). For example, Roberts et al. (2009) compare communicative VR systems to video conferencing, noting that VR was much more effective at communicating attention, but had three times the delay of video conferencing (at 150 ms).

5.2 Behavioral Measures

Like most computing systems, video conferencing is a tool that supports tasks, so conferencing effectiveness is often measured through its impact on tasks. These tasks can be simple conversation, or more applied work requiring informational exchange. Video conferencing is meant for multiple users, so tasks are often collaborative, and include finding a point on a shared object (Monk & Gale, 2002), word games (Driskell & Radtke, 2003), market trading (Nguyen & Canny, 2007), puzzle solving (Gergle et al., 2012), code refactoring (D’Angelo & Begel, 2017), and charades games (Wu et al., 2021). The goal of the evaluation is to measure the impact of a system improvement or experimental manipulation on task performance, either quantitatively or qualitatively.

5.2.1 Quantitative Measures. Traditional measures of task performance focus on efficiency. Video conferencing researchers make widespread use of both time (Wu et al., 2021; Monk & Gale, 2002; Gergle et al., 2012;

O’Malley et al., 1996; Bennett et al., 2021; Homacian et al., 2021) and accuracy (O’Malley et al., 1996; Monk & Gale, 2002; Driskell & Radtke, 2003; Wu et al., 2021). Other quantitative measures more specific to video conferencing include gaze estimation (Grayson & Monk, 2003) and counting gaze overlap with eye tracking (D’Angelo & Begel, 2017).

5.2.2 Communicative Measures. Communication researchers have devised a number of methods for assessing conversational efficiency and fluency, and these have naturally found application in studies of communicative systems like video conferencing. Typically, these communicative assessments involve recording the conversations on the system, then compiling various characterizing statistics. These include the following:

Word count: a simple count of the number of words used in the conversation (O’Malley et al., 1996; Monk & Gale, 2002). Fewer words imply more efficient conversation (and better video conferencing).

Turn count: as illustrated by Table 3, conversations can be parsed into a series of “turns,” with each participant successively responding to what the other has said. Researchers perform this parsing, then count the number of turns (Monk & Gale, 2002; Gergle et al., 2012; O’Malley et al., 1996; Olbertz-Siitonen, 2015). Fewer turns are a more direct measure of conversational efficiency than word count.

Interruption count: a count of when two or more participants are speaking at once (Monk & Gale, 2002; O’Malley et al., 1996; Schoenenberg, 2016; Tam et al., 2012; Olbertz-Siitonen, 2015). Fewer interruptions indicate more conversational efficiency.

Pause count: a count of when no participants are speaking for a significant length of time (Monk & Gale, 2002; Olbertz-Siitonen, 2015).

Deictic word count: a count of the number of words that rely on context, typically provided by a grounding object of discussion (D’Angelo & Begel, 2017).

Homaecian et al. (2021) have also recently proposed a methodology for detailed analysis of conversational grounding, utilizing a diagramming system they developed called “Joint Action Storyboards.” With this scheme, they can measure the relationship of user interfaces, user interaction and cognition during communicative grounding.

5.3 Subjective Measures

In contrast to technical and behavioral measures, subjective measures obtain direct feedback from users, sometimes in the form of interviews (Nardi & Whittaker, 2001), but more often as surveys. For example, the International Telecommunications Union (ITU) advocates measuring Quality of Experience (QoE), which is “The overall acceptability of an application or service, as perceived subjectively by the end-user” (ITU, 2007). When possible, it is usually best to use standard surveys, since they are tried and tested, easily compared, and do not require the effort of generating a bespoke survey. Standard surveys used in video conferencing research include the following:

ITU-R BT.500: The ITU (2000) describes very detailed procedures for measuring subjective quality of audio and visuals, culminating in a survey. Typically, these are relatively short lists of closed questions addressing fidelity and overall experience. Examples can be found in several papers (Ishibashi et al., 2006; Berndtson et al., 2012; Schmitt et al., 2014; Gunkel et al., 2015; Schoenenberg et al., 2014).

Trust: if users trust other conferencing participants (Butler Jr., 1991), used in Bos et al. (2002) and Nguyen and Canny (2007).

Group Belongingness: whether one feels part of the conversational group (Kraut et al., 1998), used in Bennett et al. (2021).

Interpersonal Attraction: indicates liking and attraction of conversational partners (Oh et al., 2016), used in Wu et al. (2021).

Social Presence: to capture the sense that users are connected with others through the system

(Nowak & Biocca, 2003), used in Wu et al. (2021).

Copresence: for assessing the feeling that the user is with other entities (Nowak & Biocca, 2003), used in Wu et al. (2021).

Temple Presence Inventory: for user presence when engaging with media (Lombard et al., 2009), used in He et al. (2021).

Despite the advantages of standardized assessments, many studies create bespoke surveys for their own purposes, often driven by the specific needs of their research. For example, Tam et al. (2012) created a survey capturing “naturalness” of conversation.

6 Video Conferencing Solutions

Over the years of research in video conferencing, many improvements have been suggested and prototyped. We split these into two categories, window-based solutions aimed at gaze, the most dominant solution target in the Solutions column in Table 1; and solutions addressing other nonverbal cues that are poorly supported in video conferencing environments. By reviewing such solutions, we can build on them for future work, and identify gaps that are opportunities for further work.

6.1 Window-Based Solutions for Gaze

These video conferencing solutions seek to restore gaze and other spatial cues by defining a virtual window shared by conference participants, with two screens at different locations representing opposite sides of the window. As we have already discussed in Section 4.2, increased gaze awareness can increase productivity/efficiency (Monk & Gale, 2002; Koboyashi et al., 2021), trust formation (Nguyen & Canny, 2005), and presence (Lawrence et al., 2021). Modern solutions are more likely to utilize XR (Edelmann et al., 2013; Koboyashi et al., 2021).

MAJIC (Okada et al., 1994) creates a virtual conference table shared between remote locations, viewed through a shared window. To capture and display gaze,

it aligns a video projector and camera with each participant. Conference attendees using MAJIC felt that gaze was effectively conveyed, and a feeling of togetherness created. Created roughly 30 years ago, MAJIC suffered from poor image quality caused by a new, half-mirror technology transparent to cameras, user movement limitations, and the need to create a realistically sized conference table, which ultimately allows MAJIC to support only a maximum of four users. MAJIC's approach proved effective, and echoes across subsequent gaze solutions.

GA Display (Monk & Gale, 2002; see also Section 4.2) utilizes cameras pointed at half-silvered mirrors to capture the gaze of two participants, and a translucent display in front of a monitor in order to display captured gaze over a shared object of discussion. With GA Display, users required 55% fewer turns and 949 fewer words than in an audio-only system. While an early solution to the gaze problem, GA Display's experimentally measured benefits show the potential of restoring shared gaze awareness, particularly in collaborative tasks.

Gaze-2 (Vertegaal et al., 2003) conveys gaze in small group conferences. It differs from the other solutions in this section (aside from MAJIC) in that it can serve three or more remote locations (not just two remote screens defining the two sides of a virtual window). Each user sees the videos of other users in a row of tiles. Eye trackers for each user determine who they are looking at, and this information is used to rotate the user's tile toward that other user in each user's view, just as people might rotate their heads. Gaze-2 also employs the eye tracker to choose the camera with least parallax from among several cameras pointed at each user, so that users appear to be looking "straight out" of their tile. Under testing, automated camera shifts didn't affect perception of eye contact, and weren't considered highly distracting. Gaze-2 is unique in supporting gaze outside of one-on-one conference settings, and is a worthy vein for further research, given the nearly two decades of subsequent advancement in video conferencing technology.

Multiview (Nguyen & Canny, 2005; Nguyen & Canny, 2007) allows not just two participants, but two small groups to converse through a shared virtual window. Each group has one screen, displaying the virtual window. The screen is retroreflective like traffic

signs, reflecting light primarily back in the direction from which it arrived. For each participant, there is a matching camera and projector. A participant's projector is located close to their head, and with the retroreflective screen, ensures that each participant sees a unique display. A participant's camera is also located "close" to their head on the shared virtual window (e.g., for the rightmost participant, at the rightmost location on the screen) but at the remote location, giving them a view of the remote site that closely approximates their view through the window. Under testing, Multiview users formed trusting relationships just as quickly as face-to-face conference participants.

Kuster et al. (2012) restore mutual gaze in two-way conferencing by using image warping to colocate the camera and the display. In this way, when the user looks at their conference partner, the partner sees the user looking at them. Tracking and rendering are performed with a consumer GPU and Kinect sensor. This system is unique in working within traditional, single-camera conferencing systems, but the views synthesized with its image warping were not completely convincing.

GazeChat (He et al., 2021) is a novel audio-only conferencing system that represents users as gaze-aware 3D profile pictures; the eyes in the pictures moved, providing meaningful cues about where the corresponding conversational partner was looking, facilitated by webcam-based eye tracking and neural network rendering. In a 16-person study, GazeChat outperformed simple audio and video in feelings of eye contact, while significantly outperforming audio in user engagement.

6.1.1 XR Window-Based Solutions. More modern window-based solutions supporting gaze have included tracking, 3D display, and spatial audio to create a shared 3D space, a hallmark of XR technology.

Face2Face (Edelmann et al., 2013) might be viewed as an improved version of GA Display, adding a holographic projection screen supporting 3D viewing, a more compact and flexible form factor supporting a wider range of views, and touch interaction.

Kobayashi et al.'s (2021) system improves gaze in two-way conferencing using a unique embedding of multiple cameras into a screen, rather than with Kuster et al.'s (2012) image warping. Quantitative testing

shows that users can more accurately estimate gaze than they would in a single camera system. In the long run, we expect such systems to be compelling solutions, but significant engineering challenges remain to realize them, particularly in inexpensive consumer systems.

Google's Project Starline (Lawrence et al., 2021) creates a rich virtual window supporting continuous view change, coupled with spatial audio and stereoscopic display. Rather than relying only on several discrete cameras, Starline uses a combination of high resolution 3D capture and rendering subsystems. Based on participant surveys, Starline is notably superior to standard video conferencing in creating presence, attentiveness, reaction-gauging, and engagement.

6.2 Solutions Addressing Other Nonverbal Cues

Other solutions prioritize support for nonverbal communicative cues other than gaze, including objects of discussion, location, and gesture.

For example, Jansen et al.'s (2011) system is specifically for noisy, group-to-group calls from home. The system consists of a hidden microphone array for spatial audio, and a number of HD cameras to facilitate dynamic composition based on group movement. When compared to traditional conferencing solutions, Jansen et al.'s system offers more flexibility in the kind of tasks remote groups can engage in, such as playing networked digital games.

6.2.1 XR Solutions. However, most of these solutions make extensive use of XR technology. Consider MultiStage (Su et al., 2014), which is designed largely for stage performers, allowing remote performers to interact through a CAVE-like system on a virtual stage. Stages are equipped with sensors to detect actors and large displays visualizing the connected performance. MultiStage's primary innovation is a method of masking delays that can replace high latency actors with either prerecorded video of said actor or a computable model.

ReMa or Remote Manipulator (Feick et al., 2018; see also Section 4.3) offers rich, six-degree-of-freedom conversational grounding (in location and orientation). When a user demonstrates a physical object to a remote

partner by adjusting the object's orientation, ReMa will replicate this manipulation for the remote partner with a robotic arm and a duplicate of the object. This is accomplished with a combination of tracking equipment following the user with the original object, and their manipulations being mapped onto a robotic arm with a proxy object. Evaluation finds that ReMa served as an effective collaboration tool, especially when combined with video conferencing.

Wu et al. (2021) built a camera-based tracking system for VR that captures additional nonverbal communication cues like body language and facial expressions, and maps them onto virtual avatars. The system bolsters collaborative VR environments by allowing a broader range of socialization. Compared against a VR system with body-worn trackers in a game of charades, the expressive system facilitated more social presence and attractiveness, and improved task performance.

CollaboVR (He et al., 2020; see also Section 4.3) is a framework for VR collaboration in a shared 3D environment, allowing for sharing of freehand 2D sketches, which can be converted into 3D models with procedural, real-time animations. Based on cloud architecture to reduce client-side computational load, it also offers side-by-side (integrated), face-to-face (mirrored), and projected layouts to reduce clutter. Studies showed that the face-to-face layout was preferred, as it minimized obstructions from others, while also allowing users to focus on their collaborator's response.

Yu et al. (2021) created a 3D telepresence system that allows AR, MR, and VR interactions in a shared 3D environment. Remote users, joining the scene in VR, are presented as 3D avatars, while local users were presented as either avatars, or a point cloud representation, that captured their entire bodies, although their upper face was obscured by their headset. Through a study of a teleconsultation task, the point cloud representation proved more effective, as users found it more expressive than the avatar, despite the obfuscation of upper facial details caused by point display.

While this list of conferencing solutions may seem extensive, we find it surprisingly short, given the decades of research on this topic, and especially the new importance of video conferencing. We believe many research opportunities remain, which we detail next in Section 7.

7 Overview and Future Research

Table 1 summarizes the body of research we have reviewed, describing efforts to understand Zoom fatigue, human communication during video conferencing, deficits of current video conferencing technology, and proposed solutions. Over the course of our investigation, the importance of reducing the cognitive effort of conferencing by more effectively capturing and displaying aspects of in-person interaction has become evident. Not only do improvements upon communication improve productivity, they reduce the long-term strain of video conferencing. When updated with modern XR technology, many of the solutions we surveyed may prove much more effective.

With this overview of video conferencing research, we can identify a wide range of scientific and engineering opportunities that remain underexplored.

7.1 Opportunities for Scientific Research

Zoom Fatigue: Nearly all the research we surveyed studied single video conferences, rather than long-term conferencing across several remote meetings. In today's post-pandemic, hybrid working environment, this is a very common scenario that deserves attention. How important are reducing delay, communicating gaze, and offering objects of discussion in this long-term context? Answers could reprioritize work on engineered video conferencing solutions.

Gaze: In today's hybrid working environment, larger conferences with many participants have become more commonplace. We expect that gaze will gain importance with conference size, but research should confirm this.

Delay: Further research on delay in the post-pandemic context is needed. Much of the existing research is quite old, predating such technologies such as GPUs and machine learning, and applications such as widespread non-business use, gaming, and large-scale teaching. Delays well under 100 ms are important and

possible in many other settings; it would not be unreasonable to find similar needs for conferencing, especially over the long term. Finally, little is known about the effects of variation in delay on communication.

Nonverbal cues: This category of video conferencing shortcomings hides many that have seen little or no examination. In particular, we are not aware of any research on “big face” or backchanneling, much more common today than pre-pandemic, and very little on gestures.

Neurological measures: While we intended to include a column for neurological measures in Table 1, we found no research using these measures. This is troubling, since recent research shows that the human brain responds strongly to faces and social interaction (Hoehl et al., 2008). Further work should address this deficit as soon as possible, and may reveal phenomena otherwise missed in prior work.

Communicative measures: Space did not allow us to break out these measures from their parent behavioral category, although we are confident that these measures are not being used enough to evaluate new video conferencing solutions. This should change, so that future engineering efforts can be more effectively evaluated.

Complex models of video conferencing: While a great deal of research has investigated how one shortcoming affects video conferencing, we are aware of no research that studies how they interact. Consider Riedl et al.'s (2021) posited model of Zoom fatigue in Section 3: how strong are the relationships it depicts? Which are strong, and which are weak? Answers to such questions will help prioritize future research.

7.2 Opportunities for Engineering Research

Overcoming shortcomings: While much work has been done to overcome video conferencing's shortcomings, much remains. For example,

how can some of the solutions for delivering gaze be delivered with typical, modest conferencing hardware? Are there ways of compensating for delay that do not introduce serious communicative tradeoffs? How might modern XR technologies be leveraged to improve previous solutions?

Video conferencing at scale: As we have just mentioned when discussing gaze, video conferences during hybrid work commonly have dozens of participants. Even outside of gaze, little of the research we found addressed conferencing at these scales. This is unfortunate because when conferencing takes place at this scale, it is at its worst, and the need for improvement is greatest. How can conferencing support grounding, recognition, interruption, and discussion at such scales? Researchers may find inspiration in the different types of meetings and purposes that real-life conferences support.

Heterogeneous video conferencing systems: XR technologies are still emerging, and will not be ubiquitous for many years at least. Heterogeneous systems, with different technologies used by participants in the same conference, will be commonplace (e.g., Telelife; Orlosky et al., 2021). How can the technical, health, and social asymmetries of hybrid systems be accommodated, particularly in educational environments? A few studies of these issues exist (Yoshimura & Borst, 2021; Hopkins & Benford, 1998), but more are necessary to establish a complete picture of the complex effects on fatigue, fairness, and diversity of such heterogeneity in conferencing technology.

Better-than-real conferencing: Lastly, most research strives to make video conferencing as good as face-to-face. But where might it be better? For example, could video conferencing systems keep meetings effectively summarized and on schedule? Could they permit freedom of motion? Might they support design review more effectively than in-person meetings?

8 Conclusion

This paper has reviewed video conferencing's shortcomings both before and during the pandemic, ways of measuring them, and attempts at addressing them—with an eye towards XR's potential impact on a burgeoning hybrid working environment. We paid particular attention to the ways that the legacy of video conferencing research could apply to the long-term Zoom fatigue that emerged during the COVID-19 lockdown. Despite the relative recency of the fatigue phenomenon, prior studies and solutions offer a wealth of applicable observations and methods.

Additionally, we discussed many remaining scientific and engineering research opportunities, including research employing neurological and communicative measures, which should guide future investigation; and video conferencing at scale. We hope that the next review of video conferencing research will find that many of our remaining research questions will have at least initial answers.

Acknowledgments

Our sincerest thanks to Chung-Che Hsaio and Professor David Berube for many fruitful conversations about video conferencing. This work was supported by North Carolina State University's Department of Computer Science.

REFERENCES

- Aaronson, L. S., Teel, C. S., Cassmeyer, V., Neuberger, G. B., Pallikkathayil, L., Pierce, J., Press, A. N., Williams, P. D., & Wingate, A. (1999). Defining and measuring fatigue. *Image—the Journal of Nursing Scholarship* 31, 45–50. 10.1111/j.1547-5069.1999.tb00420.x
- Angelopoulos, A. N., Martel, J. N., Kohli, A. P., Conradt, J., & Wetzstein, G. (2021). Event-based near-eye gaze tracking beyond 10,000 Hz. *IEEE Transactions on Visualization and Computer Graphics*, 27, 2577–2586. 10.1109/TVCG.2021.3067784
- Bailenson, J. (2020). Why Zoom meetings can exhaust us. *Wall Street Journal*. Retrieved from <https://www.wsj.com/articles/why-zoom-meetings-can-exhaust-us-11585953336>

- Bailenson, J. N. (2021). Nonverbal overload: A theoretical argument for the causes of zoom fatigue. *Technology, Mind, and Behavior*, 2. 10.1037/tmb0000030
- Becher, A., Angerer, J., & Grauschopf, T. (2020). Negative effects of network latencies in immersive collaborative virtual environments. *Virtual Reality*, 24(3), 369–383. 10.1007/s10055-019-00395-9
- Bennett, A. A., Campion, E. D., Keeler, K. R., & Keener, S. K. (2021). Videoconference fatigue? Exploring changes in fatigue after videoconference meetings during COVID-19. *Journal of Applied Psychology* 106, 330–344. 10.1037/apl0000906
- Berndtsson, G., Folkesson, M., & Kulyk, V. (2012). Subjective quality assessment of video conferences and telemeetings. *19th International Packet Video Workshop*, 25–30. 10.1109/PV.2012.6229740. ISSN: 2167-969X
- Bohannon, L. S., Herbert, A. M., Pelz, J. B., & Rantanen, E. M. (2013). Eye contact and video-mediated communication: A review. *Displays* 34, 177–185. 10.1016/j.displa.2012.10.009
- Bos, N., Olson, J., Gergle, D., Olson, G., & Wright, Z. (2002). Effects of four computer-mediated communications channels on trust development. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 135–140.
- Boyaci, O., Forte, A., Baset, S. A., & Schulzrinne, H. (2009). vDelay: A tool to measure capture-to-display latency and frame rate. *11th IEEE International Symposium on Multimedia*, 194–200. 10.1109/ISM.2009.46
- Butler Jr., J. K. (1991). Toward understanding and measuring conditions of trust: Evolution of a conditions of trust inventory. *Journal of Management*, 17, 643–663.
- Creswell, J. W. (2014). *Qualitative, quantitative and mixed methods approaches*. Sage Publications.
- D'Angelo, S., & Begel, A. (2017). Improving communication between pair programmers using shared gaze awareness. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 6245–6290. 10.1145/3025453.3025573
- D'Angelo, S., & Gergle, D. (2018). An eye for design: Gaze visualizations for remote collaborative work. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–12.
- Degges-White, S. (2020). Zoom fatigue: Don't let video meetings zap your energy. *Psychology Today*. Retrieved from <https://www.psychologytoday.com/us/blog/lifetime-connections/202004/zoom-fatigue-dont-let-video-meetings-zap-your-energy>
- deHahn, P. (2020). Zoom fatigue is something the deaf community knows very well. *Quartz*. Retrieved from <https://qz.com/1855404/zoom-fatigue-is-something-the-deaf-community-knows-very-well>
- Driskell, J., Radtke, P., & Salas, E. (2003). Virtual teams: Effects of technological mediation on team performance. *Group Dynamics: Theory Research And Practice*, 7, 297–323. 10.1037/1089-2699.7.4.297
- Driskell, J. E., & Radtke, P. H. (2003). The effect of gesture on speech production and comprehension. *Human Factors*, 45, 445–454. 10.1518/hfes.45.3.445.27258
- Edelmann, J., Mock, P., Schilling, A., & Gerjets, P. (2013). Preserving non-verbal features of face-to-face communication for remote collaboration. In *Cooperative design, visualization, and engineering* (Y. Luo, Ed.) Springer. Lecture Notes in Computer Science, 27–34. 10.1007/978-3-642-40840-3_4
- Fauville, G., Luo, M., Muller Queiroz, A. C., Bailenson, J. N., & Hancock, J. (2021a). Zoom exhaustion & fatigue scale. SSRN Scholarly Paper ID 3786329, Social Science Research Network, Rochester, NY. 10.2139/ssrn.3786329
- (2021b). Nonverbal mechanisms predict zoom fatigue and explain why women experience higher levels than men. *SSRN Electronic Journal*. 10.2139/ssrn.3820035
- Feick, M., Mok, T., Tang, A., Oehlberg, L., & Sharlin, E. (2018). Perspective on and re-orientation of physical proxies in object-focused remote collaboration. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–13.
- Fosslien, L., & Duffy, M. W. (2020). How to combat Zoom fatigue. *Harvard Business Review*. Retrieved from <https://hbr.org/2020/04/how-to-combat-zoom-fatigue>
- Friston, S., & Steed, A. (2014). Measuring latency in virtual environments. *IEEE Transactions on Visualization and Computer Graphics*, 20, 616–25. 10.1109/TVCG.2014.30
- Gan, Y., Greiffenhagen, C., & Reeves, S. (2020). Connecting distributed families: Camera work for three-party mobile video calls. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–12. 10.1145/3313831.3376704
- Gasteratos, K., Spyropoulou, G.-A., & Suess, L. (2021). “Zoom Dysmorphia”: A new diagnosis in the COVID-19 pandemic era? *Plastic and Reconstructive Surgery*, 148, 1073e. 10.1097/PRS.00000000000008559
- Gergle, D., Kraut, R., & Fussell, S. (2012). Using visual information for grounding and awareness in collaborative tasks. *Human-Computer Interaction*, 28(1), 1–39. 10.1080/07370024.2012.678246

- Grayson, D. M., & Monk, A. F. (2003). Are you looking at me? Eye contact and desktop video conferencing. *ACM Transactions on Computer-Human Interaction*, 10, 221–243. 10.1145/937549.937552
- Gunkel, S. N., Schmitt, M., & Cesar, P. (2015). A QoE study of different stream and layout configurations in video conferencing under limited network conditions. *Seventh International Workshop on Quality of Multimedia Experience*, 1–6. 10.1109/QoMEX.2015.7148085
- He, Z., Du, R., & Perlin, K. (2020). CollaboVR: A reconfigurable framework for creative collaboration in virtual reality. *IEEE International Symposium on Mixed and Augmented Reality*, 542–554. 10.1109/ISMAR50242.2020.00082. ISSN: 1554-7868
- He, Z., Wang, K., Feng, B. Y., Du, R., & Perlin, K. (2021). GazeChat: Enhancing virtual conferences with gaze-aware 3D photos. *34th Annual ACM Symposium on User Interface Software and Technology*, 769–782.
- Hickman, S. (2020). Zoom exhaustion is real. Here are six ways to find balance and stay connected. *mindful*. Retrieved from <https://www.mindful.org/zoom-exhaustion-is-real-here-are-six-ways-to-find-balance-and-stay-connected/>
- Hoehl, S., Reid, V., Mooney, J., & Striano, T. (2008). What are you looking at? Infants' neural processing of an adult's object-directed eye gaze. *Developmental Science*, 11, 10–16. 10.1111/j.1467-7687.2007.00643.x
- Homacian, L., Wallace, J. R., & Scott, S. D. (2021). Joint action storyboards: A framework for visualizing communication grounding costs. *Proceedings of the ACM on Human-Computer Interaction*, 5, 28:1–28:27. 10.1145/3449102
- Hopkins, G., & Benford, S. (1998). Vivid: A symbiosis between virtual reality and video conferencing. Retrieved from <https://www.semanticscholar.org/paper/Vivid%3A-A-Symbiosis-between-Virtual-Reality-and-Reynard-Benford/715c83821a70bf27662a76f0fc16cf2c2c290d04>
- Ishibashi, Y., Nagasaka, M., & Fujiyoshi, N. (2006). Subjective assessment of fairness among users in multipoint communications. *Proceedings of the 2006 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*, 69–cs. 10.1145/1178823.1178905
- ITU. (2000). BT.500: Methodology for the subjective assessment of the quality of television pictures. *International Telecommunication Union*. Retrieved from <https://www.itu.int/rec/R-REC-BT.500/en>
- (2007). Definition of quality of experience (QoE). *International Telecommunication Union*. Retrieved from <https://www.itu.int/md/T05-FG.IPTV-IL-0050/en>
- Jansen, J., & Bulterman, D. C. A. (2013). User-centric video delay measurements. *Proceedings of the 23rd ACM Workshop on Network and Operating Systems Support for Digital Audio and Video*, 37–42. 10.1145/2460782.2460789
- Jansen, J., Cesar, P., Bulterman, D., Stevens, T., Kegel, I., & Issing, J. (2011). Enabling composition-based video conferencing for the home. *IEEE Transactions on Multimedia*, 13, 869–881. 10.1109/TMM.2011.2159369
- Jiang, M. (2020). The reason Zoom calls drain your energy. *BBC Remote Control*. Retrieved from <https://www.bbc.com/worklife/article/20200421-why-zoom-video-chats-are-so-exhausting>
- Kobayashi, K., Komuro, T., Kagawa, K., & Kawahito, S. (2021). Transmission of correct gaze direction in video conferencing using screen-embedded cameras. *Multimedia Tools and Applications*, 80, 31509–31526. 10.1007/s11042-020-09758-w
- Kraut, R. E., Gergle, D., & Fussell, S. R. (2002). The use of visual information in shared visual spaces: Informing the development of virtual co-presence. *Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work*, 31–40.
- Kraut, R. E., Rice, R. E., Cool, C., & Fish, R. S. (1998). Varieties of social influence: The role of utility and norms in the success of a new communication medium. *Organization Science*, 9, 437–453. 10.1287/orsc.9.4.437
- Kuster, C., Popa, T., Bazin, J.-C., Gotsman, C., & Gross, M. (2012). Gaze correction for home video conferencing. *ACM Transactions on Graphics*, 31, 1–6. 10.1145/2366145.2366193
- Lam, H., Bertini, E., Isenberg, P., Plaisant, C., & Carpendale, S. (2011). Empirical studies in information visualization: Seven scenarios. *IEEE Transactions on Visualization and Computer Graphics*, 18(9), 1520–1536. 10.1109/TVCG.2011.279
- Lawrence, J., Goldman, D. B., Achar, S., Blascovich, G. M., Desloge, J. G., Fortes, T., Gomez, E. M. et al. (2021). Project Starline: A high-fidelity telepresence system. *ACM Transactions on Graphics*, 40(6).
- Li, R., Whitmire, E., Stengel, M., Boudaoud, B., Kautz, J., Luebke, D., Patel, S., & Akşit, K. (2020). Optical gaze tracking with spatially-sparse single-pixel detectors. *IEEE International Symposium on Mixed and Augmented Reality*, 117–126. 10.1109/ISMAR50242.2020.00033

- Lombard, M., Bolmarcich, T., & Weinstein, L. (2009). Measuring presence: The Temple Presence Inventory. Retrieved from https://www.researchgate.net/publication/228450541_Measuring_Presence_The_Temple_Presence_Inventory
- Monk, A. F., and Gale, C. (2002). A look is worth a thousand words: Full gaze awareness in video-mediated conversation. *Discourse Processes*, 33, 257–278. Publisher: Routledge eprint: 10.1207/S15326950DP3303_4
- Mota, D. D. C. F., & Pimenta, C. A. M. (2006). Self-report instruments for fatigue assessment: A systematic review. *Research and Theory for Nursing Practice*, 20(1), 49–78. 10.1891/rtnp.20.1.49
- Nardi, B., & Whittaker, S. (2001). The place of face-to-face communication in distributed work. In P. Hinds & S. Kiesler (Eds.), *Distributed work* (pp. 83–110). Retrieved from <https://psycnet.apa.org/record/2002-17012-004>
- Nesher Shoshan, H., & Wehrt, W. (2021). Understanding “Zoom fatigue”: A mixed-method approach. *Applied Psychology*. n/a. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/apps.12360>
- Nguyen, D., & Canny, J. (2005). Multiview: Spatially faithful group video conferencing. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 799–808.
- (2007). Multiview: Improving trust in group video conferencing through spatial faithfulness. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1465–1474.
- Nowak, K. L., & Biocca, F. (2003). The effect of the agency and anthropomorphism on users’ sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 12, 481–494.
- Oducado, R. M. F., Fajardo, M. T. R., Parreño-Lachica, G. M., Maniago, J. D., Villanueva, P. M. B., Dequilla, Ma. A. C. V., Montaña, H. C., & Robite, E. E. (2021). Is video-conference “Zoom” fatigue real among nursing students? *Journal of Loss and Trauma*, 0(0), 1–3. 10.1080/15325024.2021.1950987
- Oh, S. Y., Bailenson, J., Krämer, N., & Li, B. (2016). Let the avatar brighten your smile: Effects of enhancing facial expressions in virtual environments. *PLOS One*, 11, e0161794. 10.1371/journal.pone.0161794
- Okada, K.-I., Maeda, F., Ichikawa, Y., & Matsushita, Y. (1994). Multiparty videoconferencing at virtual social distance: MAJIC design. *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*, 385–393. 10.1145/192844.193054
- Olbertz-Siitonen, M. (2015). Transmission delay in technology-mediated interaction at work. *Psychology Journal*, 13(2–3), 203–234.
- O’Malley, C., Langton, S., Anderson, A., Doherty-Sneddon, G., & Bruce, V. (1996). Comparison of face-to-face and video-mediated interaction. *Interacting with Computers*, 8, 177–192. 10.1016/0953-5438(96)01027-2
- Orlosky, J., Sra, M., Bektaş, K., Peng, H., Kim, J., Kos’myna, N., Höllerer, T., Steed, A., Kiyokawa, K. et al. (2021). Telelife: The future of remote living. *Frontiers in Virtual Reality*, 2. <https://www.frontiersin.org/article/10.3389/frvir.2021.763340>
- Pachnowski, L. (2002). Virtual field trips through video conferencing. *Learning & Leading with Technology*, 29(6), EJ654047.
- Perer, A., & Shneiderman, B. (2009). Integrating statistics and visualization for exploratory power: From long-term case studies to design guidelines. *IEEE Computer Graphics and Applications*, 29(3), 39–51. 10.1109/MCG.2009.44
- Peters, C. (1938). Talks on “see-phone”: Television applied to German telephones enables speakers to see each other. *The New York Times*, 1687–1695.
- Ratan, R., Miller, D. B., & Bailenson, J. N. (2021). Facial appearance dissatisfaction explains differences in Zoom fatigue. *Cyberpsychology, Behavior, and Social Networking*. 10.1089/cyber.2021.0112
- Rommel, A. (2021). Scientists want virtual meetings to stay after the COVID pandemic. *Nature*, 591(7849), 185–187. 10.1038/d41586-021-00513-1
- Riedl, R. (2021). On the stress potential of videoconferencing: definition and root causes of Zoom fatigue. *Electronic Markets*. 10.1007/s12525-021-00501-3
- Robert, Y. (2020). Here’s why you’re Feeling Zoom fatigue. *Forbes*. Retrieved from <https://www.forbes.com/sites/yolarobert1/2020/04/30/heres-why-youre-feeling-zoom-fatigue/?sh=40eeabe02ac6>
- Roberts, D., Duckworth, T., Moore, C., Wolff, R., & O’Hare, J. (2009). Comparing the end to end latency of an immersive collaborative environment and a video conference. *Proceedings of the 13th IEEE/ACM International Symposium on Distributed Simulation and Real Time Applications*, 89–94. 10.1109/DS-RT.2009.43
- Rosenberg, S. (2020). Here’s why Zoom fatigue is real. *Axios*. Retrieved from <https://www.axios.com/2020/04/22/zoom-fatigue-coronavirus-teleconferencing>

- Schmitt, M., Gunkel, S., César, P., & Bulterman, D. (2014). Asymmetric delay in video-mediated group discussions. *Sixth International Workshop on Quality of Multimedia Experience*. 10.1109/QoMEX.2014.6982280
- Schoenenberg, K. (2016). The quality of mediated-conversations under transmission delay. Doctoral thesis, Technische Universität Berlin, Berlin. 10.14279/depositonce-4990
- Schoenenberg, K., Raake, A., & Koeppe, J. (2014). Why are you so slow? Misattribution of transmission delay to attributes of the conversation partner at the far-end. *International Journal of Human-Computer Studies*, 72, 477–487. 10.1016/j.ijhcs.2014.02.004
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2001). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin. <http://catdir.loc.gov/catdir/enhancements/fy1105/2001131551-t.html>
- Sklar, J. (2020). “Zoom fatigue” is taxing the brain. Here’s why that happens. *National Geographic*. Retrieved from <https://www.nationalgeographic.com/science/article/coronavirus-zoom-fatigue-is-taxing-the-brain-here-is-why-that-happens>
- Stotts, P., Gyllstrom, K., Miller, D., Mcc, J., & Smith, J. (2010). *Facetop: An integrated desktop/video interface for individual users and paired collaborations*. Technical Report TR05-005. Department of Computer Science, University of North Carolina at Chapel Hill.
- Su, F., Bjørndalen, J., Ha, P., & Anshus, O. J. (2014). Masking the effects of delays in human-to-human remote interaction. *Federated Conference on Computer Science and Information Systems*. 10.15439/2014F137
- Tam, J., Carter, E., Kiesler, S., & Hodgins, J. (2012). Video increases the perception of naturalness during remote interactions with latency. *CHI ’12 Extended Abstracts on Human Factors in Computing Systems*, 2045–2050. 10.1145/2212776.2223750
- Vertegaal, R., Weevers, I., Sohn, C., & Cheung, C. (2003). Gaze-2: Conveying eye contact in group video conferencing using eye-controlled camera direction. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 521–528.
- Whittaker, S. (2003). Theories and methods in mediated communication. In *The handbook of discourse processes*. Routledge.
- Wu, Y., Wang, Y., Jung, S., Hoermann, S., & Lindeman, R. W. (2021). Using a fully expressive avatar to collaborate in virtual reality: Evaluation of task performance, presence, and attraction. *Frontiers in Virtual Reality*, 2, 10. 10.3389/frvir.2021.641296
- Yan, B., Ni, S., Wang, X., Liu, J., Zhang, Q., & Peng, K. (2020). Using virtual reality to validate the Chinese version of the independent television commission, Sense of Presence Inventory. *SAGE Open*, 10(2), 2158244020922878. 10.1177/2158244020922878
- Yoshimura, A., & Borst, C. W. (2021). A study of class meetings in VR: Student experiences of attending lectures and of giving a project presentation. *Frontiers in Virtual Reality*, 2, 10.3389/frvir.2021.648619
- Yoshitake, H. (1971). Relations between the symptoms and the feeling of fatigue. *Ergonomics*, 14, 175–186. 10.1080/00140137108931236
- Yu, K., Gorbachev, G., Eck, U., Pankratz, F., Navab, N., & Roth, D. (2021). Avatars for teleconsultation: Effects of avatar embodiment techniques on user perception in 3D asymmetric telepresence. *IEEE Transactions on Visualization and Computer Graphics*, 27, 4129–4139. 10.1109/TVCG.2021.3106480