Communication Behaviors in Colocated Collaborative AR Interfaces

Mark Billinghurst
Daniel Belcher
Arnab Gupta

Human Interface Technology Laboratory, University of Washington

Kiyoshi Kiyokawa

Human Interface Technology Laboratory, University of Washington
Emergency Communication Section, Communications Research Laboratory, Tokyo

The authors present an analysis of communication behavior in face-to-face collaboration using a multi-user augmented reality (AR) interface. 2 experiments were conducted. In the 1st experiment, collaboration with AR technology was compared with more traditional unmediated and screen-based collaboration. In the 2nd experiment, the authors compared collaboration with 3 different AR displays. Several measures were used to analyze communication behavior, and the authors found that users exhibited many of the same behaviors in a collaborative AR interface as in face-to-face unmediated collaboration. However, user communication behavior changed with the type of AR display used. The authors describe implications of these results for the design of collaborative AR interfaces and directions for future research.

1. INTRODUCTION

For several decades, researchers have explored how to use computers to support collaboration. The field of computer-supported collaborative work contains numerous examples of interfaces that facilitate mediated collaboration; however, much of this work is focused on remote collaboration.

In contrast, we are interested in developing interfaces for face-to-face collaboration; in particular, collaboration on tasks that involve viewing and manipulating spatial data. Many different interfaces have been developed for supporting

---

This work was supported in part by the Advanced Research and Development Activity (ARDA) under contract number NMA201-01-C-0006.

Requests for reprints should be sent to Mark Billinghurst, Human Interface Technology Laboratory, Fluke Hall, Room 215, University of Washington, Mason Rd., Seattle, WA 98195. E-mail: gro@hitl.washington.edu
colocated collaboration. Some have been based around computer conference rooms. For example, the Collaboration Laboratory room at Xerox (Stefik et al., 1987) used a network of workstations and a shared electronic whiteboard to support small group meetings. More recently, the i-LAND interface integrated several components into a combination of real and digital work environments for creative teams (Strietz et al., 1999). This is based on a concept called roomware, whereby computer-augmented objects in a room can be dynamically reconfigured to support face-to-face collaboration. The largest of these components is an interactive electronic wall that serves as a large common display.

Although these interfaces are good for supporting document-centered tasks, they have some limitations for collaborative viewing and interaction with spatial data. For example, it is often difficult to manipulate three-dimensional (3D) data on a two-dimensional screen or with traditional input devices (Shu & Flowers, 1994). Screen-based interfaces also create a separation between the real and digital domains. Real objects and interactions within the real world play an important role in face-to-face collaboration (Harrison & Minneman, 1996); however, apart from tangible user interfaces, such as the Meta-Desk (Ullmer & Ishii, 1997) or Fjeld’s Bricks (Fjeld, Voorhorst, Bichsel, Lauche, & Rauterberg, 1999), most colocated computer-supported collaborative work systems do not support object-based collaboration or include elements of the real world.

A subtler problem is that many interfaces for face-to-face collaboration artificially separate the communication and task space. When people sit at a table, the space between them is used for sharing communication cues, such as gazes, gestures, and nonverbal behaviors. If the people are communicating about objects placed on the table, then the task space is a subset of the communication space (Figure 1); however, when users are collaborating in front of a screen, their attention is focused on the screen (Figure 2). Thus, collaborators may exhibit different communication behaviors when using a screen-based interface than when seeing each other across a table.

Alternative methods to overcome these limitations include using large screens to project a 3D virtual image into space. The Cave Automated Virtual Environment (Cruz-Neira, Sandin, Defanti, Kenyon, & Hart, 1992) and the Responsive Workbench (Kruger et al., 1995) allow a number of users to view stereoscopic 3D images by wear-
ing LCD shutter glasses. Unfortunately, in this environment it is difficult to support more than two independent viewpoints. Mechanical devices can also be used to create true volumetric displays. These include scanning lasers onto a rotating helix to create a 3D volumetric display (Soltan, Trias, Dahlke, Lasher, & McDonald, 1995) or projecting images onto a rotating plate (see http://www.actuality-systems.com/). Although these systems support multiple viewpoints, they do not allow direct interaction with the virtual content because of the rotating display surface.

In our work, we use augmented reality (AR) technology to overlay 3D virtual imagery directly onto the real world. In this way, the display space and communication space overlap. Other researchers have found that collaborative AR interfaces can enhance colocated collaboration and facilitate very natural face-to-face communication (Ohshima, Satoh, Yamamoto, & Tamura, 1998; Szalavari, Schmalstieg, Fuhrmann, & Gervautz, 1998). This is because users can view and manipulate virtual models while seeing each other in the real world, allowing natural nonverbal communication cues to regulate the collaboration (Kiyokawa, Takemura, & Yokoya, 1999). In AR interfaces, real objects can be used to interact with the virtual content, increasing the intuitiveness of the interface.

Although several collaborative AR interfaces have been developed, there have been few rigorous user studies conducted with them. In comparison, other research in teleconferencing has evaluated the effect of technology on communication behaviors. For example, Isaacs and Tang (1993) compared audio-only conferencing to video conferencing and analyzed the differences in speaker behavior, whereas Krauss and Bricker (1967) studied the effect of delay in a conferencing link on communication behavior. Similar experiments are needed with collaborative AR interfaces.

In this article, we describe two communication experiments. In the first, we compared collaboration with AR technology to more traditional, unmediated, and screen-based collaboration. In the second experiment, we compared natural face-to-face collaboration to that with three different AR displays. As we discuss in the next section, we used a variety of experimental measures. After this discussion, we present results from our user studies and suggest directions for future research.
2. EXPERIMENTAL MEASURES

For decades, studies have examined the effect of technology on collaboration. Researchers such as Monk, McCarthy, Watts, and Daly-Jones (1996) argued that a multidimensional approach to evaluating video-mediated collaboration is needed to capture all the effects of the technology used. Thus, in our work we use a variety of performance, process, and subjective measures. Performance measures are those that measure a task outcome, such as the time required to complete a task. Process measures are conversational elements that occur during the collaboration, such as the number of words spoken. Subjective measures are the participant’s own subjective impressions of the collaboration.

2.1 Performance Measures

Until recently, most mediated-communication experiments used only performance measures to differentiate among communication conditions. For example, Chapganis (1975) compared how quickly participants could communicate using 10 different modalities, such as audio only, video, writing, and unmediated face-to-face communication. Typical performance measures include how fast a task can be completed and the quality of the outcome of the collaboration. However, task outcome is often a poor measure of the effect on communication of different technology. Indeed, in many telecommunication experiments there were no performance differences between mediated conditions (Williams, 1977). This may be because participants try to protect their primary task of getting the work done (McCarthy & Monk, 1994), which they may do through increased workload, a factor not measured by performance-oriented studies. Hockey (1983) argued that this can be observed in measures that look at the process of task performance. In our user studies, we measure performance time, but only to provide a gross measure of the difference among conditions.

2.2 Process Measures

Process measures are objective communication measures that capture the process of collaboration. These are extracted from transcriptions of audio and video recordings and notes made during the collaborative task. With the right process measures, one can find considerable differences among technology conditions.

One of the difficulties with collecting process measures is deciding which metrics to use. Nyerges, Moore, Montejano, and Compton (1998) provided a good introduction to the art of coding groupware interactions and gave guidance regarding good metrics to code for. Measures that have been found to be significantly different across technology conditions include:

- Frequency of conversational turns (Daly-Jones, Monk, & Watts, 1998; O’Conaill & Whittaker, 1997; O’Malley, Langton, Anderson, Doherty-Sneddon, & Bruce, 1996).
• Incidence/duration of overlapping speech (Daly-Jones et al., 1998; O’Conaill & Whittaker, 1997).
• Number of interruptions (Boyle, Anderson, & Newlands, 1994; O’Conaill & Whittaker, 1997).
• Turn completions (Tang & Isaacs, 1992).
• Dialogue structure (Boyle et al., 1994; O’Malley et al., 1996).
• Back-channels (O’Conaill & Whittaker, 1997).

Gestures and nonverbal behaviors can also be analyzed for characteristic features. In our work, we transcribe videotapes for both low-level speech and gesture features.

2.3 Subjective Measures

Subjective measures are based entirely on the users’ perception of their experience. The typical method is to have users fill out a survey questionnaire after each experimental condition. Daly-Jones et al. (1998) provided a set of questions that have been found to be sensitive to differences in mediating technology. These questions refer to interpersonal awareness, ease of communication, and the suitability of the communication mode for the task; for example:

I was very aware of the presence of my conversational partner.
I could readily tell when my partner was concentrating on what I was saying.

These are usually answered on a scale that ranges from disagree to agree.

In our experiments, we used a modified version of Daly-Jones et al.’s (1998) survey questions as well as postexperiment interviews to capture the users’ experience.

3. EXPERIMENT 1: COMMUNICATION IN A COLLABORATIVE AR INTERFACE

In Experiment 1, we compared collaboration with an AR interface to a projection display and to unmediated face-to-face interaction. The three different conditions we tested were:

FtF: unmediated face-to-face collaboration
AR: augmented reality face-to-face collaboration
Proj: projection screen-based collaboration

The main differences among these conditions were in the viewpoints of the collaborators and the method for interacting with the shared objects (see Table 1).

By considering the innate characteristics or affordances of each condition, we can predict the impact on communication behaviors. In the FtF and AR conditions, users could more easily share nonverbal cues and use gestures to interact with objects. Thus, we should see a difference in language and behavior between the FtF
Table 1: Technology Affordances of Each Condition

<table>
<thead>
<tr>
<th></th>
<th>Face to Face</th>
<th>Augmented Reality</th>
<th>Projection Screen</th>
</tr>
</thead>
<tbody>
<tr>
<td>User viewpoint</td>
<td>Natural eyes</td>
<td>Private display</td>
<td>Public display</td>
</tr>
<tr>
<td></td>
<td>Independent view</td>
<td>Independent view</td>
<td>Common viewpoint</td>
</tr>
<tr>
<td>Interaction</td>
<td>Easy to change viewpoint</td>
<td>Easy to change viewpoint</td>
<td>Difficult to change</td>
</tr>
<tr>
<td>Natural object manipulation</td>
<td>Two handed</td>
<td>Tangible AR techniques</td>
<td>Mouse based</td>
</tr>
<tr>
<td></td>
<td>Space-multiplexed input</td>
<td>Two handed</td>
<td>One handed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Space-multiplexed input</td>
<td>Time-multiplexed</td>
</tr>
</tbody>
</table>

and AR conditions and the Proj condition. For example, we should see more deictic language used in the FfF and AR conditions and more gestures and communication spread across the verbal and nonverbal channels.

However, viewing the world through an AR display is very different from seeing it naturally. The field of view may be limited, the resolution may be poor, and the display may not support stereoscopic viewing. These factors combine to give a mediated viewing experience that is very different from natural face-to-face collaboration and may also affect the communication behaviors.

3.1 Experimental Design

Pairs of users were engaged in an urban design task. This involved placing nine real or virtual model buildings on a 3 x 3 block street layout on a table. The virtual models were exact copies of the real buildings. For each condition, the buildings needed to be placed to satisfy 10 rules, such as:

- The CHURCH is next to the THEATER.
- The FIREHOUSE and the BANK are across the street.
- The CITY HALL is in the middle of town.

The rules were designed to be complementary, and each participant was given only 5 of them.

Participants were told they could not show their partner the piece of paper on which the rules were written but that they could freely use speech and gesture to collaborate. This was to encourage them to communicate together, even if just to read out the rules they had. They were given 7 min to complete the task. Before beginning the task, they were given several minutes to familiarize themselves with the real model buildings so they could easily recall what each building was.

Participants were asked to arrange the same set of buildings under three conditions.

**FfF.** The participants sat or stood on opposite sides of a table (see Figures 3 and 4) and attempted to lay out a set of nine real model buildings subject to the constraints mentioned earlier.
AR interaction. The participants were given a set of nine cards with Japanese Kanji characters marked on them (Figure 5). They also wore the DaeYang Cy-Visor head-mounted display with a small video camera attached (Figure 6). These displays are full color, biocular rather than stereoscopic and have SVGA (800 x 600 pixel) resolution and a 30° field of view. When they looked at the marked card, a computer vision based tracking method was used to track the cards and overlay virtual buildings on them (Figure 7; see http://www.hitl.washington.edu/artoolkit/). Before beginning, participants were given several minutes to practice with the AR content.

Proj. Participants sat side by side at one end of a table, facing a large projection screen (Figure 8). On the table were two cards, each with a Polhemus Fastrak magnetic tracker and a button mounted on it (see http://www.polhemus.com). These trackers were used to find the position and orientation of the card relative to a fixed magnetic source. Shown on the screen was a graphical interface that allowed the participants to place virtual models of buildings on a flat ground plane (Figure 9).
FIGURE 5  Tracking cards.

FIGURE 6  Augmented reality condition.

FIGURE 7  Augmented reality view.
As they moved the Fastrak sensors, virtual markers on the screen moved in the same way. Buildings could be selected and placed using the button on the tracker. This interface was developed using Sense8's WorldUp software (see http://www.sense8.com). Before beginning, participants were trained in how to use the interface, and they practiced placing buildings until they were comfortable with the software.

A projection screen interface using two 6-degree-of-freedom input devices was chosen because it supports simultaneous user input, enabling users to move and place the buildings in parallel. The projection condition creates a separation between the task space (the projection screen) and the communication space (the users at the table) when compared to the other conditions.

We used a within-subject design so that each pair of participants completed all three interface conditions. The conditions were presented in a counterbalanced manner to reduce order effects.
3.2 Results

The participants in this experiment were 14 pairs of college-age adults: 6 pairs of women and 8 pairs of men. Ages ranged from 21 years to 38 years. Several of these pairs were later dropped from the study because of incomplete results, so the final results were drawn from 12 pairs of participants.

**Performance.** The participants were given 7 min to complete the logic puzzle task. They were also told that it was important to satisfy all of the 10 rules. There was a significant difference between puzzle solution times across conditions. Average solution times are shown in Table 2. We conducted a nonparametric Kruskal–Wallis test on the log of the solution times, $H(2) = 13.0, p < .002$. To make comparisons between pairs of conditions, we used a Mann–Whitney test and found that there was no significant difference in performance between the FF and Proj condition ($Z = -1.44, N = 26, p = .151$). However, there was a significant difference between the Proj condition and the AR condition ($Z = -2.21, N = 26, p < 0.05$) and between the FF condition and the AR condition ($Z = -3.51, N = 26, p < 0.01$). Thus, users solved the puzzle significantly slower in the AR condition.

**Communication process measures.** Complete transcriptions were made of the speech communication between pairs as they completed each of the conditions. These were then coded for low- and high-level speech behaviors. The video of each pair of collaborators was watched and coded for nonverbal and gestural behaviors. The communication measures collected included:

- Number and type of gestures made
- Number of deictic phrases spoken
- Average number of words per phrase
- Number of speaker turns

**Turn taking.** Table 3 shows the number of speaker turns and the average number of words per turn. Using a one-factor analysis of variance (ANOVA), we found that there was no difference in the average number of words per turn, $F(2, 30) = 1.37, p = .27$, or in the number of turns per second, $F(2, 30) = 0.23, p = .79$.

**Deictic phrases.** The number of deictic phrases used was counted. These are phrases with the words *this, that,* or *there* that cannot be fully understood by considering the speech alone. The average percentages of deictic phrases out of the total

<table>
<thead>
<tr>
<th>Table 2: Average Solution Time in Seconds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Average solution time</td>
</tr>
<tr>
<td>SD</td>
</tr>
</tbody>
</table>
number of phrases spoken in each condition are shown in Table 4. We used a Kruskal–Wallis test and found $H(2) = 7.58, p < .05$, so there was a significant difference in percentages of deictic phrases used across conditions. Using a Mann–Whitney test, we found no difference in deictic percentages between the FtF condition and the Proj condition ($p = .365$) or between the FtF condition and the AR condition ($p = .116$), but there was a significant difference between the Proj condition and the AR condition ($Z = -2.79, N = 22, p < .01$). Thus, users in the AR condition used proportionally more deictic phrases than participants in the Proj condition.

**Simultaneous speech.** We counted the numbers of turns in which simultaneous speech occurred. *Simultaneous speech* is defined as when both speakers are speaking at the same time. This includes types of speech such as interruptions that result in a change of speaker turn and completions in which one speaker completes the other’s sentence. The amount of simultaneous speech as a percentage of total speaker turns across all three conditions is shown in Table 5. There was no difference across conditions.

**Questions asked.** We counted the number of user turns that contained one or more questions and calculated a percentage of turns with questions in them relative to the total number of all speaking turns for each user. As the data in Table 6 show, in

<table>
<thead>
<tr>
<th>Table 3: Average Words Per Turn and Turns Per Second</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Face to Face</strong></td>
</tr>
<tr>
<td>Average words/turn</td>
</tr>
<tr>
<td>Average turns/sec</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 4: Average Percentage of Deictic Phrases</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Face to Face</strong></td>
</tr>
<tr>
<td>Average deictic phrases</td>
</tr>
<tr>
<td>SD</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 5: Average Percentage of Simultaneous Speech</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Face to Face</strong></td>
</tr>
<tr>
<td>Average % of simultaneous speech</td>
</tr>
<tr>
<td>SD</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 6: Average Proportion of Questions Asked</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Face to Face</strong></td>
</tr>
<tr>
<td>Percentage of questions</td>
</tr>
<tr>
<td>SD</td>
</tr>
</tbody>
</table>
the AR condition participants used questions more frequently than in the other conditions. These results were significantly different. We used a Kruskal–Wallis test and found $H(2) = 6.83, p < .05$. When we performed a pairwise comparison with a Mann–Whitney test, we found a significant difference between the AR condition and the FtF condition ($Z = -2.37, N = 22, p < .05$), and between the AR condition and the Proj condition ($Z = -2.07, N = 22, p < .05$), but not between the FtF condition and the Proj condition ($Z = 0.49, N = 22, p = .65$). Thus, participants asked proportionally significantly more questions in the AR condition than in the other two conditions.

However, many of the questions asked in the AR condition related to building identification; for example, “Where is the bank?”, “What is that building you are holding?”, and “That’s the firehouse?” The prevalence of these questions illustrates the reduction in perceptual cues caused by the AR head-mounted display (HMD). When we removed these questions (see Table 7), there was no significant difference across conditions, $H(2) = 1.46, p = .93$.

**Gestural communication patterns.** We also counted the number of gestures made and grouped them into one of four types:

- **Point:** a pointing gesture
- **Pick:** a gesture involved picking or moving an object
- **Collab:** a collaborative gesture, such as passing an object between people
- **Other:** any other gesture

Figure 10 shows the percentage breakdown for each type of gesture made out of all the gestures for each condition. In the Proj condition, the majority of gestures made (66%) were Pick gestures, whereas Point gestures made up less than 30%. In contrast, the percentages of Pick and Point gestures in the FtF and AR conditions were very similar. There was a significant difference between the percentage of Pick gestures made, $H(2) = 7.67, p < .05$, but not between the percentage of Point gestures made, $H(2) = 4.42, p = .11$. In the case of Pick gestures, there was no difference between the AR condition and the FtF condition ($Z = -0.68, N = 22, p = .49$), but there was a significant difference between the FtF condition and the Proj condition ($Z = -2.53, N = 22, p < .05$) and between the AR condition and the Proj condition ($Z = -2.15, N = 22, p < .05$). There was a significantly higher percentage of picking gestures in the Proj condition than in the other two conditions.

**Subjective measures.** After completing each condition, participants filled out a survey regarding how they felt about the interface and how well they could

<table>
<thead>
<tr>
<th>Table 7: Modified Average Proportion of Questions Asked</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Face to Face</strong></td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>Percentage of questions</td>
</tr>
<tr>
<td>SD</td>
</tr>
</tbody>
</table>
collaborate with their partner. All of the questions were answered on a scale that ranged from 1 (not very easily) to 7 (very easily). For example:

Q6: How easily could you move the buildings (1 = not very easily, 7 = very easily)?
Not Very Easily 1 2 3 4 5 6 7 Very Easily

In response to the question “How easily could you work with the other person?”, participants reported that it was easiest to work together in the FtF condition. There was a significant difference in working together between conditions (Figure 11); a one-factor ANOVA yielded $F(2, 69) = 5.96, p < .005$. Participants reported feeling that it was as difficult to work together in the AR condition as it was in the Proj condition. Similarly, when asked “How easily could you understand your partner?”, participants reported that it was significantly easier to understand each other in the FtF condition, than in the Proj condition or the AR condition (Figure 11), $F(2, 69) = 6.31, p < .005$.

When asked how easily they could tell where their partner was looking and pointing, participants reported that perceiving these nonverbal cues was far easier in the FtF condition and that there was little or no advantage in using the AR display compared with the Proj display condition (Figure 12). A one-factor ANOVA revealed a highly significant difference in the average answers to both these questions. For “How easily could you tell where the other person was looking?”, $F(2, 69) = 25.4, p < 1 \times 10^{-8}$, whereas for “How easily could you tell where the other person was pointing?”, $F(2, 69) = 14.2, p < 1 \times 10^{-4}$.

Participants were also asked how easy it was to pick and move the buildings. As shown in Figure 13, they reported feeling that it was nearly as easy to pick and move the AR objects as it was the real models and significantly easier than with the projection screen interface. A one-factor ANOVA of participant responses to the question “How easily could you pick up the buildings?” yielded $F(2, 39) = 37.8, p <$
FIGURE 11 How easy [was it] to work together and understand each other? (1 = not very easy, 7 = very easy). FtF = face to face; Proj = projection; AR = augmented reality.

FIGURE 12 How easy [was it] to recognize where your partner was looking and pointing? FtF = face to face; Proj = projection; AR = augmented reality.

$1 \times 10^{-8}$; whereas for the question “How easily could you move the buildings?”, we found $F(2, 39) = 28.4$, $p < 1 \times 10^{-7}$.

**Participant comments.** After participants completed the experimental task, they were interviewed for 15–20 min. In general, participants felt that although the AR condition was enjoyable, there were some usability issues that should be addressed. Ten of fourteen interviewees mentioned perceptual problems such as lim-
FIGURE 13 How easy was it to pick up and move the buildings? (1 = not very easy, 7 = very easy). FfF = face to face; Proj = projection; AR = augmented reality.

icted field of view, low resolution, and blurry imagery. Several participants felt that the limited field of view forced them to consciously switch modes in the AR task, for example, between reading the rules or placing buildings, or between collaborating with their partner and placing houses. Two participants made an interesting comment: They felt that the AR condition created a form of tunnel vision.

In addition to the perceptual issues, many participants commented on the limitations of the vision-based tracking. In half of the interviews the problem of the virtual buildings flicking in and out of view was mentioned. Several people mentioned that this made the buildings feel less real or said that they were reluctant to point at virtual objects, because this would make the buildings disappear. These factors taken together meant that users felt they had to “concentrate on seeing” in the AR case.

Despite these perception problems, participants felt that interaction with the virtual objects was easier in the AR condition than in the Proj condition. This was partly due to the tangible interaction metaphor we used that made the virtual buildings seem more real. For example, one participant said that

AR’s biggest limit was lack of peripheral vision. The interaction physically (trading buildings back and forth) as well as spatial movement was natural, it was just a little difficult to see. By contrast in the Projection condition you could see everything beautifully but interaction was tough because the interface didn’t feel instinctive.

Another participant said, “AR is more real since you are actually holding something just like the real thing.” In contrast, in 8 out of the 14 interviews, participants mentioned the difficulty of interacting with objects in the Proj condition. For some people this was because of the mouselike interface used, whereas in 6 interviews people mentioned how their hands kept bumping into their collaborator’s hands. Some participants said that this was due to the difficulty of seeing their real hands while focusing on the projection screen in front of them.
Participants also commented on the difference in viewpoints between the conditions. In seven interviews, participants said that in the FtF condition they were able to make eye contact and see the gaze direction of their partner sitting across the table. However, in the AR condition the narrow field of view and the bulky display made it impossible to clearly see gaze direction. In both the AR and the FtF conditions the participants said that most of their attention was focused on the task space between them, and they could easily see the gestures that were being made. As one participant said, “The AR condition was better than the Projection condition because I could point and see my partner and where she pointed.” However, in two of the interviews, participants talked about how having different viewpoints made them feel that they were working independently—or, as one person put it, “working solo together.” This was not the case with the Proj condition. Some participants commented that sharing the same viewpoint in the Proj condition helped their performance because they felt that they and their partner were working toward the same goal. In four of the interviews, participants said that they did not need to see the gaze of their collaborator, because they were looking at the scene from the same viewpoint. Others felt that not facing their partner made the Proj condition the most difficult.

3.3 Discussion

In general, these results show that although participants did not feel that the AR and FtF conditions were very similar, they used similar gesture and speech patterns in these conditions, and different behavior in the Proj condition. In their deictic speech patterns there was no difference between the FtF and AR conditions, and their picking behavior in the AR condition was more similar to that in the FtF condition than the Proj condition.

The tangible-interface metaphor seems particularly successful. Although hampered by the narrow field of view and low resolution of the AR display, participants felt that they could pick and move objects as easily in the AR condition as in the FtF condition and far more easily than the Proj condition.

On the basis of the interviews conducted after the experiment, it seems that the perceived difference between the AR condition and other two conditions can be explained by the reduction in perceptual cues due to current AR technology. For example, even though participants were sitting across the table from one another, in the AR case they had no peripheral vision, making it more difficult for them to recognize when their partner was looking at them. It is encouraging that even with reduced perceptual cues participants exhibited behaviors in the AR condition that are similar to unmediated FtF collaboration and significantly different from traditional interfaces. Higher resolution AR displays with a larger field of view should produce results even more similar to unmediated FtF collaboration.

There are many possible form factors for AR displays. In this experiment we used only one, an HMD, that was worn all the time and that reduced perception of the real world and available peripheral cues. Other display types might not have these same characteristics. To explore this further, we conducted a second experiment, in which we explored the influence of form factor on collaboration.
4. EXPERIMENT 2: EFFECT OF AR DISPLAY TECHNOLOGY

In Experiment 1, we compared unmediated face-to-face collaboration to collaboration with a projection screen and with an AR interface. In Experiment 2, we wanted to compare face-to-face collaboration with three different AR displays:

LCD: A 5-in (12.7 cm) diagonal LCD with a camera on the back (Figure 14a).
HHD: A hand-held display (Figure 14b) consisting of an Olympus FMD-150 display modified to be mounted on a handle with a camera. This is held in one hand and can be held against the face or taken away.
HMD: a head-mounted AR display (see Figure 14c) consisting of a Canon GT-270 display with a camera mounted on it. This display leaves both hands free when being used.

Characteristics of these displays are shown in Table 8.

The field of view occupied by the LCD display depends on how far away from the user's eyes it is held. For example, if the display is held at arm's length, then the field of view is about 5°.

As one can see, the resolution of the displays is similar, and two of them have almost the same field of view. However, one difference is in the amount of peripheral awareness they support. The Canon HMD is in front of the eyes at all times and gives no peripheral awareness beyond its 31° field of view. The Olympus HHD can be easily removed from in front of the face, allowing the user to easily see the real world.

**FIGURE 14** Panel A: LCD display; Panel B: hand-held display; Panel C: head-mounted display.

<table>
<thead>
<tr>
<th>Display</th>
<th>Resolution</th>
<th>Usage</th>
<th>Field of View</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCD panel</td>
<td>224,000 pixels</td>
<td>1 or 2 handed</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(320 × 234 RGB triads)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hand-held device</td>
<td>240,000 pixels</td>
<td>1 handed</td>
<td>37.5°</td>
</tr>
<tr>
<td>(Olympus FMD-150)</td>
<td>(377 × 212 RGB triads)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Head-mounted device</td>
<td>270,000 pixels</td>
<td>Hands free</td>
<td>31°</td>
</tr>
<tr>
<td>(Canon GT-270)</td>
<td>(400 × 225 RGB triads)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The LCD can be held at arm's length, allowing the user to see both the real and virtual views of the world at the same time, thus supporting the greatest peripheral vision. Previous research has shown that peripheral awareness is important in supporting colocated collaboration. Thus, we predicted that the HHD and LCD display configurations would affect communication less than the HMD; however, because the HHD has a greater field of view than the LCD panel, users may perform better using this display.

4.1 Experimental Design

To explore the effect of display affordances, we engaged pairs of participants in the same urban design task used in Experiment 1. Participants completed the task in each of four conditions:

- FtF – Two participants sat across a table with a street map marked on it. They were given a set of nine model buildings to layout on the street map and were free to place the buildings wherever they liked, subject to the same rules used in Experiment 1.
- LCD – This condition was identical to the FtF condition, but the participants each held an LCD panel and manipulated a set of nine marked tracking cards. When they looked at these cards through the LCD panel, they saw virtual copies of the real models they used in the FtF condition. They were free to lay the buildings out wherever they liked subject to the same rules used in Experiment 1.
- HHD – This condition was identical to the LCD condition, but each participant used an HHD.
- HMD – This condition was identical to the LCD condition, but each participant wore an HMD.

Each pair of participants first completed the FtF condition, then each of the three AR conditions in a counterbalanced order. As in Experiment 1, in each condition 2 participants were each given a set of 5 rules, and the goal was to place nine buildings in a way that satisfied all 10 rules. Unlike in Experiment 1, participants were allowed to take as long as they needed to complete the building layout task, although each condition was typically completed in 5-10 min.

4.2 Experimental Measures

The experimental measures were the same as those in Experiment 1, including:

- Performance measures: the time to complete the building layout task.
- Communication process measures: turn taking, deictic speech, questions, gestures, and simultaneous speech.
- Subjective measures: postcondition surveys and interviews.

The FtF condition (normal face-to-face collaboration) was used as a baseline against which the other measures were compared. We predicted that the results
from the LCD and HHD conditions would be more similar to the FtF condition than to the HMD condition. If this is indeed the case, then HMDs may have a negative impact on colocated collaboration compared with other types of AR displays.

4.3 Results

The participants were 12 pairs of college-age adults: 4 pairs of women and 8 pairs of men, ranging in age from 20 to 37 years.

**Performance.** The average times to complete the design task are presented in Table 9. The FtF condition produced the quickest average solution time, whereas the HHD condition was the fastest of the three AR conditions. However, there was no significant difference among the four conditions; a Kruskal–Wallis test yielded $H(2) = 6.14, p = .105$. As in Experiment 1, participants took longest time to complete the task in the HMD condition. When we used a Mann–Whitney test to compare the FtF and HMD conditions and found a nearly significant result ($Z = -1.848, N = 12, p = .068$).

**Communication process measures.** Participants were recorded on videotape while they performed the urban design task. However, because of the length of the solution times (up to 15 min in some cases), we did not make a complete transcription. Instead we selected 2 min of interaction in the middle of each task and transcribed this. It was possible for us to completely code the entire tape for nonverbal behaviors.

**Turn taking.** The number of speaker turns and the average number of words per turn were measured. For each pair we then normalized these relative to the results in the FtF condition. The average and normalized results are shown in Table 10. The HHD condition produced results closest to the FtF condition, but there was no difference in the normalized average number of words per turn across the three AR conditions, $H(2) = 0.796, p = .672$. There was also no difference in the normalized number of turns per second across conditions, $H(2) = 0.227, p = .893$.

**Deictic phrases.** We also counted the number of deictic phrases. The average percentages of deictic phrases out of the total number of phrases spoken in each condition are shown in Table 11. There was no significant difference across conditions, $H(2) = 1.503, p = .682$.

<table>
<thead>
<tr>
<th>Table 9: Average Performance Time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td><strong>FtF</strong></td>
</tr>
<tr>
<td>-----------------------------------</td>
</tr>
<tr>
<td>Average task time (sec)</td>
</tr>
<tr>
<td>329.1</td>
</tr>
<tr>
<td>422.8</td>
</tr>
<tr>
<td>366.3</td>
</tr>
<tr>
<td>463.9</td>
</tr>
<tr>
<td><strong>SD</strong></td>
</tr>
<tr>
<td>170.1</td>
</tr>
<tr>
<td>191.6</td>
</tr>
<tr>
<td>217.4</td>
</tr>
<tr>
<td>195.1</td>
</tr>
</tbody>
</table>

*Note. FtF = face to face; HHD = hand-held device; HMD = head-mounted device.*
Table 10: Average Words Per Turn and Turns Per Second

<table>
<thead>
<tr>
<th></th>
<th>FfF</th>
<th>LCD</th>
<th>HHD</th>
<th>HMD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average words/turn</td>
<td>13.22</td>
<td>9.79</td>
<td>11.00</td>
<td>10.51</td>
</tr>
<tr>
<td>Normalized average</td>
<td>1.00</td>
<td>0.78</td>
<td>0.88</td>
<td>0.84</td>
</tr>
<tr>
<td>Average turns/second</td>
<td>0.091</td>
<td>0.064</td>
<td>0.080</td>
<td>0.061</td>
</tr>
<tr>
<td>Normalized average</td>
<td>1.00</td>
<td>1.03</td>
<td>1.12</td>
<td>0.91</td>
</tr>
</tbody>
</table>

Note. FfF = face to face; HHD = hand-held device; HMD = head-mounted device.

Table 11: Average Percentage of Deictic Phrases Used

<table>
<thead>
<tr>
<th></th>
<th>FfF</th>
<th>LCD</th>
<th>HHD</th>
<th>HMD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average %</td>
<td>33.1</td>
<td>36.8</td>
<td>32.9</td>
<td>35.5</td>
</tr>
<tr>
<td>SD</td>
<td>17.1</td>
<td>9.1</td>
<td>9.6</td>
<td>10.2</td>
</tr>
</tbody>
</table>

Note. FfF = face to face; HHD = hand-held device; HMD = head-mounted device.

Questions asked. We counted the number of user turns that contained one or more questions and calculated a percentage of turns with questions in them. As the results in Table 12 show, there was no difference in percentage of questions across conditions, $H(3) = 0.467, p = .92$; however, as in Experiment 1, many of the questions asked in the AR conditions related to building identification. The percentage of questions asked regarding building identification is shown in Table 13. There was a significant difference in these results, $H(3) = 12.5, p < .01$. We used a Mann–Whitney test to make comparisons among AR conditions and found that there were significantly fewer questions regarding building identification in the HHD condition than in the LCD condition ($Z = -2.028, N = 13, p < 0.05$) and a near-significant difference between the HHD and HMD conditions ($Z = -1.874, N = 13, p = .064$). There was no difference between the LCD condition and the HMD condition ($Z = -0.487, N = 13, p = .65$). Thus, participants asked fewer building-identification questions in the HHD condition than the other two AR conditions.

Table 12: Proportion of All Questions Asked

<table>
<thead>
<tr>
<th></th>
<th>FfF</th>
<th>LCD</th>
<th>HHD</th>
<th>HMD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of questions</td>
<td>24.9</td>
<td>26.8</td>
<td>24.5</td>
<td>26.1</td>
</tr>
<tr>
<td>SD</td>
<td>9.2</td>
<td>9.9</td>
<td>8.9</td>
<td>10.4</td>
</tr>
</tbody>
</table>

Note. FfF = face to face; HHD = hand-held device; HMD = head-mounted device.

Table 13: Proportion of Building Identification Questions Asked

<table>
<thead>
<tr>
<th></th>
<th>FfF</th>
<th>LCD</th>
<th>HHD</th>
<th>HMD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of questions</td>
<td>4.0</td>
<td>10.3</td>
<td>6.5</td>
<td>12.3</td>
</tr>
<tr>
<td>SD</td>
<td>4.0</td>
<td>5.3</td>
<td>6.0</td>
<td>8.1</td>
</tr>
</tbody>
</table>

Note. FfF = face to face; HHD = hand-held device; HMD = head-mounted device.
Nonverbal communication patterns. We categorized the number of gestures made in each condition into one of four types:

Point: a pointing gesture.
Pick: a gesture that involved picking or moving an object.
Collab: a collaborative gesture, such as passing an object between people.
Other: any other gesture.

The percentage breakdowns for each type of gesture made are shown in Figure 15. As one can see, there was a smaller percentage of Pick gestures made in the FfF condition than in the AR conditions. There was a significant difference across conditions in the percentage of Pick gestures, $H(3) = 8.47, p < .05$. There was a larger percentage of Point gestures in the FfF condition than in the LCD condition ($Z = -2.49, N = 26, p < .05$) and the HHD condition ($Z = -2.09, N = 26, p < .05$), but not in the HMD condition ($Z = -0.74, N = 26, p = .46$). There was also a significant difference in the percentage of Point gestures across conditions, $H(3) = 8.91, p < .05$. Using the Mann–Whitney test, we found a significant difference between amount of pointing in the FfF and LCD conditions ($Z = -2.85, N = 26, p < .01$) and the FfF and HHD conditions ($Z = -2.03, N = 26, p < .05$), although not between the FfF and HMD conditions ($Z = -0.88, N = 26, p = .38$). Participants used proportionately fewer pointing gestures in the conditions in which they had fewer hands free.

We also counted the amount of time that the participants spent standing. Although participants started the experiment in a seated position, they were free to stand and lean over the table during the task. The average percentage of time that participants were standing during each of the conditions is shown in Table 14. As can be seen, participants almost never stood in the FfF condition but were on their feet over 50% of the time in the LCD condition. There was a highly significant difference between these results, $H(3) = 29.7, p < .0001$. We used a Mann–Whitney test to compare among AR conditions and found a significant difference between the

![Figure 15](image-url) Percentage breakdown of gestures. FfF = face to face; HHD = hand-held device; HMD = head-mounted device.
Table 14: Percentage of Time Participants Spent Standing

<table>
<thead>
<tr>
<th></th>
<th>FfF</th>
<th>LCD</th>
<th>HHD</th>
<th>HMD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of time</td>
<td>0.07</td>
<td>52.3</td>
<td>40.4</td>
<td>24.3</td>
</tr>
<tr>
<td>spent standing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. FfF = face to face; HHD = hand-held device; HMD = head-mounted device.

amount of time participants spent standing in the LCD condition and HMD condition ($Z = -2.66, N = 24, p < .01$) but not in the HHD condition and HMD condition ($Z = -1.25, N = 24, p = .212$) or the LCD condition and HHD condition ($Z = -1.09, N = 24, p = .276$). Participants spent more time standing in the LCD case to see more of the task space in the small field of view display.

Subjective Measures

After completing each condition, participants filled out the same survey that was used in Experiment 1. They were asked how easy it was to pick and move buildings by answering the following questions:

How easily could you pick up the buildings? (1 = not very easily, 7 = very easily)
How easily could you move the buildings? (1 = not very easily, 7 = very easily)

The average responses are shown in Figure 16. As one can see, participants felt it was easiest to pick and move objects in the FfF condition. There is a significant difference across conditions for both questions. For the picking results, an ANOVA yielded $F(3, 68) = 6.23, p < .01$. For the moving results, $F(3, 68) = 3.79, p < .05$. It is interesting that even though the LCD panels required the use of at least one hand, participants did not perceive them to be worse than the other displays for supporting interaction with the buildings. Using a paired $t$ test, we also found a significant difference between the average FfF picking value and that for the LCD condition, $t(25) = 3.64, p < .001$. Similarly, a paired $t$ test also revealed a significant difference between the average FfF moving value and that for the LCD condition, $t(25) = 2.35, p < .05$.

![Figure 16 Ease of interaction with the buildings. FfF = face to face; HHD = hand-held device; HMD = head-mounted device.](image-url)
FIGURE 17 Ease of collaboration. FfF = face to face; HHD = hand-held device; HMD = head-mounted device.

To evaluate their opinion of the ease of collaboration, we asked participants the following questions:

How easily could you work with the other person? (1 = not very easily, 7 = very easily)
How easily could you understand the other person? (1 = not very easily, 7 = very easily)

The average responses are shown in Figure 17. Participants felt that it was significantly easier to work together and understand each other in the FfF condition. For the working-together results, an ANOVA yielded $F(3, 92) = 8.53, p < .0001$. For the moving results, $F(3, 92) = 5.01, p < .01$. Across the AR display conditions, participants reported that it was easiest to work with the other person and to understand him or her in the HHD condition, although these results were not significantly different from those in the other display conditions.

Finally, we used the following questions to ask participants how easy it was to perceive nonverbal cues:

How easily could you tell where the other person was looking? (1 = not very easily, 7 = very easily)
How easily could you tell where the other person was pointing? (1 = not very easily, 7 = very easily)

The average responses are shown in Figure 18. There was a highly significant difference among the four conditions for looking, $F(3, 92) = 9.98, p < .0001$; there was also a significant difference between just the AR display conditions, $F(2, 69) = 3.81, p < .05$. As expected, the more the display covered the users’ eyes, the more difficult it was for them to tell where their partner was looking. There was also a highly significant difference among conditions for pointing, $F(3, 92) = 15.71, p < 1 \times 10^{-7}$, and a nearly significant difference between just the AR display conditions, $F(2, 69) = 2.54, p = .08$. Participants felt that the display conditions that allowed them to naturally view the real world allowed them to most easily see where their partner was pointing.
These results were confirmed when participants were asked to rank the following statements on the degree to which they agreed or disagreed with them (1 = disagree, 7 = agree):

I could tell when my partner was concentrating.
I was very aware of the presence of my partner.
I could tell when my partner was looking at me.
I could tell when my partner was listening to me.

The average rankings for each condition are shown in Figure 19. As one can be see, the FfF condition was again ranked highest for each of the questions. The ANOVA results for each of the rankings across all the conditions are shown in Table 15. There was a significant difference in average response for each of the statements across all conditions. When just the AR display conditions are considered, there was a significant difference in participants’ awareness of the presence of their partners and their ability to tell when their partner was looking at them. In response to these statements, the average score for the HMD was much lower than those in the other AR display conditions and, on average, the HMD condition produced the lowest scores across all statements. It is clear that the participants felt that the display form factor affected their ability to perceive their partner and nonverbal conversational cues.
Table 15: Analysis of Variance Scores for Presence Question Ratings

<table>
<thead>
<tr>
<th>Statement</th>
<th>All Four Conditions</th>
<th>AR Display Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$F^a$</td>
<td>$p$</td>
</tr>
<tr>
<td>Concentrating</td>
<td>3.60</td>
<td>&lt; .05</td>
</tr>
<tr>
<td>Presence</td>
<td>9.58</td>
<td>$&lt; 1 \times 10^{-4}$</td>
</tr>
<tr>
<td>Looking</td>
<td>12.82</td>
<td>$&lt; 1 \times 10^{-6}$</td>
</tr>
<tr>
<td>Listening</td>
<td>5.11</td>
<td>&lt; .01</td>
</tr>
</tbody>
</table>

Note. AR = augmented reality.

$a df = 3, 92. b df = 2, 69.$

Participant Comments

Participants felt that the biggest difference among the AR display conditions was in the ability to see their partner. In the majority of interviews, (10 of 12), participants commented on how easy it was to look at their partner in the LCD condition. They did not need to move anything away from their face, and they could even see the LCD screen at the same time as when they looked at their partner in the real world. In 3 cases, participants specifically commented on how their communication was not blocked by anything covering their face. Conversely, in 10 of 12 interviews, participants commented on how the HMD condition made their partner seem distant or made them feel detached from the real world. As one participant commented about the HMD display condition, “When I was thinking inside the box, he wasn’t in the box.” In contrast, in only 2 interviews people said that they felt confined by the HHD, whereas in another 2 cases people commented on the peripheral cues the HHD provided. These differences between the display conditions were summed up by 1 participant: “The LCD provided no barriers to understanding. The handheld display sometimes makes it hard to see where he was pointing. The HMD display made him seem distant.”

Of course, there was almost an inverse relation in participants’ ability to see their partners and their ability to see the virtual content. People commented on the narrow field of view of the LCD panel (four cases) and how the virtual models looked flat (two cases). However, for the HMD case, in two interviews people talked about how much they liked the sense of immersion they got and how they could easily see their partner at the same time as the virtual buildings.

Participants felt that the physical form factor of the different displays affected their interaction. Although most felt that using the HMD made it very difficult to collaborate with their partner, in 7 interviews participants commented on how much they enjoyed having both hands free to move the models. In 7 of the interviews participants also mentioned that the HMD had the advantage that it was always on their face, so in order to see different portions of the task space they just had to turn their head. In contrast, some participants (5 of 12 interviews) mentioned that it was awkward to keep on remembering to hold the HHD to one’s face and difficult to move the display and one’s head at the same time so that they could keep viewing the AR scene. Two participants felt that the HHD was easiest to use, whereas in 3 interviews participants mentioned that the HHD felt like a mag-
nifying glass or a pair of opera glasses. Finally, in 4 interviews participants mentioned how they could point the LCD screen at what they were interested in, whereas almost nobody mentioned that the LCD display often required them to use two hands or that it was more cumbersome (2 interviews).

One strong point of the HHD was the ease with which people could switch between the real and augmented views. In 5 of 12 interviews participants commented on how easy it was to remove the HHD from their face and swap back to the real world. For example, 1 participant wrote “The handheld display was easiest to focus and still switch back and forth between the virtual reality image and real sight.” Similarly, in 4 interviews people talked about how they liked the HHD because they could talk to each other face to face and then just use the display for quick glances at the virtual scene. As 1 participant said: “We spent most of our time talking to each other normally, using the device only when necessary.”

4.4 Discussion

In this second experiment we explored the effect of different types of AR displays on communication behaviors. In general, the AR conditions produced results that were significantly different from those of the face-to-face conditions; however, the face-to-face condition always occurred first, followed by the AR conditions in a counterbalanced order. This may have biased some of the face-to-face results.

When considering the AR results, we noted that there was little difference in speech behaviors among the LCD, HHD, and HMD conditions. Participants spoke the same number of words per turn, the same number of turns per second, and the same amount of deictic phrases. However, with an HHD they asked fewer building-identification questions than with an LCD panel or an HMD. This is because the LCD panel affords only a small field of view, whereas the HMD removes the user from seeing the world naturally. Both made it difficult for participants to see the AR buildings or the nonverbal cues of their partners. In contrast, the HHD provided the same field of view as the HMD and allowed the participants to easily see the real world.

When considering nonverbal communication, we noted a significant difference in the amount of picking and pointing gestures made between the FfF condition and some of the AR conditions; however, there was no difference among AR conditions. Although participants could use both hands in the HMD condition and only one in the other conditions, the form factor of the AR displays did not greatly affect the ease with which participants could make nonverbal gestures. The differences in AR displays did cause the users to stand up so they could compensate for the field-of-view limitations. Users of the LCD display spent on average over half their time standing, and even when using an HMD participants felt it was necessary to stand for a portion of the time so they could get a better view of the task space.

Participants felt strongly about how the different AR displays allowed them to perceive nonverbal cues and the presence of their partner. The end-of-task surveys indicated that participants felt that the differences in display affordances did not affect their ability to interact with the AR content or work together but did affect how well they could tell where their partner was looking or pointing. The results
matched the amount of peripheral cues supported by the displays; the more the participant could see an unmediated view of the real world (as with the LCD panel), the more he or she felt aware of his or her partner.

5. LESSONS FOR DESIGNING COLLABORATIVE AR INTERFACES

From these two experiments, we can propose several rules of thumb for designing AR interfaces for face-to-face collaboration.

First, interaction with the AR content should be based around tangible interaction techniques. By having our virtual models appear attached to physical objects, participants could manipulate them as easily as the real models.

Second, a large part of communication is about perception. Mediating the view of the real world will affect a user’s ability to perceive his or her collaborator and the transmission of nonverbal cues. This may change the language that is spoken as users try to compensate for limited visual perception. For example, in our experiments, participants asked more questions to identify the virtual buildings.

Finally, until HMDs have been developed with a wide field of view and high resolution (and if there is no compelling reason to use an HMD), a display form factor should be chosen that enables a user to easily perceive the unmediated real world. Our participants felt that the LCD and HHD were both more useful for collaboration than the HMD. Note that we used a video see-through HMD; an optical see-through HMD would probably produce different results, because users can see the real world directly through the HMD regardless of its resolution.

6. CONCLUSIONS

In these two experiments we began to explore communication and task behaviors with a face-to-face collaborative AR interface. We found that ease of interaction and some task performance can be unchanged when using different AR displays, but the AR interface does affect the perception of communication cues.

However, a large amount of work still needs to be done to compare how collaboration with these types of interfaces is different from more traditional computer-supported tools. Our results have been collected from only one type of task: object-based collaboration. It is likely that results for other tasks will differ. In particular, we expect that conversational tasks, such as negotiation, will be even more susceptible to differences in the AR interface. Past work on teleconferencing has found that negotiation and conversational tasks are more sensitive to differences between communication media. Experiments are needed in which object manipulation is just a small part of the task at hand.

In our work not only did we reduce perceptual cues by removing peripheral vision, but also the AR displays showed a biocular rather than stereoscopic view of the real world. Because this task involved close-in object manipulation, the lack of stereo would have made it more difficult to pick and move objects. This may also explain why participants across the table appeared to be flat and more distant. We
need to conduct another experiment in which we compare display properties such as stereo versus nonstereo and video see-through versus optical see-through.

Finally, we intend to explore how AR interfaces can be combined with projective technology and normal desktop displays to enhance face-to-face collaboration. One of the greatest benefits of AR interfaces is that they can be seamlessly combined with other, more traditional interface technology and existing workplace displays. However, there are few researchers who have presented interfaces that combine AR techniques with other approaches, and there is a large number of interface questions that need to be answered before truly useful interfaces can be developed.

REFERENCES


