HMD-Based Navigation for Ventriculostomy

Maia Stiber

CIS II Final Report

***Mentors**: Ehsan Azimi, Peter Kazanzides, Chien-Ming Huang, Dr. Judy Huang, and Dr. Camilo Molina*

Introduction

A *Ventriculostomy* is a procedure conducted by neurosurgeons bedside to relieve pressure in the brain. This procedure is done by drilling a burr hole and inserting a catheter into a ventricle to drain cerebrospinal fluid. The target of this procedure is the Foramen of Monro. Surgeons largely rely on experience and spatio-temporal reasoning, because they look at a CT scan to guide this procedure. In addition, because patients have various anatomies, the Foramen can be shifted. This results in misses in about a third of insertion attempts. That's not ideal because generally this procedure is time sensitive as pressure in the brain continues to build during failed attempts and also repeated insertion attempts can cause damage. There needs to be a way to eliminate the additional attempts.

Current HMD System

There is already a HMD system, created by my mentors on this project, that addresses this issue. See Figure 1.

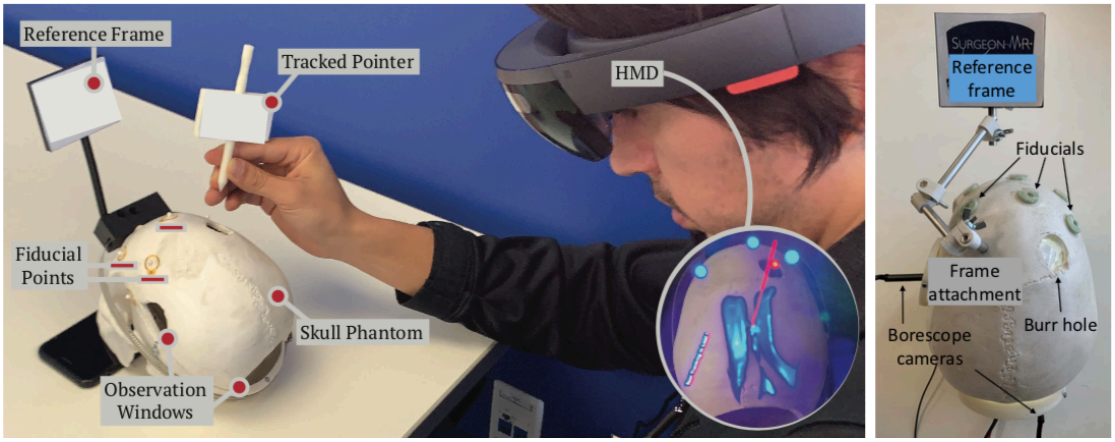

Figure 1: Set up of the AR navigational system that aids neurosurgeons in a ventriculostomy procedure.

It provides visual guidance for the catheter insertion path by overlaying the target of the procedure on the patient's head as well as the patient's anatomy and a guideline for the catheter insertion. This system uses Hololens to project all of the visualizations.

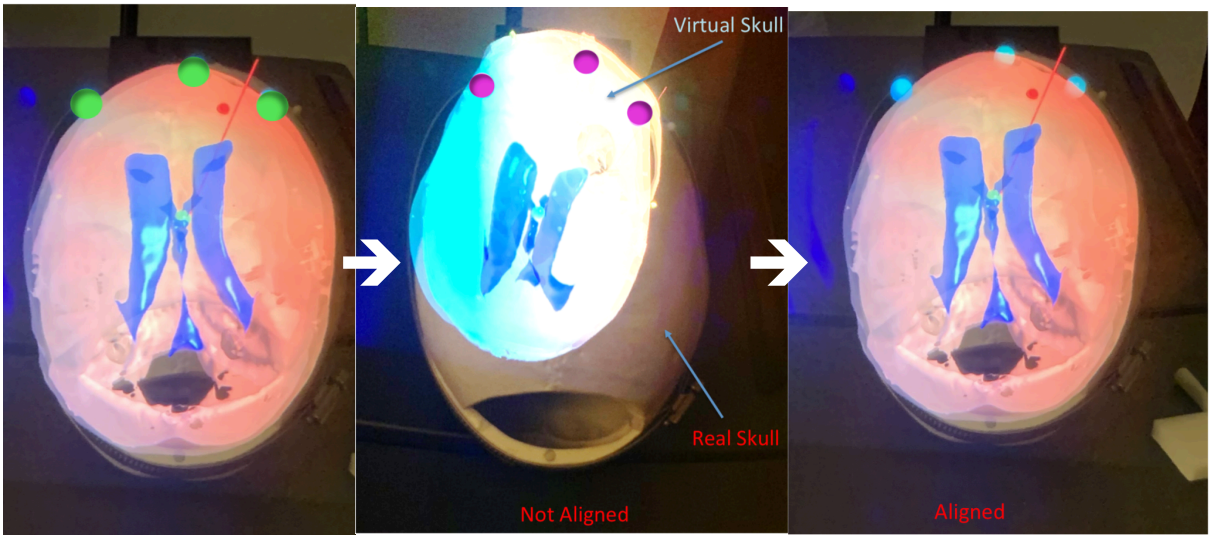

Figure 2: Workflow of the registration process as see through the Hololens.

The overall workflow of this current system is: calculating the wearer's interpupillary distance (provided by Hololens), registering the CT scan to the patient's head, and then planning the path at which the surgeon would insert the catheter along. The process of registering involves touching a tracked pointer to the fiducials and then recording the location using voice commands. The spheres in the hololens turn green to show that the fiducials are registered. The next step in the registration process is to use the commands "align" and "register" so that the visualization is an exact overlay of the patient's anatomy. This is then followed by the path planning process involving recording the desired start location of insertion and then having the system project a path from that start point to the target to insert the catheter along. The voice commands used for this step are "plan" and "complete."

Goals and Aim

The goal of this project is to evaluate the current AR-Guided Ventriculostomy System and to provide some adaptive assistance when inserting catheters. The aim of my project is to see if the AR-guided ventriculostomy is preferred by neurosurgeons, aiding in catheter alignment.

Materials and Methods

A user study was set up to evaluate the AR system and achieve the goal of the project through a mock ventriculostomy procedure.

User Study Design

Hypothesis: AR-Guided ventriculostomy increases accuracy and decreases mental task load.

Figure 3: Left: Overall set up of the phantom used in the user study. Right: "Brain" of phantom showing the locations of the three targets.

This is a within-subject study meaning that the participants conducted catheter insertions with and without the AR system. A phantom* was created as the platform for the trials, Figure 3. The phantom has registration fiducials on the front of the skull, a sample burr hole on the top of the head, and openings on the side for borescope cameras to aid in insertion accuracy calculations (specific later). Inside the plastic skull is the "brain" of the phantom. The "brain" is made out of gelatin and there are three balls embedded in it, where the possible shifts/locations of the Foramen of Monro are: one in the middle of the brain (normal location) and two shifted. These balls are the sample targets for this procedure, so each participant will do three trials per condition. In order to use this phantom, a CT scan was taken and then doctored so that the foramen is at each one of those balls.

Study Procedures

After the participants filled out the consent form, they were asked to fill out a demographics questionnaire. The order of the conditions and the targets within those conditions was randomly determined before each study. For the trials using the baseline condition (without AR) was first, the participants were introduced to what the task was. Then for each trial for the baseline conditions, they were shown the corresponding CT scan and shown how to read the scan. They were then told to not insert the catheter more than 6cm below the upper skull line. For the trials using the AR system, the participants were first asked to watch a tutorial video explaining how to use the system. The participants then followed the workflow described in the Introduction section to register the HMD system to the phantom's skull. Then for each trial, they would plan a path and insert the catheter. After both conditions and trials were done, participants were asked to fill out additional questionnaires about their experience.

* The phantom was created before I started the project.

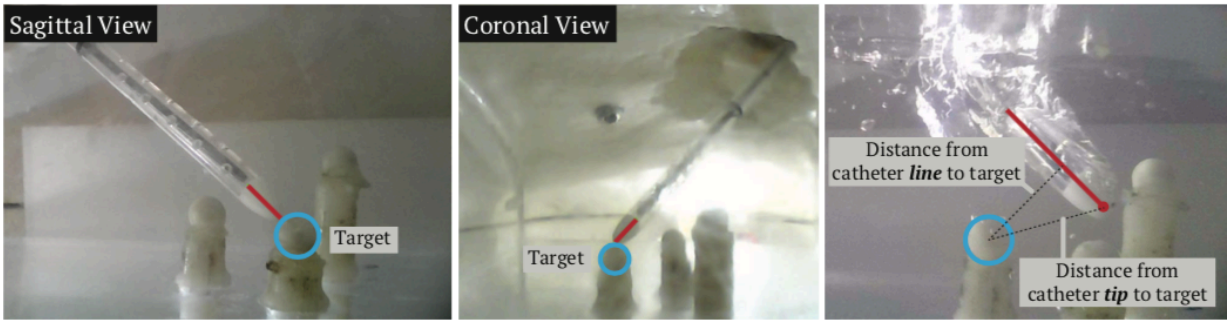

Figure 4: Visual description of the two task accuracy measures. Left two photos show the definition of "distance between catheter tip and target." Right photo shows the definition of "distance from catheter line to target."

Participants

A total of 10 participants were convenience sampled due to the virus, with ages ranging from 21 to 35 (M = 25.44, SD = 5.11). All of them had engineering or medical backgrounds and were somewhat familiar with mixed reality devices (M = 2.7, SD = 0.82 in a 5-point Likert scale with 1 being no experience and 5 being a lot of experience). Only one of the participants had done a Ventriculostomy before: the one neurosurgeon that was part of the study. Due to the virus, we weren't allowed to conduct a study with all neurosurgery residents.

Measures

Data was recorded through video of the attempts, using a camera that provided an external view of the process and two borescope cameras, and with questionnaires. The two borescope cameras were used to calculate task accuracy and the questionnaires were used to understand the usability of the system and mental workload. An expert coder annotated the external camera videos and quantitative measures (workflow timing, re-registration times) were generated from them.

Task Accuracy.[†] The two borescope cameras were used to determine these measures. They have an average error of 0.3mm. There were two measures for task accuracy: distance between catheter tip and target and the distance between catheter line and target. This means finding the closest distance between the catheter line and target; see Figure 4. The reason we have the second measure is that the HMD system did not provide any guidance of how far to insert the catheter. The participants just needed to know that they had to insert the catheter in 6cm. So, during the user study, users would insert the catheter hit the target and keep inserting past it forgetting to only insert it in 6cm. The reason that we felt this is a valid measurement is that, in the actual procedure, if the catheter hits the target, liquid will come out.

Workflow Timing. This measure illustrates the additional time that is required by the AR system. This was calculated from coding external camera footage. These measures were separated by step in the workflow: registration, planning, and insertion. The time taken to do registration was calculated from when the participants touched the tracked pointer

[†] The system was developed and task accuracies were calculated by someone else. I just did the data analysis associated with them.

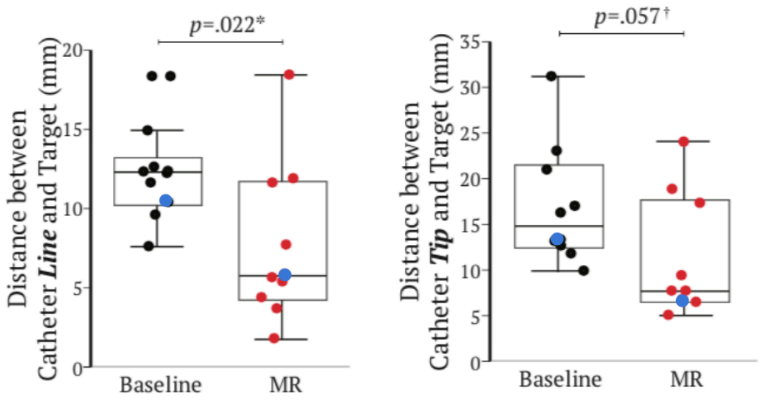

Figure 5: Results for the task accuracy measurements. The blue dots are the neurosurgeon's data points.

to the fiducials to the last verbal command of the registration process. Planning time was determined from when the tracked pointer was hovering over the start position of the catheter to the last verbal command of the planning process. The insertion time was coded from when the participants picked up the catheter and lined it up to be inserted to when the participants confirmed that they thought that they had hit the target.

Re-registration Frequency. The measure was also determined by the coding of the external videos. This is broken down into the number of total re-registrations, due to projection shifting or failure to do it properly the first time, and re-digitizing specific fiducials.

Usability. To determine the usability of the system as perceive by the participants, they were asked to fill out a System Usability Scale (SUS) questionnaire [1]. SUS is a well-established 10-question scale where the higher the score is the more usable the system is.

Mental Workload. The NASA Task Load Index was well-established scale used to measure mental load of the workflow [2]. It is important to note that the lower the score the less mental workload it takes to use the system.

Results

Task Accuracy. One-way ANOVA was used to determine the statistical significance between two groups where the fixed effect was with or without AR and the random effect was the participants. For distance between the catheter line and target, there was significant measured difference, $F(1, 18) = 6.24$, $p = .022$, between the two conditions. The participants with the mixed reality aid were able to maintain a shorter distance to the target ($M = 7.63$, $SD = 5.00$) when compared to using the baseline setup ($M = 12.21$, $SD = 2.93$). This was also reflected through the neurosurgeon's average accuracy results: for AR, it was 7.7mm ($SD = 3.08$) and for baseline it was 10.4mm ($SD = 2.17$).

The measure for the distance between the catheter tip and target showed that there was a marginal improvement, $F(1, 18) = 4.14$, $p = 0.057$, in accuracy from non-assisted to assisted. Using AR, participants were able to achieve an average accuracy of 10.96mm ($SD = 6.61$) and without was 16.93mm ($SD = 6.52$). For the neurosurgeon, the mixed reality

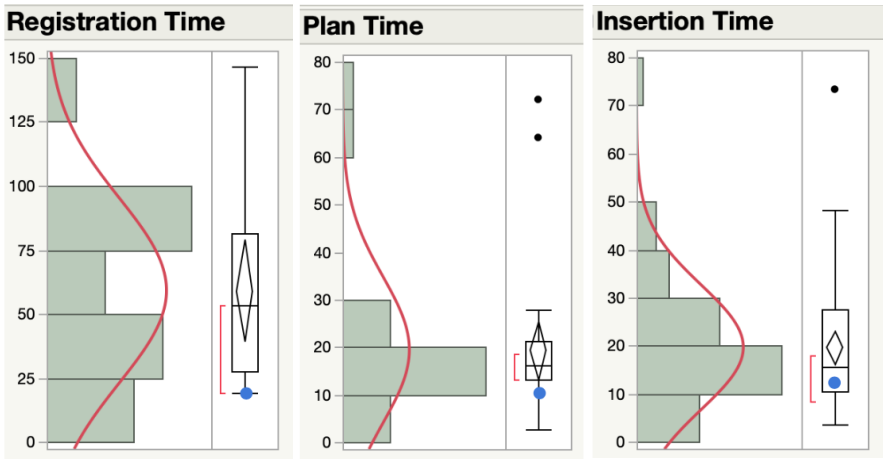

Figure 6: Distribution of workflow timing divided into the three steps of the AR guided Ventriculostomy procedure. The blue dots are the neurosurgeon's data points.

accuracy was 9.37mm (SD = 2.93) and baseline was 13.3mm (SD = 0.96). All of distribution of data points can be found Figure 5.

Workflow Timing. This measure, as mentioned above, was broken down into three parts (see Figure 6). For the registration step, between the Hololens and the phantom, the average time for registration was 52 seconds (SD = 35.39) and for the neurosurgeon was about 18.99 s. The second step was the planning part of the workflow. The average time was 19.44 s (SD = 15.35) and the neurosurgeon's time was 10.14s. Then for the insertion times for both conditions (baseline and with AR), the average time was 19.67s (SD = 12.93) and the surgeon's was 11.72 s.

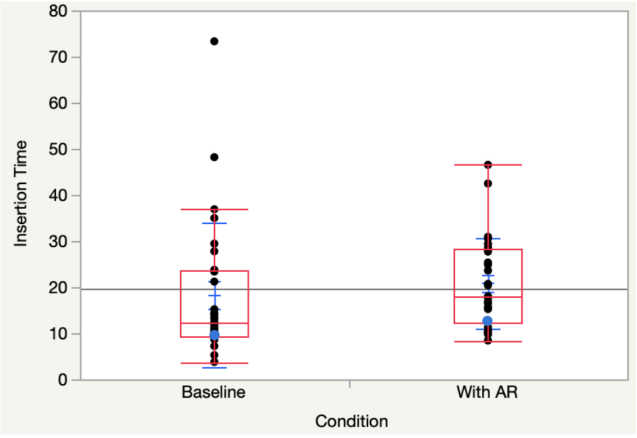

Figure 7: This shows the distribution of insertion times separated by condition. The blue dots are the neurosurgeon's data points.

If the insertion times were are separated by condition (see Figure 7), there is a noticeable difference, U = -1.76, p = 0.79, r = - 0.33, between the baseline (Mdn = 12.50), which is slightly faster, and the AR system (Mdn = 18.07). This difference was also reflected in the neurosurgeon's averages: baseline was 10.44s (SD = 3.38) and AR was 13.44s (SD = 2.75). AR insertion times were expected to take longer as the participants have to spend time trying to line the catheter up with the projected guideline. It is important to note that there

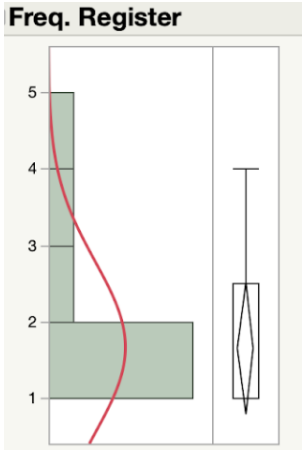

Figure 8: Distribution of the number of times participants need to re-register the Hololens to the phantom

does not seem to be any correlation between insertion time and accuracy, r(51) = -0.01, p = 0.89. This is the same case when the insertion times are separated by condition: baseline was r(24) = 0.063, p = 0.77 and r(24) = -0.059, p = 0.77.

Re-registration Frequency. Participants need to re-register the Hololens on average 1.67 times (SD = 1.12) with the averages per fiducial (3 fiducials) being around 2 times, shown in Figure 8. The general reasons for why the HMD needed to be re-registered were that the registration was done poorly in the first place: this was the first time for most participants using this system, and over time registration can shift. It is important to note that, for the person who had to re-register the system 4 times, this was due to the Hololens having to re-start. The neurosurgeon did not have to re-register at all.

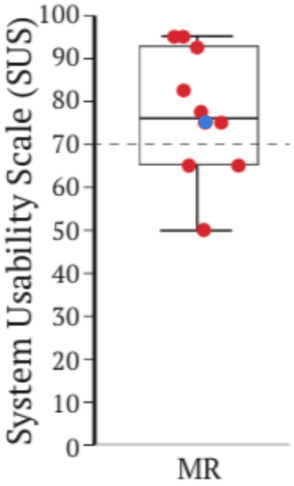

Figure 9: Distribution of how participants evaluated the usability of the system. The blue dot is the neurosurgeon's data point.

Usability. The average for this is the system was 77.25 (SD = 14.69); it's suggested that systems need a score of at least 70 to be considered usable [1]. This was also reflected in the neurosurgeon's evaluation of 75, shown in Figure 9.

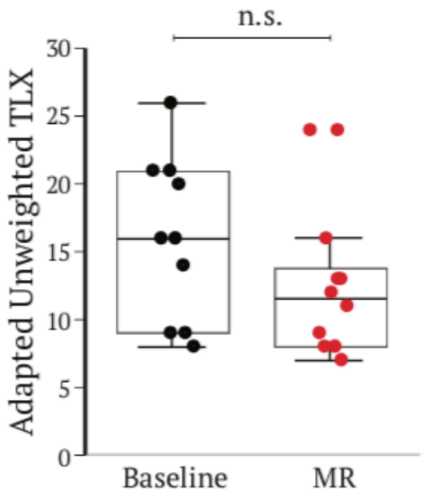

Figure 10: Distribution of how much mental workload the participants thought the two conditions took.

Mental Workload. There was no significant statistical difference, $F(1,18) = 2.44$, $p = 0.14$, in the mental workload between the AR system ($M = 12.1$, $SD = 5.04$) and baseline condition ($M = 16.00$, $SD = 6.07$).

Conclusion

In this report, the presented AR interactive navigational system for ventriculostomy was evaluated to determine if this system improved accuracy and decreased mental workload. A user study was run and showed that this system improved the accuracy of catheter insertions compared to baseline and added minimal additional time to the actual procedure. However, part of the hypothesis was disproved. The AR system did not decrease the mental workload required to do a ventriculostomy procedure. This AR system was also determined as usable by the participants, including the neurosurgeon.

Future work could consist of, first, finishing the script, which would adaptively prompt users to move their head when inserting the catheter. During the user study, we noticed that participants who moved their head around, taking advantage of the 3D guideline, appeared to be more successful than those who didn't. And so this script would prompt users to move their heads when inserting. In addition, a user study with all neurosurgery residents will need to be run; this was our original plan but we weren't able to do it because of the virus. The third one could be providing real-time catheter tracking on catheter alignment and insertion depth with real patient type scenarios. The last one could be to provide adaptive training for users, as users get better using the system they would not need as much feedback so it is closer to the real-life situation.

Acknowledgements

Thank you to Ehsan Azimi, Dr. Peter Kazanzides, and Dr. Chien-Ming Huang for helping me with the project and giving me advice. Thank you to Dr. Camilo Molina and Dr. Judy Huang for providing medical expertise and user study tutorial advice. Thank you to Zhiyuan Niu

for helping with the user study, providing the infrastructure for calculating task accuracy, and running around to finish the user study before labs shutdown. Thank you to Ruby Liu and Nicholas Greene, who worked on the system before me and also helped with the user study.

Lessons Learned

Other than the knowledge presented in this report, I learned about what Augmented Reality looks like in the field of medicine. In addition, I learned how to design and run a user study. While I have run a user study before, this was the first time I had designed one and conducted the data analysis portion of it.

Management

I met with my mentors, excluding the two neurosurgeons, biweekly. The two neurosurgeons I communicated with by email as needed.

	Original	**Revised**
Minimum:	• User Study results and Analysis of Data	• User Study results and analysis written as part of a submission to MICCAI 2020
Expected:	• Video Analysis Results • Catheter tracking in phantom	• Video Analysis Results
Maximum	• Script that provides feedback to surgeons on how well aligned and how close catheter is to the target during actual procedure	• Script that provides aid in depth perception with adaptive prompts based on wearer's behavior

Table 1: List of the deliverables for the project showing what the original ones were and what they had to be changed to due to the virus.

The results and discussion presented in this report are the minimum and expected deliverable. The revised maximum deliverable was not completed as some time was lost in the schedule due to having to revise the project because of the virus and spend additional time learning things that were initially planned for. See Table 1 for a complete list of deliverables for the project. In addition, the minimum deliverables were submitted as part of a paper to MICCAI 2020.

Bibliography

[1] Brooke, J.: Sus-a quick and dirty usability scale. Usability evaluation in industry 189(194), 4–7 (1996)

[2] Hart, S.G.: Nasa-task load index (nasa-tlx); 20 years later (2006)