Communication Behaviors of Co-located Users in Collaborative AR Interfaces

K. Kiyokawa[1], M. Billinghurst[2], S. E. Hayes[2], A. Gupta[2], Y. Sannohe[2], and H. Kato[3]
[1]*Communications Research Laboratory, Japan*
[2]*Human Interface Technology Laboratory, University of Washington*
[3]*Hiroshima City University, Japan*
kiyo@crl.go.jp, {grof, beephree, arnab, yuki}@hitl.washington.edu, kato@sys.im.hiroshima-cu.ac.jp

Abstract

We conducted two experiments comparing communication behaviors of co-located users in collaborative augmented reality (AR) interfaces. In the first experiment, we compared optical, stereo- and mono-video, and immersive head mounted displays (HMDs) using a target identification task. It was found that differences in the real world visibility severely affect communication behaviors. The optical see-through case produced the best results with the least extra communication needed. Generally, the more difficult it was to use non-verbal communication cues, the more people resorted to speech cues to compensate. In the second experiment, we compared three different combinations of task and communication spaces using a 2D icon designing task with optical see-through HMDs. It was found that the spatial relationship between the task and communication spaces also severely affected communication behaviors. Placing the task space between the subjects produced the most active behaviors in terms of initiatory body languages and utterances with least miscommunications.

1. Introduction

In face-to-face collaboration, a wide variety of verbal and non-verbal communication cues are used to establish shared understanding. The various cues presented in face-to-face collaboration can be organized according to the different communication channels used (see table 1).

Table 1: Unmediated communication cues.

Audio	Visual	Environmental
Speech	Gaze	Object Manipulation
Paralinguistic	Gesture	Writing/Drawing
Paraverbals	Face Expression	Spatial Relationships
Prosodics	Body Position	Object Presence
Intonation		

In technologically mediated collaboration, each of these cues may or may not be transmitted between the collaborators. The ability of communication media to support different communication cues is related to the affordance of the media [5]. For example, although face-to-face and video conferencing both transmit visual communication cues, it is hard for users to separate visual cues from the background in video conferencing [6]. Even co-located collaborators can also suffer from reduced communication cues. For example, the attention of collaborators in front of a screen is often focused on the shared screen, making it difficult to exchange visual communication cues.

In contrast, Augmented Reality (AR) technology can be used to develop new types of collaborative interfaces, which allow users to see each other at the same time as virtual objects. For example, AR2 Hockey allows two users to play a version of air hockey where the puck is a virtual object [11], while the AR Conferencing application superimposes live video of users over the real world [2].

In a co-located AR interfaces, users are expected to exhibit the same communication behaviors as in unmediated face-to-face collaboration. Previous work comparing collaborative AR interfaces to immersive virtual environments has found that users perceived gaze cues better [8], and performed more quickly on a simple spatial task [1] in an AR setting. However, there has been no research conducted on communication behaviors in collaborative AR interfaces. By understanding how the affordances of AR interfaces affect communication, better AR interfaces can be built.

In this paper we report on two experiments that study communication behaviors in face-to-face AR collaboration. In the first we research how different display affordances may change the nature of the collaboration. In the second, we explore how the location of the AR task space affects communication. Before describing our work in these areas we first describe the metrics we use to measure the impact of AR technology on communication.

2. Communication Metrics

To be able to understand the impact of AR technologies we need to arrive at methods for evaluating collaborative interfaces. Researchers such as Monk et al. [9] argue that a multidimensional approach is needed, so we used a variety of *performance*, *process* and *subjective* measures.

2.1. Performance Measures

Performance measures are those that measure a task outcome, such as task completion time. However, performance measures alone are inadequate. In many telecommunication experiments there were no performance differences between mediated conditions [15]. In our work, we measure performance outcomes, but only to provide a gross communication measure.

Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR'02)
0-7695-1781-1/02 $17.00 © 2002 IEEE

2.2. Process Measures

Process measures are objective measures that capture the process of collaboration and are extracted from transcriptions of video recordings and notes made during the collaborative task. Measures that have been found to be significantly different across technology conditions include:

- Frequency of conversational turns [3]
- Incidence of overlapping speech [12]
- Dialogue structure [4]
- Turn completions [14]
- Backchannels [10]

Gesture and non-verbal behaviors can also be analyzed for characteristic features. In our work we transcribe videotapes for both speech and gesture features.

2.3. Subjective Measures

Subjective measures are based entirely on the users' opinion of the collaborative interface. The typical method for gathering subjective data is to have users fill out a survey questionnaire. Daly-Jones provides a set of questions that have been found to be sensitive to the differences in mediating technology [3]. These refer to interpersonal awareness, and ease of communication. Typical questions include:

I was very aware of my conversational partner.
I could tell when my partner was concentrating

These are usually answered on a scale of *Disagree* to *Agree*. We used a modified version of the Daly-Jones survey questions and post-experiment interviews.

3. Expt. 1: Display Affordances

In this experiment we compared the communication behaviors between three typical AR configurations and a virtual reality configuration. We designed a simple task in which a pair of subjects with HMDs had to collaboratively identify a single target virtual cube among many similar cubes. In this task, two subjects sat on opposite sides of a black desk (W108cm, D82cm, H75cm) and wore an Olympus Mediamask HMD (figure 1). Subjects completed the object identification task for each of the following four conditions:

***Optical case**:* The HMDs were in see-through mode and the virtual cubes were superimposed on the real world (fig.3).
Stereo case: The HMDs were in a video see-through mode. Subjects could see a stereo live video image with stereoscopic virtual objects superimposed on it.
Mono case: Subjects saw a monoscopic live video image captured through the left camera mounted on the HMD (figure 4). Virtual objects were still seen stereoscopically.
VR case: A stereoscopic computer-generated scene composed of cubes and the partner's virtual head was shown (figure 5). The real world was invisible.

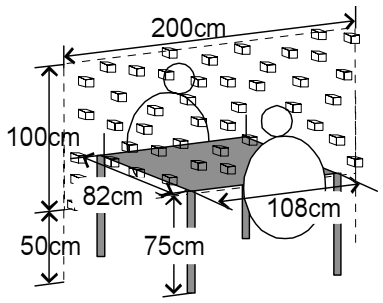

Figure 1: Layout of the experiment.

In each condition, both subjects wore an Olympus

Figure 2: Starting block.

Figure 3: Optical case.

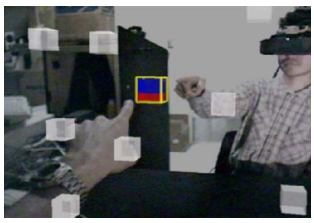

Figure 4: Mono case.

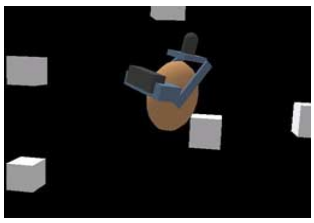

Figure 5: VR case.

Mediamask HMD, which had a 640x480 pixel resolution and horizontal and vertical fields of view of 60 and 34 degrees, respectively. Two cameras (Toshiba IM-43H) and an optical 3D sensor (3rdTech HiBall 3000) were attached to each HMD. The HMDs were see-through but a black cloth was attached to the front for the video-see through and VR conditions. Each HMD was connected to a Pentium III PC with a GeForce2 MX 32MB graphics card and two video-capture cards. The CPU speeds were 1.0GHz and 700MHz, fast enough to generate 30 fps graphics.

3.1. Expected Outcome

By considering the affordances of each condition we can predict the communication behaviors that the subjects should exhibit (see table 2).

Table 2: Condition affordances

Condition	Non-verbal cues	Stereo view	Partner
Optical	Gaze + gesture	Yes	Real
Stereo	Gaze + gesture	Yes	Video
Mono	Gaze + gesture	No	Video
VR	Gaze	Yes	Graphical

Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR'02)
0-7695-1781-1/02 $17.00 © 2002 IEEE

The Optical condition gives the subjects a real view of their partner, enabling them to see normal unmediated communication cues. Subjects should collaborate most naturally and complete the task faster in this condition. In Stereo and Mono conditions, the subject's partner is seen as a video image. This may make it difficult for them to see different communication cues, and may require them to use more speech and gestures. In the Mono condition, the virtual objects may blend into video, making it even more difficult to distinguish communication cues. In the VR condition subjects are given minimal communication cues, which may increase the performance time.

3.2. Procedure

The experimental task involved two subjects, a trainer and a trainee, both trying to look at the same virtual object. To start with, a single virtual cube, 5cm on each side, appeared 15cm above the center of the desk (figure 2). The cube appeared red to the trainer and white to the trainee. When both of them looked within 10 degrees of the cube for two seconds, it disappeared and the trial began. Subjects could now see fifty cubes randomly distributed, at least 10cm from one another, on a vertical virtual plane between them (figure 1). All cubes appeared white to the trainee, but one of them was red in the trainer's view.

The trainer's goal was to find the red cube and have the trainee see it. The trainee couldn't visually differentiate the target, so they needed to communicate, using any communication cues. In the trainer's view, the target's color changed from red to blue when the two of them looked at it, while it remained white to the trainee. The task was completed when both of them kept the target within 10 degrees of their centers of view for two seconds. When a trial was finished, the single starting block appeared once again and the roles of the trainer and trainee were exchanged. Each pair played thirty rounds for each of the four conditions for a total of 120 trials.

This was a simple selection task that encouraged subjects to use gaze, speech and gestural cues to identify an object. The participants were forced to collaborate, so the quicker they could perceive communication cues, the quicker they should have been able to finish the task.

3.3. Results

Twelve pairs of subjects were used, ranging in age from 19 to 45 years. Each round typically lasted only five to fifteen seconds, so we matched two persons of the same gender who were strangers to each other to avoid desultory conversation. After the rules were read, every pair first practiced the Optical case for ten times then experienced the conditions in a counterbalanced order.

3.3.1. Performance Measures

As expected, subjects completed the task quickest in the Optical case and slowest in the fully immersive VR mode.

Task completion times

Figure 6 shows the average task completion times, which we divided into two parts. The *search* time was the time it took the trainer to locate the target within 10 degrees. The *direction* time was the time taken to direct the trainee to the target position. A one-factor ANOVA gave no significant difference in the search stage ($F(3,1436)=1.52$, $p=0.21$). However, the direction time was by far the shortest in the Optical case ($F(3,1436)=19.14$, $p<1x10^{-11}$).

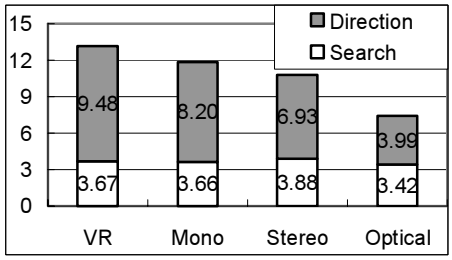

Figure 6: Averages of search and direction times (sec).

Amount of trainees' looking away from target

We counted numbers of occurrences when a trainee looked away from the target (figure 7). A trainee would look away from the target more often to see her trainer's gaze and gesture if she was having difficulty seeing them. There was no significant difference in the searching stage ($F(3,1436)=1.42$, $p=0.24$). However, for the direction stage, trainees looked away least often in Optical case (e.g. $T(450)=5.60$, $p(T<t)<1x10^{-7}$ when compared to Stereo). There was also a significant difference between the VR and Stereo cases ($T(511)=2.38$, $p(T<t)=0.018$).

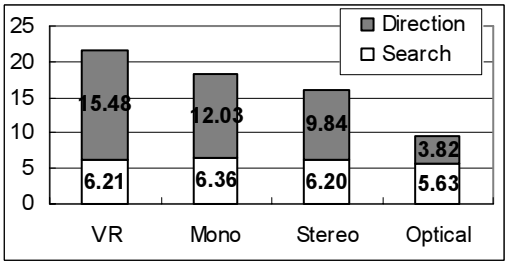

Figure 7: Average number of trainees' looking away

Findings in performance measures

The results show that the differences in the perception of communication cues produce differences in direction times. Subjects were best able to see non-verbal cues in the Optical case, which produced the fastest average task time and the least number of miscommunications.

3.3.2. Process Measures

We videotaped each experimental session and counted the pointing gestures and spoken phrases.

Pointing gestures

Making pointing gestures more than once in a condition could be considered as a signal of misunderstanding. Figure 8 shows the average numbers of *extra* pointing gestures

Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR'02)
0-7695-1781-1/02 $17.00 © 2002 IEEE

exhibited in those cases where subjects used at least one pointing gestures. A two-way t-test found trainers used significantly fewer pointing gestures in the Optical case compared to the Mono case ($T(315)=2.18$, $p(T<t)=0.03$), and nearly significant ($p<0.20$) when compared to VR ($T(62)=1.57$, $p(T<t)=0.12$) and the Stereo cases ($T(458)=1.42$, $p(T<t)= 0.16$). Note that 19 of 24 subjects used pointing gestures in the VR case at least once, implying that pointing is a natural behavior.

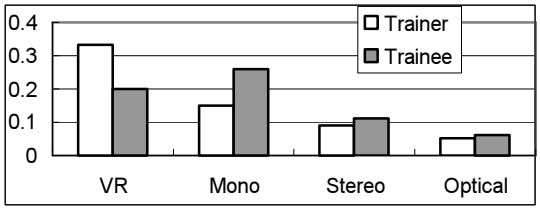

Figure 8: Average number of extra pointing gestures.

Deictic phrases

Deictic phrases contain the words "this", "that", "here" or "there" and cannot be fully understood by speech alone. Figure 9 shows the average number of deictic phrases spoken per trial. There was no significant difference in the number of deictic phrases spoken by trainers ($F(3,1436)= 1.98$, $p=0.12$). On the other hand, trainees spoke the fewest deictic phrases in the Optical condition. There was a highly significant difference among the four conditions ($F(3,1436)=9.77$, $p<1x10^{-5}$), but no difference between the Mono, Stereo and VR cases ($F(2,1077)=0.28$, $p=0.76$).

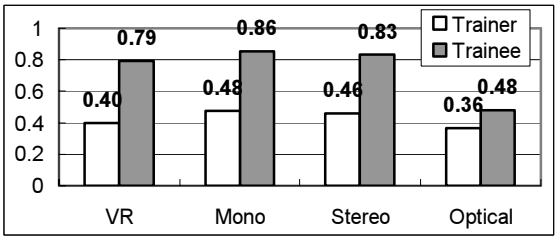

Figure 9: Average number of deictic phrases.

Positional phrases

Positional phrases were those that included words like "on the edge" and "third row from top". Figure 10 shows the average numbers of positional phrases spoken per trial. Trainees used positional phrases most often in the Mono condition ($F(2,1077)=1.20$, $p=0.30$, no significant for the rest of the three cases, while $F(3,1436)=12.0$, $p<1x10^{-7}$, highly significant for four cases). In this case it was harder for a trainer to point to the target precisely in 3D space so the trainee may have needed to use positional phrases to ensure its position. Trainers used them least often in the Optical case ($F(2,1077)=1.21$, $p=0.30$, no significant for the rest of the three cases, while $F(3,1436)=9.29$,$p<1x10^{-5}$, highly significant for four cases).

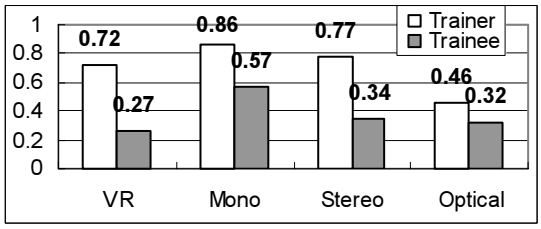

Figure 10: Averages of number of positional phrases.

Findings in process measures

Subjects used least extra pointing gestures in the Optical case. Deictic phrases were mainly used when a speaker specified a position that she thought the listener could see. Positional phrases were used when the listener couldn't see the position, compensating for degraded visual cues.

3.3.3. Subjective Measures

Questionnaires

After each condition, subjects filled out a survey (table 3). All of the questions were answered on a scale of one ("Disagree") to seven ("Agree"). Subjects preferred the Optical case the most. They felt the VR case was easy to see and to complete the task compared to other cases.

Table 3: Subject Questionnaire for Exp 1.

Questions about visibility
Q1-1 The view of the real world was very natural.
Questions for trainers
Q1-2 It was very easy to see my trainee.
Q1-3 I could very easily tell where my trainee was looking.
Q1-4 I could very easily tell where my trainee was pointing.
Questions for trainees
Q1-5 It was very easy to see my trainer.
Q1-6 I could very easily tell where my trainer was looking.
Q1-7 I could very easily tell where my trainer was pointing.
Questions about overall preferences
Q1-8 It was very easy to perform the task.
Q1-9 I like the condition very much.

Subjects felt that the Optical case provided more natural view than Mono and Stereo (e.g., $T(45)=3.22$, $p(T<t)=0.002$ for Stereo) (Q1-1, figure 11). Surprisingly, there was absolutely no significant difference, between Mono and Stereo cases ($T(46)=0.08$, $p(T<t)=0.94$).

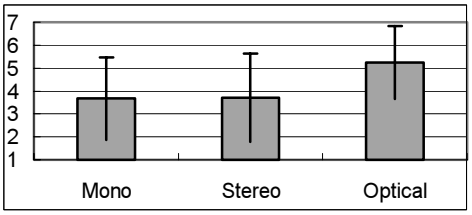

Figure 11: Real world visibility (1=disagreed,7=agreed).

Subjects felt it easier to see the partners in Optical case than in Mono and VR cases (Q1-2 & Q1-5, figure 12, e.g., $T(37)=2.94$, $p(T<t)=0.006$ for Q1-5 for Mono).

Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR'02)
0-7695-1781-1/02 $17.00 © 2002 IEEE

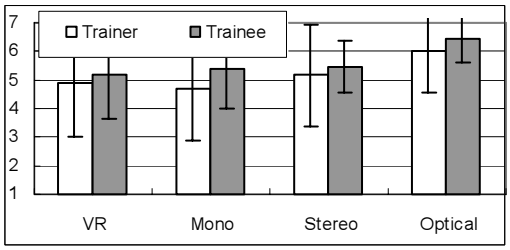

Figure 12: Ease of seeing partner.

The Optical case was the easiest to tell where the partners were looking (Q1-3 & Q1-6, figure 13). No significant difference was found across the rest (e.g. $F(2,67)=0.16$, $p=0.85$, while $F(3,89)=3.04$, $p=0.03$ for all cases in Q1-3).

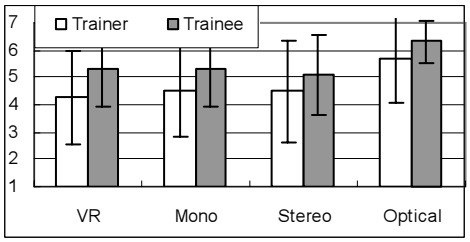

Figure 13: Ease of telling of partner's looking position.

The Optical case was also easier to tell where the partners were pointing (Q1-4&Q1-7, figure 14) than Mono ($T(45)=3.06$, $p(T<t)=0.004$), but the difference was less clear when compared to Stereo ($T(45)=1.75$, $p(T<t)= 0.087$).

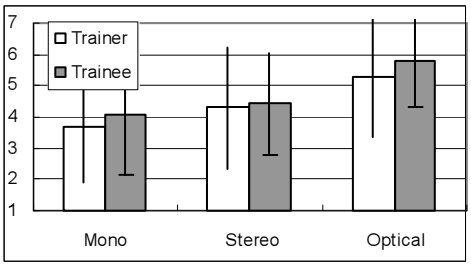

Fig. 14: Ease of telling of partner's pointing position.

Figure 15 shows the results of the overall impressions (Q1-8 & Q1-9). The Optical case was significantly more favored than the Mono and Stereo cases, but the difference was less significant when compared to VR case. Also, there was weak indication that Stereo was slightly more favored than Mono (e.g., $T(45)=1.46$, $p(T<t)=0.15$ for Q1-8). Mono was even less favored than VR (e.g., $T(45)=1.53$, $p(T<t)=0.13$ for Q1-8).

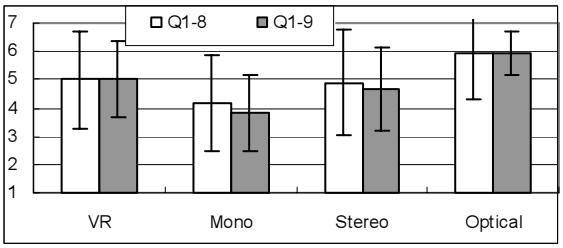

Figure 15: Preferences in performance and liking.

Extraneous comments and Observation

We transcribed all extraneous comments in the experiment. In the VR case, subjects immediately noticed that they couldn't see their hands and some felt discomfort (*"I need my hand! I cannot talk without it"*). Then they tried positional phrases (*"We have to find out a way to say the row and columns"*). After that, some started to use their face direction (*"It is easier when I look at your face"*).

In the Mono case subjects commented on the unnatural view (*"Whoa, it's a very distorted view"*). Trainers first tried to point but it was inaccurate and difficult without depth perception (*"Your finger is about four inches off"*, *"This is pretty hard"*).

In the Stereo case, the stereoscopic view seemed rather natural (*"This is really good"*). Compared to the Mono case, it was less difficult to point to the target (*"It actually works when you point"*). However, their visual spaces were not exactly the same mainly due to the camera offsets (*"Oh, my visual space and yours are not the same"*), causing pointing gestures inaccurate sometimes.

In the Optical case, subjects mostly talked about how easy the condition was (*"This is easy"*) and how well their visual spaces were registered (*"Cool, your finger's exactly on the cube"*).

Findings in subjective measures

Subjects felt the Optical case was most natural, and easier to tell where their partner was looking or pointing, when compared to the video see-through cases. Subjects felt few differences between the monocular and binocular video see-through conditions, though the monocular case was even less favored than the immersive VR condition.

3.4. Discussion

The Optical case produced the fastest average time and the least number of miscommunications. Generally, the more difficult it was to use non-verbal communication cues, the more subjects resorted to speech cues to compensate. Subjects also felt the Optical case was the easiest to tell where their partner was looking or pointing. When the virtual scene is well registered to the real world, optical see-through approach is the best in order for natural and smooth communication. Note that subjects only practiced in the Optical condition, which might have biased the results. Nevertheless, the Optical condition seems so superior that these results are unlikely to have changed with additional practice in the other conditions.

The Stereo video see-through interface was more favored by subjects than the Mono condition. The Stereo condition also reduced the need for extra pointing gestures and positional phrases. However, the difference was less clear than expected. With a task that requires direct manipulation, the difference may increase. To evaluate characteristics of video see-through approaches in more detail, we will also need to use a camera with a smaller viewpoint offset [13].

Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR'02)
0-7695-1781-1/02 $17.00 © 2002 IEEE

4. Expt. 2: Communication Space Separation

In this second experiment, we explored how the separation between the task space and the communication space affected collaboration, using the same hardware as in the first experiment. We designed a two-dimensional icon-drawing task in which a pair of subjects had to collaboratively draw icons for given design themes. Subjects sat opposite each other wearing the Mediamask HMDs in the optical see-through mode. They were able to see stereoscopic virtual objects including a virtual 31x31 grid. They completed the task in each of the following three conditions (see figure 16):

Wall case: The virtual grid appeared on a black wall off to the subject's right or left hand side. Subjects needed to turn their heads to see the collaborator. Compared to other conditions, they saw the grid from similar viewpoints.
Table case: The grid appeared on the table between the subjects. Subjects needed to lift their heads to see their collaborator. Depending on the design, one of the subjects might have to see the grid upside down.
Floating case: The virtual grid appeared floating in space between the two subjects. Subjects were able to see their partners at all times while completing the task. However, they were viewing the grid from opposite sides.

4.1. Expected Outcome

In each of the three cases, the space between subjects is used for sharing gaze, gesture, and non-verbal behaviors. In the Floating case, the task space is a subset of the communication space. In this case, they can see each other at the same time as the objects, and should exhibit most natural communication. On the other hand, in the Wall case, the subjects are focused on the wall and cannot easily see each other. Thus, they should exhibit the least natural communication. The Table case should come in-between, but subjects will be viewing the grid from the opposite sides, which may disturb them.

4.2. Procedure

Each of the subjects had a mouse and was able to draw on the grid. Holding the left mouse button down, they could drop white cubes onto the grid. The middle button changed the brush size, while the right one erased.

For each condition, subjects were told that they had five minutes to draw a 2D icon for film themes ("Action & Adventure", "Mystery & Suspense", and "Science Fiction & Fantasy" for 1st, 2nd, 3rd condition, respectively). They were given a theme name and then told to use the full five minutes to make the most "professional" icon possible. Subjects were forced to collaborate because there was only one drawing surface and only one subject's mouse was active at a time. A subject could draw for 30 seconds before mouse control swapped to the other person.

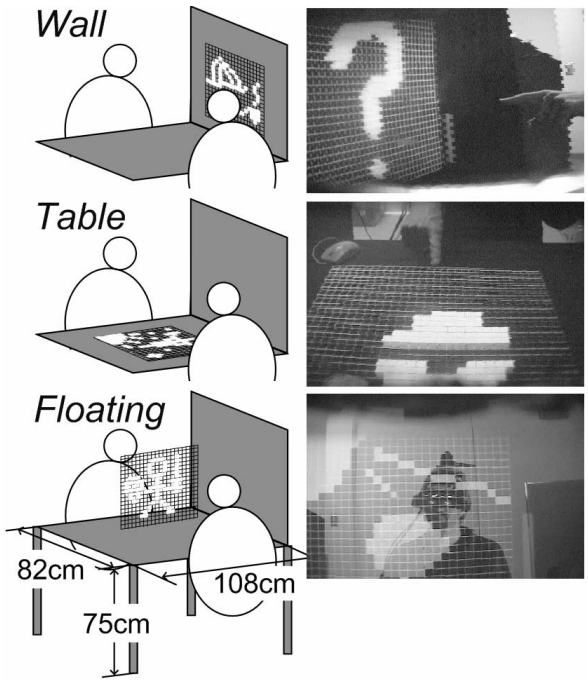

Figure 16: Conditions in Experiment Two.

4.3. Results

The subjects in this experiment were 13 pairs of people ranging in age from 19 to 50 years. They knew each other well to stimulate rich and smooth conversation. After the instructions were read, every pair first practiced all of the three interface conditions, and then experienced the three cases in a counterbalanced order.

4.3.1. Performance Measures

Design outcome

A variety of designs were created in the experiment (figure 17). There was no clear difference in design quality across the three conditions, but some common images did emerge for the same design themes. For example, for the "Mystery & Suspense" theme, subjects tended to draw question marks or magnifying lenses.

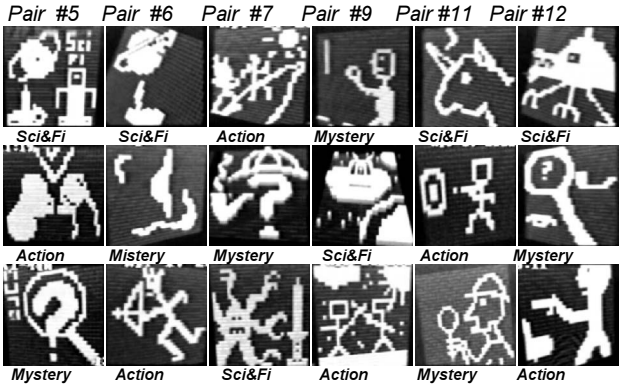

Figure 17: Examples of design outcome. From the top row, wall case, table case, and floating case.

Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR'02)
0-7695-1781-1/02 $17.00 © 2002 IEEE

Mouse motion

The average velocities of a mouse cursor on the grid in the three conditions were 5.1, 5.5, 5.2 (cm/s) for the Wall, Table and Floating cases, respectively. We found no significant difference here ($F(2,75)=0.78$, $p=0.46$).

Head motion

Figure 18 shows the average head angular velocity. A two-way t-test found subjects rotated their head 19% faster in the Floating case than in the Table case ($T(45)=2.56$, $p(T<t)=0.014$). This may be simply because subjects saw the grid squarely in the Floating case so that they needed to rotate their heads more often to see the whole task space.

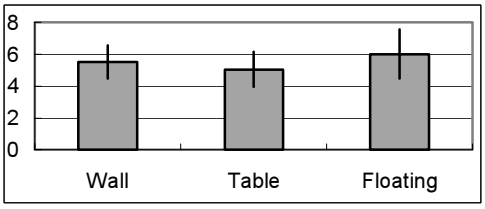

Figure 18: Averages of head angular velocity (deg/s).

Findings in performance measures

We see little difference in performance measures across conditions. Differences among conditions do not seem to affect the outcome of the design.

4.3.2. Process Measures

We videotaped each experimental session and counted different types of non-verbal and verbal behaviors subjects exhibited.

Exhibited Gestures

Figure 19 shows the average number of *pointing*, *design* (showing shape, size or movement), and *expressional* (e.g., shrugging for "I don't know") gestures exhibited by subjects per trial. Subjects tended to make more pointing gestures in the Floating case than in the Wall case ($T(49)=1.39$, $p(T<t)=0.17$). Subjects made design gestures 107% more in the Floating case than in the Wall case ($T(42)=2.57$, $p(T<t)=0.014$). Due to huge deviations, we don't see any significant differences in expressional gestures across conditions ($F(2,75)=0.15$, $p=0.86$).

Perceived Gestures

Figure 20 shows the average ratios of each of the three types of gestures captured by the partners' head mounted cameras to those made by subjects. These ratios should show ease of perceiving those gestures for a subject.

Almost 97% of pointing gestures are seen by the partner's camera in the Floating case, which is significantly more than that of the Wall case (82%, $T(20)=2.18$, $p(T<t)=0.04$), and near-significantly more than that of the Table case (90%, $T(26)=1.67$, $p(T<t)=0.11$). This is because arms were usually reached out toward the task space when subjects made pointing gestures.

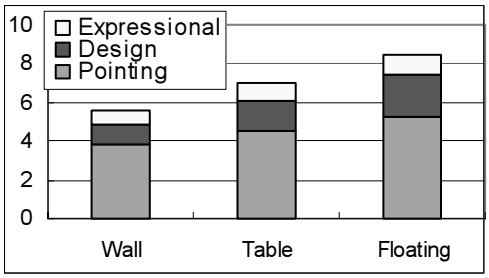

Figure 19: Average numbers of each type of gestures.

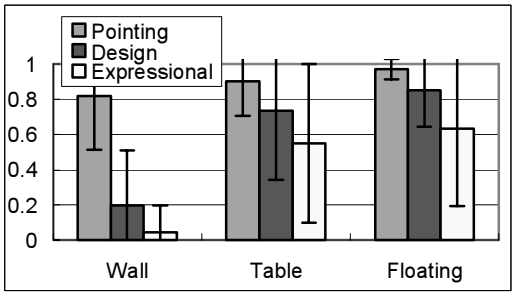

Figure 20: Ratios of each type of gestures seen by partners' cameras to those made by subjects.

Only 20% of design gestures are seen by the partner's camera in the Wall case, which are much fewer than those in the Table (74%, $T(30)=4.33$, $p(T<t)=0.00015$) and the Floating cases (85%, $T(21)=6.86$, $p(T<t)<1.0\text{x}10^{-7}$). This shows that design gestures are usually made near the body.

Only 4.6% expressional gestures are seen in the Wall case, which is much fewer than those in the Table (55%, $T(14)=3.66$, $p(T<t)=0.0026$) and the Floating cases (63%, $T(18)=4.79$, $p(T<t)=0.0001$). This shows that express-ional gestures are almost always made near the body.

Words and sentences

Subjects spoke almost the same amount of words and sentences in all conditions. We counted all uttered words and sentences and normalized them using results in the Floating case for each group and condition. The normalized numbers of words in the Wall and Table cases are 1.04 and 1.00, and there was no significant difference ($F(2,33)=0.12$, $p=0.88$) across the three condition. The normalized numbers of sentences in the Wall and Table cases are 0.95 and 0.96, and we found no significant difference either ($F(2,33)=0.28$, $p=0.76$). But, we did find a tendency that the average number of words per a sentence in the Floating case was smaller (i.e., a sentence is shorter) than that in the Wall case ($T(11)=1.77$, $p(T<t)=0.10$).

Classified sentences

We modified Doherty-Sneddon's coding scheme [4], and classified all uttered sentences into four different categories: *initiations* (sentences, commands, suggestions or questions), *Y/N-responses* (simple responses that could be summarized as yes or no), *what-responses* (responses that could not be summarized as yes or no, often include new information),

Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR'02)
0-7695-1781-1/02 $17.00 © 2002 IEEE

and *clarifications*. Figure 21 shows the normalized average number of each type of sentences.

Subjects made initiations 20% and 29% more in the Floating case than in the Wall and Table cases, respectively (e.g., T(11)=4.25, p(T<t)=0.001, for the Table case). Since subjects could see the partners all of the time, they were motivated to involve each other more by making initiatory utterances. There was no significant differences in Y/N-responses (F(2,33)=0.29, p=0.75) and What-responses (F(2,33)=0.68, p=0.52).

Subjects made clarifications 45% and 52% less in the Floating case than in the Wall and Table cases, respectively (e.g., T(11)=3.16, p(T<t)=0.009 for the Table case), implying that verbal responses were not as necessary since they could understand each other easily by seeing visual communication cues. In the Table case, the angle between the task and communication spaces seemed enough to impede visual communication cues.

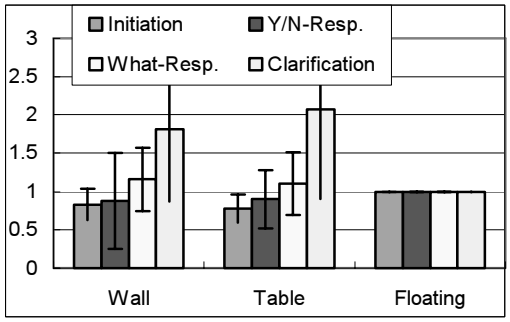

Figure 21: Normalized average number of classified sentences.

Overlaps

Figure 22 shows the normalized average number of speech overlaps (simultaneous speeches or interruptions). We found overlaps are 85% more in the Floating case than that in the Wall case (T(10)=3.38, p(T<t)=0.007).

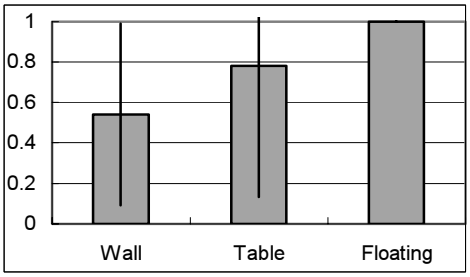

Figure 22: Normalized average number of overlaps.

Deictic phrases

Figure 23 shows the normalized average numbers of deictic phrases uttered by a subject whose mouse was active and inactive. We found no significant difference (F(2,30)=0.27, p=0.76) for active mouse holders. However, when subjects' mice were inactive, they used deictic phrases most often in the Floating case (117% and 43% more than those of the Wall and Table cases, respectively, e.g., T(10)=4.64, p(T<t)=0.0009 for the Table case). A tendency was also found that inactive mouse holders used deictic phrases more in the Table case than in the Wall case (T(20)=1.56, p(T<t)=0.13).

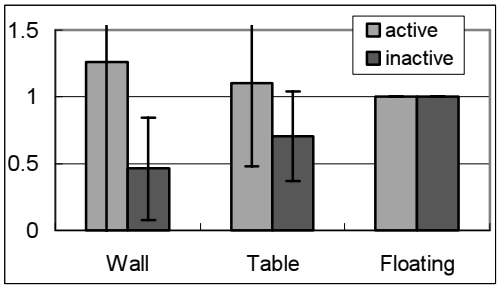

Figure 23: Normalized average numbers of deictic phrases of active and inactive mouse holders.

Laughter

Interestingly, subjects made laughter 65% more often in the Floating case than in the Wall case (T(45)=2.21, p(T<t)=0.033, figure 24) and 100% more often in the Wall case than in Table case (T(39)=2.58, p(T<t)=0.014). The Floating condition seemed to be most social.

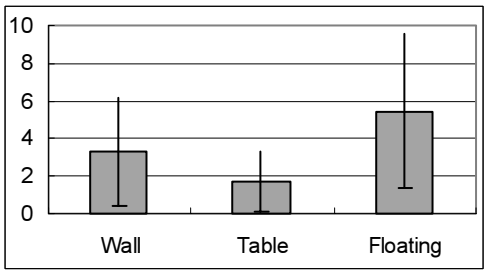

Figure 24: Average number of laughter.

Findings in process measures

Compared to the Wall and Table cases, the Floating case tended to produce more pointing and design gestures. With the Floating case, such gestures were automatically in the field of view hence they were easier to perceive. Although the total amount of utterances were almost the same across conditions, the ratio of initiations was higher, and that of clarifications was lower, in the Floating case. Ratios of speech overlaps and occurrences of laughter were also higher in the Floating case. The communication in the Floating case was more natural and subjects were motivated to make initiatory body motions and utterances.

4.3.3. Subjective Measures

Questionnaires

After each condition, subjects filled out questions on a scale of one ("Disagree") to seven ("Agree") (Table 4), and also gave rankings on overall impressions after all of the three conditions.

Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR'02)
0-7695-1781-1/02 $17.00 © 2002 IEEE

Table 4: Subject Questionnaire for Exp 2

Questions about communication richness

Q2-1 I looked at my partner very often
Q2-2 I used pointing gestures very often
Q2-3 It was very easy to see my partner
Q2-4 I could very easily understand what my partner was doing
Q2-5 I could easily tell when my partner was looking at me
Q2-6 I could very easily tell where my partner was looking

Questions about overall impressions

Q2-7 In which condition was it the easiest to work together
Q2-8 Which condition did you like the best
Q2-9 In which condition was it the easiest to communicate with your partner

They felt they looked at the partners more often in the Floating case than in the Wall case (Q2-1, $T(49)=2.10$, $p(T<t)=0.04$), but there was no significant difference between the Wall and Table cases (figure 25). We found no significant difference in the frequency of pointing gestures (Q2-2, $F(2,75)=1.22$, $p=0.30$).

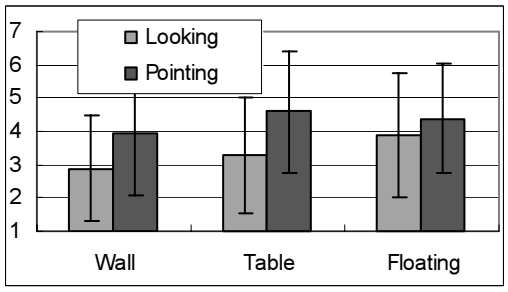

Figure 25: Frequency in looking at partners and use of pointing gestures

Subjects felt it easiest to see the partners in the Floating case (e.g., Q2-3, $T(37)=3.09$, $p(T<t)=0.004$ for Wall) but there was no significant difference between the Wall and Table cases ($T(49)=0.94$, $p(T<t)=0.35$) (figure 26). Subjects also felt it was easier to understand what the partners were doing in the Floating than in the Table case (Q2-4, $T(50)=2.19$, $p(T<t)=0.034$).

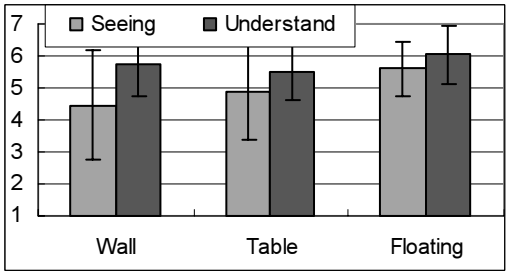

Figure 26: Ease of seeing partners and understanding of partners' doing.

When asked how easy it was to tell if they were being looked by the partners (Q2-5), subjects didn't feel any difference ($F(2,75)=0.24$, $p=0.78$, figure 27). Furthermore, they didn't feel any difference in ease of telling where the partners were looking (Q2-6) ($F(2,75)=0.92$, $p=0.40$).

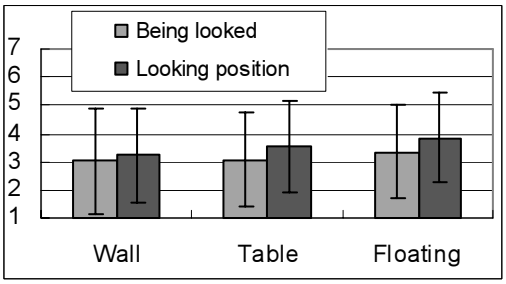

Figure 27: Ease of telling of being looked by partners and partners' looking position.

Subjects felt it easiest to work together in the Wall case than in other cases (e.g., $T(48)=4.54$, $p(T<t)<1.0\text{x}10^{-4}$ for Table, figure 28). The Floating case is also easier for collaboration than the Table case ($T(49)=2.98$, $p(T<t)=0.0045$). They favored the Wall and Floating cases more and felt that the communication was easier in those cases than in the Table case (e.g., for liking, $T(50)=4.95$, $p(T<t)<1.0\text{x}10^{-5}$ for Wall and $T(47)=3.34$, $p(T<t)=0.0016$ for Floating). In both questions, no significant difference was found between the Wall and Floating cases (e.g., $T(48)=1.08$, $p(T<t)=0.28$ for liking).

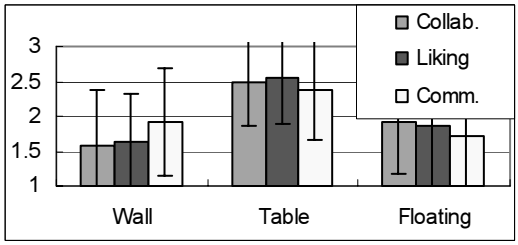

Figure 28: Average rankings in ease of collaboration, liking, and ease of communication.

Subjects comments

When writing about the Wall condition, subjects mentioned it was very easy for communication because they had the same visual perspective. Some said it was hard to see the partner, which they generally did not feel they needed to.

For the Table case, most subjects said it was hard to see and work together with the upside down 2D design. However, a few subjects said it was fun and natural since working on a table was a natural activity. It is interesting that a few subjects pointed it was hard to move a cursor while a mouse was in their view, since they usually use a mouse diagonally on a table without seeing it.

For the Floating case, many subjects commented on the intuitiveness and usefulness of the condition; it was fun, easy to know what the partner was going to do, easy to communicate and collaborate, and easy to point. Several subjects also mentioned they needed to look at the picture from the back. This required mental rotation, and made the task confusing [7]. A few subjects commented it was distracting to see someone through the grid.

Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR'02)
0-7695-1781-1/02 $17.00 © 2002 IEEE

Findings in subjective measures

In The Floating case, it was easy to see and understand the partner. However subjects felt no difference in how easy it was to tell when they were being looked or where the partners were looking. The Wall case was easiest to work together, as they could see the 2D task space from the same viewpoints. Because of the upside down view, they disliked and felt it hard to communicate in the Table case.

4.4. Discussion

Large differences were found in process measures. In the Floating case, the communication between subjects became more natural, social and easier. Being able to see the partner all of the time, they were motivated to involve each other by pointing and design gestures and initiatory utterances. They also made less clarification and made more laughter. Placing the task space between users in space is useful for natural communication.

However, subjects didn't seem to benefit from this ability for better outcome or working efficiency. What mattered more here was the orientation of the shared object. They liked the Wall case just because they could see the 2D grid from similar perspectives. Many subjects commented the preference rankings would be different if it were a 3D task, not a 2D one, since people are used to see 3D objects from different points of view. We need to follow this experiment by using a 3D task with spatial manipulations, to further investigate the impact of AR interfaces.

5. Conclusions

In order to reveal the technical and social impact of collaborative AR interface, we conducted two experiments on users' communication behaviors in face-to-face AR collaboration. In the first experiment, we found the real world visibility affects communication, using an object identification task. The optical see-through display was the best in terms of words and gestures needed to complete the task. The video see-through displays seemed suffered from camera offsets [13].

In the second experiment, we found the location of the AR task space affects communication, using a 2D icon drawing task. When the task space was placed between them in space, the communication became more natural, social and easier. In this condition, subjects made more initiatory body motions, utterances and laughter. However, the orientation of the task space mattered. Subjects liked to have it on the wall most, as they could see it from the same viewpoints [7].

In summary, we found that for co-located AR interfaces, optical see-through HMDs viewing a task space that is between the participants may produce the most natural collaboration. In the future, we will focus on using a 3D task, remote AR collaboration and so on.

6. References

1. Billinghurst, M., Weghorst, S., Furness, T., Shared Space: An Augmented Reality Approach for Computer Supported Cooperative Work. Virtual Reality: Research, Development and Application, 1998.
2. Billinghurst, M., Kato, H. Out and About: Real World Teleconferencing, British Telecom Technical Journal (BTTJ), Millenium Edition, Jan 2000.
3. Daly-Jones, O., Monk, A., Watts, L. Some Advantages of Video Conferencing Over High-quality Audio Conferencing: Fluency and Awareness of Attentional Focus. *Int. J. Human-Computer Studies,* 49, 21-58, 1998.
4. Doherty-Sneddon, G., Anderson, A., O'Mally, C., Langton, S., Garrod, S., Bruce, V. Face-to-Face and Video-Mediated Communication: A Comparison of Dialogue Structure and Task Performance, *Journal of Experimental Psychology: Applied*, Vol.3, No.2, pp.105-125, 1997.
5. Gaver, W., The Affordances of Media Spaces for Collaboration, in *Proc. CSCW'92*, Toronto, Canada, ACM Press, pp.17-24, 1992.
6. Heath, C., Luff, P. Media Space and Communicative Asymmetries: Preliminary Observations of Video-Mediated Interaction. *Human-Computer Interaction,* Vol. 7, pp.315-346, 1992.
7. Ishii, H., Kobayashi, M. and Arita, K., "Iterative Design of Seamless Collaboration Media," Communications of the ACM (CACM), Special Issue on Internet Technology, ACM, Vol.37, No.8, pp. 83-97, 1994.
8. Kiyokawa, K., Takemura, H., Yokoya, N. "SeamlessDesign for 3D Object Creation," *IEEE MultiMedia, ICMCS '99 Special Issue*, Vol.7, No.1, pp.22-33, 2000.
9. Monk, A.F., McCarthy, J., Watts, L. & Daly-Jones, O., Measures of process, in Thomas, P.J. (ed) CSCW requirements and evaluation, Berlin: Springer-Verlag, pp.125-139, 1996.
10. O'Conaill, B., and Whittaker, S., Characterizing, predicting and measuring video-mediated communication: a conversational approach. In K. Finn, A. Sellen, S. Wilbur (Eds.), *Video mediated communication*. LEA: NJ, 1997.
11. Ohshima, T., Satoh, K., Yamamoto, H. and Tamura, H. "*AR2 Hockey: A Case Study of Collaborative Augmented Reality*," Proc. IEEE VRAIS '98, pp.268-275, 1998.
12. Sellen, A. Remote Conversations: The effects of mediating talk with technology. *Human Computer Interaction*, Vol.10, No.4, pp.401-444, 1995.
13. Takagi, A., Yamazaki, S., Saito, Y., Taniguchi, N. "Development of a Stereo Video See-through HMD for AR Systems," Proc. IEEE & ACM ISAR 2000, pp.68-77, 2000.
14. Tang, J., Isaacs, E., Why Do Users Like Video? Studies of Multimedia-Supported Collaboration. Sun Microsystems Laboratories, Technical Report, SMLI TR-92-5, 1992.
15. Williams, E., Experimental Comparison of Face-to-Face and Mediated Communication: A Review. Psychological Bulletin, 84, (5), pp.963-976. 1977.

Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR'02)
0-7695-1781-1/02 $17.00 © 2002 IEEE