Exploring the Effects of Audience Visibility on Presenters and Attendees in Online Educational Presentations

Full paper

Jennifer Marlow
FXPAL
Palo Alto, California
United States
marlow@fxpal.com

Jason Wiese
University of Utah
Salt Lake City, Utah
United States
wiese@cs.utah.edu

Daniel Avrahami
FXPAL
Palo Alto, California
United States
daniel@fxpal.com

ABSTRACT

Video conferencing is widely used to help deliver educational presentations, such as lectures or informational webinars, to a distributed audience. While individuals in a dyadic conversation may be able to use webcam streams to assess the engagement level of their interlocutor with some ease, as the size of the audience in a video conference setting increases, it becomes increasingly difficult to interpret how engaged the overall group may be. In this work, we use a mixed-methods approach to understand how presenters and attendees of online presentations use available cues to perceive and interpret audience behavior (such as how engaged the group is). Our results suggest that while webcams are seen as useful by presenters to increase audience visibility and encourage attention, audience members do not uniformly benefit from seeing others' webcams; other interface cues such as chat may be more useful and informative engagement indicators for both parties. We conclude with design recommendations for future systems to improve what is sensed and presented.

CCS CONCEPTS

• **Human centered computing → Human Computer Interaction (HCI)**

KEYWORDS

Video conferencing; Presentation; Engagement

ACM Reference format:

J. Marlow, J. Wiese, and D. Avrahami. 2017. Exploring the Effects of Audience Visibility on Presenters and Attendees in Online Educational Presentations. In *Proceedings of the 8th International Conference on Communities and Technologies* (C&T '17), 10 pages.
DOI: 10.1145/3083671.3083672

C&T '17, June 26-30, 2017, Troyes, France

ACM ISBN 978-1-4503-4854-6/17/06…$15.00
http://dx.doi.org/10.1145/3083671.3083672

1. INTRODUCTION

Advances in video conferencing tools are making it easier for people around the world to advance their education by viewing online lectures, training, and webinars. In distributed educational settings such as these, a presenter often delivers information to a large group of listeners, who may be connecting in real-time or viewing an asynchronous recording of the presentation. In addition to streaming video feeds of presenter and audience webcams, videoconferencing tools often also support chat, screen sharing, slides, and polls as optional plugins.

One oft-cited benefit of video conferencing (compared to other distributed communication media) is the ability to access visible cues about how a conversation partner is reacting and what they are doing [6]. Furthermore, since attendees may need to perform multiple tasks during an online meeting [27], being seen on video can be helpful to promote focus. According to a 2014 survey [1], most professionals (76%) feel attendees are less likely to pay attention during a presentation-centric Web conferencing meeting than during a video-centric all-in-one meeting (this belief has been represented in the literature, such as in [4]).

Prior research suggests that seeing the video of meeting participants provides further benefits; for example, during a one-on-one video call, nonverbal and verbal cues can be used to detect or infer another person's activity and attitudes, often quite accurately [24]. However, an important question is to what extent this method can scale to larger group settings with people in various states of attention. In such settings, getting a read on the "average" group engagement may be hard based on an assessment of each individual member's perceived activity.

The goal of the work presented in this paper is to understand how webcam streams of audience members are used by presenters to assess engagement, and whether seeing other audience members affects attendees' comprehension and engagement in the presentation. Specifically, we investigated the following research questions:

- What do leaders of web-based presentations perceive as the advantages and disadvantages of seeing their audience?
- Is seeing other audience members harmful or beneficial to attendees of web-based presentations?

To explore these questions, we conducted a mixed-methods study that began with interviews to understand how presenters currently assess their audience's engagement and what tool features they deem to be useful in doing so. We then used this information to design and conduct an online experiment investigating how different interface features and characteristics of the audience influenced a viewer's perception of others' attention and interest in the presentation.

The contributions of this work include an understanding of the role of web-streaming cameras and other functionalities in video-mediated presentations. We found that while presenters believed there was benefit in seeing the webcam streams of audience members listening to their presentations, they were primarily used to encourage accountability through visibility. Particularly as audience size increased, requiring webcam feeds from all was either not allowed by the limitations of video conferencing tools or occupied most available screen space. Based on our examinations, it also emerged that alternate approaches (such as using a backchannel) may provide more reliable and useful cues of audience engagement. We conclude with proposed alternative design parameters to better serve both leaders and viewers of online video-mediated presentations.

2. RELATED WORK

Past research in the field of CSCW has revealed the potential benefits of sharing and viewing webcam livestreams among distributed colleagues in a work context. For example, being able to see others in real time can help with community building, presence, and availability awareness (e.g. [3; 11]). As videoconferencing technology has improved, its use has expanded into other areas and use cases. Understanding audience engagement has been applied in a variety of domains, both in asynchronous educational settings (e.g. MOOCs), and in face-to-face meetings or classrooms. Little has been done in the context of synchronous, video-mediated communications. Nevertheless, relevant findings and approaches in these other domains informed our hypotheses and can be applied to future solutions.

2.1 Assessing audience engagement

Prior work has looked at different approaches for assessing audience reactions such as understanding, confusion, or engagement in a variety of contexts. These approaches generally take two forms: Explicit (requiring individuals to take some concrete action to signal a reaction to material being presented) or Implicit (inferring individual attitudes/behaviors through other signals). This work has primarily been applied to education (synchronous and asynchronous settings) and in-person, face-to-face meeting settings. Table 1 gives a summary of related projects in each area.

In the asynchronous, explicit category, prior work has looked at students watching videos for educational purposes, such as in the context of a MOOC, and taking some action to tag or annotate sections of content with questions or to indicate points of confusion [10, 12].

In the asynchronous, implicit category, prior work has used eye gaze as an implicit indicator of attentional focus [30], or as a tool for note taking on lecture videos [25]. Other implicit metrics have also been explored; for example, using pulse [35] or EEG activity [32] to signal viewers' internal states and reactions to asynchronous, online content.

Other research has focused on providing awareness of audience reactions in a synchronous setting. Here, different approaches have been taken to allow viewers to provide real-time feedback over remote presentations via tags [14] or through a mobile phone app for in-person meetings [34]. Implicit assessments of attention in synchronous settings include, again, using eye gaze or affect/face detection for either remote students in an online educational situation [2, 16, 29], or in the classroom [26]. Other related work, not exclusively limited to a meeting situation, also considered the possibilities of being more transparent about activities individuals are engaged in on their mobile devices [15, 19].

	Examples of explicit approach to inferring audience engagement	**Examples of implicit approach to inferring audience engagement**
Asynchronous presentations (MOOCs, educational videos)	Signaling confusing parts of a video [10, 11]; Leaving time-anchored comments [21]	Eye gaze; pulse; EEG activity to indicate reactions to watching a video [24, 29, 30, 31, 33]
Synchronous, in-person presentations	Asking questions or providing feedback to presenters via mobile apps [8, 34]	Eye gaze; Providing ambient information about individual activity [2, 16, 25, 26]
Synchronous, distributed video-mediated presentations	SynTag [14]; Adobe Connect tools like raising hands	Measuring and visualizing how much participants speak (based on text chat) [22, 33]

Table 1. Categorization of approaches to infer engagement

The issue of understanding audience reactions is useful in the educational context, because teachers online often don't have access to typical classroom signals of confusion such as a raised hand [12]. Currently-available video conferencing tools like Adobe Connect offer some features to help presenters engage their audience more, including "raise hand," polls, a chat bar, and so on. However, these are up to the individual presenter to use and it is unclear how and why people rely on these different features and how much they help.

While a couple of examples exist, little work has explored the role of implicit cues in a synchronous, video-mediated group setting (or how others can assess audience engagement/reactions without the audience members doing anything on their own). Related work has focused on visualizing contributions to an online text-based discussion (e.g. [22, 33]), primarily to other group members within small group discussions. In large group settings where a presenter does most of the talking, one additional non-intrusive way of quickly understanding audience engagement could be to use cues present in the streaming webcam feed.

In the context of an online presentation setting, better audience visibility (through webcam streaming) has the potential to benefit both presenter and audience alike. On the presenter side, having more detailed insight into audience behavior and reactions can help them adapt the delivery of material accordingly (e.g. make it more understandable or be more engaging). Prior work on showing presenters or interviewees feedback about their behavior suggests that it can help people adapt and improve what they do [13], so it is likely that a similar adaptation can occur in an online setting. In work looking at the role of visualizing an instructor's face in online lectures, Kizilcec *et al.* [18] provided mixed hypotheses for the effects of sharing the instructor's face, including the tension between the benefits of increased social presence and the potential negative effects of

distraction and cognitive load. We discuss these constructs as they may apply to audience visibility as well.

2.2 Assessing engagement of other audience members

Much as people in a real-world presentation may notice the activities of those around them, including potentially being distracted by movement of images and light coming from others' laptop use [28], audiences in a video-mediated presentation may be positively or negatively affected by the visible presence of others. Here we present prior work supporting both potential outcomes.

2.2.1 Social presence

On the audience side, a better sense of "social presence" can improve shared attention. Shared attention can be defined as "the perception of in-the-moment attention to an object from a first-person plural perspective" [31]. The mere perception that others are also paying attention to the same thing we are has been shown to affect memory, motivation, judgment, and emotion [31]. The presence of others can also influence one's own exertion of mental effort: Prior work has suggested that performing a task next to a person exerting effort on a task will make you do the same [9].

It therefore stands to reason that seeing other people can help the presenter judge what people are doing and adapt content accordingly. Seeing other people (who are perceived to be paying attention) can also be beneficial to an isolated viewer by giving them a greater sense of community [20]. However, it could be that seeing other people will not be beneficial if those people are perceived to be bored, distracted, or otherwise disengaged. This is because emotional contagion often happens when shared attention is felt, meaning that a feeling can be amplified by the presence of others [31].

2.2.2 Distraction

Another potential problem with showing or viewing more webcams is that it may be distracting. Visual dispersion, or needing to attend to a variety of visual information sources has been linked to increased difficulty in a classroom setting [5]. This issue can affect both presenter and listener alike. However, results of an experiment looking at the effects of showing a video of the instructor's face showed mixed effects. The video of the instructor's face was shown to increase a feeling of the instructor's social presence, but it also was viewed as distracting and hindered comprehension of the material being presented [17]. Displaying embedded chat during a MOOC was deemed to be distracting by around half of participants who viewed it, but there was no evidence that the presence of chat hindered learning [7]. Thus, understanding the interplay between the social presence of other listeners provided by webcam streams and the potential distraction introduced by doing so was the focus of our research.

3. UNDERSTANDING PRESENTERS' USE OF A VIDEO CHANNEL IN ONLINE PRESENTATIONS

To understand presenters' use of a video channel in online presentations, we conducted interviews with eight professionals who led regular online meetings and presentations with groups of ten or more attendees. Each interview lasted about 30 minutes and focused on interviewees' experiences using online meeting tools (including video conferencing) to teach, share and exchange information. We asked them to describe how they used video conferencing tools to present or lecture to their remote audiences. Two researchers conducted each interview, one as the primary interviewer and the other as note-taker. Additionally, interviews were recorded and referred back to for exact quotations.

3.1 Participants

All eight interviewees (two female, six male) were in the educational domain and regularly gave lectures or presentations at the middle/high school level (P1), the university level (P2, P3, P4, P5), or in the context of online training/education for professionals (P6) or to a large research group (P7, P8).

4. INTERVIEW FINDINGS

Following the interviews, we used an open coding approach to categorize and organize the responses into three themes pertaining to our research question: 1) presenters' perceived need and benefits of using a video channel, 2) the role of back-channel features as a complement to video, and 3) drawbacks and challenges of using a video channel.

4.1 Perceived need and benefits of using a video channel

The interviewees stated that assessing audience attention remotely is much more difficult than in person, particularly due to the lack of visual cues about viewer engagement. The interviewees considered the video channel a necessary tool for assessing their audience's engagement and for enforcing accountability.

4.1.1 Assessing engagement and enforcing accountability

Five interviewees mentioned that having viewers pay attention was important and felt that using attendees' webcam feeds was a reliable source of information about engagement. For example, P7 described assessing attendees who share their webcams in this way:

"We can totally tell that these people are very carefully listening…responding to questions is a different thing, but I'm pretty sure they're not doing anything else…You can pretend a lot of things, but I feel like if it's pretty natural you can tell if this person is paying attention or doing something else" (P7).

These interviewees also expressed a belief that all attendees should be showing their webcam at all times. They prioritized this highly, often making it a requirement, because of a sense that this was important to make people pay attention. As P1, a teacher of an online middle school program, mentioned:

"For the 13 year olds, again, they're 13. If I don't have them on video all the time, some of them are definitely going to be there engaged and paying attention…but a third are going to be off somewhere, I don't know. I mean they may physically be present but they won't be mentally present so I have to do that…" (P1).

Finally, one interesting reason for using a video channel, expressed by one of our interviewees involved in an online university degree program, was the need to "verify" the identity of the students attending the class:

"How do you know it's them? We know it's them. We can see them."… "The accreditors want you to really know who it is." (P5).

4.2 Backchannel as a complement to video

While presenters saw webcams as a valuable and useful way to both understand and encourage viewer engagement and activity, there were also alternate methods of doing so (such as optionally using "poll" or Q&A pods, or directly addressing individuals with questions). However, the most commonly used complement to the video feed was a chat functionality that ran alongside the meeting.

For all eight interviewees, the video meeting was augmented with a backchannel tool, such as a chat feature within Adobe Connect, 2u, Zoom, or Hangouts video conferencing (P1, P2, P3, P4, P5, P6, P8) or a parallel chat platform such as a Slack channel where people could type in comments and questions (P7). This backchannel was often seen as an added feature for engagement that is not possible in a traditional meeting or classroom setting:

"One of the things that was sort of unexpectedly rich about it, much better than a regular classroom, is, the teacher's talking, the kids are talking, and then there's a sidebar which is a chat screen which is going non-stop..." (P4).

For our interviewees, these backchannels were either used on an as-needed basis, with listeners asking questions spontaneously, or used as a means of getting everyone to participate and respond at the same time, for example, by typing in an answer and hitting "send" on the count of three (P2, P5).

Chat was seen as a useful indicator of engagement because it clearly linked relevant comments to an individual participant:

"If people have links, they can contribute them to the ongoing chat window, and that is how I can tell if they are engaged or not" (P3).

Chat was also useful to allow active participation from audience members in public or semi-public environments. For example, a person in P7's group regularly participated from his workplace and was able to share his video, but not talk aloud.

"[The attendee is] actually working, and there's like people walking in the background. He never talks, he just comes and participates on Slack because he's in the office and cannot speak." (P7).

4.2.1 Combining signals from chat and video

Combining the signals from a chat backchannel and a video stream helped presenters get a richer view of what their audience was doing as well as promote engagement. For example, interviewees described matching the visual cues of typing (seen from the webcam) with the appearance of text in the chat.

"I can tell that if this person is fast typing something but nothing is coming on Slack, it means that this person is typing somewhere else" (P7).

4.3 Drawbacks and challenges of using a video channel

Using webcam views of the audience was regarded by our interviewees as a desirable; however, interviewees also described several drawbacks and challenges associated with using video channels.

4.3.1 The challenge of assessing engagement from video

Similar to descriptions in [4], our interviewees acknowledged the difficulty of precisely assessing engagement through video.

"*It's much harder to read a class online -- to tell when students are distracted and not focusing, and therefore need more support / encouragement.*" (P1).

When attendees engage in other activities, those remain hidden from the presenter. For example, it is often impossible to know what kinds of applications attendees are using at a given time. One instructor addressed the audience in the following way:

"*I don't know how you're taking notes, but if you have a Word doc open, write down customer goals for 30 seconds.*" (P6).

4.3.2 A tradeoff of screen real estate

One tradeoff of prioritizing individual video streams in a presentation with many attendees is that it takes away screen real estate that could be used to show other content.

"15 people on video, that takes up a huge amount of the screen and it's a choice I continue to make because it does mean that I can see them, and that's so important. I can't have very big images up with so many faces." (P1).

This tradeoff raises an important question: While a presenter may wish to see the entire audience, what would benefit each audience member most? Do all viewers need to see each other? Is it enough for them to see only some other viewers? Or would it be better for them not to see each other at all?

4.3.2 Factors inhibiting the ability to view all audience members

While some of our interviewees required attendees to share their webcams, four of our interviewees said that they did not set rules about whether attendees should or should not stream their webcam. This led to a mixed setting in which some participants had webcams on and some did not.

"Usually what happens is that from 20 students I have four that they don't have the webcam [on] and it's hard for me to interact with them when I have the faces of the other students looking at me." (P2).

Furthermore, in a couple of interviews, participants described challenges associated with tool restrictions of the number of attendees they could see. For example, P2 initially used GoToMeeting web conferencing software, but found that, *"One important limitation was that it only showed five students with camera. So we couldn't see each other."* He then switched over to a platform (Zoom) that allowed all students to show their webcams. He said, *"For me, this makes a real big difference because I am teaching to persons that I don't know if they are paying attention, if they are falling asleep. That, for me, is kind of an extra stress while doing the lecture."*

Similarly, the meetings led by P7 had as many as 35 people attending, but the tool used (Google Hangouts OnAir) only allowed ten attendees to stream their videos at a time. This, however, may place those who share their video at a disadvantage compared to those who don't, given that their behavior is observed. This can quickly make sharing videos significantly less desirable. P7 came up with an interesting solution to this problem by turning the ability to stream the webcam into a reward or incentive awarded each time to 10 motivated students who want to see and be seen by the professor.

5. ONLINE STUDY: THE EFFECTS OF SEEING OTHER AUDIENCE MEMBERS ON ATTENDEES

Our interviews with presenters highlighted presenters' desire to see their audience (even at the expense of precious screen real estate) and their belief in in their ability to use webcam views to gauge audience engagement, and to ensure attention through accountability. Prior work suggests, however, that given the choice attendees themselves would prefer not to share their video. This could be due to desire for controlling self-presentation or privacy concerns [4], or simply due to it making multitasking more obvious and socially inappropriate [23, 24]. Indeed, while presenters want to see their audience, it is unclear whether audience members themselves benefit from seeing other attendees. In particular, could observing the video of others in the presentation affect an attendee's attitudes towards the material presented, and affect their attention and comprehension?

The study presented next aimed to answer the following specific questions:

- Would viewing other attendees' webcam streams affect comprehension?
- Would viewing other attendees lead to distraction?
- Would viewing other attendees increase interest in the presentation?

5.1 Method

To better understand whether viewing the webcam streams of other meeting participants affects meeting attendees, we designed a controlled between-subjects online experiment on Amazon Mechanical Turk. In the experiment, each participant viewed a segment of a pre-recorded presentation (webinar). The presentation was accompanied by videos of others who had previously watched the presentation. We manipulated the number of other attendees each participant saw. While relying on pre-recorded presentations and audience videos eliminated participants' ability to take active part in the presentations, it enabled us to reach a very high number of participants and to maintain experimental control.

After watching the segment of the presentation, participants answered a set of questions about the content of the presentation, and rated their overall interest in and attention paid to the presentation. They then rated their perceptions of other attendees' engagement in the meeting (we chose the term 'engagement' as it was commonly referred to by presenters in our interviews), whether they liked seeing other attendees, whether seeing other attendees was distracting, and finally expressed their (hypothetical) willingness to sharing their own webcam with the presenter and attendees.

5.2 Independent variables

5.2.1 Presentation content

The clips watched by participants were 5-minute segments extracted from 1-hour-long presentations publicly available on the internet. The two presentations were chosen to represent extreme points on several dimensions including topic, presenter engagement and audience participation. Specifically, the first was a presentation on The Ph.D. program at McGill University's Information Studies department[1] and included presentation slides and a view of the presenter. The second was a presentation on generational differences in the workplace[2] and included slides, a view of the presenter, and recorded chat between presentation attendees. Additionally, the presenter in this latter video referred explicitly to content discussed in the chat. We anticipated different levels of participant interest in the two presentations, and wanted to explore the effect of webcam views in the context of these extremes.

5.2.2 Views of presentation attendees

To manipulate the availability of webcam views of presentation attendees, we recruited 10 people (5 men, 5 women) and recorded them via their webcam as they individually watched the two presentation segments. We then merged each webcam view and the corresponding presentation videos, as illustrated in Figure 1. Specifically, we created the following instances for each segment: all 10 webcam views showing (*Full webcams*, see Figure 1a, 1d), 5 webcam views showing and 5 static avatars (*Mixed webcams*, see Figure 1b, 1e), and 0 webcam views and 10 avatars (*None*, see Figure 1c, 1f). The Mixed condition simulates a common state described by our interviewees where some, but not all of attendees share their video. To control for effects of our pre-recorded attendees, we created two versions for the Mixed condition, each using a different set of five pre-recorded attendees.

To summarize, the full experimental design was as follows:

× 2 Presentation (*McGill* vs. *Generational*)
× 4 Webcam views
(*Full webcams, Mixed(a), Mixed(b), None*)

= 8 conditions in total

5.3 Dependent variables

Before watching the video, the topic of the presentation was described to participants and they were asked whether they expected the content of the presentation would be interesting (with a simple Yes/No reply). After watching the video, the survey prompted participants to give open-ended responses describing the content of the presentation and to answer eight true/false questions about the content of the presentation. We combined the eight comprehension questions into a single *Comprehension* score (0-8), and used the open-ended responses to confirm that participants had paid attention to the video clip.

Participants then rated their agreement with a set of statements on a five-point Likert scale ranging from 1 ("Completely Disagree") to 5 ("Completely Agree").

1. *[Interest]:* The presentation was interesting to me.
2. *[Attention]:* It was easy to pay attention to the presentation.
3. *[Audience Engagement]:* The other people were engaged with the presentation.
4. *[Confidence]:* How confident are you in your response to the previous question?
5. *[Presenter]:* The presenter was able to keep the audience interested in the material.
6. *[Liked Webcams]:* I liked seeing the other viewers' streaming webcams
7. *[Distraction]:* The other viewers' streaming webcams were distracting.

[1] https://connect.mcgill.ca/p11ko93j2om/

[2] http://meetingone.adobeconnect.com/p7ox0wjt78w/

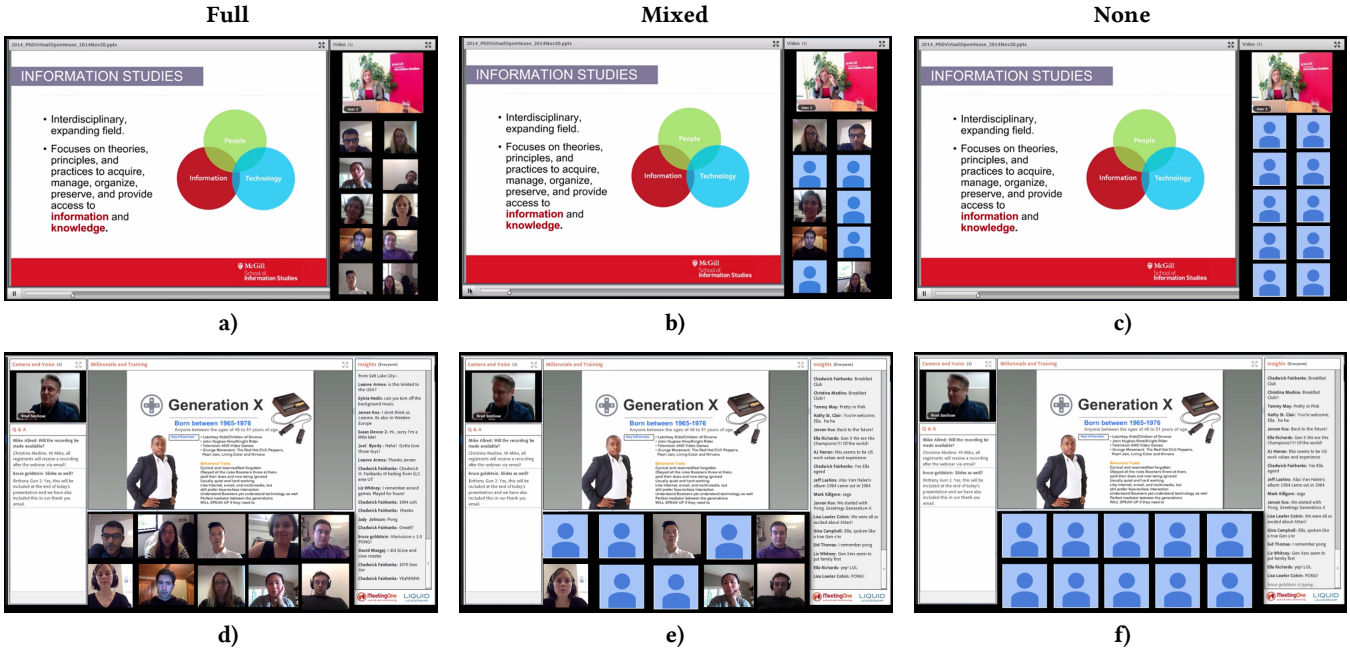

Figure 1. Screenshots from the presentation on McGill University's Information Studies department (top) and on generational differences (bottom). Showing webcam conditions: *Full* (left), one of the two *Mixed* (center), and *None* (right).

8. *[Sharing Own Webcam]:* I would have been OK sharing my webcam with the presenter during the presentation.

(Note that questions 6 and 7 were shown only to participants in the Full and Mixed webcam conditions.)

Finally, we collected demographic information, including participants' age (bucketed), gender, and prior experience with watching web-based presentations such as webinars.

5.4 Procedure

The experiment was set up as an online experiment using Amazon Mechanical Turk. Participation in the task was limited to individuals in the United States. Upon accepting the task, participants were randomly assigned to one of the eight conditions described above. Participants indicated their interest in the topic of the presentation and then watched the clip for a total of five minutes. Video controls were disabled, and a timer was used on the page to prevent participants from advancing past the video before the video was finished. Participants then answered the open-ended questions, rated the behavior they saw in the video, and finally provided demographic information.

We recorded the time participants took to interact with each page of questions in order to filter participants who quickly answered all the questions or neglected the task for very long durations. We paid participants $1.20 for the overall task, which took an estimated 10 minutes to complete.

5.5 Participants

238 participants completed the task. We excluded 4 participants whose open ended responses indicated they did not at all attend to the content of the presentation. We also excluded 14 participants whose completion time was too short/long based on a Mahalanobis Outlier analysis. Thus, our full dataset for analysis included 220 participants (112 in the Generations video condition, 108 in the McGill condition). 40% of participants were women. Age was reported in bands with 45% between the ages of 25-34 and 28% between the ages of 35-44. 70% of participants indicated that they had previously watched a webinar.

5.6 Limitations

The main limitation of this study, mentioned above, is that in order to reach a large number of participants, presentations were viewed asynchronously and participants could not take active part in the presentation. As a result, participants did not have to share video from their own webcam, only viewing the videos of others. Indeed, seeing one's own video can, in itself, be distracting. Another smaller limitation is that participants did not choose the topic of presentation to attend. However, our setup nonetheless replicates the case of students or employees who may have missed a required class or meeting and need to view a recording of it to catch up on the content that was covered.

6. RESULTS

Before commencing the full analysis, we compared the data from the Mixed webcams (a) and Mixed webcams (b) conditions for each presentation to identify any differences that may be caused by the subset of attendees shown. Since no significant differences were present between Mixed (a) and Mixed (b), we combined the data from these two conditions for each clip into a single "*Mixed*" condition. The data for the following analysis thus includes two videos and three levels for webcam views (Full, Mixed, and No webcams).

Before watching the videos, more participants expressed interest in watching the Generations video than the McGill video (76% vs. 64%; $X^2(1)=3.8$, p=.05).

The data were analyzed using a repeated-measures ANOVA. The model included video type (McGill vs. Generations), number of webcam streams (Full, Mixed, and None), expected interest in watching the video (Yes vs. No), gender (Female vs. Male), all two-way interactions, and the 3-way interaction (Video, Webcam Views, Gender). Likert-scale ratings were align-rank transformed [36].

Overall, women rated the videos as slightly more interesting than men did (M=3.2 vs. M=2.8; F[1,219]=5.8, p=.017) and scored slightly better on the comprehension questions (M=6.9 vs. M=6.5; F[1,219]=6.6, p=0.01). As we anticipated, participants found the Generations video more interesting than the McGill video (M=3.5 vs M=2.4; F[1,219]=60.2, p<.001) – this serves as a successful manipulation check for our choice of video clips. Participants also rated the Generations presenter's ability to keep the audience interested significantly higher (M=3.8 vs. 2.9; F[1,219]=28.3, p<.001). Yet participants' comprehension score was higher for the McGill video than the Generations video (M=7.3 vs. M=6.1; F[1,219]=98.9, p<.001). We suspect that this resulted from the content of the McGill presentation being described completely in the presentation slides.

6.1 Webcam views

Participants in the Full and Mixed webcam conditions rated how much they liked seeing the views of other attendees and how distracting seeing other attendees was. 48% of participants in the Full and Mixed webcam conditions indicated that they liked seeing the webcam views of other attendees; versus 19% who did not, (the remaining provided a neutral rating). 37% indicated that seeing other attendees was distracting; versus 58% who did not. Distraction was also one of the common issues raised in participants' open-ended responses. For example:

"It was difficult to stay focused after reviewing the slides and I found myself wandering looking at the other attendees, etc." (s148).

However, the number of webcam views (Full or Mixed) did not show an effect on ratings of distractions or liking.

To explore whether being distracted by webcam views negatively impacted participants, we tested the effect of reported distraction on comprehension score (by adding distraction ratings as a factor into our model). The test shows an inverse correlation between ratings of webcam views as distracting with comprehension scores (F[1,219]=4.4, p=.038); participants who performed worse on the comprehension questions found the webcam streams more distracting. Open-ended responses further illuminated this finding:

"I felt like being able to see the other people watching was also distracting. I started watching them and what they were doing and a few times it pulled me away from the thread of the talk" (s24).

"I was able to pay attention to the presentation, but I was also focused on what the other individuals were doing. Being able to see everyone is kind of distracting to the overall presentation and makes me feel self-conscious knowing that other people can see me" (s42).

6.2 Interpreting others' engagement from video

Adding participants' own ratings of how interesting they found the presentation reveals a surprising effect, whereby the more a participant finds the presentation interesting, the more likely they were to rate other attendees as engaged in the presentation (F[1,219]=137.2, p<.001). For example, one participant in the McGill 5 Views condition wrote,

"I found the topic and speaker very interesting." and *"The other people looked interested as well."* (s139).

Another, in the same condition wrote, *"The speaker was really boring. The information was not too interesting but the speakers tone and way of talking made it much worse. Plus the people watching looked totally bored so that did not help either"* (s106).

Relationship between variables			p-value
Rating of others' webcams as distracting	- ⟷	Comprehension	.038
Interest in presentation	+ ⟷	Rating others as engaged in the presentation	.001
Interest in presentation	+ ⟷	Willing to share own webcam	.001

Table 2. Summary of significant relationships

These examples illustrate a potential challenge for using the video of other attendees as a mechanism for increasing students' engagement in a presentation. However, it is important to remember that the direction of the effect could also be the reverse – that perceptions of others' engagement influence one's own ratings of interest. While we controlled for the audience seen in each condition, the experimental design does not establish the causality of this phenomenon.

6.3 The value of backchannel chat

Recall that in the Generations video, along with slides and view of the presenter, we also showed a chat ongoing between the original presentation attendees (See right side of Figures 1d, 1e, 1f). The active chat served as a strong cue of audience engagement in the presentation to the study participants.

This was highlighted in participants' open-ended responses (60 of the 112 Generations video viewers referred to the chat bar in their comments). While a handful (4/60) of these participants described the active chat as distracting (e.g. *"all the text chatter on the right was a little distracting*" (s9), the majority of comments (47/60) referencing the chat indicated that it was seen as valuable.

6.4 Willingness to share own video

Turning to participants' willingness to share their webcam video with the presenter and other attendees (asked hypothetically at the end of the study), we observe a couple of noteworthy effects. While men were slightly more willing than women to share their webcam video (M=2.9 vs. M=2.5; F[1,219]=6.4; p=.01), overall, participants were less likely to want to share their webcams if they expected the presentation to be uninteresting (M=2.3 vs. M=2.9; F[1,219]=12.8; p<.001). This could be because participants do not want to be seen when bored or because they expect to want to multitask [4, 24].

7. DISCUSSION

In the interviews, we saw that presenters didn't always have access to individuals' webcams to assess what they were doing. However, they perceived the value of a streaming webcam to be high, and as such often required participants to share their camera or used it as an incentive for hard workers. At the same time, webcams still didn't give a full picture of what people were doing, and were most useful when visual cues (such as a person seen to be typing) were augmented with evidence of that typing appearing in the relevant presentation's chat window.

In the online study, we found that webcam information from other audience members was not interpreted in a uniform way, with viewers' own attitudes towards the content shaping what they thought others were doing. There was also some indication that the constant motion coming from webcams (i.e. because it is impossible to sit completely still) could in fact be distracting and

hinder comprehension. These results raise an important question of the extent to which presenters' interpretations of audience engagement are similarly biased based on their own preconceptions.

Our use of Mechanical Turk for the second study clearly has its advantages and disadvantages. On the one hand, it allowed us to capture the attitudes of a large number of participants in a range of conditions that would be most difficult to do in another method. On the other hand, it precluded exploring live interaction between student, instructor, and content. Nonetheless, we believe our investigation serves an important step in the effort to understand the effect of audience visibility on online learners.

Based on findings from both studies, one design implication would be to explore ways of visually altering the user's video feed or avatar to (1) correspond with chat activity and/or (2) reduce the amount of fidgeting or motion that others are exposed to.

While webcams are fairly commonplace and easy for participants to share and view, our work suggests they are imperfect information sources. Video conferencing tools for large group presentations might consider providing ambient awareness of different cues about the audience both to presenters and to the other viewers (perhaps in different ways.) Here we discuss several possible dimensions of future audience-awareness visualizations that could be considered in the design of future platforms to support video-mediated presentations.

7.1 Design considerations for augmenting video-communication platforms with engagement cues

7.1.1 Timing of presentation of awareness information

Awareness information can be provided in real-time (such as a streaming webcam) or it can be presented afterwards, as a form of summary. The cognitive load involved in trying to monitor engagement levels while simultaneously delivering a presentation may be too high for presenters, unless such information is greatly abstracted and presented in a simple, ambient visualization (for example, as a fluctuating meter or using color codes to indicate different levels of attention).

7.1.2 Individual vs. group level

An additional question is whether engagement information should be provided about each individual or about the group as a whole. The relative utility of group vs. individual information may vary depending on the type of presentation or meeting. If it is a one-time presentation, like a webinar, the presenter may only care about aggregate information to help inform how they deliver the same content next time. However, instructors may be interested in following individual student engagement over time, in order to identify people who might need additional support. Alternatively, in accordance with our findings, it may be most beneficial to have a different UI layout for audience members where they see only their own webcam stream, along with the stream of the presenter (to minimize distraction caused by others). However, the UI could also dynamically display the webcams of other audience members when they speak, for example, to ask a question. How to best allow flexibility and customization in the collection and reporting of information for users in different roles is an important design consideration for future work.

7.1.3 Nature of information being visualized

Currently, most tools use basic and high-level inputs to provide awareness about viewers' activity (if they do so at all). In addition to streaming webcam, tools like the Zoom video conferencing platform have recently introduced an "attention tracking" feature which simply *"Lets the host see an indicator in the participant panel if a meeting/webinar attendee does not have Zoom in focus during screen sharing."* However, there are multiple other ways of detecting and presenting audience engagement, which could include using either the webcam or knowledge about active apps and computer activity to sense and provide details about:

- whether each participant is doing something related or unrelated to the presentation
- what apps or sites participants are using [note taking, Facebook, etc.]
- participants' emotional state [happy, bored, etc.]
- where participants are focusing on their screen [eye gaze behavior]

However, collecting more detailed types of sensed activity such as these have the possibility to create privacy concerns and may give viewers a sense of being "monitored" by the presenter. Some of these privacy concerns could be alleviated only displaying sensed information at a very high and anonymized level. An overall "engagement metric" could take into account several different cues and average them out into a vague, aggregate score that still obfuscates the details an individual's behavior. In the future, we plan to explore the effects of visualizing different types of engagement activity in different ways on presenters' and viewers' reactions.

Finally, we note that our experiment of audience's attitudes and comprehension looked only at two relative extreme points on the spectrum of presentation content, presenter engagement, and side-channel communication. Teasing out the effects of each of these dimensions is left for future work.

8. CONCLUSION

Online presentations will continue to be important ways of learning and disseminating information as distributed education and work continue to grow. In this work, we saw how commonly used functionalities such as webcams and chat still only provide an incomplete and imperfect assessment of others' engagement, both from the presenter and attendees' perspectives. Our results suggest a tension between what is seen as beneficial by presenters and what can be detrimental to students. Future work can explore the value and effects of presenting (possibly different) information visualizations to both parties to improve the experience for both sides.

REFERENCES

[1] https://bluejeans.com/press-releases/recording-large-meetings (Retrieved January 18, 2016).
[2] Nigel Bosch, Sidney D'Mello, Ryan Baker, Jaclyn Ocumpaugh, Valerie Shute, Matthew Ventura, Lubin Wang, and Weinan Zhao. 2015. Automatic detection of learning-centered affective states in the wild. *Proc. IUI*, 379–388.
[3] Jed R. Brubaker, Gina Venolia, and John C. Tang. 2012. Focusing on shared experiences: moving beyond the camera in video communication. *Proc. DIS*, 96–105.
[4] Michael Boyle, Christopher Edwards, and Saul Greenberg. 2000. The effects of filtered video on awareness and privacy. *Proc. CSCW*, 1–10.
[5] Anna C. Cavender, Jeffrey P. Bigham, and Richard E. Ladner. 2009. ClassInFocus: enabling improved visual attention strategies for deaf and hard

of hearing students. *Proc. SIGACCESS*, 67–74.
[6] Herbert H. Clark and Susan E. Brennan. 1991. Grounding in communication. *Perspectives on socially shared cognition* 13, 1991: 127–149.
[7] Derrick Coetzee, Armando Fox, Marti A. Hearst, and Bjoern Hartmann. 2014. Chatrooms in MOOCs: all talk and no action. *Proc. L@S*, 127–136.
[8] Ionut Damian, Chiew Seng Sean Tan, Tobias Baur, Johannes Schöning, Kris Luyten, and Elisabeth André. 2015. Augmenting social interactions: Realtime behavioural feedback using social signal processing techniques. In *Proc. CHI*, 565–574.
[9] Kobe Desender, Sarah Beurms, and Eva Van den Bussche. 2015. Is mental effort exertion contagious? Psychonomic bulletin & review: 1–8.
[10] Brian Dorn, Larissa B. Schroeder, and Adam Stankiewicz. 2015. Piloting TrACE: Exploring Spatiotemporal Anchored Collaboration in Asynchronous Learning. *Proc. CSCW*, 393–403.
[11] Paul Dourish and Sara Bly. 1992. Portholes: Supporting awareness in a distributed work group. *Proc. CHI*, 541–547.
[12] Elena L. Glassman, Juho Kim, Andrés Monroy-Hernández, and Meredith Ringel Morris. 2015. Mudslide: A Spatially Anchored Census of Student Confusion for Online Lecture Videos. *Proc. CHI*, 1555–1564.
[13] Mohammed (Ehsan) Hoque, Matthieu Courgeon, Jean-Claude Martin, Bilge Mutlu, and Rosalind W. Picard. 2013. MACH: My Automated Conversation Coach. *Proc. Ubicomp*, 697–706.
[14] Yen-Chia Hsu, Tay-Sheng Jeng, Yang-Ting Shen, and Po-Chun Chen. 2012. SynTag: A Web-based Platform for Labeling Real-time Video. *Proc. CSCW*, 715–718.
[15] Pradthana Jarusriboonchai, Thomas Olsson, and Kaisa Väänänen-Vainio-Mattila. 2015. Social Displays on Mobile Devices: Increasing Collocated People's Awareness of the User's Activities. *Proc. MobileHCI*, 254–263.
[16] Pooya Khorrami, Vuong Le, John C. Hart, and Thomas S. Huang. 2014. A system for monitoring the engagement of remote online students using eye gaze estimation. *Proc. ICMEW*, 1–6.
[17] René F. Kizilcec, Kathryn Papadopoulos, and Lalida Sritanyaratana. 2014. Showing face in video instruction: effects on information retention, visual attention, and affect. *Proc. CHI*, 2095–2102.
[18] René F. Kizilcec, Jeremy N. Bailenson, and Charles J. Gomez. 2015. The instructor's face in video instruction: Evidence from two large-scale field studies. *Journal of Educational Psychology* 107, 3: 724.
[19] Lisa Kleinman. 2010. Physically present, mentally absent? Technology multitasking in organizational meetings.http://repositories.lib.utexas.edu/handle/2152/ETD-UT-2010-05-800.
[20] Chinmay Kulkarni, Julia Cambre, Yasmine Kotturi, Michael S. Bernstein, and Scott R. Klemmer. 2015. Talkabout: Making distance matter with small groups in massive classes. *In Proc. CSCW*, 1116–1128.
[21] Yi-Chieh Lee, Wen-Chieh Lin, Fu-Yin Cherng, Hao-Chuan Wang, Ching-Ying Sung, and Jung-Tai King. 2015. Using time-anchored peer comments to enhance social interaction in online educational videos. In *Proc. CHI*, 689–698.
[22] Gilly Leshed, Diego Perez, Jeffrey T. Hancock, et al. 2009. Visualizing real-time language-based feedback on teamwork behavior in computer-mediated groups. *Proc. CHI*, 537–546.
[23] Kelly Lyons, Henry Kim, and Saggi Nevo. 2010. Paying attention in meetings: Multitasking in virtual worlds. *First Symposium on the Personal Web, Co-located with CASCON*, 7.
[24] Jennifer Marlow, Eveline van Everdingen, and Daniel Avrahami. 2016. Taking Notes or Playing Games?: Understanding Multitasking in Video
[25] Communication. *Proc. CSCW*, 1726–1737.
Cuong Nguyen and Feng Liu. 2016. Gaze-based Notetaking for Learning from Lecture Videos. *Proc. CHI*, 2093-2097.
[26] Catharine Oertel and Giampiero Salvi. 2013. A gaze-based method for relating group involvement to individual engagement in multimodal multiparty dialogue. *Proc. ICMI*, 99–106.
[27] Malin Pongolini, Johan Lundin, and Lars Svensson. 2011. Global online meetings in virtual teams: from media choice to interaction negotiation. In *Proc. C&T*, 108–117.
[28] Faria Sana, Tina Weston, and Nicholas J. Cepeda. 2013. Laptop multitasking hinders classroom learning for both users and nearby peers. *Computers & Education* 62: 24–31.
[29] Bertrand Schneider and Roy Pea. 2013. Real-time mutual gaze perception enhances collaborative learning and collaboration quality. *Intl. Journal of Computer-Supported Collaborative Learning* 8, 4: 375–397.
[30] Kshitij Sharma, Hamed S. Alavi, Patrick Jermann, and Pierre Dillenbourg. 2016. A Gaze-based Learning Analytics Model: In-video Visual Feedback to Improve Learner's Attention in MOOCs. *Proc. LAKS*, 417–421.
[31] Garriy Shteynberg. 2015. Shared attention. *Perspectives on Psychological Science* 10, 5: 579–590.
[32] Daniel Szafir and Bilge Mutlu. 2013. ARTFul: adaptive review technology for flipped learning. *Proc. CHI*, 1001–1010.
[33] Yla R. Tausczik and James W. Pennebaker. 2013. Improving Teamwork Using Real-Time Language Feedback. *Proc. CHI*, 459-468.
[34] Jaime Teevan, Daniel Liebling, Ann Paradiso, Carlos Garcia Jurado Suarez, Curtis von Veh, and Darren Gehring. 2012. Displaying mobile feedback during a presentation. *Proc. MobileHCI*, 379–382.
[35] Xiang Xiao and Jingtao Wang. 2015. Towards Attentive, Bi-directional MOOC Learning on Mobile Devices. *Proc. ICMI*, 163–170.
[36] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. *Proc. CHI*, 143–146.