Googly Eyed: A comparative study of HUDs vs. mobile phone using video conferencing for remote collaboration

Steven Johnson
Computer Sciences
UW Madison
sjj@cs.wisc.edu

Madeleine Gibson
Industrial Engineering
UW Madison
mcgibson2@wisc.edu

Ying Zheng
Computer Sciences
UW Madison
fivefive@cs.wisc.edu

ABSTRACT

Heads-up displays (HUDs) such as Google Glass have provided a novel approach for remote real-time visual and audio information exchange. Thus, comparing to the conventional remote collaboration methods such as mobile phone audio communication and/or video conferencing, HUDs may suggest a more effective and efficient way to collaborate. To determine how HUDs affect task performance during remote collaboration for novice mobile users we conducted a two by two between subjects controlled lab study where 16 participants, divided into 4 groups, collaboratively performed different physical tasks (a dynamic/static robot assembly task) with different devices (HUD/mobile phone video conferencing with the Google Hangout app). According to our experimental results, compared to mobiles phones, HUDs provided better collaboration efficiency (i.e. shorter task completion times) in both dynamic and static tasks. As we expected, HUDs exhibited a larger improvement in dynamic tasks than static tasks in terms of the rate of errors. In addition to the quantitative results, on average, participants indicated that they had a better overall remote collaboration quality with HUDs.

Author Keywords

Remote collaboration, HUDs, dynamic and static physical tasks,

INTRODUCTION

Collaboration on physical tasks is a common phenomenon in our everyday life, and thus is a very important study subject in plenty of fields, such as education [1], anthropology [2], and human-computer interfaces [3]. With the rapid development of remote information sharing technologies, remote collaboration on physical tasks is becoming more and more common where side by side collaboration is not feasible. Normal remote collaboration technologies include audio communication (e.g. phones) and video sharing (e.g. Facetime on iPhone). Fussell, S.R., *et. al.* [4] conducted an experiment to determine the most effective approach of remote collaboration among audio, scene video, head-mounted video, and combined video of scene and head-mounted by having participants perform a collaborative static physic task in pairs. Based on previous research cited [5], the shared visual space is one of the most important resources for effective collaboration. The results of their experiments indicated that a scene camera works better than a head-mounted camera as it provides a more enhanced shared visual space. However, the authors may have missed some major points in their claims: 1) static tasks (e.g. assembling of a robot in a stationary position) cannot represent all kinds of physical tasks, such as tasks which involve moving and may introduce new problems in visual space sharing, and 2) in everyday life, installation of a scene camera may not be feasible. Furthermore, if a person needs help to perform some physical tasks while no one is around, the simplest thing they can do is to call someone for instructions, or, if allowed, to share his/her visual space and other information with any device in his/her pocket. In this sense, portable devices for video sharing, such as HUDs or real-time video sharing/conferencing on a smart phone, would be a more likely approach for remote collaboration.

Smart phones are not the only convenient approaches to sharing visual spaces. With the limelight of Google Glass, heads-up displays, or HUDs, have come to the forefront of public attention. The concept of HUDs can be dated back to 1960s [6], however, breakthroughs mainly occurred in the last 2 decades, as technologies including various micro-display devices, miniature modulated laser light and associated scanners, and miniature projection optics have emerged [7]. Combining with a wearable computer, Google Glass fulfilled the concept of HUD by providing functionalities such as audio communication and visual space sharing [8]. There are a few potential advantages of using it for remote collaboration on physical tasks compared to using a smart phone. For one, the collaborator who is performing the task (or the "worker" as defined in [4]) do not need to hand-hold the device to take a video and share the visual place, thus increasing their mobility. Also, the HUD will be mounted on the head of the worker, thus may be able to continuously provide usable visual information about the shared visual space [9]. However, these advantages come with a cost, as the shared visual space may be easily adjusted when using the smart phone video sharing, but it would be hard to adjust the space when using a HUD. Thus, it is of great interest if we can determine which approach would be more effective in remote collaboration on physical tasks: HUDs or video sharing on mobile phones.

BACKGROUND

In this study, we want to determine which technology type, between HUD and video sharing on a smart phone, is more effective and efficient for remote collaboration on physical tasks.

Inspired by Fussell's work in [4], we decided to focus on construction tasks falling within a general class of "mentoring" collaborative physical tasks, involving a person (known as the "worker") who directly manipulates objects and seeking help from another person (known as the "helper"). In our tasks, the worker is trying to assemble a toy cart, while the helper will provide the worker with the instructions to guide the worker to completion. Instead of using only a static construction task (assemble the cart at a stationary position) in [4], we also utilized a "dynamic" task, in which the cart parts are separated among different workstations, and the worker needs to move around to fetch the proper parts to complete the assembly task.

The workers and the helpers are allowed to communicate through one of two different channels: **1.** HUDs (Google Glass), with which the worker's field of view is transferred to the helper, and the helpers' instructions are provided to the user through the an audio and video channel embedded in Google Glass. **2.** Video sharing/conferencing through an app (Google Hangouts) on an iPhone. The workers need to choose the visual space that they want to share with the helper, and angle the phone's camera accordingly.

	HUDs	Mobile Phone
Worker's head and face	No	Yes, but dependent on worker
Worker's body and action	Yes, close-up of hand actions	Yes, probably from a distance
Task objects	Yes, close-up of focal objects of attention	Yes, but dependent on worker
Work environment	Only when it is the focus of attention	Yes, but dependent on worker
Worker's relative position	Only when the focus of attention reveals the position	Yes, but dependent on worker

Figure 1. Types of visual information provided by the two technology types.

There are four main types of visual information that are transferred during collaborative processes: worker's head and face, workers' body and action, task objects and the work environment [4], and for a dynamic task, the worker's relative position. Figure 1 shows how the approaches match up to the types of visual information. The HUD provides a close-up view of the worker's hands and focus of attention, partial views of task objects and the work environment, and the relative position of the worker. It does not provide a view of the worker's head or face. The video conferencing through a mobile phone provides a less detailed and less stable view, and requires the workers to define the visual space that they want to share, implying more effort is needed at the worker side. However, this also grants the freedom for the worker to choose the information they want to share.

In our study, we hypothesized that for a static task, the phone will outperform the HUD in efficiency (task completion time) and effectiveness (error rate). Given a dynamic task, we hypothesized that the HUD will achieve better performance.

METHOD

We conducted a 2 (HUD vs. mobile phone) x 2 (static vs. dynamic environment) fully between participant design which created four different conditions: **1.** mobile phone/static **2.** HUD/static **3.** mobile phone/dynamic **4.** HUD/dynamic.

Participants

The participants of this study were undergraduate and graduate students from the University of Wisconsin-Madison of various fields of study from engineering to education to design studies (81.6% males). A total of 32 participants were recruited. All but two participants owned a smart phone and every participant owned a laptop computer. Three participants reported that they felt uncomfortable using video conferencing while all other participants reported some level of comfort. None of the participants had prior experience building the toy cart used in the experiment. Their participation was completely voluntary.

Technology Type

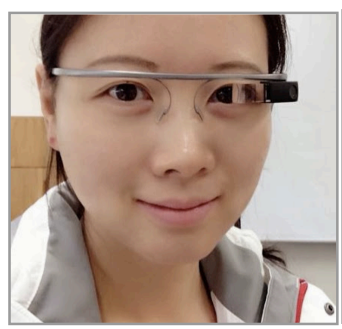

Figure 2: Technology types: (a) Google Glass and (b) iPhone

For the experiment, two different types of technology that facilitate remote collaboration were used (as shown in Figure 2a and 2b): Google Glass and the iPhone. Google Hangout, an app available for both technologies, was utilized to support the video conferencing. This is important, as it kept the interface somewhat standardized for participants in both technology conditions. The worker was equipped with one of these technologies while the helper always used a laptop for video conferencing.

Environment Type

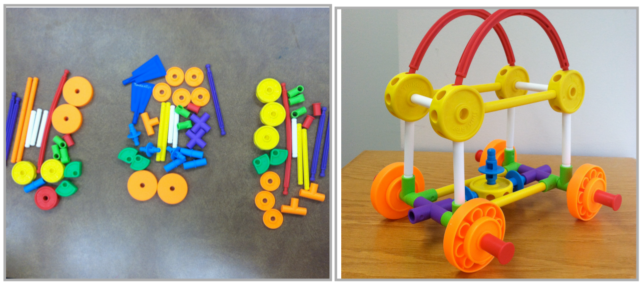

Figure 3: (a) Parts given to worker and (b) Completed task (toy cart)

In the static tasks, the parts seen in Figure 3a were gathered into one pile on a single workstation. For the dynamic tasks, the parts were separated into the groups shown in Figure 3a. These piles were placed on three workstations around the room. For both environments, participants aimed to build the toy cart pictured in Figure 3b.

Procedure

Participants were scheduled in pairs and were randomly assigned to the helper or worker role. The pairs were also randomly assigned to the technology and environment type. The helper was situated in a separate room from the worker and was given a sheet with pictures of the completed task. Before the task began, the helper and worker were briefed on and how to use the given technology type and were told the regulations associated with the environment type. For the static environment, the worker needed to remain at the same workstation. For the dynamic environment, workers were not allowed to move parts from workstations unless the parts had been assembled. Participants were also given a pre-experiment questionnaire collecting demographic information. The task (building the toy cart) was made up of two subtasks. The first task was to build a set of wheels while the second task included building the carriage and assembling everything together. Participants were given ten minutes to complete each subtask. They completed questionnaires after each subtask.

Measures

Three measurements were collected during the experiment: completion time, error rate, and perception of collaboration.
Completion Time: This measurement was recorded from the time the helper began giving instructions to the time they indicated that they had completed the task. Completion time was measured for both subtasks and total completion time was computed.
Error Rate: Errors were counted when a piece was incorrectly assembled or the wrong part was used. If the participants did not finish building the cart within the given time any remaining pieces that had not been properly assembly were considered errors.
Perception of Collaboration: After each subtask, participants were given a questionnaire to rate the perception of their collaboration on a five point Likert scale. Participants were asked to rate how well they worked together, how easy it was to be understood by their partner, and how easy it was to communicate and to assess the overall success of the collaboration.

RESULTS

Completion Time

The first measure of task performance we studied was the total task completion time for our participants. We hypothesized that phone users would outperform the HUD users during a static task and that the HUD users would outperform the phone users in a dynamic task. We conducted a two-way analysis of variance (ANOVA) to test whether the technology type and environment type affected total task completion time for the pairs and found a significant main effect of environment type $F(1,12) = 6.502$, $p = .03$ on total task completion time. This is to be expected, as participants will inevitable be slower in a dynamic task where they must move about the room. There was no significant or marginal effect of technology type or the interaction term between technology type and environment type on total task completion time.

Figure 4 shows a box and whisker representation of the completion time data for each of the four task conditions. Looking at the median scores, it appears that HUD users were slightly faster than phone users in the dynamic task, and slightly slower in the static tasks, just as we predicted. Note that the plots for the phone tasks have much greater variance than those of the HUD. Due to our small dataset, a few outliers in the phone tasks caused the variance in their completion times to be very large, which may explain why we did not find significant results for this measure. We attempted to correct for this by comparing the effects of our conditions on the logarithm of task completion time but still found no significant results. The issue here is primarily the large variance for the phone dynamic tasks. Out of our four data points, phone dynamic contains the two slowest pairs and the other two groups are among the top five or so in terms of speed. Two workers in this condition were efficient at carrying their helper around on the phone and giving them context properly, while the two slower groups left the phone at one station and spent much more time running around searching for components. This caused enormous variability for this group, which we believe is likely the cause for insignificant results for completion times in regards to technology types.

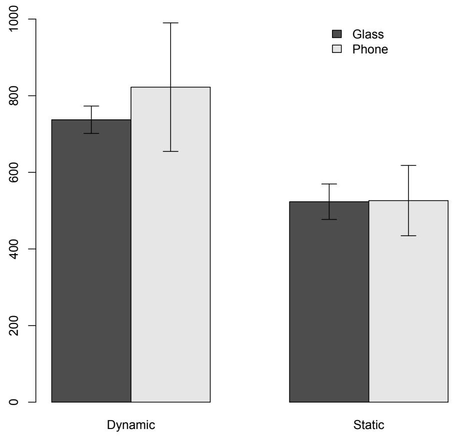

Figure 4: Task completion time given condition

Error Rates

The other piece of objective data on which we examined the effects of our conditions was the error rate during construction of our participants. For each subtask, participants told the test administrator when they believed they had finished the task. At that time the task timer was stopped, and the number of pieces on the model that were misplaced was counted to determine the number of errors. The errors found in the first subtask were not re-counted in the error count of the second subtask. With this data, we conducted a two-way analysis of variance (ANOVA) to test whether the technology type and environment type affected the rate of errors for the participant pairs and found a marginal interaction effect of technology type and environment type on error rate $F(1,12) = 3.261$, $p = .09$. Participant pairs using a phone to collaborate averaged 1.5 errors, while those using the HUD averaged 1.0. In a static environment, there were no errors by participant pairs using a phone, while the HUD participants averaged 2.0 errors per task. In the dynamic environment, however, HUD users made no errors, while phone pairs averaged 3.0 errors per task.

Subjective Measures

Ease of Understanding

On a Very Difficult (1) – Very Easy (5) scale for ease of understanding, both workers and helpers rated the HUD as superior to the phone. For the static task, helpers rated the phone on average 4.0 and the HUD 5.0. For the dynamic task, helpers rated both conditions on average 4.25 for ease of understanding. Workers rating of the phone averaged 3.75, while their rating of the HUD averaged 5.0.

We conducted a two-way analysis of variance (ANOVA) to test whether the technology type and environment type affected ease of understanding for the helpers and found a marginal main effect of technology type $F(1,12) = 3.429$, $p = .09$ and a marginal interaction effect of technology type and environment type on ease of understanding $F(1,12) = 3.429$, $p = .09$. We performed a similar test for the workers, and found a significant main effect of technology type $F(1,12) = 5.556$, $p = .04$ on ease of understanding.

Ease of Communication

On a Very Difficult (1) – Very Easy (5) scale for ease of communication, both workers and helpers rated the HUD as superior to the phone. For helpers, the phone averaged a score of 4.125 and the HUD averaged 4.75. For workers, the phone also averaged 4.125, while the HUD averaged 4.875.

We conducted a two-way analysis of variance (ANOVA) to test whether the technology type and environment type affected ease of understanding for the helpers and found a marginal main effect of technology type $F(1,12) = 4.412$, $p = .06$ on ease of communication. A similar analysis for workers found a significant main effect of technology type $F(1,12) = 6.000$, $p = .03$ on ease of communication.

Perception of Overall Success

On a Very Unsuccessful (1) to Very Successful (5) scale for the overall success of the collaboration, workers rated the HUD as superior to the phone. The phone averaged a score of 4.5 for overall success, while the HUD averaged a perfect 5.0. We conducted a two-way analysis of variance (ANOVA) to test whether the technology type and environment type affected ease of understanding for the workers and found a significant main effect of technology type $F(1,12) = 8.000$, $p = .02$ of technology type on success of collaboration.

DISCUSSION

Our results show that the environment type and technology type together had an influence on participants' rate of errors. Given a static task, the phone outperformed the HUD in terms of errors with an average error rate of 0.0 versus 2.0, while in a dynamic task the HUD far outclassed the phone with an average error rate of 0.0 versus 3.0. Although our results suggest an effect of technology type on completion time, we did not find a significant correlation.

Our subjective results were largely unaffected by the environment type – the HUD was simply rated better than the phone regardless of the environment in the majority of the measures for both workers and helpers. This is probably partially due to the novelty effect of participants getting to interact with Google Glass before it is available to the public. However, the helpers, who did not get to use Glass for the task, still gave it higher ratings than the phone, on average, for ease of understanding and ease of communication. This implies that while the scores may have been inflated due to this novelty effect, there is still solid evidence supporting the HUD's superiority over the phone for remote collaboration. In sum, results for completion times were inconclusive, the phone outclassed the HUD in terms of error rate for static tasks, and the HUD was superior to the phone in all of the subjective measures.

CONCLUSION

Our results show great support for the superiority of newly developed HUDs in both task environments. Other than an increase in the error rate for static tasks, the HUD outperformed the phone technology in all subjective measures, and the results suggest that it also reduces the total task completion time for dynamic tasks while maintaining the completion time for static tasks. These results conflict with those found in Fussell, S.R., *et. al.* [4], and the likely reason for this is the development of the HUD technology itself. Their study was conducted ten years ago with a very clunky HUD used for the experiment. The HUD we used is lightweight and efficient, and seems to promote effective remote collaboration in both static and dynamic task environments.

ACKNOWLEDGMENTS

Thank you to Professor Mutlu and our classmates in CS 770 for their valuable feedback and suggestions throughout the project. We would also like to thank our participants and the HCI lab for giving us access to Google Glass.

REFERENCES

1. Ford, C. E. Collaborative construction of task activity: Coordinating multiple resources in a high school physics lab. *Research on Language and Social Interaction*, 32 (1999), 369-408.
2. Goodwin, C. Professional vision. *American Anthropologist*, 96(1996), 606-633.
3. Anderson, R.E. Reed, K.B. and Peshkin, M.A. Physical Collaboration of Human-Human and Human-Robot Teams. *IEEE Trans. on Haptics,* 1, 2 (2008), 108-120.
4. Fussell, S. R., Setlock, L. D., & Kraut, R. E. Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks. In Proc. SIGCHI 2003, ACM (2003), 513-520.
5. Daly-Jones, O., Monk, A. & Watts, L. Some advantages of video conferencing over high-quality audio conferencing: fluency and awareness of attentional focus. *International Journal of Human-Computer Studies,* 49 (1998), 21-58.
6. Sutherland, I.E. A head-mounted three-dimensional display. *Fall Joint Comput. Conf. AFIPS Conf. Proc.* 33(1968), 757–764.
7. Rolland, J., Hua, H. Head-Mounted Display Systems, http://3dvis.optics.arizona.edu/publications/pdf/EOE_HMDsystem_Rolland_Hua_05.pdf
8. Goldman, D. Google unveils 'Project Glass' virtual reality glasses, *Money CNN*(2012)
9. Thomson, R., & Lynn, J. The benefits of using head mounted displays and wearable computers in a military maintenance environment. In *Education and Management Technology (ICEMT), 2010 International Conference on, IEEE* (2010, November), 560-564.
10. Stapleton, C., Hughes, C., Moshell, J.M., Micikevicius, P., Altman, M. Applying mixed reality to entertainment. *IEEE Comput.* 35, 12 (2002), 122-124.
11. Gurevich, P., Lanir, J., Cohen, B., & Stone, R. TeleAdvisor: a versatile augmented reality tool for remote assistance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2012, May), 619-622, ACM