Towards Ideal Window Layouts for Multi-Party, Gaze-Aware Desktop Videoconferencing

Sasa Junuzovic, Kori Inkpen, Rajesh Hegde, and Zhengyou Zhang

Microsoft Research

ABSTRACT

In high-end desktop videoconferencing systems, several windows compete for screen space, particularly when users also share an application. Ideally, the layout of these windows should satisfy both (a) layout guidelines for establishing a rich communication channel and (b) user preferences for window layouts. This paper presents an exploration of user preferences and their interplay with previously established window layout guidelines. Based on results from two user studies, we have created five recommendations for user-preferred window layouts in high-end desktop videoconferencing systems. Both designers and end-users can use these recommendations to setup "ideal" layouts, that is, layouts that satisfy both user preferences and existing layout guidelines. For instance, we have developed an application that utilizes the recommendations to guide users towards ideal layouts during a videoconference.

KEYWORDS: Videoconferencing; telepresence; CSCW; window layout; user preferences; user study; guidelines.

INDEX TERMS: H.4.3 [Information Systems Applications]: Communications Applications – Computer conferencing, teleconferencing, and videoconferencing; H.5.2 [Information Interfaces and Presentation]: User-interfaces – screen design, user-centered design.

1 INTRODUCTION

High-end desktop videoconferencing systems, such as the one pictured in Figure 1, strive to provide users with an experience that rivals that of face-to-face communication. In these systems, users can hear high-quality audio and see high-quality video of each other, as well as, share an application. Together, these three sharing mediums combine to create a rich communication channel. However, as Figure 1 shows, several windows compete for limited screen space and an appropriate layout must be chosen to fit them onto the display. The choice of window layout can impact the user experience for two reasons. First, previous work has shown that layouts can impact important aspects of a communication channel, including the sense of presence [10] and gaze awareness [3][6][15]. Second, these layouts can also lower user satisfaction if they do not satisfy users' personal layout preferences, which have been shown to differ greatly from person to person in non-videoconferencing scenarios [1][7][9].

In this paper, we explore user preferences for window layouts in high-end desktop videoconferencing systems and their interplay with previously described layout guidelines. Although previous research has examined presence and gaze awareness parameters for videoconferencing [3][10], these have primarily been explored in controlled experiments in absence of real tasks. Moreover, these studies focused on parameters of video windows in absence of shared application windows, so window layout management was less of an issue.

Given the growing use of desktop videoconferencing systems for small group meetings, we wanted to explore the interplay of user window layout preferences with existing layout guidelines for high-end desktop videoconferencing systems during small group tasks. The results of this work enabled us to define window layout recommendations that help optimize the user experience. We also built a prototype system that uses these rules to evaluate arbitrary window layouts and guide users towards ideal layouts.

The rest of this paper is organized as follows. In the next section, we consider background work relevant to our research. In the following two sections, we describe two user studies that explored user preferences for window layouts. Next, we derive the recommendations for ideal layouts. Then, we briefly describe the application that uses these rules to guide users towards ideal layouts. We end with conclusions and directions for future work.

2 BACKGROUND WORK

As mentioned above, previous work has studied window layouts, both generally and for videoconferencing specifically. In this section, we first give the results of these studies. Then, we show how the study findings can be applied in high-end desktop videoconferencing systems in small group collaborations.

2.1 General Window Layout Preferences

Studies of general window layouts have discovered that users follow certain window layout styles. For example, Hutchings and Stasko [9] studied twenty users and found that they used three layout styles: 1) maximize all windows; 2) do not maximize any window and create a layout that keeps all windows visible; and 3) dedicate some screen space for secondary applications and use the rest of the space to make other windows as large as possible. Moreover, Grudin studied dual-monitor window layouts used by eighteen people [7] and found that the additional monitor usually

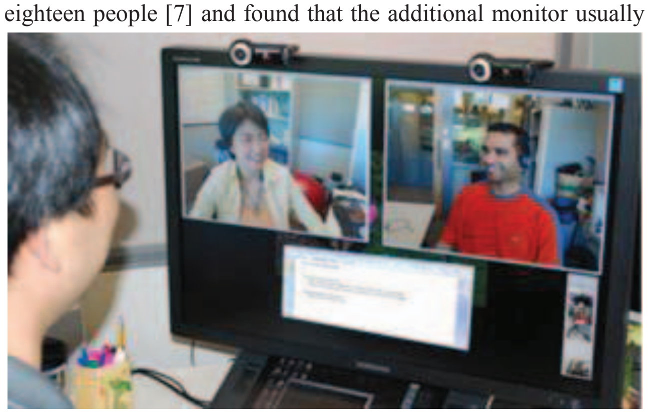


Figure 1. A high-end desktop videoconferencing session with three participants who are working on a document together.

Redmond, WA 98052, USA
{sasajun, kori, rajeshh, zhang}@microsoft.com

Graphics Interface Conference 2011
25-27 May, St. John's, Newfoundland, Canada

displayed secondary applications. Badros et al. [1] argued that users' window management requirements are complex and that satisfying them with manual window size and position adjustments is tedious. Thus, they created an intelligent window manager system that accepts layout constraints from users and then automatically enforces these constraints as users open, close, move, and resize windows.

2.2 Videoconferencing Window Layout Guidelines

In addition to studying window layouts for general applications, previous work has also investigated layouts for videoconferencing systems. While the former has focused on user preferences for satisfactory layouts, the latter takes a more prescriptive approach by focusing on what layout steps are needed to establish a rich communication channel. Several window layout guidelines have been proposed. They focus on preserving gaze awareness and establishing a sense of presence.

Gaze awareness has been shown to be an important aspect of non-verbal communication [3][6][15]. Gaze awareness can be classified into two types: mutual and directional. Mutual gaze is preserved if users can tell from the videos on their screens that another user is looking at them. Chen [3] has shown that mutual gaze is preserved if the angle between the eyes of the participant in the video and the camera lens is less than (i) one degree in the upward and horizontal directions and (ii) five degrees in the downward direction. Directional gaze is preserved if users can tell that a user is looking at another user or the shared application. Sellen [15] found that it can be preserved by using a horizontal offset among the video windows. The offset can preserve head-turn cues, which, in turn, can provide directional gaze as in face to face scenarios.

Presence is defined as the feeling of being in the same room as others. Ichikawa et al. [10] showed that the sense of presence in a videoconference is preserved when the size of the remote participants makes it appear as if they are sitting a typical distance away from the local user (e.g. across a table) and the background behind them is seamless.

Window layouts can also have a social impact. Huang et al. [8] have shown that artificially tall people are perceived as more dominant than artificially short people in a videoconferencing discussion. While these impacts are also important, our focus was on gaze and sense of presence.

One way to ensure that gaze and sense of presence are preserved is to create a system in which users have no choice of window layouts or camera placement. Nguyen and Canny [12] took this approach in designing a non-desktop videoconferencing system that preserves gaze between groups of collaborators. Unlike their work, our work focuses on desktop scenarios where users can choose arbitrary window layouts and camera positions.

From the perspective of exploring user preferences for window layouts in desktop videoconferencing systems, prior work provides two main results. One result is that when users have to manage multiple windows, their management styles vary. Since high-end videoconferencing systems have several windows, users will likely have preferred window styles. The other result is that window layouts in videoconferencing systems must follow the established guidelines in order for users to experience a rich communication channel. Thus, the question is how these guidelines interact with the users' window management styles.

2.3 High-End Desktop Videoconferencing

To explore user preferences and their interplay with previously established guidelines, we used an in-house videoconferencing prototype (see Figure 1). The system provides users with high quality audio and video of each other, both of which are important for a rich communication channel [13]. To support gaze awareness, the system uses multiple cameras, one for each remote participant. To provide the users with a shared workspace, we used Microsoft OneNote, which adds multi-user text editing support to the videoconference.

Using Chen's [3] angle requirements, the system can provide mutual gaze awareness by positioning each user's video close to the camera sending video to that user. Our system can also provide directional gaze awareness. First, the system flips one of the video images to make it appear as if the users are all sitting around a table. Second, a large horizontal offset between video windows is used to provide awareness of head-turning. The prototype system also supports resizing of video windows but does not provide a seamless background between the videos.

Like most high-end desktop videoconferencing systems, the prototype was designed to provide an experience equivalent to that of a face-to-face meeting for a small, distributed group of people. The typical number of people in such conferences has not been reported, but one can make some implications from studies of collocated meetings. Panko and Kinney [14] profiled face-to-face meetings of real knowledge workers and found that 65% of meetings had two participants, 11% had three, and 80% of meetings had five or fewer participants. Monge et al. [11] also studied face-to-face meetings and found that 57% of meetings had five or fewer participants. Based on this data, we expect that a typical number of participants in a desktop videoconference would be between two and five. In this initial exploration of window layout preferences, we focused on a three-way videoconference.

3 Preferred Layouts

We conducted a lab experiment to explore user preferences for window layouts when using a desktop videoconferencing system.

3.1 Participants and Procedure

Eighteen participants were recruited and divided into six groups of three. We did not control for participant familiarity; however, the recruits all worked in the same building with around three hundred other people, so some participants knew each other.

Each participant within a group was placed in a separate office equipped with a computer running the videoconferencing prototype. Of the three computers used for this study, two had dual 21" monitors while the third had a single 30" monitor. We chose to explore a multi-monitor and a large display configuration because both are becoming increasingly more popular. For example, in 2003 Czerwinski et al. [4] reported that at least 20% of information workers were using dual monitors.

The participants first performed five training exercises to experience the full power of the prototype and to learn about

1. the impact of video and camera placement on mutual gaze
2. the impact of video size on mutual gaze
3. the impact of video size on the sense of presence
4. the impact of video and camera placement on directional gaze
5. the real-time multi-user editing capabilities of OneNote

Each exercise was performed interactively by the entire group. The participants were free to consult with each other about what each of them was doing and needed to do. After each exercise, the participants were free to adjust the windows and cameras. For instance, when the participants were learning about the impact of video and camera placement on mutual gaze, they started by reading about mutual gaze. Then, they were asked to position the cameras and windows in a manner that did not preserve mutual gaze and then discuss their feeling of it. Following this, they were asked to position the cameras and windows in a manner that preserved mutual gaze and again discuss their feeling of it. They were then allowed to position the cameras and video windows any way they wanted. Other exercises proceeded similarly.

By the end of training, the participants had effectively chosen the window layouts and camera locations that they felt were useful. These layouts were used at the start of the collaborative task. To remove any bias from the default window and camera settings, the settings that the participants saw on arrival were randomly selected: the cameras locations, whether on top of the monitors or below them on the table, and video and application window sizes and positions varied among participants.

Following the training exercises, the participants completed the study task. The study task was chosen to "mimic" a brainstorming meeting in which participants first list as many useful ideas as possible and then rank the ideas from best to worst. We chose to use a brainstorming task because it is an interactive task that people frequently do in small group meetings. Moreover, brainstorming has been used in previous research [5]. Brainstorming also benefits from a rich communication channel (e.g., verbal exchanges, facial expressions, mutual and directional gaze) as people have to explain ideas to each other and convince each other why the ideas are good. To keep the participants engaged, we selected a brainstorming topic that would be interesting to many people: designing a new superhero.

The brainstorming task had two stages. During the first stage, the participants brainstormed for ten minutes. Even though the participants were free to adjust the camera locations and window layouts at any point, it is possible that the brainstorming activity could have kept them too busy to do so. Thus, at the end of the first stage, they were given up to ten minutes to make any desired adjustments to the camera locations and window layouts. As in the training exercises, they were free to consult each other about the changes they were making. Following this, they performed the second stage of the task, during which they continued to brainstorm and grouped ten key ideas that described their superhero for another fifteen minutes.

At the end of the brainstorming task, participants completed a questionnaire and then took part in a debrief interview. We video recorded the participants, their dialogue, and their displays. In addition, we logged all window layout events, such as size adjustments, moves, and z-order (i.e. stacking order) changes.

3.2 Results

From the video data and data logs, we extracted all camera and window settings that the participants used. Video data for one participant did not record correctly, so we have video data for only seventeen participants.[1] The initial window layouts and camera positions did not seem to influence the participants as only one kept the initial settings. Of the remaining sixteen users for who we have video data, four modified window but not camera settings, while the rest modified both.

3.2.1 Common Window Layouts

The window layouts used by the participants can be generalized into three common layouts for each monitor configuration. The common dual-monitor layouts, shown in Figure 2 (top row), are

A. medium side-by-side videos on one monitor;
B. medium videos with one in the middle of each monitor;
C. medium side-by-side videos on the inside corner of each monitor.

The three common single-monitor layouts, shown in Figure 2 (bottom row), are

D. large side-by-side videos each half the width of the monitor;
E. medium videos touching neither each other nor the edge of the monitor;
F. small side-by-side videos in the middle of the monitor.

In all of the general layouts, the OneNote window is excluded as it did not appear to follow any pattern. The video windows were always as high up on the screen as possible. We also chose to center the cameras above the video windows because that is what the participants did in 62.5% (25/40) of the settings they used.

The questionnaire results provided additional support for the choice of general layouts. Participants ranked diagrams of eight different settings from best (rank 1) to worst (rank 8). The settings are shown in Figure 3: participants in the dual-monitor condition ranked the dual-monitor diagrams (D1 through D8) on the top, while participants in the single-monitor condition ranked the single-monitor diagrams (S1 through S8) on the bottom. The dark window in each diagram is the shared application window. The mean ranks of the diagrams are shown in Figure 4. The best ranked dual-monitor layouts were D3, D5, and D6, which correspond to Layout B in Figure 2. The best ranked single-monitor layouts were S2, S3, and S5, which correspond to Layouts E and F in Figure 2.

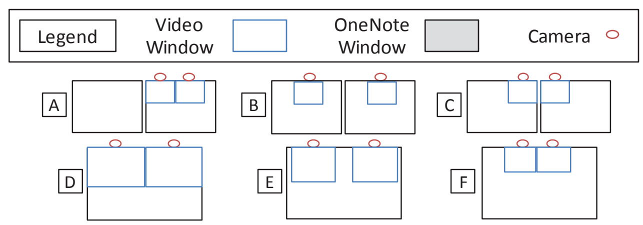


Figure 2. Preferred (top row) dual-monitor and (bottom row) single-monitor camera locations and window layouts.

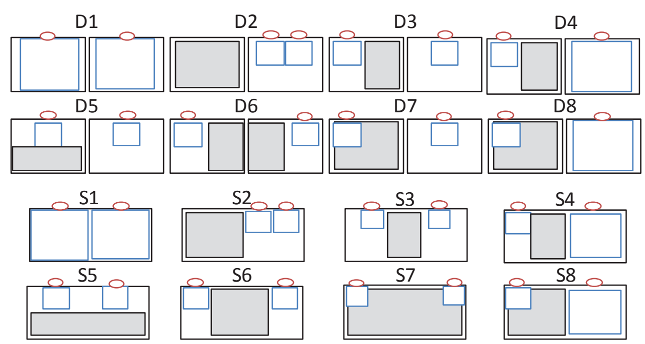


Figure 3. Settings from the questionnaire: (top) dual-monitor (D1, …, D8) and (bottom) single-monitor (S1, …, S8).

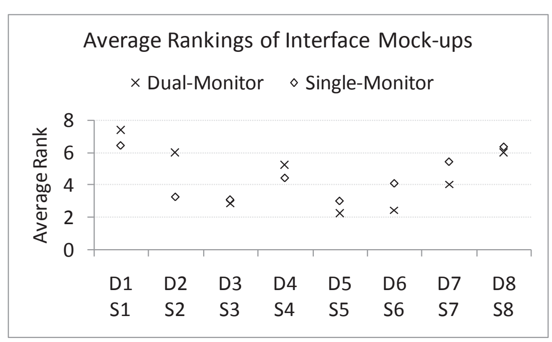


Figure 4. Average rankings of dual-monitor (x) and single-monitor settings (o) in Figure 3 (lower is better)

[1] When calculating participant percentages, the denominator is seventeen for video data and eighteen for questionnaire data.

Next, we explore window layout preferences that are embodied in these common layouts.

3.2.2 Mutual Gaze Support

The perceived value of mutual gaze was mixed as reported in the questionnaires. Eight participants (44%) felt it was important to know when a team-member was looking at them, seven (39%) felt that it was not important, and three (17%) neither agreed nor disagreed that it was important.

In our prototype system, mutual gaze is best supported when the cameras are placed on top of the monitors and the video windows are placed and centered directly below the cameras. Examining the window layout data, we found that all but two participants (88%) placed their cameras on top of the monitors. Ten of these participants (59%) selected layouts that encompassed all mutual gaze factors. Of the remaining participants, four (24%) did not center the video windows below the cameras, one (6%) did not place the video windows immediately underneath the cameras, and two (11%) placed their cameras below the monitors.

We also examined the vertical distance between the eyes of a person in each video window and the top of the screen (i.e., the camera lens) for the fourteen participants who placed the cameras on top of the monitors and centered the videos immediately below them. These results revealed that 39% of the time the eyes were within three inches of the lens, and 78% of the time they were within four inches of the lens. For a user sitting three to four feet away, these measurements result in vertical angles of 4.76 to 5.94 degrees, which are consistent with Chen's [3] vertical angle threshold for mutual gaze.

3.2.3 Directional Gaze and Head Turn Support

The perceived value of directional gaze was also mixed. Only five participants (27%) felt that it was important to know when a team-member was looking at the other team member, while six (33%) felt that it was not (seven neither agreed nor disagreed).

In our prototype system, directional gaze is best supported when the video windows are offset horizontally from each other, and head turn is supported when this offset is large. Examining the window layout data we found that seven participants (41%) chose to offset their video windows horizontally while the rest positioned the video windows side-by-side, with no offset.

Examination of the horizontal distance between the eyes of the people in the videos revealed that it was never less than four inches. As a result, the horizontal angle when looking at one team-mate was always more than one degree away from the camera for the other team-mate, which broke mutual gaze and supported directional gaze [3]. This worked for groups of three; however, with larger groups, it would be difficult to tell which other team-mate someone was looking at.

Being able to tell when users are looking at the OneNote document is another aspect of directional gaze. The questionnaire results revealed that six participants (33%) felt that it was important to know when the other team members were looking at the OneNote document.

In our prototype system, directional gaze for the shared OneNote document is best supported when the document is offset vertically from the video windows. Examining the window layout data we found that eleven participants (64%) chose to offset their shared document vertically while the remaining six (35%) positioned it either beside or overlapping the video windows.

3.2.4 Spatial Localization

The questionnaire results revealed that sixteen participants (89%) felt that it was important to be able to see both team-members' videos at the same time, and fourteen felt that it was also important to see the shared document at the same time.

This result has the potential to conflict with directional gaze and head turn. As Sellen [15] notes, head turning is important for regulating the flow of conversation. However, placing the video windows closer together reduces head movement. Additionally, positioning the video windows far apart to enhance directional gaze and awareness of head turns will make it difficult for users to see both team-members' videos at the same time. Analysis of the window layout data revealed that most of the time, the video windows were not as far apart as they could have been and measurements revealed that the distance between the eyes in the videos was never more than twenty-three inches.

3.2.5 Video Size

As mentioned previously, the size of the video representing the remote team member can impact users' sense of presence. The questionnaire results did not reveal a strong preference for large videos. Only five participants (28%) indicated that large videos were important. Examining the window layout data, we found that eight participants (47%) chose to have large video windows.

3.2.6 Overlap

As screen real estate becomes limited, users sometimes need to overlap windows. We examined the window layout data to see if the participants ever chose a layout in which windows overlapped. Seven participants (41%) chose to overlap windows, but in all cases, the overlap never obscured an important area (e.g., eyes or faces in the videos or active area of the document).

3.2.7 Multi-Monitor Issues

In the multi-monitor conditions, we examined how participants organized windows across the two monitors (we have data for eleven multi-monitor participants). Eight of the participants (73%) chose to place each video window on a separate monitor. The remaining three participants (27%) placed both video windows on the same monitor, and the OneNote document on the other monitor. Additionally, only five participants (45%) had windows that spanned the bezels, and these were all OneNote windows. For four of these five participants, the bezels were dividing the active work area (e.g., line of text or a drawing) on two screens.

4 VALIDATING LAYOUT RESULTS

The previous study results show that users' preferences for window layouts generalize to three sets of preferences on single-monitor and three different sets of preferences on dual-monitor desktops. One important question is whether these preferences, which were discovered during a brainstorming task, generalize to other collaborative scenarios. Another important question is whether any of the common layouts are more preferred than the others. Finally, it is important to evaluate if the training exercises the participants performed in the previous study biased them to any particular layouts. To further explore these three issues, we performed another user study.

4.1 Participants and Procedure

Nine new participants were recruited, and as in the first study, they were assigned into groups of three. Also as in the first study, each participant within a group was placed in a separate office equipped with a computer and videoconferencing equipment. Unlike in the first study, in which some participants knew each other, none of the participants knew each other. Also, while in the first study, one participant in a group had a single-monitor setup while the other two had dual-monitors, in this study, all participants within a group had the same monitor setup. Two groups had dual 21" (regular size) monitors, while the third group used a single 24" (large) monitor. One recruit did not show up, however, so one of the dual-monitor sessions had only two users.

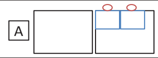
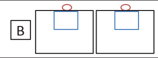
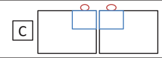
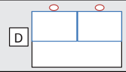
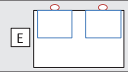
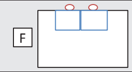
	Setting						
	5 Participants						
	3 Participants						
#	Question	Avg	Std	Avg	Std	Avg	Std
1	The game and video window sizes and positions were well selected.	3.8	1.3	2.8	1	4	0
		3.0	1.7	3.3	1.2	3.3	1.2
2	Overall, I was not satisfied with the interface settings.	2.5	1.7	2.8	1.0	2.0	0.8
		2.0	1.0	2.3	1.5	2.3	1.5

Table 1. The questionnaire completed after each poker mini-game. Scale: 1 (strongly disagree) to 5 (strongly agree).

The participants communicated with each other through a videoconferencing session, which was established before the study began using the prototype described in Section 2. To remove any training bias, participants did not perform any training exercises. They were, however, informed that there were two cameras connected to their computers, and that each one sent its video feed to one of the other participants. A participant was also allowed to ask others to adjust their setup if it was difficult to see them. Then, they began the collaborative task.

Our goal was to explore a qualitatively different task than the one in the first study. The task still had to emphasize non-verbal communication and be realistic and engaging. The task we chose was multi-player poker. More specifically, the participants competed against each other in a mini multi-player poker tournament. To reduce task learning effects, we recruited people who had played poker before. All recruits ranked their skills on a scale of none, beginner, average, and advanced. Most of them ranked their skill levels as average or advanced, while the rest ranked them as beginner.

Poker is qualitatively different from brainstorming – the former is competitive while the latter is cooperative. Moreover, thousands of people play online poker every day [17] so the task is realistic and engaging. To make sure that participants took the game seriously, we gave away $50 gift cards for the first and second place prizes. As a result, we expected the players to utilize the videos in order to “read” what the other players were thinking, similar to how they would interact with each other in a face-to-face poker game. Since poker is a competitive task, the participants could have attempted to position the cameras and windows in a manner that provides the least information to others. For this reason, they were required to cooperate with requests by another participant to adjust camera and window positions.

The tournament had two rounds. The first round consisted of three mini games of poker, each of which lasted twenty minutes. At the start of each mini-game, we set the participants’ layouts to one of the general layouts identified in the previous study. The order in which the participants experienced the settings was randomly assigned to counter-balance any learning effects between games. Also, players’ chip totals were reset after each game so that players who got knocked out could continue to play.

After each mini-game, the participants completed a questionnaire, shown in Table 1, in which they evaluated the setting they had used during the game. After the third mini-game, they ranked all of the settings they used from best (rank 1) to worst (rank 3) and drew their favorite size and position of the poker window for each general setting.

The winner from each group from the first round moved onto to the second round. We were interested to see whether participants would chose to use one of the general layouts from the first round, or a completely different layout. Therefore, participants were given the same monitor setup (dual or single) from the previous round but were free to use any layout they desired. To force the participants to choose a setting, we placed the video windows in random locations and the cameras lens-down on the table. However, they were not required to attempt to choose settings that provided mutual or directional gaze. Once they chose a setting, the players played poker until only one of them had chips remaining, which took forty minutes.

When the second round finished, the players were asked to draw and comment on the settings they chose to use. Their choices of settings and the comments provide us with the explicit preferences for the participants in this study.

4.2 Results

The results of the questionnaires completed after each mini-game in the first round are shown in Table 1. We excluded answers from one of the dual-monitor participants as the participant indicated that he never used the videos. All of the other participants indicated that they used the videos often. As Table 1 shows, Layout B was least liked of the dual-monitor layouts, while Layouts A and C were ranked similarly. All of the single-monitor layouts were ranked similarly. The rankings from the final questionnaire, shown in Table 2, are consistent with these results. The average rank of the three single-monitor settings was the same (2.0). For the dual-monitor case, Layout A in Figure 2 received the best rank (1.3), followed by Layout C (1.8), while Layout B received the worst rank (3.0). In addition to the questions shown in Table 1, the questionnaire after each mini-game gave the participants an option to draw the settings they would have preferred over the settings they used. Three players, all of whom used dual-monitors, chose to draw the settings they would have liked, which are displayed in Figure 5. One participant additionally drew a preferred single-monitor layout, which is why there are four drawings in Figure 5. All of the sketches were drawn by participants after the round in which they used Layout B from Figure 2. In every sketch, the video windows were drawn closer together than they were in Layout B. Hence, these drawings add further support for a preference for layouts which exhibit the spatial localization characteristic.

The data collected after the first round is consistent with the data collected after the second round of the tournament. The scans of the preferred settings the players drew after the second round are shown in Figure 6. As the drawings show, the dual-monitor setting Layout A from Figure 2 was preferred. Moreover, the single-monitor drawing resembles the single-monitor Layouts D and F. Finally, the users’ free form comments indicated that they preferred having the video windows close to each other because that made it easier to see their competitors without having to turn their heads far.

The results from the first round also confirm the controlled window overlap characteristic. In particular, we asked the participants to draw their two favorite locations of the poker window in each general layout. In five drawings, the participants overlapped the poker and video windows. In all five of these drawings, the OneNote window covers the bottom of the videos without covering the faces of the people in the videos.

Finally, the natural interface characteristic was also exhibited in the results. In all seven of the drawings from the first round, the

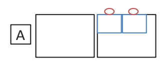
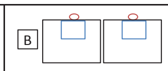
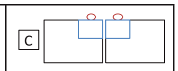
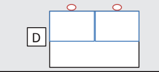
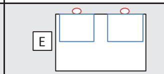
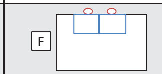
5 Participants						
3 Participants						
	Avg	Std	Avg	Std	Avg	Std
	1.3	0.5	3.0	0.0	1.8	0.5
	2.0	1.0	2.0	1.0	2.0	1.0

Table 2. The average rank of the interfaces – the lower the rank value, the more the users liked the interface.

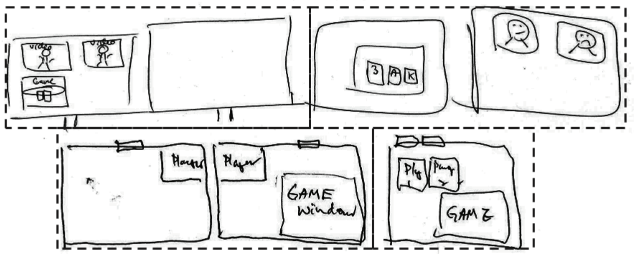


Figure 5. Desired settings drawn by players in the first round.

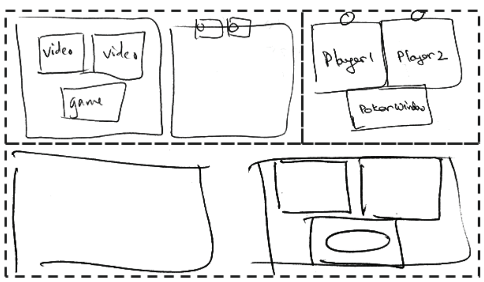


Figure 6. Settings used by players in the second round.

sizes of the video windows were the same and the video windows were drawn at the top of the monitors. Moreover, in two of the drawings, the cameras were placed centered above the videos. However, the camera placement was not universal. In two of the drawings, the cameras were not centered above the videos, and in the other drawings, the cameras were not drawn.

In summary, the results of the second study support the layout preferences identified in the first study. Hence, it appears that users' preferences are applicable to at least two types of collaborative tasks.

5 Discussion and Guidelines

Our results suggest that although many of our participants did not feel that mutual gaze was important, most eventually settled on a setup that helps preserve mutual gaze (i.e., positioning the video windows directly below the cameras). The results for directional gaze need to be examined in conjunction with users' preference for spatial localization of windows. Many users felt that directional gaze was not important. However, there was a tendency to position windows in a way that preserves directional gaze (i.e., eyes of the participants in the two videos offset by at least four inches and OneNote document placed either below the videos or on a second monitor). On the other hand, the participants also preferred to be able to see all windows at once, so they tended to not spread the windows out as far as possible. This likely reduced the awareness of head turning, but felt more comfortable to the users.

Video size did not seem to be an issue for our participants, but they did exhibit a strong preference towards keeping the sizes of the video windows equal. Also, users generally tried to not overlap windows. When windows did overlap, the faces of the participants in the videos and the active area in the shared document were never covered. Finally, similar to results found in previous multi-monitor research [7][9], we found that our participants tried to not spread documents across the bezels: they partitioned windows semantically across the monitors, either placing a video on each monitor or using one monitor to view the videos and the other monitor to view the shared document.

Based on these behaviors, we define five layout recommendations:

1. The vertical distance between the eyes of a user in the video and the lens of the camera sending video to that user should be less than four and never more than five inches.
2. The horizontal distance between the eyes of participants in the videos should be between eight and sixteen inches and never more than twenty-three or less than four inches.
3. A window that overlaps a video window should never cover the eyes or the face of the person in the video.
4. Video windows sizes should be equal.
5. In multi-monitor configurations, do not place video windows so that they span bezels.

These recommendations can help create ideal window layouts, that is, layouts that both satisfy user preferences and window guidelines for establishing an optimal communication channel. By themselves, they may not necessarily lead to ideal layouts as observed behavior is not necessarily a good basis for deriving ideal behavior. On the other hand, the guidelines for establishing an optimal communication channel by themselves are not sufficient for creating ideal layouts as something that is optimal may not be actually ideal if it does not satisfy user preferences. For example, studies of keyboard layouts have found that Dvorak layouts are more optimal than Qwerty layouts, where optimality is measured by how fast users can type. However, users prefer the Qwerty layout and are unwilling to learn the new Dvorak layout. Thus, even though the Dvorak layout is optimal, it is not ideal. As these studies have shown, user preferences (and resistance to change) can outweigh the advantages of the optimal design. Thus, it is important to study design from both directions. Our results are meant to augment previous work by examining window layouts for videoconferencing from the user perspective.

5.1 Generalizability

The layout recommendations should be treated as rough guidelines. They were derived from actions of a relatively small sample of users in two controlled studies. In order to refine the recommendations into precise and general rules, additional evaluation is needed. For instance, in our studies, participants only had the videoconferencing and shared application windows open but not those of other applications. Ideally, layout preferences should be studied when windows of other applications are also competing for screen space.

While our shared application had only one window, shared applications in general may have multiple windows or it is possible to divide their single window into multiple windows using systems such as Metisse [2] and WinCuts [16].

Users' familiarity with each other may affect layout preferences. Previous research has shown that a rich communication channel is more important when users are not familiar with each other compared to when they know each other. Whether collaborator familiarity also has an effect on window layout preferences remains to be studied. Finally, our study sample consisted of twenty-seven participants. A larger number of users need to be sampled to derive precise layout rules. To summarize, our layout recommendations should be treated as rough layout guidelines rather than precise rules.

6 Guiding User Layouts

The five layout guidelines can be used to guide users toward layouts that maximize the user experience. Such guidance is important as a badly chosen layout by one participant can result in a poor experience for all of the participants. For instance, one participant in the first study said *"something is wrong. I can always tell when she is or is not looking at me. You, on the other hand, I can never tell!"* The reason was that the participant who was the target of the remark had positioned his video windows far from the cameras. He could not understand what was wrong because the other participants had positioned their cameras and windows in a way that preserved mutual gaze. To him, everything seemed fine. As this example illustrates, what is needed is a mechanism that identifies layout problems and guides users toward layouts that maximize the user experience.

We developed a prototype application that evaluates the layout recommendations at runtime and guides users toward layouts that result in a better experience for all participants. To evaluate the recommendations, the application needs the locations of the eyes and faces in the videos and the locations of the camera lenses. The locations of the eyes in the videos can be automatically detected using vision algorithms. In its current version, our application assumes that the eyes and faces occupy rectangles centered in the video with a width and height that are twenty and forty percent of those of the video, respectively. Ideally, the camera location should also be detected automatically, but we are not aware of an algorithm that could do this. Instead, our application assumes that the camera is placed on top of the monitor in a manner that puts the camera lens immediately above the screen. Many cameras, including the Logitech Pro 9000 we used in our studies, can be positioned in such a fashion. Our application also assumes that the participant in a video window is centered below the camera sending the video of the local user to that participant.

The application works as follows. First, it instructs the users to place the cameras on top of the monitors and keep them vertically centered above the video windows. Then the application evaluates the recommendations, each of which has a "good" and "tolerable" threshold. Each user can see a total of fifteen user experience indicators, one for each rule for each window layout (one local and two remote). A layout indicator for a rule is colored "green" and "yellow" when the layout satisfies the "good" and "tolerable" thresholds; otherwise, it is "red."

7 Conclusions and Future Work

Videoconferencing system issues, such as providing high quality audio and video to the users, have greatly improved in recent years. However, for these systems to be successful, other details must be done correctly. In this paper, we investigated one of these details, the window layouts.

Our work presents three contributions that can help designers of videoconferencing systems. The designers can 1) use one of the general layouts we found as the default window layout in their systems, 2) evaluate layouts in their systems using the preferred layout characteristics and the layout recommendations we have identified, and 3) incorporate in their systems the application we developed that dynamically evaluates the window layout users select and guide the users toward a layout that provides a good user experience. In addition, the layout recommendations are stated in non-technical terms that users can understand. As a result, the recommendations can be used to educate users how to arrange the windows and cameras. We have also created TUXI, a system that applies these recommendations at runtime to guide users towards ideal layouts.

It is important to use our results as rough guidelines rather than precise rules. To refine the guidelines into precise rules, we plan to study window layout preferences for a larger sample of users. We also plan to consider the effect on layout preferences of users' familiarity with each, having other applications open, and sharing applications that have multiple windows. We also plan to explore preferred window layouts and their layout characteristics for systems that support larger groups of people. While we explored the window layout issues in three-way videoconferencing, we expect that our results should scale to slightly larger numbers of users. For example, in a conference with four or five people, all of the videos should be able to fit side-by-side in such a manner that they satisfy the horizontal and vertical distance thresholds we identified. In addition, our prototype videoconferencing system assumed that each user's computer had multiple cameras attached to it. This was necessary in order to preserve mutual and directional gaze among the participants. Typical users, however, may not have multiple cameras attached to their computers. For example, laptops typically have one integrated webcam. Hence, we plan to investigate window layout preferences for systems with a single camera. However, for a high-end videoconferencing experience, users will need to have multiple cameras. Finally, we are planning a formal evaluation of the system we developed to guide users toward a good layout.

References

[1] Badros, G.J., Nichols, J., and Borning, A. SCWM: an intelligent constraint-enabled window manager. *AAAI Spring Symposium Series on Smart Graphics 2000*.

[2] Chapuis, O. and Roussel, N. Metisse is not a 3D desktop! *UIST 2005*.

[3] Chen, M. 2002. Leveraging the asymmetric sensitivity of eye contact for videoconference. *CHI* 2002.

[4] Czerwinski, M., Smith, G., Regan, T., Meyers, B., Robertson, and Starkweather, G. 2003. Toward characterizing the productivity benefits of very large displays. *INTERACT* 2003.

[5] Dourish, P. and Bellotti, V. Awareness and coordination in shared workspaces. *CSCW* 1992.

[6] Grayson, D. and Monk, A. Are you looking at me? Eye contact and desktop video conferencing. *TOCHI 10*, 3 (2003), 221-243.

[7] Grudin, J. Partitioning digital worlds: focal and peripheral awareness in multiple monitor use. *CHI* 2001.

[8] Huang, W., Olson, J.S., and Olson, G.M. Camera angle affects dominance in video-mediated communication. *CHI EA 2002*.

[9] Hutchings, D.R. and Stasko, J. Revisiting display space management: understanding current practice to inform next-generation design. *GI* 2004.

[10] Ichikawa, Y., Okada, K., Jeong, G., Tanaka, S., and Matsushita, Y. MAJIC videoconferencing system: experiments, evaluation and improvement. *ECSCW* 95.

[11] Monge, P.R., McSween, C., and Wyer, J. A profile of meetings in corporate America: results of the 3M meeting effectiveness study. Los Angeles: University of Southern California, Annenberg School of Communications, 1989.

[12] Nguyen, D. and Canny, J. MultiView: spatially faithful video conferencing. *CHI 2005*.

[13] Olson, J.S., Olson, G.M., and Meader, D.K. What mix of video and audio is useful for small groups doing remote real-time design work? *CHI* 1995.

[14] Panko, R.R. and Kinney, S.T. Meeting profiles: size, duration, and location. *HICSS* 1996.

[15] Sellen, A.J. Speech patterns in video-mediated conversations. *CHI* 1992.

[16] Tan, D., Meyers, B., and Czerwinski, M. WinCuts: manipulating arbitrary window regions for more effective use of screen space. *CHI 2004*.

[17] Wikipedia: Online Poker. http://en.wikipedia.org/wiki/Online_poker.