ARcall: Real-Time AR Communication using Smartphones and Smartglasses

HEMANT BHASKAR SURALE, University of Waterloo, Canada and Snap Inc., USA

YU JIANG THAM, Snap Inc., USA

BRIAN A. SMITH*, Snap Inc., USA and Columbia University, USA
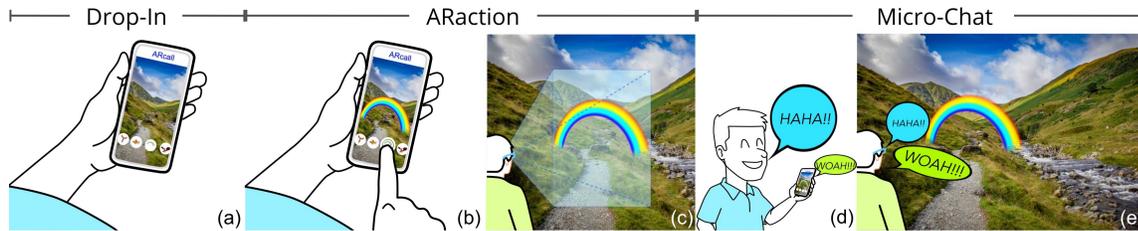
RAJAN VAISH*, Snap Inc., USA

Fig. 1. ARcall's main components, which occur in real-time, including Drop-In, ARaction and Micro-Chat: (a) Drop-In allows the Friend to see the Wearer's view on a smartphone; (b-c) ARaction allows the Friend interact with the Wearer by sending AR content; (b) The Friend selects an AR rainbow to send; (c) The Wearer sees the rainbow augmenting their reality; (d-e) Micro-Chat allows the Friend and the Wearer to talk to each other; (d) The Friend says "Hahaha!!" after sending the rainbow; (e) The Wearer reacts to the rainbow by saying, "Whoa!!" as the Friend remains dropped in.

Augmented Reality (AR) smartglasses are increasingly regarded as the next generation personal computing platform. However, there is a lack of understanding about how to design communication systems using them. We present ARcall, a novel Augmented Reality-based real-time communication system that enables an immersive, delightful, and privacy-preserving experience between a smartphone user and a smartglasses wearer. ARcall allows a remote friend (*Friend*) to send and project AR content to a smartglasses wearer (*Wearer*). The ARcall system was designed with the practical limits of existing AR glasses in mind, including shorter battery life and a reduced field of view. We conduct a qualitative evaluation of the three main components of ARcall: *Drop-In*, *ARaction*, and *Micro-Chat*. Our results provide novel insights for building future AR-based communication methods, including, the importance of context priming, user control over AR content placement, and the feeling of co-presence while conversing.

CCS Concepts: • **Human-centered computing** → **Ubiquitous and mobile computing systems and tools**.

Additional Key Words and Phrases: augmented reality, immersive communication, calling

*Co-Principal Investigators.

1 INTRODUCTION

As AR-based smartglasses are evolving, they are unlocking exciting new forms of immersive communication. Prior research has explored the utility of AR mainly as a means of annotating environments [22, 35, 43, 54], communicating with avatars [68], relaying remote instructions and guiding [30, 79], visualizing biosignals [69], and rendering visual alterations [72]. However, there is a lack of research in understanding how AR serves in the context of informal communication using modern AR smartglasses. [60]. Specifically, Pfeil et al. [66] highlighted the need for telepresence systems to consider both streamers' and viewers' experiences when communicating with each other via devices such as smartglasses and via device pairings such as smartglasses and smartphones. Given the recent surge in AR smartglasses in the consumer market [12, 24, 33, 61, 73], it is evident that everyday communication practices will extend to AR smartglasses [66]. Therefore, we explore the design of ARcall, an Augmented Reality (AR) based communication system, in the context of informal communication between a smartphone user and a smartglasses wearer.

We introduce ARcall, a native AR-based real-time communication system designed to support the one-to-one paradigm [66]. Our system consists of two software applications: a smartglasses application for the smartglass user (the "Wearer") and a smartphone application for the smart phone user (the "Friend"). ARcall enables sending and projecting AR content to the Wearer *natively* in AR format. For instance, a Friend can send AR content like making it snow or dropping a rainbow into the Wearer's environment (Figure 1 (b–c)). When designing ARcall, we considered the benefits of wearable and ready-to-project AR smartglasses, in addition to socio-technical constraints such as privacy, interruptions, and mixed-device configuration necessary for interactions between smartphones and smartglasses [1].

ARcall consists of the following main design components:

- *Drop-In*: Drop-In allows the Friend to see the Wearer's current context (i.e., where they are and what they are doing) by "dropping in" and viewing a video feed of the Wearer's point of view (see Figure 1 (a)). Drop-In is an invite-only feature in which the Friend is granted access to time-limited *Drop-In sessions* only when the Wearer goes online and invites them. Note that the Friend can decide to Drop-In at their convenience during the session.

- *ARaction*: ARaction enables interaction with the Wearer using AR. While dropping in on the Wearer from the smartphone, the Friend can browse, select, and send AR content to the Wearer (see Figure 1 (b–c)). The AR content is then displayed on the Wearer's glasses, augmenting their reality in real time.

- *Micro-Chat*: A Micro-Chat is a time-bounded voice call between the Wearer and Friend. Micro-Chats enable the Friend to add meaning to the AR content they send and experience the Wearer's reaction in real time (see Figure 1 (d–e)). While talking, the Friend can continue to send AR content and see the Wearer's point of view (with AR content they've just sent superimposed onto it) in real time. The Micro-Chat is capped at one minute and happens as part of the Drop-In session. It runs for as long as Drop-In session lasts, which preserves the battery life of the AR glasses. The Wearer can extend the Drop-In session in 30-second increments by tapping a button on the temple of the smartglasses during the session.

We conducted a qualitative evaluation of ARcall with 14 participants to probe overall experience, initial perceptions of the system, and feelings of connectedness and understand the limitations of native AR communication. We learned that both Friends and Wearers found ARcall to be fun and immersive and that ARaction added an element of surprise to their interactions. Friends felt that they were augmenting Wearers' reality, and Wearers felt that the AR content that their Friends sent was relevant and personalized to their environment. Wearers felt connected to their Friend by giving the Friend the ability to project AR content to their display. ARcall's Drop-In component helps Friends learn about the Wearer's context (helping the Friends select and send meaningful, personalized AR content), seeing the Wearer's point

of view also increased the users' feeling of togetherness. Finally, ARcall's Micro-Chat component created intimacy and added meaning to the interaction. As soon as Wearers heard their Friends talk, they felt they were sharing a moment together. Friends highly valued the ability to experience the Wearer's reaction in real-time.

The main contributions of this paper are as follows:

- Understanding user behaviour and the affordances created by our augmented-reality-based real-time communication method. Our design approach takes into account the practical limitations of modern smartglasses.
- Our qualitative evaluation of the key components of our proposed system provides novel insights for building future AR-based communication methods. For instance, our evaluation revealed the importance of context priming using Drop-In, engagement using ARaction, and creating meaningful conversations using Micro-Chat.

2 BACKGROUND AND RELATED WORK

We examined several threads of communication-related research from the perspective of designing communication systems for AR smartglasses: AR view annotation, pre-call context assessment, and time-bounded calling.

2.1 AR View Annotation

To make video calls more interactive, researchers have explored how AR-based annotations can assist users engaged in collaborative tasks [10, 23, 31, 62]. Jo et al. [36] investigated a smartphone-based system that allows a user to control a remote scene and augment it with drawings during a live video call. Similarly, Ryskeldiev et al. [74] introduced a system capable of sharing the smartglasses wearer's point of view (POV), including photospherical imagery and the wearer's current orientation. The system helped reduce the cognitive load of peers involved in a remote collaboration task. Further, Nassani et al. [55] found that anchoring text on live video in a spatialized way makes them more effective than displaying static text. Past research focused on creating shared experiences for collaborative tasks using either smartphones or desktop computers [44, 47, 52], but it is unclear how these results would translate to smartglasses-based communication. Users exhibit different behaviors when video calling on smartglasses than when they use smartphones. For instance, Kun et al. [46] show that smartglasses wearers do not spend time looking at callers' video feeds. Similarly, in virtual reality environments, He et al. [29] show that the communication typically goes beyond conducting face-to-face communication or collaborative tasks to projecting shared artifacts with remote audiences. In summary, most of the past research focuses on *collaboration*; the extent to which these results can translate to native AR *communication* is unknown.

Furthermore, current AR-based communication practices are limited in four ways [13, 26, 77]. First, the sender can send AR content, often in the form of a pre-recorded video with superimposed AR content, without seeing the Friend's current context, so the AR may not be relevant to the Friends' surroundings. Second, these experiences are less mobile-friendly since users have to hold their smartphones facing themselves. Third, they are suitable only for smartphones or desktop computers rather than smartglasses. ARcall, on the other hand, focuses on providing users with the ability to remotely augment the reality of the smartglasses wearer in real-time, as opposed to sending pre-recorded AR content or streaming a 3D model of the caller using a complex assembly to emulate face-to-face communication [21, 42, 64].

2.2 Context Assessment before Calling

Allowing friends or family members to bring unique experiences to users' smartglasses in a satisfying way without any explicit triggers from them is a challenge. It is important for callers to know their friends' availability before calling and potentially interrupting them. Ames et al. [4], however, show that family members often do not schedule video calls in advance and consider it socially strange or unnatural to do so. Unplanned audio and video calls typically start with questions such as "Are you busy?", "Where are you?", and "Is this a good time to talk?"; however, in the AR context, it may be too late to ask such questions once the call has been placed. Nardi et al. [53] highlighted this fundamental asymmetry in conversation, referring to limits on the receiver's availability, especially when the person is engaged in another task or conversation.

One way to tackle such issues is by implementing a two-step calling process. In the first step, a caller gauges the context of the callee and in the next step, initiates a call. For instance, Cramer et al. [14] investigated the "check-in" model of location sharing, in which a user pre-defines a list of friends or publicly shares their check-in location to help others determine an appropriate time to call. A similar study by Schildt et al. [75] reported that the participants often checked the location of someone they were about to contact to determine if the person they wanted to call was available for further communication. Pfleging et al. [67] proposed a solution that shares a driver's live camera feed; this enables the caller to become a "virtual passenger" in the vehicle, thereby empowering the caller to gauge the current situation and act accordingly. Further, modern video chatting applications address this challenge by notifying others when the callee is available or online. For instance, apps like Houseparty app [32] and Clubhouse [3] send users notifications when their friends are available for a video or audio chat. Facebook Rooms [20] adopts a similar approach.

Procyk et al. [70] prototyped a wearable video chat experience in which loved ones participated in a collaborative task over distance. Participants used smartphones and cameras mounted on either a hat or a normal pair of glasses. However, to date, no past work has explored the role of context assessment prior to initiating a conversation with a smartglasses wearer, an aspect of AR communication we investigate with the ARcall's Drop-In feature.

2.3 Time-Bounded Calls for Mobility

Video chat as a means of experience sharing has received significant attention [9, 27, 58]; the use of AR glasses to support video-based exchanges is a particularly attractive extension, as smart eyewear can support hands-free experience sharing with great ease. Typically, such scenarios involve the smartglasses wearer broadcasting their video and the remote user viewing the video and participating virtually. Increasing the remote user's sense of participation is key to making such interactions engaging. Procyk et al. [70] showed that audio played a central role in creating a strong sense of presence and connection with the remote partner during a shared remote activity. Similarly, Neustaedter et al. [57] examined the role of audio during a passive always-on video call in which family members simply wanted to be aware of what another member was doing without explicitly conversing during the call. Depending on the situation, an AR glasses wearer may not always want to see a video of the caller [46].

Despite the advantage of AR smartglasses allowing others to share the wearer's viewpoint anytime and having a screen that can project anytime, they also have a significant disadvantage in terms of experience sharing. Specifically, voice and video calls drain their batteries quickly, rendering them unusable until they are recharged. One potential solution to this problem is to support time-bound calls on AR smartglasses. Limiting call length may also help the wearer to focus, as past studies emphasize that users' cognitive resources are very limited when they are mobile [65]. There is a lack of significant research on time-bounded video calls using AR glasses.

We attempt to provide a feasible solution to the three core problems outlined above — enhancing interaction, enabling context assessment, and implementing time-bounded calling to limit power consumption — in designing for a wearable form factor. With ARcall, our goal is to reinvent traditional communication practices for wearable smartglasses, Keeping in mind the challenges such as shorter call duration (due to limited battery), context assessment, and AR-based interactions, Pfeil et al. [66] emphasized the need to develop a new set of design considerations for use with modern wearable technologies. Therefore, we will analyze a unique set of design considerations in the following section.

3 DESIGN CONSIDERATIONS

Our design considerations focus on the benefits of the mobile, wearable, and ready-to-project AR smartglasses form factor while also accounting for socio-technical limitations [1] such as privacy, interruptions, control, and so forth.

D1. *Giving control to Wearers and freedom to Friends:* An AR communication system must balance the interests of both the Wearer and the Friend: it must give Friends (callers) the ability to initiate calls [65] while preserving the Wearer's ultimate control over when they can be called [15]. Through ARcall's *Drop-In* component, we explore two techniques for addressing these competing interests. First, Drop-In is invite-only, meaning Friends can only drop in when the Wearer starts a session and invites them. Second, once invited, the Friend can drop in at any time to gauge the Wearer's context — where they are and what they are doing — *before* sending AR content and speaking to the Wearer in a Micro-Chat. Once invited, the Friend can drop in as many times as they want during the ARcall session.

D2. *Respecting the Wearer's privacy:* Sharing one's real-time camera feed with a remote Friend can make smartglass communication more compelling, but it also raises privacy concerns. Often, users have concerns over whether they might be bothering the person they want to reach before placing a call [38]. Our design for ARcall explores three techniques for mitigating privacy concerns. First, ARcall's *Drop-In* component is invite-only, as described in D1, and is designed only for interactions with a single Friend. Second, invitations automatically expire after one hour. Third, ARcall gives Wearers the option of blurring their video feeds. Naively, one might be tempted to include additional privacy controls, but it is critical to balance explicit efforts to fine-tune the privacy controls against the risk of degrading the user experience [2].

D3. *Facilitating relevant and well-placed AR content:* Native AR communication is potentially compelling for users because it can directly augment the Wearer's reality. For example, a Friend might send AR balloons to the Wearer on their birthday, and the balloons might appear to float in front of them. Two key determinants of whether or not the AR is compelling, however, are whether it is well-suited to the Wearer's current environment (i.e., whether it is relevant to the Wearer's context [50]), and whether it is well-placed within that environment [71]. For instance, in ARcall, Drop-In allows the Friend to see that the Wearer is celebrating their birthday, and ARaction allows them to augment the Wearer's reality with a floating balloon in the same way they could in the real world.

D4. *Field-of-view accommodation:* Effective AR experiences must take into account the limited field of view afforded by the displays currently built into smartglasses. With this in mind, in designing ARcall, we explored the effectiveness of restricting AR content to allow 3D objects (virtual models) only rather than entire "world reskins" [16, 45]. In our exploration, Wearers found the AR content to be more compelling when it augmented *part* of their environment rather than the *entire* environment.
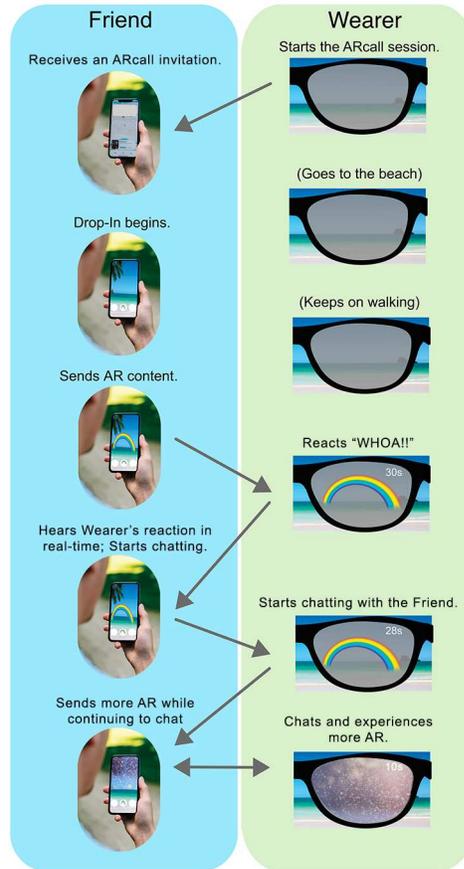
Fig. 2. ARcall interaction sequence. The Wearer starts the ARcall session, which invites the Friend to Drop-In. Later, the Friend drops in and surprises the Wearer with an AR rainbow. The Wearer reacts by saying "WHOA‼", and the two start chatting while the Friend sends even more AR content.

D5. *Handling power constraints:* Power constraints are a major hurdle for current smartglass devices [28, 51]. This is especially true for AR communication applications: prolonged and simultaneous operation of smartglass device displays, microphones, and live camera feeds (for instance, during audio or video calls) mean that it is critical to carefully manage power consumption. To constrain power consumption while giving users flexibility, Drop-In sessions are kept short: the Wearer can configure them to last between 30 seconds and one minute. However, if desired, the Wearer can extend active Drop-In sessions (while the Friend is currently dropped in) by 30-second increments by tapping a button on the temple of the smartglasses, and they can do this as many times as they want.

4 ARCALL

ARcall is an mixed-device calling system that facilitates AR-based real-time calls between smartphones and smartglasses. It comprises a smartglasses app (Glass App) for the Wearer and a smartphone app (Friend App) for the Friend (see Figure 3). The Wearer also has a companion smartphone app (Wearer App) to start and end ARcall sessions and configure
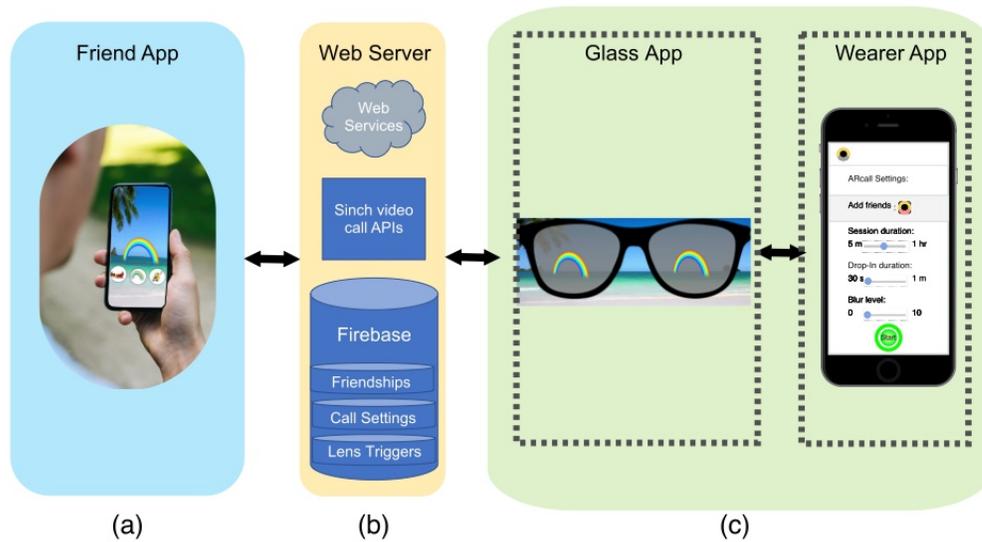
Fig. 3. ARcall System Diagram: (a) The 'Friend App' enables a Friend to drop in, send AR content, and talk to the Wearer; (b) The Web server handles video streaming, Friendship database, ARcall preferences, and connects the Wearer with the Friend remotely in real time; (c) The 'Wearer App' lets the Wearer control their preferences and start ARcall sessions. The 'Glass App' enables the Wearer to experience projected AR content.

other settings. This is in line with most consumer camera glasses [12, 33, 78], which have companion smartphone apps to control their settings.

First, the Wearer starts an *ARcall session* to invite their Friend to drop in. The ARcall sessions can last for up to an hour and expires automatically. During this time, the Friend can drop in anytime, initiating *Drop-In sessions*. Drop-In sessions are moments of co-presence between the Wearer and Friend that last for up to a minute each. During these Drop-In sessions, the Friend can engage with the Wearer using AR and voice in real-time. The end-to-end latency is around 100ms.

As an example of a typical interaction, the Wearer, who is going to the beach, might create an ARcall session to invite a close friend to drop in. The Friend can then drop in anytime in the next hour. This allows the Friend to learn the Wearer's context — where they are and what are they doing — before sending AR content and interacting with the Wearer. The AR content augments the Wearer's reality. The Friend can experience their reaction, and the two can talk to each other in real time. If the Wearer wants, they can tap on their smartglasses to extend the Drop-In session by 30-second increments, or the call ends, and the Friend can drop in again anytime during the ARcall session. These short Drop-In sessions preserve the battery in the smartglasses, but give the Wearer control to engage with the Friend for as long as they want. ARcall makes it possible for the smartglasses Wearer and their Friend to share a native AR-based experience in real-time.

Wearer Side:

Our *Wearer app* is written in Swift and runs on iOS. The Wearer app runs on a companion smartphone. This app provides controls to set the duration of an ARcall session, to start and end the session, and to select a Friend to send the

ARcall invite to. The Wearer sets the ARcall parameters before inviting their Friend. These parameters are then pushed to a cloud database server, Firebase [25].

Our *Glass app*, which runs on the Wearer's smartglasses, is written in Java and runs on Android operating system version 5.0 and API level 21. ARcall was implemented on Snap's 4th Generation AR Spectacles (2021) [34]. We used the Sinch video calling API [76] for both the Glass app and the smartphone app used by the Friend. The Glass app pulls the ARcall session parameters from the cloud database server. Note that these parameters are set in the Wearer app before starting the ARcall session. The Glass app also listens to the taps on the right temple of the smartglasses.

Friend Side:

Friend App is written in Swift and runs on iOS. The Wearer app runs on a smartphone. This application used ARKit [7], SceneKit [6], and SpriteKit [8] to render AR content on the smartphone. It also employs the Sinch video calling APIs [76] to support high-quality video calls with the Wearer. This application supports three tasks: scrolling through the list of AR content, sending the AR content to the Glass App, and the ability to mute microphone audio.

5 MAIN DESIGN COMPONENTS

We introduce three main components of ARcall, namely *Drop-In*, *ARaction*, and *Micro-Chat*. We describe these components along with the corresponding application features available to both the Wearer and the Friend.

5.1 Drop-In

Drop-In allows the Friend to see the Wearer's live context — where they are and what they are doing — by "dropping in"and seeing a video feed of the Wearer's point of view (see Figure 1 (a)). This is an invite-only feature between the Wearer and the Friend.

Wearer Side:

Inviting the Friend to the ARcall session (Wearer App): The Wearer specifies the Friend they are inviting before beginning the ARcall session. The information about the Friend is maintained in the cloud database server, Firebase [25]. The Wearer can add or remove the Friend as desired.

Setting the blur level (Wearer App): The Wearer can control the way their point of view (POV) is shared with the Friend. The blur level can be set to a value from 0 (high quality video stream) to 10 (completely blurred video stream). Blur control gives the Wearer a flexible way to maintain their privacy and to control their visual feed prior to inviting the Friend to drop in (D2).

Setting an ARcall session duration (Wearer App): The ARcall session's duration (i.e., the period during which the Friend is allowed to drop in) can be set for 5 minutes to an hour. The Wearer has to explicitly tap the "start" button to begin the ARcall session.

Setting a Drop-In session duration (Wearer App): The Wearer can also set how long each Drop-In session (period of co-presence with the Friend) can last within the greater ARcall session. We call this the Drop-In session duration. We conducted empirical tests to assess how long the smartglasses could support video streaming before becoming too hot (D5). Informed by this test, we restricted the possible duration to between 30 seconds to a minute. Regardless of the chosen duration, the Wearer can extend an active Drop-In session (one in which their Friend is currently dropped in) by 30-second increments, and they can do this as many times as they want.

Friend Side:

Fig. 4. (Left) The Friend selects an AR phoenix from a carousel on the Friend App. (Right) The AR phoenix is immediately projected onto the Wearer's point of view.

Reacting to an invitation (Friend App): When the Wearer begins the ARcall session, the Friend will receive a Drop-In invitation on their smartphone. The Friend can then drop in anytime by tapping on the invitation.

Seeing the Wearer's POV (Friend App): When the Friend participates in the Drop-In session, they can see and hear what's happening in the Wearer's environment via a camera feed. This camera feed is placed at the center of the smartphone display, as shown in Figure 4. The remote Friend can gauge the Wearer's context — where they are and what are they doing — in order to send them AR content (D1).

5.2 ARaction

ARaction component enables interaction with the Wearer using AR content. The Friend augments the Wearer's reality by projecting AR content onto their POV. As illustrated by Figures 2 and 4, the Friend can send AR content in real-time after dropping in.

Friend Side:

Selecting the AR Content (Friend App): We carefully selected AR content that fits the display and can be seen clearly by the Wearer. With D4 and D3 in mind, we chose two types of AR content. The first type consists of *animated 3D virtual objects*, such as a flying bird or a dragon, that are placed in the middle of the smartglass display when sent by the Friend (see Figure 5). Animated content can improve the quality of expression while conversing [5]. These 3D objects will follow the Wearer's head position and will remain visible, staying in the field-of-view of the smartglasses. The second AR content type consists of *particle objects*, such as snow flakes, that completely cover the smartglass display. In total, we employed ten pieces of AR content. The AR content was organized as a horizontal carousel at the bottom of the screen (see Figure 4), similar to popular photo- and video-sharing apps such as Snapchat [59] and Instagram [19]. Only one piece of AR content can be displayed to the Wearer at a time, but the Friend can send as many pieces of AR content as they wish.

Sending AR content (Friend App): After the Friend drops in to gauge the Wearer's context, they can choose a piece of AR content to project to the Wearer's view. While Drop-In lets the remote Friend *peek* into the Wearer's context, ARaction lets them *project* AR content at appropriate times (D1).

Fig. 5. Sample AR content for the Friend. We show (left to right) a phoenix, a dragon, holiday ornaments, and snow.

5.3 Micro-Chat

A Micro-Chat is a time-bounded voice call between the Wearer and the Friend that happens alongside the ARaction and adds meaning to it. It starts as soon as the the Drop-In session starts, specifically, once the Friend drops in to see the Wearer's point of view.

Wearer Side:

Extending a Drop-In session (Glass App): A countdown timer shows the time remaining in the Drop-In session. The timer's state is synced with the Glass App. The Wearer can extend the Drop-In session by 30-second increments during the session by pressing a button located on the right temple of the smartglasses.

Friend Side:

Setting audio controls (Friend App): The Friend's smartphone app lets them 'mute' or 'unmute' their microphone, so they can decide when they would like to start talking to the Wearer (see Figure 4 (left)).

6 FIRST-USE STUDY

This study qualitatively evaluated the following four aspects of the ARcall: (1) the overall user experience, (2) initial perceptions of using ARcall, (3) the feeling of connectedness, and (4) the friction points of using ARcall.

6.1 Participants

In total, 14 volunteers participated in the study. Our study group consisted of four females and ten males, with two participants in their 20s and 12 in 30s. The participants had diverse backgrounds and included a 3D artist, a product designer, a hardware engineer, and a software engineer. We asked participants to participate in pairs with their friends, so in total, we had seven sessions. Six participants had prior experience with smartglasses, seven participants had experience using AR filters, and nine participants were used to communicating with their friends using photos or videos using a smartphone. All Wearer participants had normal or corrected vision and did not wear prescription glasses.

6.2 Procedure

The participants went through a five-step procedure: a demographic questionnaire, ARcall onboarding, ARcall experience testing, a post-experience questionnaire, and a semi-structured interview. The ARcall experience testing took 10-15 minutes, and the study took roughly 60 minutes.

First, each pair of participants completed a demographic survey that asked them to indicate their age, gender, and relationship. Next, we conducted a 20-minute onboarding session in which we showed the participant pairs how to use

ARcall and explained the available AR content. This try-it-yourself onboarding session ensured that we could mitigate the novelty effect of using smartglass. Since we conducted our investigation during the holiday season, we themed the AR content around holiday festivities and fantasy.

Next, for a portion of the ARcall experience testing, we asked the pairs to imagine that they lived in different places. We asked the Friend participant to use ARcall to send the Wearer holiday cheer, and we asked the Wearer to take a walk outside and invite the Friend to drop in. The Friend stayed in a conference room while the Wearer took a walk outside. Over 70 percent of informal calls in daily life consist of small talk and showing people things from our immediate environments that we would like to discuss [63]. These informal calls are typically short in duration, spanning 10-15 minutes each. Considering the primary use of camera glasses when wearers are engaged in outdoor activities [11, 56] and the limited cognitive resources at the Wearer's disposal [65], the ARcall experience was designed to run for 10-15 minutes.

All of the AR content available for the Friend to send was festivity or fantasy themed (see Figure 5), as the study took place around the holiday season. The available content included a bird, a dragon, holiday ornaments, holiday gifts, snow, a Star of David, a whale, a mistletoe sprig, a unicorn, a reindeer, and a Santa on a sleigh. To observe whether the Wearer's context played a role in the AR content selected, we placed six visual artifacts along the Wearer's walking path. Each artifact was themed to subtly correspond to a piece of AR content that the Friends could select. For example, a news story poster with a banner image of a dragon. Figure 6 (b) top highlights a sample of the visual artifacts that we placed on the walking path, along with their associated AR content.

Next, we asked the pair to complete a post-experience questionnaire that probed the users' experience with Drop-In, ARaction and Micro-Chat, which included open-ended questions with responses on a 7-point Likert scale (i.e., "1" indicated Strongly Disagree and "7" indicated Strongly Agree). Lastly, we conducted a semi-structured interview to capture remarks that the questionnaire may not have elicited.

6.2.1 Data collection and Analysis. The software application logs were analyzed to derive quantitative metrics like the median duration of the Drop-In session, the median number of AR content items sent, and the total number of times the session was extended. Further, a thematic analysis of the post-experiment responses and the semi-structured interview responses was conducted. Two separate researchers iteratively coded the collected data, then merged their findings to identify the common themes that emerged on further analysis. We discuss these themes for each ARcall design component in the next section. Our evaluation protocol and analysis method are similar to Judge et al.'s investigation [38].

7 RESULTS

The median Drop-In session duration was 1.9 minutes, and the median number of pieces of AR content sent per Drop-In session was 11. Over 200 pieces of AR content were sent in total, and Wearers extended the duration of the Drop-In sessions 48 times. Overall, the participants provided positive feedback on the experience of using ARcall and felt that it would add great value to smartglasses (MED=6).

7.1 Experience with ARaction

Wearers reported that the AR content they got from Friends was highly relevant to their context (MED=6) and felt that it augmented their reality (MED=6). Moreover, they were able to understand Friends' intent behind the AR content (MED=6). This shows that ARcall's three components worked together effectively, allowing Friends to drop in to gauge
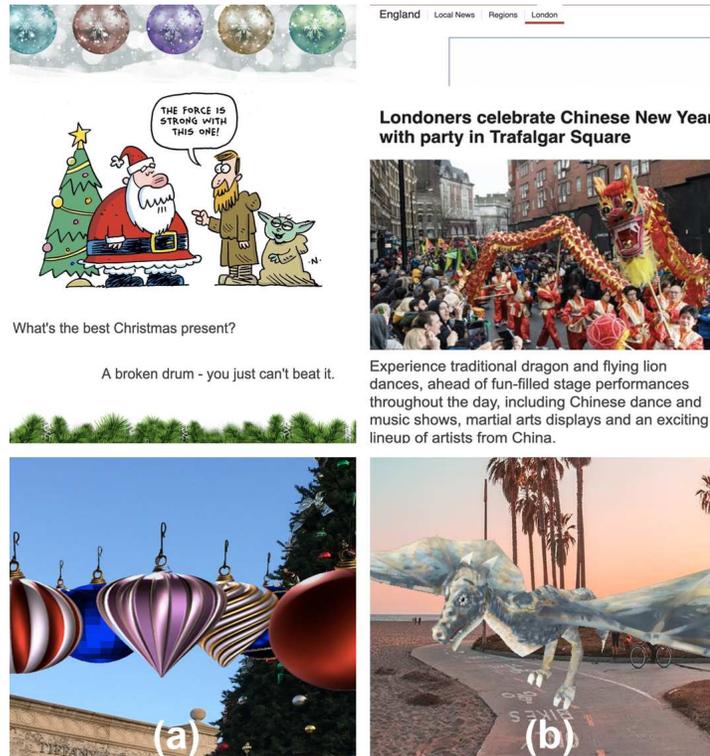
Fig. 6. Sample visual artifacts and corresponding AR content. (a) A Santa surrounded by ornaments and AR content of holiday ornaments; (b) A Dragon Festival news excerpt and AR content of a dragon.

the Wearer's context, send AR content relevant to that context, and use their voices to convey their intent — all simultaneously, in real-time.

7.1.1 Both Wearers and Friends found that ARaction could enable moments of surprise. On the Wearer side, participants were often pleasantly surprised by the AR content. For instance, W13 noted, *"I was surprised [and] I felt closer to [the Friend.]"* Also, W9 added, *"[ARaction] adds an element of surprise, which is really cool to see[.]"*

On the Friend side, projecting the AR content and getting the Wearer's reaction was perceived positively, and seeing the Wearer's reaction acted as a key motivator for sending them content. Friends felt very comfortable sending as much AR content as they wished (MED=6) and did not feel hesitant about sending AR content (MED=6), likely as a result of being specifically invited to the ARcall session by the Wearer. F10 commented, *"[ARaction was] unique because of the ability to shock people with the AR content."*

7.1.2 Both Wearers and Friends felt connected and found ARaction fun to use. On the Wearer side, W7 observed, *"Adding the AR content makes the interactions fun. I can see this being a great way to have fun with friends without having the event be too formal. Simply messing around with AR content [...] from a Friend [while I'm on a stroll] makes me feel special that someone reached out to me."* W3 added, *"They add somewhat of a gamification feel to my world which was [...] unique, and I wanted to try it out more and with more friends."* This participant went on to add, *"I think it's a superior experience [compared to livestreaming and video calls]. Life suddenly feels gamified, and you have your friend with you —*

there's more opportunities to share [...]. And the fact that you are not looking at your phone, you are [freer] to explore [the world around you]."

Friends felt that getting the Wearer's reaction was highly important to the AR content they projected (MED=6), and they had fun seeing the Wearer's reactions firsthand (MED=6). Friends also felt that hearing the Wearer heightened their sense of togetherness (MED=6). F16 remarked, *"It was an awesome way to have fun with my friend and mess with them. :) Sending AR content jokingly or situationally was interesting."*

7.1.3 Both Wearers and Friends found ARaction to be personalized and expressive. The Friend's ability to gauge the Wearer's context before projecting AR content, including their ability to pick up on the subtle visual artifacts that we put in place, made Wearers feel that Friends had hand-picked the AR content for them. For instance, W3 observed, *"Based on what I was looking at, they augmented the AR content to my reality, and it was pretty spot on."* W6 pointed out the way the context was utilized by their Friend: *"Pretty neat to have telepresence with another, and for them to be able to interact with what I was seeing."*

From the Friend's perspective, F4 felt, *"[This whole] experience make[s] our conversation [...] closer to each other, and it will blur the boarders and on more step[s] to brain-to-brain conversation."*

7.1.4 Both Wearers and Friends found ARaction to be immersive. ARaction enabled the AR content to be immersive for both parties. W3 noted, *"[I was s]urprised. I felt like my friend was somewhere in the heavens looking at me. [...] I think it did feel like my friend augmented my reality."*

From the Friend's perspective, Friends felt that they had the power to change the Wearer's reality. F12 said, *"With the snow I was able to make [W12] feel like he was in a different place, which to me is definitely augmenting his reality."* Moreover, F12 highlighted the value of ARaction: *"I think it adds quite a lot because it's more than just talking. It makes you feel like you are almost experiencing the same thing as the person."* F12 continued, *"I think it is a benefit because it makes it a shared experience for the two of you, instead of one person joining another person's experience."*

7.1.5 Friends wished to be offered a wider selection of AR content. Our participants had ten pieces of AR content to choose from, but they felt they needed more. W11 said that, for them, compelling AR meant *"anything that is relevant or contextual to the environment or feels more blended with the reality in front of me,"* strengthening the case for a wider selection of AR content. Moreover, studies have shown that a series of multiple emoticons can help convey new meanings [41] and, at times, convey meanings that go beyond the original intended meanings [40]. W5 noted, *"The AR content w[as] decorative, but [it] didn't necessarily capture how I was feeling."* F14 suggested additions such as *"Dog, cat, cars, [and] real world objects like tree[s]."* W5 suggested *"ones that feel relevant to the current view, my feelings, or my friend's feelings."* This suggests a need to develop more sophisticated AR content that can convey an understanding of a scene and be tailored to users' emotions.

7.1.6 Both Wearers and Friends wished to control the placement of AR content. Another direction for improving the experience is letting the Wearers move the AR content that is projected. Our current selection of AR content consisted of either particle objects (e.g., snow) or 3D virtual objects (e.g., a bird) that were automatically positioned and could not be moved. For instance, a projected bird would always fly in the center of the screen, and the snow would always take over the screen to simulate a snowfall. W11 lamented, *"I did feel a loss of control. Once the AR content appeared, it took over a lot of my field of vision, so I was no longer able to read the [news story that I was reading]."*

Friends acknowledged the issue of occlusion as well. F10 commented, *"It was fun to surprise them with different AR content although I felt like I was blocking their view the whole time."*

7.2 Experience with Drop-In

Most of the Wearers set the session duration to the maximum value of one minute and the camera feed to the "No Blur" setting. We suspect that the blurring felt unnecessary because the participants trusted each other during the study. Yet, the blurring feature might be necessary for a different usage context [17, 49] or if the degree of trust between the Friend and the Wearer varies [37, 48]. The Wearers indicated that ARcall's Drop-In feature protected their privacy (MED=6) and that they were also comfortable sharing their POV with their chosen Friend (MED=6). The ability to extend or end the call at any time is highly important to the Wearers (MED=6).

When they considered design aspects D1 and D2, the Friends echoed these sentiments in their responses to the following prompts: "Since my friend [the Wearer] invited me to their ARcall session, I did not feel that I was invading their privacy" (MED=6) and "I felt comfortable seeing the Wearer's POV" (MED=6). Drop-In elevated the sense of togetherness between the two parties (MED=6). W15 was skeptical initially, setting the blur option to "Full blur"; however, this participant later changed the setting to "No Blur" once they began experiencing ARcall.

7.2.1 Drop-In not only helped Friends determine the Wearer's context, it made them feel a sense of togetherness in relation to the Wearer. Our semi-structured interview feedback shed light on the value of Drop-In. W3 felt that *"Drop in was a very interesting experience. I felt [like I was] being guided by my friend and interestingly felt closer [to them] in that moment."* W9 stated, *"It is really cool that I can allow a friend to drop in because I am waiting for [the] surprise of them joining."* W7 felt that *"[it] makes it fun that friends can pop in and have fun with your worldview."* W3 echoed this sentiment: *"I think the Drop-In adds [to] the augmented reality part of the experience. I feel like it's very important to see what the Wearer is seeing."* Friends also felt that Drop-In was necessary for effective AR. F8 remarked, *"I think this adds a lot of value [to AR]. I really like[d] being able to see what my friend was seeing."* F4 highlighted the convenience associated with the ability to drop in: *"[I]t's fun and it [was a] really convenient way to show me what [W4] [was feeling and seeing] at that moment."* F10 and F4 observed, *"I think it's a unique experience seeing through the eyes of the other person and being able to interact, not face-to-face, but through their eyes."*

7.2.2 Drop-In did not seem to create privacy concerns, but Wearers wanted an indicator that showed when their Friend was present. The Wearers strongly agreed that Drop-In's invite-only nature was an important feature for controlling their privacy (MED=7). This supports design considerations D1 and D2, which recommend that allowing the Wearer to invite only one Friend would make ARcall a trustworthy and safe experience. In our ARcall implementation, there was no indication of when a Friend had dropped in; we designed it that way to make it more likely for the Wearer to be surprised by the Friend's AR in a fun way. This setup worked great for most of the Wearers, but some expressed the desire for an indicator. W7, for example, expressed the need for *"a better way to know when [F7] starts watching what I am sharing,"* adding that, *"[...] unless my friend [projected] an AR experience, I couldn't tell if they had joined."*

7.3 Experience with Micro-Chat

Wearers found several aspects of Micro-Chat highly engaging. They responded positively to the following prompts: "I felt together with my Friend as soon as I heard them talk" (MED=6) and "It was fun to [Micro-Chat] with my friend within my chosen time duration" (MED=6). On the Friend side, participants felt it important to chat with the Wearer right after projecting AR content to them (MED=6). They also found it fun to experience the Wearer's reactions using Micro-Chat (MED=6).

7.3.1 Both Wearers and Friends found that Micro-Chat added intimacy. Wearers felt that Micro-Chat provided instant context to their experience. W9 commented, *"It was my favorite thing—hearing my friend right away and knowing that*

they [could] see what I [projected]." W3 said, *"I think more than the AR experience, the voice chat was fantastic. I think this feature really builds [on] the human factor of [smartglasses]: the communication side."* For Friends, experiencing the Wearer's reaction to their AR content gave the Friend a sense of being together with the Wearer (MED=6). Friends found it important to get the Wearer's feedback on their chosen AR content (MED=6) using Micro-Chat. F6 remarked, *"[Micro-Chat] adds context to the AR content."* F16 asserted, *"ARcall would be nothing without its interaction aspects. Voice and AR content are the only way to interact with [a friend wearing AR glasses]. Critical."*

7.3.2 Friends felt that the duration of Micro-chat was short. There were a few exceptions where the Friends felt Micro-chat was limiting. F14 stated, *"Most video calls are longer for me, and I would prefer anything more than a sound byte."* We also saw contrary remarks, for instance, F6 pointed out, "I can see it being helpful for quick questions and fast interactions." Specific to F14's concern about the call length, ARcall does provide the wearer with a way to extend the call. So, in a way, it depends on the the mutual decision of the parties in the ARcall experience. The Wearers indicated no concerns regarding the Micro-chat.

8 DESIGN IMPLICATIONS AND LIMITATIONS

Drop-Ins enable surprise, but Wearers should be able to know about Friends' presence as soon as they drop in: While most Wearers enjoyed being surprised by AR content, some wanted more context about their Friend's presence (see 7.2.2). In future AR communication apps, Friends' information can be presented more explicitly along with the projected AR content (e.g., "John just dropped-in!"). There may be times when a surprise is undesirable, and the wearer may benefit from increased control over incoming call requests. For example, for non-subtle AR lenses, it would be useful to snooze the incoming call or have the option of seeing a basic LED-like notification when AR content arrives rather than the content itself.

AR communications systems should feature a wide variety of AR content and intelligent scene-based recommendations: Results (7.1.5) show that Wearers felt that AR content was highly personalized when they noticed its relevance to a visual artifact that they were looking at. Thus, future AR communication systems should provide a gallery of AR content for Friends to search through so that they can match the AR content they send with the Wearer's situation. Alternatively, the app could be designed as an AI-based system that can scan the Wearer's scene and recommend relevant AR content to the Friend. For instance, the Friend App could recommend a beach ball or a sandcastle if the Wearer were at the beach. However, designers should carefully craft such experiences, balancing automation and freedom of expression [5].

AR content should be chosen to minimize the difference between the way Friends see the content and the way Wearers see it: We observed that Friends' perception of how the Wearer would see the AR content they sent did not always match the way the Wearer actually saw it (see 7.1). For instance, a Friend mentioned *"With the snow AR content, I was able to make [the Wearer] feel like he was in a different place, which to me is definitely augmenting his reality."* However, the Wearer observed that *"[..] some content like the snow and the orcas [felt] strange [because they were not] full-screen."* This difference in perception between the Friend and Wearer occurs whenever the AR content is displayed edge-to-edge on the Friend's smartphone screen. It was detrimental to the Wearer's experience. They felt that the display became obvious and that it disrupted their sense of immersion. In the future, designers should take into account such rendering differences when designing for AR communication.

AR content should be movable and anchored to real world objects and surfaces: Results (7.1.6) revealed two important principles for how AR content should be presented to smartglass wearers. First, the Wearers consider AR more convincing

when it integrates with their environment realistically. For example, a palm tree model should be pinned to the ground and a bee model should be floating in the air, in the same way they would appear in the real world. The concept of native AR communication centers on immersion, so future AR communication designers should aim to anchor AR content to real-world objects or surfaces to make it more convincing. Second, both Wearers and Friends found that AR content sometimes occluded their view, so future AR communication systems should allow them to control the position of AR content within the environment.

The short duration of Micro-Chats should be seen as a means of enabling new social experiences rather than simply prolonging battery life: While we put a limit on the duration of Drop-In sessions to preserve the battery life of the smartglasses, this limit offers a valuable benefit (see 7.3): it can enable low-commitment, short bursts of interactions between friends, similar to a "hallway chat." The Drop-In concept, together with Micro-Chat, has the potential to create an "office-hour"-like experience in which, once the Wearer indicates their availability (i.e., "props open their office door"), their Friend can drop in and have a brief interaction before moving on to their business. Studies have found that short-lived interactions, even with strangers, can induce positive feelings [18]. While we did not investigate these, there are additional ways to reduce power consumption, such as lowering display brightness in low-light situations or using only one side of the display. Future designers can explore this space further and use these new design components to create compelling experiences, including components that involve more than one Friend or allow functionality when bandwidth is limited [39].

Although the proof-of-concept system demonstrated in this research could be improved further, for instance, by adding scene understanding to stylize AR lenses to the Wearer's surroundings, we feel that pursuing these improvements is a promising avenue for future research. Furthermore, because cameras are becoming more powerful, future systems could employ vision-based techniques, without bothering the user [80], to intelligently situate the AR content in order to offer a more contextual AR experience. We hope the design implications of ARcall will help designers craft novel AR communication methods and allow researchers to investigate further improvements.

9 CONCLUSION

The idea of using AR as a core medium of communication is an essential step in pushing the boundaries of technology-mediated communication. ARcall is a new type of calling system that explores native AR-based communication. Our investigation with 14 participants found that ARaction can enable moments of immersion, playfulness, surprise, and fun. We found that Drop-In helped Friends gauge Wearers' real-time context and allowed them to personalize the ARaction for the Wearer. Additionally, Micro-Chat creates intimacy and makes the moments of co-presence more effective. Our findings about user behaviour and the design insights derived from this study can inspire designers and researchers to explore a new design space for native AR communication systems. We hope AR-based communication systems like ARcall can enable people to express themselves in ways that were not possible before.

REFERENCES

[1] Mark S Ackerman. 2000. The intellectual challenge of CSCW: The gap between social requirements and technical feasibility. *Human–Computer Interaction* 15, 2-3 (2000), 179–203.
[2] Mark S Ackerman and Scott D Mainwaring. 2005. Privacy issues and human-computer interaction. *Computer* 27, 5 (2005), 19–26.
[3] Alpha Exploration Co. 2020. Clubhouse. https://www.joinclubhouse.com.
[4] Morgan G Ames, Janet Go, Joseph'Jofish' Kaye, and Mirjana Spasojevic. 2010. Making love in the network closet: the benefits and work of family videochat. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*. 145–154.

[5] Pengcheng An, Ziqi Zhou, Qing Liu, Yifei Yin, Linghao Du, Da-Yuan Huang, and Jian Zhao. 2021. VibEmoji: Exploring User-authoring Multi-modal Emoticons in Social Communication. *arXiv preprint arXiv:2112.13555* (2021).

[6] Apple Inc. 2019. SceneKit | Apple Developer Documentation. https://developer.apple.com/documentation/scenekit/.

[7] Apple Inc. 2020. ARKit - Augmented Reality - Apple Developer. https://developer.apple.com/augmented-reality/arkit/.

[8] Apple Inc. 2020. SpriteKit - Apple Developer. https://developer.apple.com/spritekit/.

[9] Uddipana Baishya and Carman Neustaedter. 2017. In Your Eyes: Anytime, Anywhere Video and Audio Streaming for Couples. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 84–97.

[10] Mark Billinghurst, Alaeddin Nassani, and Carolin Reichherzer. 2014. Social panoramas: using wearable computers to share experiences. In *SIGGRAPH Asia 2014 Mobile Graphics and Interactive Applications*. 1–1.

[11] Taryn Bipat, Maarten Willem Bos, Rajan Vaish, and Andrés Monroy-Hernández. 2019. Analyzing the use of camera glasses in the wild. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–8.

[12] Blade 2020. Vuzix Blade. https://www.vuzix.com/products/blade-smart-glasses.

[13] Ashley Carman. 2019. Snapchat launches another aging AR lens to lure people back to the app. https://www.theverge.com/2019/11/21/20974367/snapchat-time-machine-baby-aging-lens-ar-launch.

[14] Henriette Cramer, Mattias Rost, and Lars Erik Holmquist. 2011. Performing a check-in: emerging practices, norms and'conflicts' in location-sharing using foursquare. In *Proceedings of the 13th international conference on human computer interaction with mobile devices and services*. 57–66.

[15] Edward B Cutrell, Mary Czerwinski, and Eric Horvitz. 2000. Effects of instant messaging interruptions on computing tasks. In *CHI'00 extended abstracts on Human factors in computing systems*. 99–100.

[16] Fabien Danieau, Antoine Guillo, and Renaud Doré. 2017. Attention guidance for immersive video content in head-mounted displays. In *2017 IEEE Virtual Reality (VR)*. IEEE, 205–206.

[17] Tamara Denning, Zakariya Dehlawi, and Tadayoshi Kohno. 2014. In situ with bystanders of augmented reality glasses: Perspectives on recording and privacy-mediating technologies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2377–2386.

[18] Nicholas Epley and Juliana Schroeder. 2014. Mistakenly seeking solitude. *Journal of Experimental Psychology: General* 143, 5 (2014), 1980.

[19] Facebook, Inc. 2020. Instagram. https://instagram.com/.

[20] FBRooms 2020. Facebook Rooms. https://about.fb.com/news/2020/04/introducing-messenger-rooms/.

[21] Henry Fuchs, Gary Bishop, Kevin Arthur, Leonard McMillan, Ruzena Bajcsy, Sang Lee, Hany Farid, and Takeo Kanade. 1994. Virtual space teleconferencing using a sea of cameras. In *Proc. First International Conference on Medical Robotics and Computer Assisted Surgery*, Vol. 26.

[22] Steffen Gauglitz, Benjamin Nuernberger, Matthew Turk, and Tobias Höllerer. 2014. In Touch with the Remote World: Remote Collaboration with Augmented Reality Drawings and Virtual Navigation. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology - VRST '14*. ACM Press, Edinburgh, Scotland, 197–205. https://doi.org/10.1145/2671015.2671016

[23] Steffen Gauglitz, Benjamin Nuernberger, Matthew Turk, and Tobias Höllerer. 2014. In touch with the remote world: Remote collaboration with augmented reality drawings and virtual navigation. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology*. 197–205.

[24] Google Glass 2019. Glass. https://developers.google.com/glass/develop/gdk/touch.

[25] Google Inc. 2020. Firebase. https://firebase.google.com/.

[26] Graffity 2020. Graffity App. https://www.youtube.com/watch?v=DVy8SHpTxsA.

[27] Saul Greenberg and Carman Neustaedter. 2013. Shared living, experiences, and intimacy over video chat in long distance relationships. In *Connecting families*. Springer, 37–53.

[28] Daniel A Hashimoto, Roy Phitayakorn, Carlos Fernandez-del Castillo, and Ozanan Meireles. 2016. A blinded assessment of video quality in wearable technology for telementoring in open surgery: the Google Glass experience. *Surgical endoscopy* 30, 1 (2016), 372–378.

[29] Zhenyi He, Ruofei Du, and Ken Perlin. 2020. CollaboVR: A Reconfigurable Framework for Creative Collaboration in Virtual Reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 542–554.

[30] Steven J Henderson and Steven K Feiner. 2011. Augmented reality in the psychomotor phase of a procedural task. In *2011 10th IEEE International Symposium on Mixed and Augmented Reality*. IEEE, 191–200.

[31] Tobias Höllerer, Steven Feiner, Tachio Terauchi, Gus Rashid, and Drexel Hallaway. 1999. Exploring MARS: developing indoor and outdoor user interfaces to a mobile augmented reality system. *Computers & Graphics* 23, 6 (1999), 779–785.

[32] Houseparty 2020. Houseparty. https://houseparty.com/.

[33] North Inc. 2019. Focals by North. https://support.bynorth.com/.

[34] Snap Inc. 2021. Next Generation Spectacles. https://www.spectacles.com/ca-en/new-spectacles/.

[35] Hyungeun Jo and Sungjae Hwang. 2013. Chili: Viewpoint Control and on-Video Drawing for Mobile Video Calls. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13*. ACM Press, Paris, France, 1425. https://doi.org/10.1145/2468356.2468610

[36] Hyungeun Jo and Sungjae Hwang. 2013. Chili: viewpoint control and on-video drawing for mobile video calls. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. 1425–1430.

[37] Simon Jones and Eamonn O'Neill. 2011. Contextual dynamics of group-based sharing decisions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1777–1786.

[38] Tejinder K Judge and Carman Neustaedter. 2010. Sharing conversation and sharing life: video conferencing in the home. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 655–658.

[39] Joseph'Jofish' Kaye. 2006. I just clicked to say I love you: rich evaluations of minimal communication. In *CHI'06 extended abstracts on human factors in computing systems*. 363–368.

[40] Ryan Kelly and Leon Watts. 2015. Characterising the inventive appropriation of emoji as relationally meaningful in mediated close personal relationships. *Experiences of technology appropriation: Unanticipated users, usage, circumstances, and design* 2 (2015).

[41] Sujay Khandekar, Joseph Higg, Yuanzhe Bian, Chae Won Ryu, Jerry O. Talton Iii, and Ranjitha Kumar. 2019. Opico: a study of emoji-first communication in a mobile social app. In *Companion Proceedings of The 2019 World Wide Web Conference*. 450–458.

[42] Kibum Kim, John Bolton, Audrey Girouard, Jeremy Cooperstock, and Roel Vertegaal. 2012. Telehuman: effects of 3d perspective on gaze and pose estimation with a life-size cylindrical telepresence pod. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2531–2540.

[43] Seungwon Kim, Gun Lee, Weidong Huang, Hayun Kim, Woontack Woo, and Mark Billinghurst. 2019. Evaluating the Combination of Visual Communication Cues for HMD-Based Mixed Reality Remote Collaboration. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. ACM Press, Glasgow, Scotland Uk, 1–13. https://doi.org/10.1145/3290605.3300403

[44] Seungwon Kim, Gun Lee, Weidong Huang, Hayun Kim, Woontack Woo, and Mark Billinghurst. 2019. Evaluating the combination of visual communication cues for HMD-based mixed reality remote collaboration. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–13.

[45] Grzegorz Krawczyk, Karol Myszkowski, and Hans-Peter Seidel. 2005. Perceptual effects in real-time tone mapping. In *Proceedings of the 21st spring conference on Computer graphics*. 195–202.

[46] Andrew L Kun, Hidde van der Meulen, and Christian P Janssen. 2019. Calling while driving using augmented reality: Blessing or curse? *PRESENCE: Virtual and Augmented Reality* 27, 1 (2019), 1–14.

[47] Hideaki Kuzuoka, Toshio Kosuge, and Masatomo Tanaka. 1994. GestureCam: A video communication system for sympathetic remote collaboration. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work*. 35–43.

[48] Mikko Kytö, Ilyena Hirskyj-Douglas, and David McGookin. 2021. From Strangers to Friends: Augmenting Face-to-face Interactions with Faceted Digital Self-Presentations. In *Augmented Humans Conference 2021*. 192–203.

[49] Scott Lederer, Anind K Dey, and Jennifer Mankoff. 2002. *A conceptual model and a metaphor of everyday privacy in ubiquitous computing environments*. Computer Science Division, University of California.

[50] Ryan Louie, Kapil Garg, Jennie Werner, Allison Sun, Darren Gergle, and Haoqi Zhang. 2021. Opportunistic Collective Experiences: Identifying Shared Situations and Structuring Shared Activities at Distance. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–32.

[51] Kodai Matsuhashi, Toshiki Kanamoto, and Atsushi Kurokawa. 2020. Thermal model and countermeasures for future smart glasses. *Sensors* 20, 5 (2020), 1446.

[52] Jörg Müller, Tobias Langlotz, and Holger Regenbrecht. 2016. PanoVC: Pervasive telepresence using mobile phones. In *2016 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 1–10.

[53] Bonnie A Nardi, Steve Whittaker, and Erin Bradner. 2000. Interaction and outeraction: instant messaging in action. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*. 79–88.

[54] Alaeddin Nassani, Hyungon Kim, Gun Lee, Mark Billinghurst, Tobias Langlotz, and Robert W. Lindeman. 2016. Augmented Reality Annotation for Social Video Sharing. In *SIGGRAPH ASIA 2016 Mobile Graphics and Interactive Applications on - SA '16*. ACM Press, Macau, 1–5. https://doi.org/10.1145/2999508.2999529

[55] Alaeddin Nassani, Hyungon Kim, Gun Lee, Mark Billinghurst, Tobias Langlotz, and Robert W Lindeman. 2016. Augmented reality annotation for social video sharing. In *SIGGRAPH ASIA 2016 Mobile Graphics and Interactive Applications*. 1–5.

[56] Carman Neustaedter, Erick Oduor, Gina Venolia, and Tejinder K Judge. 2012. Moving beyond talking heads to shared experiences: the future of personal video communication. In *Proceedings of the 17th ACM international conference on Supporting group work*. 327–330.

[57] Carman Neustaedter, Carolyn Pang, Azadeh Forghani, Erick Oduor, Serena Hillman, Tejinder K Judge, Michael Massimi, and Saul Greenberg. 2015. Sharing domestic life through long-term video connections. *ACM Transactions on Computer-Human Interaction (TOCHI)* 22, 1 (2015), 1–29.

[58] Carman Neustaedter, Jason Procyk, Anezka Chua, Azadeh Forghani, and Carolyn Pang. 2020. Mobile video conferencing for sharing outdoor leisure activities over distance. *Human–Computer Interaction* 35, 2 (2020), 103–142.

[59] Casey Newton. 2018. You can now build your own face filter for Snapchat. https://www.theverge.com/2018/4/17/17245008/snapchat-face-filter-creation-lens-studio-giphy-discover.

[60] Next-Gen Spectacles 2021. Snap Spectacles. https://www.spectacles.com/ca-en/new-spectacles.

[61] Nreal 2020. NREAL AI. https://www.nreal.ai/.

[62] Ohan Oda, Carmine Elvezio, Mengu Sukan, Steven Feiner, and Barbara Tversky. 2015. Virtual replicas for remote assistance in virtual and augmented reality. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. 405–415.

[63] Kenton O'Hara, Alison Black, and Matthew Lipson. 2006. Everyday practices with mobile video telephony. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. 871–880.

[64] Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L Davidson, Sameh Khamis, Mingsong Dou, et al. 2016. Holoportation: Virtual 3d teleportation in real-time. In *Proceedings of the 29th annual symposium on user interface software and technology*. 741–754.

[65] Antti Oulasvirta, Sakari Tamminen, Virpi Roto, and Jaana Kuorelahti. 2005. Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 919–928.

[66] Kevin P Pfeil, Neeraj Chatlani, Joseph J LaViola Jr, and Pamela Wisniewski. 2021. Bridging the Socio-Technical Gaps in Body-worn Interpersonal Live-Streaming Telepresence through a Critical Review of the Literature. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–39.

[67] Bastian Pfleging, Stefan Schneegass, and Albrecht Schmidt. 2013. Exploring user expectations for context and road video sharing while calling and driving. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. 132–139.

[68] Thammathip Piumsomboon, Gun A Lee, Jonathon D Hart, Barrett Ens, Robert W Lindeman, Bruce H Thomas, and Mark Billinghurst. 2018. Mini-me: An adaptive avatar for mixed reality remote collaboration. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13.

[69] Thammathip Piumsomboon, Youngho Lee, Gun A. Lee, Arindam Dey, and Mark Billinghurst. 2017. Empathic Mixed Reality: Sharing What You Feel and Interacting with What You See. In *2017 International Symposium on Ubiquitous Virtual Reality (ISUVR)*. IEEE, Nara, Japan, 38–41. https://doi.org/10.1109/ISUVR.2017.20

[70] Jason Procyk, Carman Neustaedter, Carolyn Pang, Anthony Tang, and Tejinder K Judge. 2014. Exploring video streaming in public settings: shared geocaching over distance using mobile video chat. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2163–2172.

[71] Xukan Ran, Carter Slocum, Maria Gorlatova, and Jiasi Chen. 2019. ShareAR: Communication-efficient multi-user mobile augmented reality. In *Proceedings of the 18th ACM Workshop on Hot Topics in Networks*. 109–116.

[72] Jan Ole Rixen, Teresa Hirzle, Mark Colley, Yannick Etzel, Enrico Rukzio, and Jan Gugenheimer. 2021. Exploring Augmented Visual Alterations in Interpersonal Communication. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–11.

[73] Rokid Inc. 2019. Rokid Glass. https://glass.rokid.com/en/.

[74] Bektur Ryskeldiev, Michael Cohen, and Jens Herder. 2018. Streamspace: Pervasive mixed reality telepresence for remote collaboration on mobile devices. *Journal of Information Processing* 26 (2018), 177–185.

[75] Emily Schildt, Martin Leinfors, and Louise Barkhuus. 2016. Communication, coordination and awareness around continuous location sharing. In *Proceedings of the 19th International Conference on Supporting Group Work*. 257–265.

[76] Sinch. 2020. Video Calling - Get face to face with our simple APIs and SDKs. https://www.sinch.com/products/apis/calling/video/.

[77] Smoodji 2018. Smoodji Augmented Reality Application. https://www.octosense.com/project/smoodji-augmented-reality-messaging/.

[78] Spectacles 2019. Snap Focals. https://www.spectacles.com/ca-en/learn/.

[79] Balasaravanan Thoravi Kumaravel, Fraser Anderson, George Fitzmaurice, Bjoern Hartmann, and Tovi Grossman. 2019. Loki: Facilitating remote instruction of physical tasks using bi-directional mixed-reality telepresence. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. 161–174.

[80] Qianli Xu, Michal Mukawa, Liyuan Li, Joo Hwee Lim, Cheston Tan, Shue Ching Chia, Tian Gan, and Bappaditya Mandal. 2015. Exploring users' attitudes towards social interaction assistance on Google Glass. In *Proceedings of the 6th Augmented Human international conference*. 9–12.