Methodology and Study Procedure

For the purpose of evaluating our new hand gesture-based pointing method, we conducted a Fitts' Law study, comparing it with three other pointing methods. The first was a mouse, the second was a touchpad, and the third was a mouse with a 150 ms delay. The experiment was conducted as a within-subject study. Additionally, we had two more variables: the target size and the distances between the targets. The possible sizes were 30px, 40px, and 50px, and the possible distances were 100px, 200px, and 300px. Each combination of distance and size was repeated with 33 targets. The target positions were arranged in a circle with eleven points. Each new target appeared on the opposite side of the circle to ensure even spacing between targets. To that end, we also disregarded the first target of any new condition since it had a different distance from the last target compared to the rest of the condition targets.

The program opens in fullscreen mode on the screen, with a big circle on screen that, if clicked, starts the experiment. Each clicked target disappears, and the next one appears. When the last target is clicked, the program closes and saves the experiment logs in a CSV file. Each combination of test subject and input device gets its own CSV. Each entry in the CSV file consists of the entry ID, distance between the last and new target, size of the target, whether the log entry is valid, how long it took to click the target, and whether it was clicked successfully.

After the test with one input method was done, the program can be restarted with different parameters. The hand gesture recognition input is a separate program that needs to be started before the test is started.

Discussing Issues that occured:

There are a few problems that occurred during this assignment. The first was deciding which input method to use. It was given that it would be pose-based, but choosing which poses and gestures and what models to use turned out to be more difficult. Initially, we wanted to use the chest as the input area, one hand as the pointer, and the other as the mouse click. However, when occlusion occurred or tracked body parts got close to the edge, tracking became erratic, and sometimes an imagined second hand would flicker on the screen. Therefore, we abandoned this idea and instead utilized hand pose tracking with a pinching motion as a click.

This is where the second issue popped up. It took a lot of fine-tuning to get the Mediapipe hand pose detection model to work. Additionally, it turned out that, although it seemed logical for the index finger to be the position of the cursor, it was advantageous for the thumb to perform that role instead, since it moves less during a pinching motion.

The next issues were noticed when we tried out the input method in action. The first was a propensity for our input method to sometimes perform a rapid number of clicks. We took several steps to solve this. The first was that the finger distance needed to be under the threshold for a few frames before the click was registered. The second was to add a second, bigger threshold for releasing the click so that jitter at the edge of the click threshold would not be counted as multiple clicks.

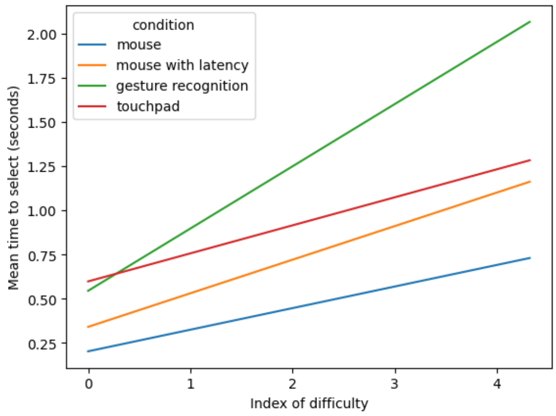
The threshold presented another issue: we could not track the distance from the camera, so what would not be a click up close would be one farther away. This locks users into a certain distance depending on how big their hands are. We first thought of factoring in the distances to other tracked points on the hand, but this proved not to be feasible, as rotating the hand would change that factor dramatically. In the end, we added a calibration sequence where the spread and pinched distances between the fingertips are measured. This allows users to set the distance they want to use the device at.

During testing, we noticed the issue of fatigue. The input method requires holding out a hand in front of oneself without any form of support, leading to significant strain over prolonged usage.

A final problem was that unconscious gestures of one's hand are also tracked. For example, scratching oneself would lead to unwanted repositioning of the cursor and clicks.

Results

Using the data collected during the study, we calculated the Index of Difficulty (ID=2D/W) for each datapoint, then performed linear regression on the data from each condition. These linear regressions yielded a linear function of the form t=a+b ·ID, modeling the relationship between selection time t and index of difficulty ID. These linear models are depicted in the figure below, and the constants a and b are reported in the table.



condition	a	b
mouse	0.202	0.122
mouse with latency	0.341	0.190
gesture recognition	0.544	0.352
touchpad	0.598	0.159

As is visible from the graph, gesture recognition generally lead to the highest selection times out of all tested input modalities. Notably, touch input has a slightly higher y-intercept, as target selection in the final selection phase necessitates lifting the finger from the touchpad and tapping down again, meaning that each “click” takes more time than any other condition. This disadvantage is quickly compensated by its superior speed in the initial selection phase. This high y-intercept prevented the touchpad from surpassing the “mouse with latency”-condition, which was otherwise clearly inferior to the lag-free mouse. With higher indices of difficulty, the slower movement speed and finnickier target selection of the mouse with latency might make it a slower input modality than the touchpad. Other than that, the mouse is by all metrics the fastest input device tested – which is not surprising. Its wide-spread adoption means that it was a device all of our participants were already highly skilled in using, but is also a testament to the popular recognition of its input efficiency.

There are several factors that might explain the poor performance of our implementation of mid-air gesture input. First of all, our participants were simply not used to using arm movement as a form of input, contributing to a lack of precision with the unfamiliar input modality.

Second of all, the position of the cursor on the screen was directly related to the position of the thumb in the camera image, meaning that the participants had to move their hand quite far to move the cursor compared to other forms of input.

Third of all, the position of the camera and its field of view meant that the participants had to lift up their whole arm from the table for most inputs. This not only lead to noticeable (and loudly communicated) fatigue and pain in the upper arm, but also to less stable hand movement, which had to be compensated by the participants being more careful and deliberate about their motions.