Quantitative Evaluation of Media Space Configuration in a Task-Oriented Remote Conference System

Kyle Koh
Dept. of Computer Science and Engineering
Seoul National University
1 Gwanak-ro, Gwanak-Gu, Seoul 151-744, Korea
kyle@hcil.snu.ac.kr

Jinwook Seo
Dept. of Computer Science and Engineering
Seoul National University
1 Gwanak-ro, Gwanak-Gu, Seoul 151-744, Korea
jseo@snu.ac.kr

ABSTRACT

While remote conference systems have been extensively studied and developed in the past with various user scenarios, many people still rely on simple tools like messengers or video chats that deliver only visual and auditory information of each remote participant as their primary methods of real time remote communication on their computers and tablets. With the simple tools, people still perform variety of tasks. This paper analyzes the tasks performed in remote conference tools running on general purpose PCs or tablets, and categorizes them into different types based on their characteristics. We performed a controlled user experiment to discover behavioral differences observed from each type of the tasks using eye trackers. The study revealed that users showed different behavioral patterns for different task types in both subjective reporting and the eye gaze data. Based on the results, we also provide a general guideline for the screen configuration of a remote conference tool.

Author Keywords

Media space; Eye-tracking; Shared workspace; Video-mediated communication; Remote collaboration tasks.

ACM Classification Keywords

H.5.3. Group and Organization Interfaces: Computer supported cooperative work.

INTRODUCTION

When the advances in computing had an impact on our lives by unloading our cognitive burden, so did the advances in communication media by bridging the gap among people in physically separated locations. Particularly, real time communication among remote audiences opened a whole new set of research and commercial opportunities for remote collaboration needs. The ultimate goal of remote conference systems is to emulate all tasks done in local meetings as naturally as possible through technological intermediaries, thus eliminating the geo-locational restraint that prevents people from interacting with each other directly.

In the past, many remote conference systems have been designed and developed to support diverse types of tasks; some of which utilize rather expensive or specially designed hardware pieces, while others exploit the capability of readily available devices like general purpose computers equipped only with a webcam and a microphone. The former usually enriches the media of communication by providing multiple channels of data ranging from touch centric tabletop displays to physical embodiment of remote personnel, whereas data channels of the latter are often limited to only video and voice with possible extension to support simple tasks like file sharing, collaborative drawing, playing games, etc. Many people still rely on the latter for their remote communication needs due to the difficulty of adaptation and installation of the systems with specialized hardware with limited budget and office/house real estate. This is exemplified by growing popularity of online messengers and VoIP services like Skype [2] and Microsoft Live Messenger[1] [3] (Figure 1).

Though limited compared to systems using specialized hardware, the simple video conference can still provide visual feedback to evoke socio-emotional interactions and interpersonal interaction [21] by delivering non-verbal communications like facial expression to supplement the limitation of voice-only communication. Video can also be used to perform basic tasks of sharing information by

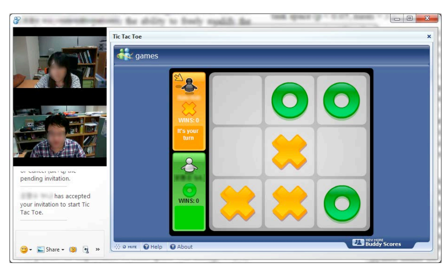

Figure 1. An example of a commercial messenger tool with videoconferencing. Task space is shown on the right-hand side and person space (for both remote person and the self-image) is shown on the left-hand side. The faces are blurred.

[1] As of Oct. 2013, Microsoft Live Messenger[3] has been merged into Skype[2].

directly showing the object of interest such as photos printed on papers within the viewing angle of the camera [20]. However, when users want to do more than just having a verbal conversation; that is, when they want to perform common tasks or share data with the others, pure video conference systems often do not satisfy the needs of sharing high fidelity images or data due to perceived distance between a local user and a remote user created by the camera and the projected image on the screen.

Such restrictions can be alleviated by providing separate task oriented workspaces shared among the participants (Figure 1). This additional channel of interaction can also supplement the video to boost even richer interactive experiences. The video and the workspace are comparable to the concept of person space and task space, the terms coined by Buxton [7]. When designing a remote conference system, designers have to take the relative utilization of the spaces into consideration, because all tasks have unique characteristics that distinguish themselves from one another. For example, task space might take a prominent role for certain tasks while person space plays more important role for some other tasks (with the extreme case of casual face-to-face conversation needing no task space at all.) If the relative utilization of each space can be measured for different types of tasks, determining the right balance among the spaces boils down to figuring out a contextual configuration of each space such as the right amount of screen real estate and an appropriate position for each space within a screen.

This paper makes three contributions: first, we introduce three different task types that are recurring themes in the media space research on readily available hardware; Collaborative Content Creation, Cooperative Problem Solving, and Competitive Gameplay. While they are briefly discussed in [16], we provide further physiological analysis of each task type, which leads to hypotheses for the relative importance of each space; second, we confirm the hypotheses by verifying them with different scenarios of screen configurations and tasks in a controlled user experiment using an eye tracker; third, we provide a guideline for designing remote conference tools based on our findings.

In the next section, we outline background and related work of our study. Next, we discuss three types of tasks and how their distinctive characteristics lead to a hypothesis that potentially explains the differences in user behavior. We then present a controlled user study to confirm our hypothesis. Finally, we discuss implications, limitations, and future direction for this work.

BACKGROUND AND RELATED WORK

Person Space, Task Space, and Reference Space

Buxton coined the terms person space and task space [7]. The person space is explained as "the collective sense of copresence between/among group participants." That is, people participating in a remote conference can feel the presence of other people that are currently in the same remote session (hence the term 'telepresence') through channels that deliver visual and auditory information of remote locations. While there have been some attempts to achieve telepresence with different types of person space with various technologies like avatars [8] or holograms [5], many researches embody only the video as a primary channel for the visual element of person space. Users can use it to sense facial expressions, gaze, body languages of remote people given enough bandwidth to carry sufficient information [12, 17, 22].

The task space is defined as "a copresence in the domain of the task being undertaken" [7]. Figure 2 shows an example of two people using a shared drawing board as task space, which was demonstrated in [24] and its successor, Videowhiteboard [25]. The task space can also be used for entertainment purposes like solving puzzles [19], or serious discussions such as brainstorming [26]. The task space may not always stay within the boundary of monitors. For example, a surgery can be performed by doctors at a distant location as demonstrated in [4], in which case, the task space is an actual human body!

The concept of Reference Space was introduced to bridge the gap among the participants on task space [6, 23]. It gives a frame of reference to "the space within which the remote party can use body language to reference the work". While it can be represented in many different forms, a simple telepointer resembling a human hand was used in our study.

Psychology for Collaboration, Cooperation, and Competition tasks

There have been psychological perspective to analyze the human behavior for tasks in CSCW (e.g. [10]). Deutsch provided an early study on the conflict from cooperation and competition to understand perceptions of conflict [9]. Jermann looked into the problem solving scenario to find how feedbacks from metacognitive tools affects performance [15]. Our work looks closely into the three recurring themes in CSCW to get insight on human behavior in the scenario.

Figure 2. An example of a fire safety poster created by study participants for a Collaborative Content Creation task. It depicts bushfire ignited by a cigarette butt wiping out mountains. The faces are blurred.

Behavior and Eye Gaze Patterns in Remote Collaboration

During tasks with two or more people involved, people often share their works in a visual form. It ranges from a simple drawing to a full document written by professionals. The physical environment offers multiple channels of interactions: voices, touch, and visual information on the state of conversation [11]. Similar things happen when the tasks are moved onto remote conference systems. Jaizhi et al. studied the behavioral variations of people according to their role while using a remote collaboration system as well as how the complexity of the task affected the usage pattern [19]. Perhaps, the most closely related work to this paper is [18] where Jaizhi et al. explored a way to automatically predict focus of attention during a task using data collected from an eye tracking device during a problem solving scenario, whose result then can be used to manipulate video feeds to show what helpers want to see it, when they want to see. By understanding the users' behavior reflected by the eye gaze can we design a tool that reflects their needs to maximize the utilization. In our work, we focused on how users behave differently on three different scenarios (discussed in the following sections) and how the result can be used in designing such tools.

TASK TYPES

We wanted to focus on a real office/home environment where no specially designed hardware is installed. Excluding a simple video chat scenario where users of the system are solely interested in informal interchange of thoughts or information (such as chitchats between friends), in which case there is no/little need for separated data feed other than video, the remote conference usually requires some sort of shared task space through which users can exchange information. Since design of screen space needs to reflect the purpose of the conference to provide an optimal condition under which users perform their tasks, it is important to understand what kind of tasks are performed. We considered how important it is to understand intention of a remote party (**low to high**), primary method of communication (**verbal**, **facial expression**, **reference on task space**), and frequency of use of the reference space to distinguish their characteristics. With this, we came up with three major categories; Collaborative Content Creation with open-end (CC), Cooperative Problem Solving with known-end (PS), and Competitive Asynchronous Gameplay (GP). The classifications are summarized in Table 1 and the details are described in their respective section. It should be noted that there are many other types of tasks performed during remote conference and the task types we chose for this study can be broken down further into different types.

Collaborative Content Creation with open-end (CC)

Many remote conference systems are often referred to as "remote collaboration tools" or "online collaboration systems," because one of the highlights of studies in media space is to enable the task of content creation across remote locations, which otherwise occurs in a physically co-located environment such as an office or a studio. It can range from a simple drawing to a full production of movie, books, etc., that requires creative thinking and ideation of products. The shared task space can be used to draw, annotate, and share images as shown in the examples like Clearboard [14], Videodraw [24] and Videowhiteboard [25]. During the creative process in remote locations, the awareness of other people comes from visual cues like videos, or person space. Therefore it is important for users to constantly refer to it to understand identity, presence, opinion, and reaction of remote parties. Nonetheless, the face value of person space might not be as significant as in GP described in more detail later. While person space provides useful information of the remote parties, it is seldom used to 'judge' their actions in CC, whereas in the GP, the facial image in person space is often in the middle of the whole game mechanic of the competition, therefore, to be 'judged.' The primary methods of communication are verbal communication and verbal reference to the task space. We hypothesized that the relative importance of person space in CC is less than that of the GP.

Cooperative Problem Solving with known-end (PS)

Many characteristics of CC are corroborated in PS in the sense that, unlike GP, participants have to work together. The only difference between them is that, for the former, the users do not start out with a clear picture of what end product might come out to be, because the creation task may result in unexpected outcomes, whereas, for the latter, the form of end result are often known if the specification (or form) of the solution is known to both parties. One example of where the specification is known and the parties need to come up with a solution is when they cooperatively try to solve a difficult math problem to enhance their skills, in which case the solution are often numbers and formulas. In the classic river-crossing problem, the end result is clear: everyone moved to the other side. In these cases, the importance of person space in PS is not on par with that of CC because any reaction, personal opinion, or facial

Task Type	Utilization of Person Space	Method of Communication
Collaborative Content Creation with open-end	Medium	Verbal Communication, Verbal Reference to Task Space
Cooperative Problem Solving with known-end	Low	Verbal Communication, Verbal Reference to Task Space
Competitive Asynchronous Gameplay	High	Facial Expression

Table 1. Task Type Classification

expression that once delivered vital clues in the creative process now carries less useful information in the problem solving scenario. Therefore, we hypothesized that, for PS, person space is less utilized than for CC, while the utilization of task space is higher. However, verbal communication and verbal reference to the Task space are still important.

Competitive Asynchronous Gameplay (GP)

As mentioned above, the GP has a different nature when it comes to how users may interact with each other in remote conference systems. They do not work *with* each other, instead, they work *against* each other. In this scenario, person space can have two important purposes: it can be used to understand and predict other people's intention [22] and also to provoke or scoff the opponent. While the latter purpose can be misused and lead to offensive and aggressive behavior (thus many online games deliberately omit video chat functionality to prevent possible conflict and violence among gamers), we intentionally included the video in our experiment, in order to measure its utilization. We assumed that participants are comfortable with video feed while playing competitive games with their close friends with whom they can use casual jeering to bring a fun element to the game. An extreme example where person space is much more important than task space is games like Poker where task space (the cards on the table) can be glimpsed once and the rest of the game mechanics rely on sensing other players' intention from their bodily or facial expressions. While task space is still a central element of the game play, it can be hypothesized that the relative importance of person space is greater in GP than that of both CC and PS. It should be noted that we only account 'competitive' gameplay as a separate type of task, because the games that require cooperation can fall into the category of PS. Also, we consider asynchronous games like board games because utilization of person space is expected to drop significantly when playing fast-paced games such as Ping Pong. In GP, the primary methods of communication is facial expression. The verbal communication might not be crucial unless the rule of the game specifically requires it.

To sum up, we hypothesized that the relative utilization of each space is different depending on the type of task performed in remote conference tool. Users involved in the GP emphasize relatively more on person space compared to the other two (H1), while those involved in PS will emphasize relatively more on task space than both CC and GP (H2). The hypotheses are summarized in Table 2.

STUDY DESIGN

In order to verify our hypotheses and to compare various screen configurations for different types of tasks, we conducted a controlled user study. The study involved participants performing a set of tasks followed by subjective reports on each configuration and task. The participants' behavior was logged with an eye tracker to measure their actual behavior, which were later used to find a screen configuration more suitable for each task type. We made an assumption that if a certain part of the screen was both consciously and unconsciously gazed by users more than the other parts, it is more important for the task than to the others, because users' gaze patterns reflect the semantics of the tasks and the area that are more focused by users are gazed more [18]. It can be important for designers of remote conference tools to put more of their resource into the area where users spend more time. A pilot study in a simplified setting was performed and revealed promising results along with some issues in with the study design.

Hypotheses	Space	The relative importance rating for the Tasks
H1.	Person Space	GP > CC > PS
H2.	Task Space	PS > CC = GP

Table 2. The hypotheses

Pilot Study

A pilot study was conducted to check the feasibility of experiment and to identify any issues with the study design. It was similar to the main study with a few exceptions of screen size (46" LCD Display) and a limited number of screen configurations. All the requirements were kept the same as the main study described in the following section. A number of issues was recognized, fixed, and reflected on the final study.

First, the screen size was too large. We originally wanted to render faces in a large enough setting so that the details of facial expression are kept over the remote image. However, this lead to noises in eye-tracking data due to wide viewing angles as well as some reported physical tiredness caused from wide angle saccade. Also, the large screens did not reflect the usage pattern in most office environments where the screens are usually much smaller than 46". Second, the pilot study did not include the video feed of self-image. This turned out to be rather inconvenient because users needed to check if their faces were being captured correctly by the camera. Therefore, self-image was included in the final study and the location of self-image became a separate

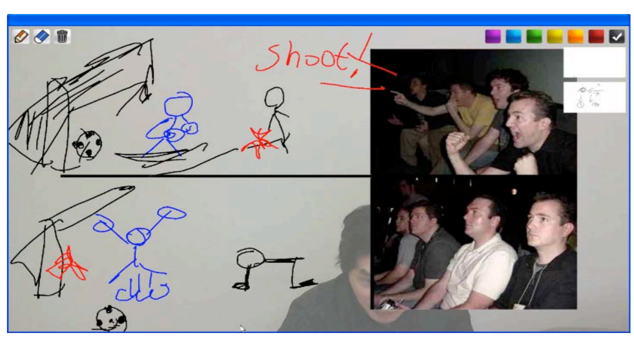

Figure 3. An example of where the comic strip can hinder the face of the remote person in Overlap condition.

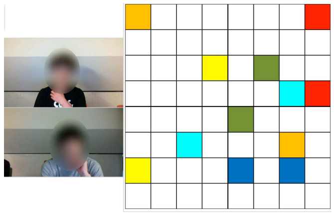

Figure 4. Participants playing a game of Flow Free. The level 13 of 8x8 is shown here. The faces are blurred.

independent variable of our study. Third, the tasks performed for each task type showed some issues during the pilot study. For CC, we provided a comic strip with four scenes with two scenes of contrasting pictures filled in for them (one with a group of men watching something indifferently and the other with the same group excited.) We asked participants to fill out the rest of two scenes to complete a story arc. While the task was adequate and ignite constant conversation (thus interaction) between two participants, two scenes provided actually occluded the viewport of person space in the Overlap screen configuration (Figure 3). Also since the images were famously quoted by online communities in the past, many people came with a similar story which was recollection of what they have seen before. Consequently, we had to change the comic strip to something on which participants had to use more of their creative thinking. For PS, we provided two river crossing games, that were chosen by previous research as a task [14]. However, the difficulty of two river crossing games used for PS was felt different by the participants, resulting in participants finishing the tasks too soon before enough data were collected or too late resulting in tiredness. Lastly, we originally made the screen configuration to be freely modifiable by the participants during the tasks to see if they would prefer configurations other than what we provided by default. This turned out to be rather intemperate because no participants cared to custom tailor the screen configuration. Instead, they maintained the initial given condition with few exceptions of “accidental” modifications. Thus, for the main study, we decided not to provide the ability to freely modify the configuration. When asked, participants “didn’t find it necessary to modify” the current configuration.

Because of these issues, no solid conclusion about screen configuration could be drawn from the results of the pilot study and the main study was redesigned to compare a wider range of screen configurations in a more controlled way. Nevertheless, the pilot study still revealed some interesting findings that verify some of the hypotheses. Participants reported significantly higher importance ratings for person space in GP than in CC. Also, we found cases where the reported importance of each space is not consistent with their actual behavior logged by the eye tracking devices. For example, the ratio of gaze on task space correlated positively with the actual screen real estate, whereas their reported importance did not show any clear patterns.

Materials and Methods

When conducting a controlled experiment involving a remote conference system, it is important to provide an identical environment to all participants in order to increase internal validity of the study. Therefore, we arranged two identical looking meeting rooms separated by a hallway. Both rooms were soundproof from each other to ensure that two participants could not detect physical presence of the other person in the other room. Also, we did not use multiple webcams to do volumetric captures nor did we consider other gaze correction method to simulate the average configuration of office/home where only one webcam is present. Since a webcam is usually installed on an upper part of the screen, we placed our cameras on top of each monitor as well. Also for this experiment, we only considered the scenario in which only two users are involved in performing tasks.

Apparatus and System Description

We prepared two quad core PCs with 22” LCD display running at 1920x1080 pixel resolution, which were connected through a TCP network. A 7” Wacom Bamboo tablet was installed on each device as an input device. Tobii x60 eye tracker was installed on both systems to log eye movements of participants at the rate of 60Hz with Tobii Studio 2.2.8. Microsoft LifeCam Cinemas was setup on each PC for both video and audio capturing. The video feed was captured in 1280x720 resolution. A remote conference tool was built on XNA framework 4.0, which rendered and managed different screen configurations. Both video and audio feeds were near delay free under the test environment. Also a telepointer resembling a hand was rendered to give a frame of reference when participants were discussing the ideas.

Participants

We did not recruit participants individually and randomly matched them for a session. We speculated that this could negatively affect the result due to awkwardness between two strangers while interacting using a remote conference system. So we decided to narrow down our candidates into pairs of people who are close to each other enough to enjoy the tasks assigned to them. Consequently, we strictly required participants to be either friends or significant others to each other. Nine pairs of people were recruited (seventeen male, one female). All had normal or corrected-to-normal vision. Seventeen people were in their 20’s and one in his/her 30’s. All were either undergraduate or graduate students with twelve majoring in Engineering and six in Art. Two of them used video chat a few times a week, one did once a week, eight did a few times a year, and seven never did. As compensation for their participation, we gave each participant two vouchers for a school cafeteria and a ballot to enter to win 128 GB SSD. Two people were randomly chosen to win the SSDs.

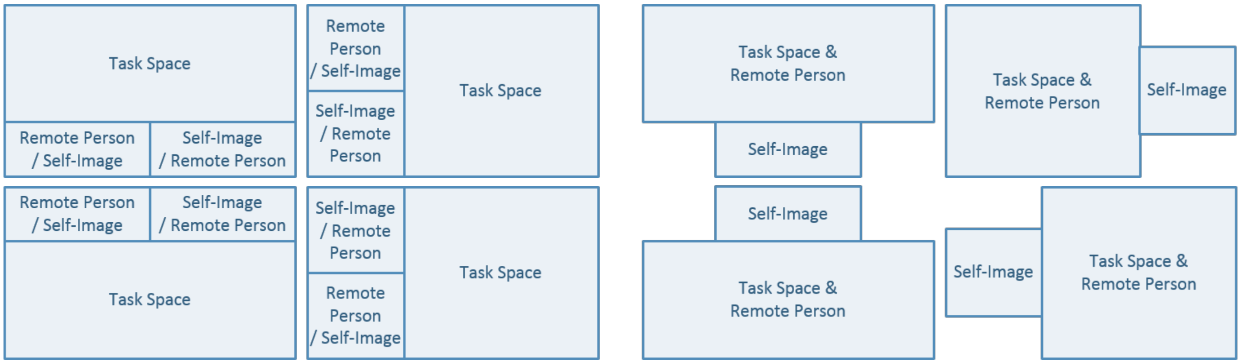

Figure 6. Left: Four main formations for screen location. The order of Remote Person and Self-Image can change. Right: Configuration of Overlap. Please note that the proportion of each space drawn here are not accurate portrait of actual formation used during the study

Procedure

Participants were sent to two separate rooms. After initial calibration of the eye-tracker using 5 point calibration provided by Tobii Studio, they were shown a simple drawing board application to get used to the tablet and its basic interface. Then, they were asked to perform two tasks per each of 3 task types, a total of 6 tasks per participants, making it 108 tasks from 18 participants. The order of the task type was shuffled using a Latin square making a within-subjects comparison. Also, different screen configuration was given to each user making it a between-subjects comparison. While all-out within-subjects designs would be generally more preferred, the decision was made because the test would have been too long (more than 3 hours) for a participant to be tested against all 36 screen configurations. Instead, screen configurations were tested on different users; i.e. each person was tested with 2 screen configurations per each of 3 task types, making 108 different combinations from 18 participants. For each task, they were told that they had 5 minutes to finish it. They were given verbal warnings when they had 2 and 1 minute left respectively. However, the time limit was not strictly enforced and participants were given a few more seconds if they desired to finish their tasks. Upon a completion of each task, they were given a set of post-task questionnaires. A post-experiment interview was also performed to collect demographics and any last minute comments. Participants spent 4 mins 25 secs (σ_{time} = 87 secs) for each task on average.

Tasks

Three tasks were designed to cover the three task types. It should be noted that the tablet was intentionally used as an input device exclusively because we custom designed all tasks to be drawing friendly to give freedom when doing CC and to avoid the effect of various input devices.

Collaborative Content Creation with open-end: Finish a comic strip with blank scenes & make a fire safety awareness poster

For one task, participants were asked to fill in blank scenes in a short comic strip with four scenes in total. To give them a kick start, we included a small drawing of a guy waking up in the morning realizing he just had a dream at the corner of the last scene. While it was up to participants to decide how the story was to be concluded with the last scene, they were allowed to freely modify the last scene as well. For the other task, we showed a blank page where participants were asked to make a simple fire safety awareness poster (Figure 2). Participants had a total freedom of how the message was delivered. We chose the two tasks that can spark active discussion between two participants. For both tasks, a whiteboard style drawing board was provided with 16 different colors for pen and an eraser. The participants were asked to collaboratively discuss their idea for a creative and coherent narrative. Participants spent 5 mins 10 secs (σ_{time} = 54secs) on average.

Cooperative Problem Solving with known-end: Flow Free

We asked participants to solve the problems from a game, 'Flow Free' developed by BigDuckGames [1]. The custom version of Flow Free was implemented to support an online cooperative play (Figure 4). Flow Free ends when all colors on the grid are connected to their respective colors with lines. No lines may overlap and only one color of a line may occupy a cell. The first two levels on 5x5 was presented to them to explain the rule of the puzzle and the level 13 on 8x8, one of more difficult levels introduced in the developer's website, was presented for practice. Then the level 1 and 4 on 14x14 were used for actual tasks. Since the purpose of the study was to analyze the usage of the spaces rather than each participant's capability to solve the puzzle, we asked them (though not forced) to finish up after 5 minutes even if the puzzle was unsolved. However, many people still went over time. They spent 5 mins 9 secs (σ_{time} = 52 secs) on average.

Competitive Asynchronous Gameplay: Five-in-a-row (Gomoku)

Participants were asked to play Five-in-a-row, also known as Gomoku, with their partner. The Five-in-a-row ends when one of the players gets an unbroken row of five stones horizontally, vertically, or diagonally. We limited game time to 5 minutes to prevent fatigue, although all participants finished their games under 5 minutes. We showed a stone grid on task space and they played the game by drawing stones on crosses (Figure 5). Due to the nature of Five-in-a-row, many local rules exist to prevent the first player from playing certain moves because he/she usually has much more advantage. Since these rules can vary depending on the participants' origin/cultural background, we asked them to discuss any local rules that should be applied to their games before the match began. They played a total of two games. Participants spent 2 mins (σ_{time} = 74 secs) on average.

Screen Configurations

We tested 4 different main formations of the spaces (Figure 6 Left).

- Task space is on the top / person space is on the bottom
- Task space is on the right / person space is on the left
- Task space is on the bottom / person space is on the top
- Task space is on the left / person space is on the right

Also, each formation had three different subformations:

- Self-image appears first (to the left or top), and remote person appears later (to the right or bottom).
- Remote person appears first (to the left or top,) and self-image appears later (to the right or bottom.)
- Remote person appears in place with task space. This is an Overlap configuration where the basic metaphor loosely follows the concept of Clearboard [14] (Figure 6 Right). This configuration was included in the test because it can potentially be used to save screen real estate if no self-image was rendered.

Also we tested 3 different ratios for the height of person space and task space: The height of person space compared to that of task space was 20% (small), 30% (medium: similar to other commercial applications), and 60% (large) respectively. The incremental numbers were chosen to find any correlation between the size of spaces and gaze patterns. This resulted in 4 (formations) × 3 (sub formations) × 3 (screen ratio) × 3 (task types) = 108 (tasks) performed in total. These configurations were chosen to emulate all possible feasible permutations of screen formation one can achieve in a single screen environment. One important rule was that the task space always took the largest proportion of the screen, because it was important to provide a sufficient room for the tasks. The independent /dependent variables are summarized in Table 3.

RESULT

There were a total of 36 different screen configurations and 3 different types of tasks as independent variables. We looked for effect on subjectively reported importance of each space along with the objective gaze pattern data collected by eye tracking devices.

Perceived Importance

The perceived importance of each space was measured with subjective reports. All importance ratings were measured on 7-point Likert scale. We found some significant effect of the task types. Tested on a Friedman test, participants reported that it was more important to see remote person's face during CC and GP than PS task (p < 0.05, mean = 3.167, 2.944, 2.111 respectively). There was no significant difference between CC and GP. Likewise, they reported that they saw remote person's face more often for CC than PS (p < 0.05, mean = 3.667, 2.167 respectively).

We also looked into individual sizes for configuration. Participants who used the 20% condition reported that they looked at remote person's face significantly less often for PS than both CC and GP (p < 0.05, mean = 1.83, 3.50, 4.33 respectively). Also for the 30% condition, people reported it was more important to see remote person's face for CC than PS (p < 0.05, mean = 3.33, 1.67 respectively). For the 60% condition, they reported that it was more important to see remote person's face for CC than PS (p < 0.05, mean = 3.67, 2.50 respectively) and that they saw remote person's face significantly more in CC than both PS and GP (p < 0.05, mean = 4.33, 2.67, 2.67 respectively). When looked into the effect of the size of person space, the participants reported that it was more important to see remote person's face for the 60% condition than the 20% condition (p < 0.05, mean = 3.67, 2.50 respectively). These results are evidences that different characteristics of each task affected perceived importance of the spaces as described more in the discussion section. This leads to the guideline of how we can improve our system in order to better reflect users' importance rating for each space by allocating more space for them.

The participants gave significantly higher importance rating for their self-image for CC than PS (p < 0.05, mean = 2.167, 1.944, respectively), although the difference in average was too small to make clear judgment. They also thought the size of task space was smaller for PS than both CC and GP when the same size was used for both tasks (p < 0.05, mean = 3.944, 4.278, 4.167). Again the difference in average is quite small to draw any implications. Also, the mechanics of Five-in-a-Row and Flow Free may have affected the result. It remains to be seen if other game types would produce different results in the future. The results are summarized in Figure 7.

Independent Variables	Dependent Variables
Screen Configurations	**Importance/Utilization of Each Space**
4 Main Formations	Subjective: Perceived
3 Sub formations	Objective: Gaze Ratio of Each Space
3 Screen Size Ratio	
Tasks	
3 Task Types	

Table 3. Independent/Dependent Variables

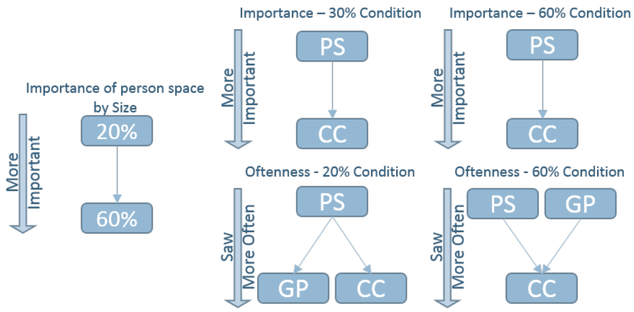

Figure 7. A summary of Perceived Importance/Perceived Oftenness of Remote Person Space based on subjective reports. Each arrow represents significant difference (p<0.05)

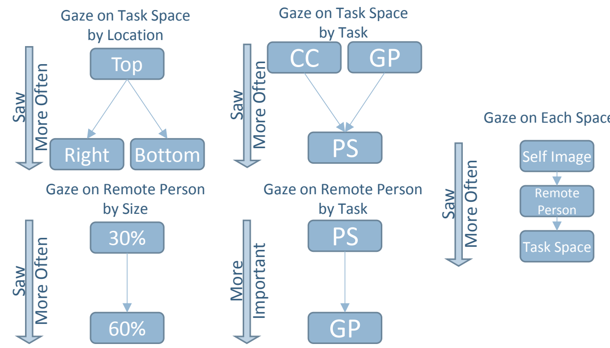

Figure 8. A summary of Actual Gaze utilization on each space based on the Eye Tracking Data. Each arrow represents significant difference (p<0.05)

Eye Tracking Data

Unlike subjective reporting, eye tracking data give us objective view on the actual usage pattern of users. Please note that, while the gaze fixation on remote person, self-image, and task space can easily be separated for the configurations except for Overlap, it is problematic for the case of Overlap. Although the region of remote person in the screen can be estimated because participants were verbally asked not to move their heads too much during the test for better tracking of the eye pupil, their faces still overlap with task space and there is not any reliable ways to detect if the participants were looking at task space or remote person. Thus in our analysis, they were counted for the fixation on both task space and remote person.

We first checked for the effect of the location of the task space. We found that people, on average, looked at task space significantly less when it was located on top, compared to when it was located to right or bottom ($p < 0.05$, mean = 72.35%, 89.76%, 88.01% respectively). When we looked into the effect of size of task space at each location, we only found a general trend; the larger remote person space was, the more it was gazed. However, we did not find any statistical significance.

We also checked for the effect of size for each space. We only found that remote person was gazed significantly more for the 60% condition than for the 30% condition ($p < 0.05$, mean = 12.36%, 9.26% respectively). However, we did not find any significant result from the gaze data on task space.

When the effect of task type was concerned, we found that remote person space was gazed significantly more for GP than PS ($p < 0.05$, mean = 16.63, 8.86% respectively), whereas task space was gazed significantly more for PS than both CC and GP ($p < 0.05$, mean = 90.85%, 85.66%, 76.10% respectively). This partially supports our hypothesis that remote person space in GP carries more relative emphasis than the others and task space in PS carries more relative emphasis than the others.

Although the self-image that was omitted in the pilot study was included in the final test, people gazed at it on an average of less than 1% of the whole task time for all screen configurations and task types. The self-image was gazed significantly less than remote person which was gazed also significantly less than task space ($p < 0.05$, mean = 0.98%, 11.12%, 84.20% respectively). While this is consistent with the subjective reporting, we wonder if this could have been affected by the relationship status of participants. Two persons said that he/she “would have had looked at self-image more often if he/she was carrying out the same task with his girlfriend/boyfriend.” Also, the order of remote person space and self-image did not affect the gaze pattern in a meaningful way mainly due to the fact that the self-image was not looked at much after all. However, remote person space was gazed at significantly more when it was presented after the self-image (i.e. to the right or the bottom) than before (to the left or the top) ($p < 0.05$). Yet, the difference in mean gaze ratio is less than 1% making it difficult to draw any real world effect. The results are summarized in Figure 8.

DISCUSSION

Our hypothesis (H1) was partially accepted according to both subjective reports and the eye tracking data. The result showed the relative emphasis on remote person space is greater in CC and GP than in PS. Eye tracking data also supported our hypothesis that the relative emphasis on remote person space is greater in GP than in PS (H1) and the relative emphasis on task space is greater in PS than in the others (H2). We did not find any strong evidence that the relative emphasis on remote person space is greater in GP than CC although hypothesized so in H1. We now wonder if the particular tasks chosen for our study may have affected the result. For example, if we had chosen a game like Poker where the face value would become even more important, GP could have dominated CC in terms of the importance of person space.

With regard to the screen configuration, the subjective importance rating and the eye gaze frequency tend to increase as the size of person space on the screen gets larger. This conforms to the conventional wisdom; the more important the space is, the more screen real estate it should be allocated with. Also, this supports the idea that the more perceptually important the space is, the more it is likely to be gazed. We suspect we can find further significant links between the importance and the type of tasks by performing experiments with larger audience for different conditions in the future.

In terms of the location of task space, putting task space on top (therefore putting person space at the bottom) negatively affected the utilization of task space according to the gaze data, although there is no evidence that the location of task space (consequently the location of person space) affects the importance of person space. We can speculate that person spaces at the bottom were distracting users. Thus, the top mounted task space should be avoided to maximize the value of each space. Also, the order of which of the self-image and person space is presented first is not as important.

Some people pointed out that using Overlap "felt like unintentional scribbling prank," therefore did not prefer it, while others pointed that is it still convenient, because the physical proximity between the spaces are closer.

From these findings, we can come up with design guidelines for remote conference tools. First, although the location of each space does not affect much, the top-mounted task space should be restrained in order to make the best use of task space. Second, the screen configuration has to reflect the type of task performed in order to maximize the utilization of each space. When the tasks involve much interaction through the video feed such as in CC or GP, more screen real estate has to be allocated to remote person's face, whereas task space should be given more priority when the task involves goal oriented setup such as PS demonstrated by the result supporting H1. Developers of remote conference systems must always keep in mind that they need to understand audiences and put more of their resources on more highly utilized area. In the case of remote conference tools, the screen real estate of readily available devices has to follow the usability guideline depending on the task they want to support in order to maximize the utility, especially when one screen devices become predominant in the wake of mobile computing era.

CONCLUSION AND FUTURE DIRECTION

In this paper, we presented a study of the use of different spaces in two way remote conference scenario in a common office environment where there are only readily available devices like PCs with an ordinary webcam. We hypothesized that users of such systems will show different usage patterns for different task types and screen configurations, which we were able to learn with both subjective reporting of study participants and their eye tracking data. We suggested three different types of tasks according to their characteristics. We also found that for CC and GP, the utilization of person space is more than that of PS. Also, we investigated a possible optimization strategy for the screen configuration. We suggest that more utilized tasks take the larger portion of the screen and that designers of such systems should avoid putting task space on top of person space to maximize its utilization.

This study opened several interesting directions for future research. We strictly limited our participants to be well acquainted with each other to eliminate awkwardness and promote active verbal communication between them. However, their relationship status could have been another important factor. Also, although it was not intentional, the participants recruited through the college bulletin board was heavily gender biased (seventeen male, one female). It will be interesting to compare a variety of social relationships to explore the impact on interaction. (e.g. friendship vs. love interest, business/professional vs. casual relationship, same-sex friend vs. opposite-sex friend vs. no acquaintance, etc.) We may also consider participants from non-academia for our future study.

We provided customizable screen interfaces in the pilot study in order to see any potential optimization by users, which was taken out for the final study due to lack of people interested in the feature and the difficulty of statistical analysis. However, it will be interesting to focus solely on this issue to compare fixed vs. customizable screen layouts. In this case, the tasks will have to be selected to encourage users to customize the layout to get meaningful result. Also, we can compare different sizes of the remote person and self-image. Since we did not find many significant differences in behavior using different task space location (except for one case where the person space was placed at the bottom), we wonder how each screen configuration can affect the performance of the actual test, which was not part of consideration in this study. Also, different environmental settings like home or office might produce different results. Once it is done, the designers will be able to narrow their choices further down. Lastly, similar experiment can be performed with three or more participants at once to scale up to a multiple user scenario. While we speculate the utilization of task space will remain the same, it will be interesting to see in which order the remote audiences must be placed on screen. It could be static or dynamic according to the conversation floor or the recency of interaction on task space by each person.

ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea(NRF) grants funded by the Korea government(MSIP) (No. 2011-0030813 and No. 2010-0002266).

We kindly thank folks at BigDuckGames[1] for letting us use their game design for our study.

REFERENCES

1. BickDuckGames, http://www.bigduckgames.com/
2. Skype^{TM}, http://www.skype.com
3. Windows Live Messenger^{TM}, http://windows.microsoft.com/en-US/messenger/home
4. Ballantyne, G.H., Moll, F.: The da Vinci Telerobotic Surgical System: The Virtual Operative Field and Telepresence Surgery. Surgical Clinics of North America, vol. 83, no. 6, pp. 1293--1304 (2003).
5. Blanche, P.A., Bablumian, A., Voorakaranam, R., Christenson, C., Lin, W., Gu, T., Flo-res, D., Wang, P., Hsieh, W.Y., Kathaperumal, M., Rachwal, B., Siddiqui, O., Thomas, J., Norwood, R. A., Yamamoto, M., Peyghambarian, N.: Holographic Three-dimensional Telepresence Using Large area Photorefractive Polymer. Nature, vol. 468, issue 7320, pp. 80--83 (2010).
6. Buxton, B.: Mediaspace – Meaningspace – Meetingspace. Media Space: 20+ Years of Mediated Life, pp. 217--231. Springer, London (2009).
7. Buxton, W.: Telepresence: Integrating Shared Task and Person Spaces. In: Proceedings of Graphics Interface '92, pp. 123--129 (1992).
8. Deng, Z., Busso, C., Narayanan, S., Neumann, U.: Audio-based Head Motion Synthesis for Avatar-based Telepresence Systems. In: ACM SIGMM 2004 Workshop on Effective Telepresence (ETP). pp. 24--30. ACM Press, New York (2004).
9. Deutsch, M. An experimental study of the effects of co-operation and competition upon group process. Human relations. (1949).
10. Finholt, T. A., & Teasley, S. D. Psychology The Need for Psychology in Research on Computer-Supported Cooperative Work. Social Science Computer Review, 16(1), 40-52. (1998).
11. Ford, C. E.: Collaborative Construction of Task Activity: Coordinating Multiple Resources in a High School Physics Lab. Research on Language and Social Interaction, vol. 32, no. 4, pp.369--408 (1999).
12. Inoue, T., Okada, K., Matsushita, Y.: Integration of Face-to-face and Video-mediated Meetings: HERMES, In: Proceedings of GROUP '97, pp. 405--414. ACM Press (1997).
13. Isaacs, E. A., Tang, J. C.: What Video Can and Cannot Do for Collaboration: A Case Study, In: Proceedings of Multimedia '93, pp. 199--206. ACM, New York (1993).
14. Ishii, H., Kobayashi, M., Grudin, J.: Integration of Interpersonal Space and Shared Workspace: ClearBoard Design and Experiments. ACM Transactions on Information Systems (TOIS), vol. 11, issue 4, pp.349--375. ACM, New York (1993).
15. Jermann, P. Computer support for interaction regulation in collaborative problem-solving. Unpublished Ph. D. thesis, University of Geneva, Switzerland. (2004).
16. Koh, K., Song, H., Jeong, D., Kim, B., Seo, J.: What Do You See When You Interact with Friends Online? Face, Hand, or Canvas?. In: Proceedings of CHI EA '11, pp.2131--2136. ACM, New York (2012).
17. Lombard, M., Ditton, T.: At the Heart of It All: The Concept of Presence. Journal of Computer Mediated Communications, vol. 3, No. 2 (1997).
18. Ou, J., Oh, L.M., Fussell, S.R., Blum, T., Yang, J.: Analyzing and Predicting Focus of Attention in Remote Collaborative Task. In: Proceedings of the 7th international conference on Multimodal interfaces (ICMI '05), pp. 116--123. ACM, New York (2005).
19. Ou, J., Oh, L.M., Yang, J., Fussell, S.R.: Effects of Task Properties, Partner Actions, and Message Content on Eye Gaze Patterns in a Collaborative Task. In: Proceedings of CHI '05, pp. 231--240. ACM, New York (2005).
20. Sellen, A., Buxton, W., Arnott, J.: Using Spatial Cues to Improve Videoconferencing. In: Proceedings of CHI '92, pp. 651--652. ACM , New York (1992).
21. Short, J., Williams, E., Christie, B.: The Social Psychology of Telecommunications. John Wiley & Sons, London (1976).
22. Tammelin, M.: From Telepresence to Social Presence: The Role of Presence in a Network-based Learning Environment. In: Tella, S. (Ed.) Aspects of Media Education: Strategic Imperatives in the Information Age. Media Education Center. Department of Teacher Education. University of Helsinki. Media Education Publications 8 (1998).
23. Tang, A., Pahud, M., Inkpen, K., Benko, H., Tang, J.C., Buxton, B.: Three's Company: Understanding Communication Channels in Three-way Distributed Collaboration, In: Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW '10), pp. 271--280, ACM Press (2010).
24. Tang, J., Minneman, S.: Videodraw: A Video Interface for Collaborative Drawing. ACM Transactions on Information Systems (TOIS), vol. 9, issue 2, pp. 170--184 (1991).
25. Tang, J., Minneman, S.: Videowhiteboard: Video Shadows to Support Remote Collaboration. In: Proceedings of CHI '91, pp. 315--322. ACM, New York (1991).
26. Wang, H., Cosley, D., Fussel, S.R.: Idea Expander: Supporting Group Brainstorming with Conversationally Triggered Visual Thinking Stimuli, In: Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW '10), pp. 103--106 (2010).