Multimedia Showdown: A Comparative Analysis of Audio, Video and Avatar-Based Communication

Joseph Tu∗
joseph.tu@uwaterloo.ca
Stratford School of Interaction Design and Business,
University of Waterloo
Canada

Arielle Grinberg
a2grinberg@uwaterloo.ca
Department of Management Science and Engineering,
University of Waterloo
Canada

Mark Hancock
mark.hancock@uwaterloo.ca
Department of Management Science and Engineering,
University of Waterloo
Canada

Lennart E. Nacke
lennart.nacke@acm.org
Stratford School of Interaction Design and Business,
University of Waterloo
Canada

Figure 1: The virtual space for this study was a mocked-up place of the American television sitcom Friends.

ABSTRACT
Our new work culture relies heavily on online meetings and computer mediated communication (CMC). However, making an online meeting engaging while keeping communication productive is a major challenge. We collected quantitative data from the user engagement scale (UES) and qualitative data from semi-structured interviews to investigate how user engagement differed. Using the gamified web-conferencing platform Gather, we compared four communication channels: (1) audio-only, (2) audio and video (no avatar), (3) audio and avatar (no video), and (4) audio and video and avatar. We began qualitative data analysis using reflexive thematic analysis. Although the UES results did not reveal significant differences, the preliminary results from the thematic analysis such as people prefer communication platforms designed for specific use cases because video makes them feel more self-conscious, while avatars make them feel more represented. Lastly, we provide a work-in-progress applied definition of user engagement in communication channels with their perspective on individual engagement constructs.

CCS CONCEPTS
• Human-centered computing → Empirical studies in collaborative and social computing.

KEYWORDS
engagement, communication channels, communication tools, virtual environment

ACM Reference Format:

1 INTRODUCTION
A productive workplace is based on effective communication. We now use online tools to communicate more than we used to, for education [9, 46, 51], online conferences and workshops [6, 33, 38], and remote work [5, 15, 58] because of the COVID-19 pandemic. However, online communication can feel constrained. Bukker et al.

∗Author is affiliated with the Department of Systems Design Engineering, University of Waterloo
found that engagement affects online work. Most online communication tools limit interactions (and, with it, engagement). People may be more interested in interacting online if these technologies incorporate game elements. In both Computer-Mediated Communication (CMC) and Human-Computer Interaction (HCI), understanding user engagement in virtual environments is essential [31, 62]. For example, Baker et al. [2] experimented with avatar-mediated communication to help shy, introverted users cope with social isolation. Ratam et al. [53] have explored how avatar and social presence can affect one’s decision via text-based communication, also known as the Proteus effect [61]. In addition, Oh et al. [48]’s systematic review explains how and when it improves virtual user experience (UX). However, we need more information about how engagement affects UX across different communication channels, particularly in a gamified communication tool such as Gather (mainly used for education, online conferences, workshops, and remote work).

In this paper, we look at how user engagement differs across various communication channels as new types of interactions and web tools become more widespread: verbal communication (Audio), non-verbal communication (Video), and a graphical image representation (Avatars) in Gather because user engagement is critical to creating effective online communication [34].

We examine engagement in online meetings in a specific virtual environment as an intersection of gaming and web conferencing through these research questions:

- **RQ1** How does user engagement differ in each communication channel (Channel A: Audio, Channel B: Audio + Video, Channel C: Audio + Video + Avatars, Channel D: Audio + Avatars)?
- **RQ2** How do Audio, Video, and Avatar communication channels compare to each other regarding benefits and trade-offs?

To answer these research questions, we used a mixed-methods approach. We administered the User Engagement Scale (UES) to assess standard engagement constructs across each communication channel. A semi-structured interview is used to dive more in-depth on specific engagement constructs, which is accessed by a hybrid thematic analysis (TA). We also collect participant demographics. In summary, we highlight how combinations of communication channels can affect one’s engagement, whether it increases, decreases or remains neutral (no change). Based on O’Brien et al. [50]’s definition of the User Engagement Scale constructs, we use communication channel elements to assess engagement and collaboration. We plan to identify three collaborator roles (facilitators, observers, space architects) that affect engagement. Combinations of communication channels play a crucial role in promoting engagement and collaboration, particularly in how the communication tool is used [31, 62]. For example, using the three combinations of communication channels (Audio, Video, and Avatar) can create a sense of presence, which is particularly useful for events like online Hackathons or the Global Game Jam [20, 40]. The purpose of categorizing roles is to determine the effectiveness of collaboration in professional settings. For example, companies may want to investigate how co-located events affect engagement and collaboration.

### 2 RELATED WORK AND GATHER

The shift in technology has made online communication tools indispensable. Research on online communication tools has become increasingly relevant in Human-Computer Interaction (HCI). For example, Bos et al. [8] studied the effects of four communication channels (face-to-face, video, audio, and text chat) and how effective video and audio conferencing groups were. Virtual teams require engagement, especially when collaborating remotely [4, 43]. For avatar communication, different communication modalities may involve different levels of information richness according to media richness theory [17]. In the last century, we have moved from text-based computer-mediated communication to multi-modal communication (e.g., including audio and video). Researchers have mostly studied how computer-mediated communication may affect interpersonal relationships because of the lack of nonverbal cues like facial expressions, body language, and gestures [56, 57] or multi-modal communications, as well as how it may affect online communication, which is affected by gender [29]. Although interactive aspects of online communication tools can aid in increasing online learning [22], Cao et al. [12] argue that we need to know whether these findings will hold across various communication channels. For example, Fish et al. [23] found that video calls are a good way to talk to people remotely.

Game elements are building blocks that are characteristic of games and, respectively, gamified systems [24] which are essential for engagement [1, 11, 39]. However, it is important to determine what elements could affect engagement [41]. For example, Bonsignore et al. [7] have looked at platforms for collaborative learning by examining how Alternate Reality Games (ARG) facilitate engagement through narrative and game mechanics. Research has shown how game elements like badges can act as a powerful motivator in educational contexts [18], with evidence demonstrating that points, levels, and leaderboards are effective means to increase short-term performance and engagement [42, 45]. Latulipe et al. [37] have discussed how gamification elements increase user engagement despite challenges presented to the user. On the contrary, adding game elements to a system without a specific purpose has raised criticism in gamified applications used for education [14, 32, 60]. Chee and Wong [14] state: “The mere inclusion of meaningless points, badges, and bright colours, which serve as the catalysts to engagement without full comprehension of their purpose or reason of attainment, fails to make a gaming experience fun and engaging.” In short, incorporating gamification (such as avatars) in an online virtual environment is a crucial consideration owing to its potential to elevate participants’ involvement and motivation, culminating in enhanced contentment and efficacy of the meeting.

Avatars are not exclusive to video games, however, they are a form of mimicry play and embodiment [25, 30, 54]. The importance of an avatar is not about the playable characters but the focus on how players engage and act as agents in a fictional world [36]. In other words, the avatar becomes an extension of the player’s body via the interface of a screen, speakers, and controllers [52]. A literature gap exists on how avatars affect engagement in an online communication tool (as part of a virtual world). Nowak and Fox [47] notes the importance of considering the inferences and attributions for avatars as long as the researchers provide clear conceptual
definition, or “it difficult to generalize or fully understand the effects of avatars on communication processes.” Our study focuses on 2D game elements, which Gather provides on their platform. The virtual environment can attribute certain mental states (e.g., beliefs, values, goals, feelings, attitudes) to the virtual character [26]. Social interaction is a form of embodied thinking in the concept of modern video games [27]. However, more information about how interaction changes in an online communication tool like Gather is needed. The ability to build specific game-like representation spaces could change how the users interact with one another. For example, proximity chat could create new engagement opportunities in online interactive art conferences [38]. As stated by Gee [27], the virtual environment of Gather could cause “us to act in the real world in ways that change it to resemble or model simulation better.”

Figure 2: Various conditions in Gather

2.1 Platform: Gather

In this paper, we consider Gather to be a gamified web-conferencing platform: a virtual environment used for communication with two or more people within the virtual space. Gather is chosen as the main platform of the study because we can isolate each communication channel (audio, video, and avatar). We created an empty virtual space for Channel A: Audio and Channel B: Audio + Video conditions (similar to a Zoom meeting). Gather has four particular features within its virtual space: (1) the spotlighting feature allows the users to broadcast throughout the virtual world; (2) proximity chat is a feature in which a user can only hear other users within an allocated space; the audio gets louder if users are within the vicinity of one another; (3) private spaces act similarly to a breakout room, which allows users to be in a small meeting while isolating communication outside of that space; (4) Gather allows a space’s owner to create interactive elements.

3 METHODOLOGY

User engagement is often defined by the depth of a user’s cognitive, temporal, affective and behavioural investment when interacting with a digital system [49]. To measure user engagement in Gather, we decided to use the User Engagement Scale (UES), a 31-item experiential questionnaire, because “the UES (or items derived from it) has been used to evaluate engagement in a range of settings” such as social networking systems and video games within the HCI domain [59]. A semi-structured interview was then conducted because it provided a flexible structure allowing us to collect new exploratory data relevant to our research topic. Open-ended questions were asked to explore the participant’s thoughts, presence, engagement, and concerns about online communication tools [16].

Before the interview, participants were also asked to complete a demographic survey (see Open Science Framework)2 that covered their age, gender, ethnicity, education, as well as their online communication habits (before COVID-19 and during COVID-19). This study was approved by the University of Waterloo Research Ethics Board (REB no. 43455). Following the ethical research guidelines, the consent ensured that participants were free to end the interview at any given time or not to answer specific questions (neither occurred during the interview). At the end of the study, participants were given a thank-you letter and were informed to contact the researchers if there was any sensitive information that needed to be retracted before publication.

A general power analysis program (GPOWER)3 was used to determine the appropriate sample size for our study [19]. We assumed an effect size of $f \approx 0.43, \alpha$, based on previous literature reported from other studies using the UES scale [50], error probability $= 0.05$, and a power of a statistic test to be at least 0.8. Based on our power analysis assessment, 64 participants between-subjects were required (16 participants recruited per four conditions), non-centrality parameter $(\lambda) \approx 11.83$, critical $F \approx 2.76$, numerator $df = 3$, and denominator $df = 60$, and power $= 0.81$. Chandler [13] recommends activities within a breakout room should be designed to accommodate small groups of 3–5 participants, focusing on acclimating them to the online platform and providing clear directives for the session. Therefore, each session had a range of 3–5 participants. Our participants were from different nationalities, including Argentina, Australia, Canada, France, Lithuania, South Africa, the United Kingdom, the United States, and Venezuela.

There were two tasks to answer RQ1 and RQ2: (1) Charades: This task remained the same for all channels. For example, in Channel A, participants were only allowed to use sounds (without visual aid) to guess. (2) Storytelling: Participants were asked to tell a short two-minute personal story. We allowed participants to make up stories on how they pleased. For example, if the theme of the story was “Horror,” participants in Channel A only used sound to enhance the storytelling experience, whereas, in Channel B, they could use facial gestures, conversely in Channel C and D, use their avatars such as mimic running away, or have others be involved in their story with a sense of space. As this is a work-in-progress, we had a third task to understand how collaboration works with avatars (in terms of video on and off). We do not report this as the main result (MR), but we do plan to build upon engagement constructs in section 5. Participants were asked to collaborate by designing a “Halloween Theme Space” together. After the participants completed the design, they were asked to share a story about the space they created collaboratively.

2OSF Link: https://osf.io/69vd5/
3GPOWER: https://stats.idre.ucla.edu/other/gpower/
3.1 Analysis

The interview data were analyzed using hybrid reflexive thematic analysis (RTA), which draws from reflexive and codebook TA types [10]. The statistical tests were conducted in RStudio4. We conducted a one-way ANOVA across the four conditions as this allowed us to see the differences between the means and distributions of each group [35]. The normality and homogeneity of variance were checked. Three or more researchers reviewed all interview transcripts. The audio recordings were transcribed in Dovetail, and were edited by at least three researchers to check for potential errors. This allowed the researchers (coders) to familiarize themselves with the data. Participant transcriptions were all under 30 minutes. At the time of this study, we consider all authors to have strong knowledge of game user research. One of the authors has worked directly with Gather’s API system. We had four researchers (coders) do line-by-line coding on the first 12.5% of the data set to create the first initial codebook (two interview transcripts from each condition). Differences in the tags created in the initial codebook stemmed from researchers having different opinions on the codes. Thus, in 16 sessions averaging around 1 hour, all researchers had to collaboratively discuss, mediate, decide, merge, and resolve all the conflicts that emerged throughout the line-by-line coding to finalize the first codebook. Consequently, the second codebook was derived from the next 12.5% of the data set (two interview transcripts from each condition), while creating a flat coding and hierarchical coding framework model [21]—relevant to engagement and researchers’ respective communication modality. Researchers had to dispute all conflicts similar to the first initial codebook to finalize the codebook. Any new relevant tags were merged when possible while keeping new entries open to interpretation under the flat coding and hierarchical coding framework model. Any codes irrelevant to the research question were labelled as “miscellaneous.” To capture the essence of each theme, affinity clustering was created on a Gather’s whiteboard. To generate these themes, we needed to ensure they fit thematically (i.e., no contradictions) and were not too broad. We decided on initial themes based on the frequency of recurring codes. After creating an initial theme, the researchers returned to each recording to check for missing important points. The themes were refined through multiple iterations; the final themes are presented in the findings below. Our theme development and reporting focused on data directly related to the research question.

3.2 Recruitment and Participants

We recruited participants using a combination of the User Interviews’ platform and social media posts by our research labs. In total, we recruited 67 participants between the ages of 18–61 (M = 35.56 years, SD = 11.14 years), and all participants were familiar with at least one or more online communication tools (e.g., Microsoft Teams, Zoom, WebEx). The questionnaire had a test question to ensure participants filled out the survey correctly: “Please select strongly agree.” Participants who did not select ‘strongly agree’ were excluded from the statistical analysis. In total, the following participant’s data were omitted for statistical analysis: Channel A (P16, P35, P38), Channel B (P1, P46), Channel C (P39), Channel D (P23), withdrawn (P52), and dropped (P64). We acknowledge that additional six participants were not recruited to replace the omitted data for the statistical analysis because it would be very difficult to maintain an even distribution across each group. We did consider bootstrapping. However, the displacement of six participants did not yield many changes and thus was not included in this paper. The omitted data is kept on the Open Science Framework2 for credibility purposes. Qualitative data from these participants were still analyzed because common themes were derived from the semi-structured interview data. All participants were awarded $10 USD gift card on the User Interviews platform. Participant’s data were saved in Dovetail2 for thematic coding; their ID preceded with a ‘P’ (e.g., P2) for reference purposes.

4 RESULTS

In this works-in-progress paper, we will only focus on the preliminary results (PR).

<table>
<thead>
<tr>
<th>Condition</th>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
<td>0.053</td>
<td>0.018</td>
<td>0.122</td>
<td>0.947</td>
</tr>
<tr>
<td>Residuals</td>
<td>53</td>
<td>7.663</td>
<td>0.145</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: No significant results across the conditions in terms of UES

4.0.1 One-Way ANOVA. We conducted a one-way ANOVA across the four conditions as this allowed us to see the differences between the means and distributions of each group [35]. The normality and homogeneity of variance were checked. Bartlett’s $K^2 = 0.5889$, df = 3, p = 0.8989 and from the Table 1, looking at the p-value we cannot reject the null hypothesis because there are no significant differences in global engagement between experimental conditions. From the ANOVA results, we do not report the Tukey Post Hoc in this paper as it is insignificant. Therefore, the rejection of the null hypothesis often leads to further investigation to determine the practical significance and implications of the observed difference. We continue to analyze the four subscales of engagement, however, the preliminary results (PR1) showed that none of the sub-scales showed significant differences between conditions. It is important to note that the rejection of the null hypothesis in ANOVA does not indicate which specific groups are different from one another, thus we focus on our thematic analysis findings to answer (RQ2) to understand the benefits and trade-offs of using one or more than one communication channel.

4.1 Thematic Analysis Preliminary Findings

4.1.1 (PR 2): Video Is Good For Engagement But Avatars Are Better For Presence. A majority of participants voiced a preference for enabling their video during communication, citing feelings of speaking into a vacuous space or being unsure of the existence of others. Such uncertainties of interlocutor presence can prompt a sense of loss, leading to lowered engagement. Participants expressed that the avatar fostered a sense of co-presence, instilling a feeling that another person was present alongside them. Interestingly, participants

---

4RStudio: https://www.rstudio.com
5User Interviews: https://www.userinterviews.com/
6Dovetail: https://dovetailapp.com/
who preferred not to use video feeds reported feeling self-conscious about their appearance. Proximity chat, privacy rooms, and interactive elements with the ability to move created a sense of space for participants. The ability to move freely within the virtual environment instilled a greater sense of connectedness to the virtual space, while still enabling them to observe ongoing interactions within the vicinity. Participants highlighted the potential for a spatial awareness that can emerge through the provision of dedicated workspaces [28]. For example, the experience of encountering other avatars and being unable to pass through them imbued a feeling of spatial presence within the virtual environment.

4.1.2 (PR 3): Video Makes People Feel More Self-Conscious While Avatars Make Them Feel More Like They Are Being Represented. Participants emphasized the significance of being prepared and attentive to self-consciousness. Concerns regarding personal presentation and environmental factors that may be captured on camera were noted to hinder conversation and engagement. While they acknowledged the potential for increased engagement, they also expressed discomfort arising from uncertainties surrounding their appearance and unfamiliarity with other participants because they stressed the idea of ‘first impression counts’. Participants in both Channel C and D (where avatars were present), reported experiencing a sense of self-representation, as they could visually perceive a virtual resemblance of themselves or how they wished to be represented, alleviated the idea of being less self-conscious as they were moved focused on the avatar (even with the video-on in Channel C). Participants reported that extended periods of having their video on could be taxing and draining, leading to discomfort and reduced engagement. During the semi-structured interviews, we observed that some Channel B and C participants would turn off their videos due to fatigue. However, some participants indicated they would leave their video-on to contribute valuable research contributions and considered it a form of remuneration.

4.1.3 (PR 4): Each Communication Channel Has A Specific Use Case. Collaboration occurs mainly in screen sharing or private rooms/spaces. Participants reported that screen sharing is a key aspect of their collaborative work, typically in the third task. Screen sharing facilitates a shared visual experience, enabling all participants to view the same content simultaneously [28, 55]. However, some participants suggested that the current screen-sharing capabilities could be improved by incorporating a zoom feature for participants to view the same content simultaneously [28, 55]. However, some participants suggested that the current screen-sharing capabilities could be improved by incorporating a zoom feature for participants to view the same content simultaneously [28, 55]. However, some participants suggested that the current screen-sharing capabilities could be improved by incorporating a zoom feature for participants to view the same content simultaneously.

5 DISCUSSIONS
As this is a work-in-progress, we cannot conclude that the communication channels differ in engagement, based on our preliminary results (PR1) from the statistical analysis. We plan to look deeper into the thematic analysis to identify what contributes to each individual construct of engagement (RQ1). Using game elements, particularly avatars, can positively impact engagement and collaboration during virtual interactions. Participants in channels where avatars were present reported feeling a sense of self-representation, which can enhance engagement by providing a feeling of co-presence and spatial awareness (RQ2).

In addition, avatars can reduce feelings of disconnection from the environment and enhance the ability to perceive other individuals in the vicinity, potentially leading to increased engagement. However, some participants in the study reported feeling a sense of reduced attention and engagement when using Channel B, despite being able to see other participants on their screens. In latent terms, looking directly at the screen (eye gaze) is not the same as making eye contact with someone online. From our preliminary results, participants reported that the presence of background noise could be a source of distraction during conversations, thus impeding their ability to comprehend and follow the discussion [44] effectively. Because of this, all people in a conversation must ensure that their environment is good for talking. Furthermore, participants expressed dissatisfaction regarding their limited ability to participate in group discussions, resulting in them muting their microphones during the conversations, leading to isolation and loneliness. As we continue analyzing our results in this work-in-progress, it is important to balance the benefits of self-representation through avatars and the potential negative effects of video fatigue on engagement levels.

5.1 Applied Definition of User Engagement in Communication Channels
We plan to observe the individual constructs and map them to O’Brien et al. [50]’s work in relation to thematic analysis results and observational data (currently still work-in-progress). We hypothesize the following in Table 2:

<table>
<thead>
<tr>
<th>Constructs</th>
<th>Engagement Definition Applied in Communication Channels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aesthetic Appeal (AE)</td>
<td>The users’ perception of the visual appearance of Gather: Audio: Minimal to some visual appeal of the user’s profile (lighting up, etc.) Video: Visual appeal of the video frames of participants on screen Avatar: Visual appeal of characters, objects, and the virtual environment</td>
</tr>
<tr>
<td>Endurability (EN)</td>
<td>Users’ overall evaluation of the experience, its perceived success and whether users would recommend the communication channel to others. All: Depends on the task or objective of the communication Avatar: No involvement, unless it’s audio task related Video: Involvement comes from gestural cues, e.g., smiles, hands, and etc. Avatar: Involvement mainly derived from interactions being embedded within the virtual space</td>
</tr>
<tr>
<td>Felt Involvement (FI)</td>
<td>Users’ feelings of being drawn in, interested, and having fun during the interaction. Audio: No involvement, unless it’s audio task related Video: Involvement comes from gestural cues, e.g., smiles, hands, and etc. Avatar: Involvement mainly derived from interactions being embedded within the virtual space</td>
</tr>
<tr>
<td>Focused Attention (FA)</td>
<td>Users’ concentration of mental activity in Gather: Audio: Listening or speaking Video: Being comfortable on video (self-conscious) Avatar: Virtual Environment, users moving around, design space, reactions, embedded tools (whiteboards, pianos, Tetris), and presence</td>
</tr>
<tr>
<td>Novelty (NO)</td>
<td>User’s level of interest in the task and curiosity evoked by the system and its contents Audio and Video: None Avatar: Being able to move (pace around), embedded interactions, customization</td>
</tr>
<tr>
<td>Potential Usefulness (PU)</td>
<td>All: Only effective if the communication channel supports the purpose of the meeting</td>
</tr>
</tbody>
</table>

Table 2: Applied constructs of UES in Gather
6 LIMITATIONS AND FUTURE WORK

There is a possibility that participants could have compared their experience with other online communication tools they have used before because this is a study designed between subjects, thus potentially adding a ceiling effect to the UES values from our preliminary results. In channels A and B, participants had access to a virtual space, which means that they could still see the researcher move around to spaces (even with the instructions to stay still). This could have biased their UES scale because they might have rated it with an avatar condition in mind (rather than just specifically audio-only or audio+video).

We acknowledge we did not disaggregate the participants by gender or geographical location, any of which could be useful and pertinent to the experience of their participants because it is extremely difficult to schedule groups of 3 or more participants at different timezone. Further analysis could examine how these factors could affect the experience of genders and social constructs. We disabled white noise (e.g., television static, fireplaces) within work possibility is studying how technology helps people get involved in different kinds of communication, especially when they are far away, to find ways that technology can be used to improve engagement and collaboration among participants. Lastly, future studies should focus on the idea of time and how collaboration stages vary between each communication modality. We plan to continue to investigate instances when a user changes from a facilitator to a listener, and how engagement varies within hybrid (co-located meetings) meetings.

ACKNOWLEDGMENTS

This work was made possible by the NSERC CREATE SwaGUR grant, Social Sciences and Humanities Research Council, Lennart Nacke’s NSERC Discovery Grant, the Canada Foundation for Innovation John R. Evans Leaders Fund “The seamless user-centered conscious experience study system (SUCCESS)”. We thank all the participants who participated in our study. Special thanks to Derrick Wang and Dr. Stuart Hallifax for assisting in the study and video work. Thanks to Dr. Neil Randall, Alexander Glover, and Ekaterina Durmanova for being radiant models in the video. Thanks to Charlie (doggo) for being an incredible source of emotional support throughout the writing process. We also thank Gather.Town for allowing us to access their API system.

REFERENCES

A Comparative Analysis of Audio, Video and Avatar-Based Communication

CHI PLAY ’23 Companion, October 10–13, 2023, Stratford, ON, Canada


