Project 2: Head-Mounted Displays Design – Bystander Perception

NATHAN DASS, Georgia Institute of Technology
ARBER MUHAREMI, Georgia Institute of Technology
TYLER WHITFIELD, Georgia Institute of Technology
MALCOLM HAYNES, Georgia Institute of Technology
THAD STARNER, Georgia Institute of Technology

Our team has developed a user study to test how notifications to head-mounted displays impact conversations and interactions. Specifically, we considered how the horizontal displacement of the notification and the length of the notification affect the flow of a conversation. This paper will highlight the key features of our finalized experimental design as well as discuss the results from the many experiments conducted.

Additional Key Words and Phrases: head-mounted displays, horizontal displacement, bystander perception

1 INTRODUCTION

For part one of this project, a major goal of was to investigate the optimal configuration for head-mounted displays. The different configurations we are most interested in are the horizontal displacement of notifications and how the user might receive notifications (e.g., always allow notifications or explicitly determine if they want to view a notification).

We set out to design and conduct a user study with the various horizontal displacements for head worn displays. During the user study, our goal was to consider both the user's perspective and the bystander's perspective and gain insight into how head-mounted displays impact interactions.

For part 1 of this project, we concluded that there are multiple avenues for conducting the study. Therefore, for part two of the project, we knew we were going to have to iterate through various designs of the experiment to get an experiment that would produce good results to analyze the bystander's perception of distraction.

Part 2 of the project focuses on conducting multiple experiments and try to determine the optimal angle for notifications on head-mounted displays. The major takeaway after conducting these experiments is that it's not the angle that makes the biggest difference to perceived distraction, but rather length of the notification or how long the person must look at a given angle.

2 BACKGROUND

DOI: 10.1145/1234

The unique part of our project is that there has not been any published work on a study of this nature. Head-mounted displays are still a new field. There have been some studies on vertical displacement of notifications, but not on horizontal displacement. However, the need for this type of study is very important and could potentially help shape the direction of the development of head-mounted displays in the future.

2.1 Other Studies

There have been a couple of studies done on bystander perception that our team has read about. The first is titled “Facecard: Avoiding interruption when discovering names in small groups” [1]. The goal of the study was to “assess bystanders’ perception of using name identification systems during a meeting” [1]. The study consisted of a small group of people talking about past internships. The experimenter in the group was using Google Glass and a phone to use the identification system to try and identify the people in the conversation. The results were that the bystanders felt more comfortable when “Facecards”, which includes a picture and information of person as seen in Figure 1, on the phone were used, but when it came to using facial recognition it was more natural when the experimenter used Glass.

Fig. 1. A prototypical example from the perspective of a Google Glass user of what information the “facecard” would show for an individual [2].

The other experiment is titled, “Documentation in a Medical Setting: Effects of Technology on Perceived Quality of Care” [3]. This study sought to determine the perception of using different technologies in a doctor’s office setting. When there was no explanation of how the doctor was using the different technologies, people were least likely to go to a doctor using a desktop computer to record information. However, when participants were told that the doctor was using each technology to record information the participants were much more satisfied with the doctor’s use of wearables (like Glass) and desktop computers.

Both studies provide interesting insight into the bystander's perception of wearable technology, but the studies don't answer the question of horizontal displacement and don't provide a good measure of how receiving notifications affects casual conversation and interactions.

2.2 Our Study

Our solution reaches into these areas of studies where no other studies have reached. Our project provides a way to easily test how notifications viewed at different angles affects the perception of distraction for a bystander. Our experimental design also provides a quantitative way of determining bystander perception that allows for quick results that can be analyzed to help develop future iterations of the experiment. This rest of this paper will go into detail about how our experimental design achieves these outcomes.

3 TECHNOLOGIES AND IDEAS EXPLORED

To simulate the head-mounted display experience, we used an array of tablets placed at different angles. This allowed us to control the horizontal displacement of notifications received. With this configuration in mind, we had to develop a system to send a message to a specific tablet for a certain duration. We decided to build a website[1] based on the Flask microframework. Figure 2 shows how the website can send a message to a specific device during the experiment.

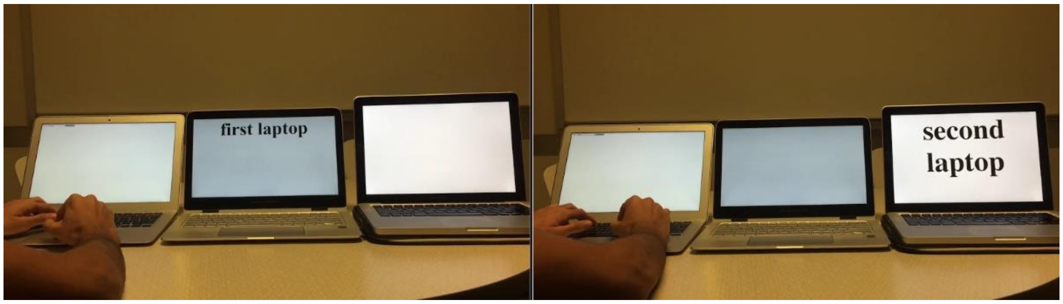

Fig. 2. The laptop on the far left is the master/admin device that can decide which display will receive a specific message in real time as well as clear currently displayed messages.

3.1 Message Length

Initially, we included messages with varying lengths in the experiment since we wanted to test for the optimal combination of message length and horizontal displacement. We had three categories of message lengths: messages that take about 1 second to read, messages that take 2-3 seconds to read, and messages that take 4-6 seconds to read. To make the experiment simpler, we eventually standardized the message length at messages that take about 1 second to read. This allowed us to focus on the optimal horizontal displacement of notifications.

3.2 Tablet Angles

First, we used 5 degrees, 10 degrees, and 15 degrees as the angles of the tablets that received notifications. We realized that putting a tablet at 5 degrees would lead to the tablet being blocked by the bystander at times, so we changed the tablet angles to 10 degrees, 20 degrees, and 30 degrees. After we standardized the message length, we realized that 10 degrees, 20 degrees, and 30 degrees did not show a difference in bystander perception. Once we realized this, we increased the tablet angles to 20 degrees, 30 degrees, and 40 degrees.

3.3 Eye tracking

We explored the idea of tracking the eyes of both people talking. By tracking the eyes of the bystander, we might be able to determine when the bystander gets a feeling that the user is not paying attention to them based on facial expressions and eye movements. On the other hand, tracking the eyes of the user would ensure knowing that the user read the message of the notification.

3.4 Video splicing

To get quantitative results, we decided to split the full recorded video into smaller video clips that are around 35 seconds long. We could play back to the bystander and gauge their perception of each video clip. Since the video clips were not the same duration, finding a solution was slightly more complicated. The duration of the video clips depended on the speed of the server we were using to host the Flask website, so we added a log file which we could use to determine the start and stop time of each of the 24 video clips relative to the original, full video.

3.4.1 Manual splicing. Initially, we would manually split the fully recorded video into the video clips we needed. This was a reliable way of obtaining the clips, but was obviously very tedious and time consuming. With this method, we could not gauge the bystander's perception right after the experiment since it would take too long to obtain the video clips.

3.4.2 Splicing with MoviePy. Eventually, we found a Python library called MoviePy [4] that had functions that would take in the path to a video file, a start time in seconds, and an end time in seconds and output a new video file that was a clipped version of the original video from the start time to the end time. MoviePy had wrappers to FFMPEG [5] functions, which made the actual process of clipping the video very fast. By using MoviePy, we could figure out a way to programmatically create all the video clips we needed in a timely manner. This allowed us to measure the bystander's perception within 5 minutes of the original conversation ending.

4 EVALUATION/DEMONSTRATION

The main goals for this project were to iterate through our experimental design and obtain initial results. We went through five iterations of our experimental design and collected qualitative and quantitative results at each iteration, which we used to create the next iteration. Below are the iterations of our experimental design as well as the initial results.

4.1 Iterations

During the experiment, there will be two people involved, the user and the bystander. During the experiment, the user and the bystander will be having a normal conversation that will be recorded. Afterwards, this conversation will be judged based on how interrupted the conversation felt. Before the experiment starts, the user will be told that they will be receiving notifications during the experiment. There will be an array of tablets surrounding the user and the bystander through which the experimenter will receive notifications during the actual experiment, which will correspond to the horizontal displacement of the notifications, as shown in Figure 3 below. For the sake of simplicity, we only placed tablets to the right side of the bystander from the user's perspective.

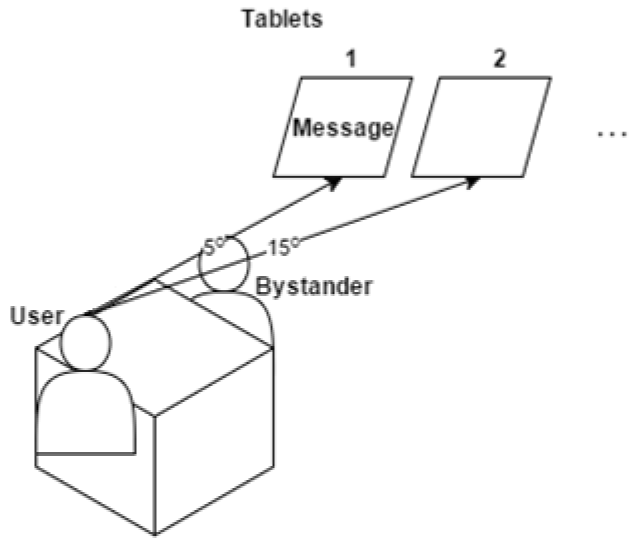

Fig. 3. Low Fidelity Experimental Design Mockup

4.1.1 Iteration 1. In the first iteration, we placed the tablets at 5 degrees, 10 degrees, and 15 degrees. We used variable message length and had 30 clips with 6 clips with messages in the 1 second category, 13 clips with messages in the 2-3 second category, 6 clips in the 4-6 second category, and 5 clips with no messages. In this iteration, the user and the bystander knew each other. The user was not one of us, but knew that there would be messages appearing on the tablets. The bystander did not know that there would be messages appearing on the tablets, but we also did not tell the bystander what the tablets were for. The conversation was recorded with a camcorder that was placed off to the side around 25 degrees to the right of the user.

We conducted one user study in this iteration. We did not implement video splicing, so we only obtained qualitative data from this iteration. We found out that long messages at 15 degrees was very noticeable. When one of the long messages appeared on the tablet at 15 degrees and the user was reading the long message, the bystander interrupted the conversation and said “Why are you looking at the camera?” which is a reasonable question because the bystander knew the approximate placement of the tablets and camcorder.

4.1.2 Iteration 2. In the second iteration, we kept the tablets at 5 degrees, 10 degrees, and 15 degrees. We used one message length and reduced the number of clips to 20 clips with 10 clips with 1 second messages and 10 clips with no messages. In this iteration, we added a Latin square for the order of the tablets displaying messages. This would reduce the impact that order might have on our results. We changed the user to one of us so we could assume that the message was read instead of having to test it. The conversation was recorded with a GoPro mounted on the bystander’s head. We also used the cover story that we were recording the conversation from different angles on the tablets to explain what the tablets were for.

We conducted one user study for this iteration as well. Qualitative results indicated that the interviewer did not notice the interviewer was distracted or receiving notifications even after we revealed the actual purpose of the tablets.

4.1.3 Iteration 3. In the third iteration, we changed the placement of the tablets to 10 degrees, 20 degrees, and 30 degrees. We changed the content of the messages to come from the MacKenzie Soukoreff phrase set [6] which is a well-known text message dataset used in other HCI studies and contains phrases that have no punctuation symbols and just a few instances of uppercase characters. We maintained the number of clips from the previous iteration. We implemented MoviePy video splitting in this iteration.

We conducted five user studies for this iteration. In the first three user studies, the user did not know the bystander, the user was one of us, and the bystander did not know the real reason for the tablets. In the fourth user study, the bystander knew the user. In the fifth user study, the bystander knew the user and the bystander was our faculty mentor, so he knew exactly what was going on in the experiment. For user studies 1-4, qualitative results showed that the bystander did not notice the user was distracted or getting messages. Even in the 5^{th} user study when the bystander knew exactly what the tablets were being used for in the experiment before the experiment started, the bystander only noticed the user looking at 4 out of the 10 messages. After revealing the actual intention of the study, bystanders told us that the cover we were using was not very convincing and still left them guessing what the tablets were for. Unfortunately, there were technical difficulties in each of the user studies, so we were not able to obtain quantitative results in this iteration.

4.1.4 Iteration 4. In the fourth iteration, we changed the placement of the tablets to 20 degrees, 30 degrees, and 40 degrees. We changed the number of messages from 10 to 12, so the number of clips increased to 24 total (12 with messages and 12 without messages). In this iteration, we used two cover stories. When the bystander first walks in, we tell them that we are using the tablets to record the conversation from different angles and we're going to analyze the videos later. After doing the initial conversation, we tell the bystander that we're actually doing a study on the correlation between memory and distraction. We initially decided to use two layers of deception because we didn't want the initial conversation to be less natural and we thought that using the original cover story in the beginning would keep it natural.

We conducted two user studies for this iteration. In the both user studies, the user did not know the bystander, the user was one of us, and the bystander did not know the real reason for the tablets. In the first user study, we used an older version of the survey which had two questions on a scale of 1-5 per video clip instead of one question on a scale of 1-7. The two questions were "How distracted did the person seem?" and "How confident are you in your answer?". We realized this mistake in the second user study and changed the survey questions before the survey portion of the second user study. We changed the questions to one question which asked the bystander to rate each clip on a scale of 1-7 on the statement "I believe my interviewer was engaged in the conversation.".

4.1.5 Iteration 5. The fifth iteration is the current iteration of the experimental design. The only difference from the fourth iteration and the fifth iteration is that we decided not to use the double deception and instead just use the correlation between memory and distraction cover from start to end. After the bystander fills in the survey, we tell them the actual point of the survey and debrief them. During the debrief, we asked people if the cover was believable and they all said it worked fine for this experiment and didn't give away what was going on.

We conducted two user studies for this iteration. The first user study went fine and used the one question per clip on a 1-7 scale survey. In the second user study, we misplaced the tablets and the angles were not actually at 20 degrees, 30 degrees, and 40 degrees so we had to throw that data out.

4.2 Evaluation

The experiment itself is designed to give us enough raw data to successfully discern the optimal horizontal displacement for a head-mounted display. This data will include which screens (corresponding to a certain horizontal displacements) received notifications, what notification was displayed, as well as whether the bystander detected the user was responding to notifications, or more generally, whether the bystander perceived the user lost focus in the conversation. To properly evaluate bystander perception, we must, at some time during the experiment, gather data corresponding to the bystander's opinion on whether the user received notifications. Figure 4 shows a graph of initial survey results from the two user studies that used the 1-7 scale.

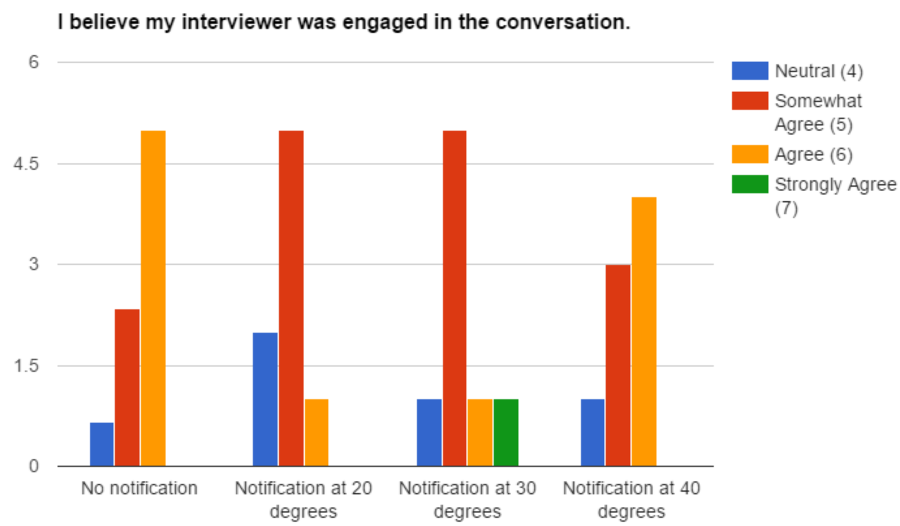

Fig. 4. Initial Survey Results

In the graph above, the x axis corresponds to the type of notification, the y-axis corresponds to the number of times a clip received a certain rating, and the bars correspond to the rating received. We think that these initial results do not show a significant difference in the distractibility of receiving a notification versus not receiving a notification. However, this graph shows the results obtained from only two user studies so we need more data before we can make any inferences.

5 CONCLUSION

5.1 Project Achievements

In Part 2 of this project, our team has iterated through our experimental design and we have reached an experimental design that we think has with high internal, external, and ecological validity. To use tablets and handheld devices to simulate horizontal displacement, our team created a website that we can use to send messages in real time.

With the data collected so far, we believe that messages that only take a second to read do not affect the distractibility of a conversation. This is interesting because it means that people perceive looking away for short durations as a normal thing and don’t associate it with being distracted to the point where it would interrupt the conversation. At the same time, we did find that if the message is too long, then the bystander will think the user is not engaged. This would imply that there is some sweet spot for message length in which the most information can be transferred without interrupting normal conversation.

5.2 Lessons Learned

A key lesson learned throughout the second phase of the project was how to rapidly iterate through a design. In this case, we rapidly iterated through an experimental design, but this process could easily be extended to software, hardware, prototyping, and more.

5.3 Next Steps

The next steps for this project will be to automate most the admin side of the user studies. This will allow us to rapidly scale up the number of user studies we perform, so we can get results from enough people to see if the horizontal displacement of notifications influences bystander perception.

5.3.1 List of High Level Tasks to Complete.

- Automate the admin side of user studies
- Collect data from at least 24 user studies
- Analyze results
- Compile results into a formal paper

REFERENCES

[1] Starner, Thad. "Facecard: Avoiding Interruption When Discovering Names in Small Groups." (2016): 1-7. Web. 9 Mar. 2017
[2] *Paul Smith Facecard*. Digital image. *Google.com*. Google, 5 July 2013. Web. 10 Mar. 2017.
[3] Deblasio, Julia, and Bruce N. Walker. "Documentation in a Medical Setting: Effects of Technology on Perceived Quality of Care." *PsycEXTRA Dataset* (2011): Web.
[4] "User Guide." User Guide – MoviePy 0.2.3.2 documentation. Web. 17 April 2017.
[5] "FFmpeg." FFmpeg. Web. 17 April 2017.
[6] "Phrase Sets." Phrase Sets for Evaluating Text Entry Techniques. Web. 17 April 2017.