# A Comparison of Gaze-Based and Gesture-Based Input for a Point-and-Click Task

Dominic Canare, Barbara Chaparro[(✉)], and Jibo He

Wichita State University, Wichita, KS, USA
{dominic.canare,barbara.chaparro,jibo.he}@wichita.edu

**Abstract.** Alternative input devices to the computer mouse are becoming more affordable and accessible. With greater availability, they have the potential to provide greater access to information for more users in more environments, perhaps while also providing more natural or efficient interaction. However, most user interfaces are built to be mouse-driven, and the adoption of these new technologies may depend on their ability to work with these existing interfaces. This study examined performance with gesture-control and gaze-tracking devices and compared them to a traditional mouse for a standard Fitts' point-and-click task. Both gesture-controlled and gaze-tracking proved to be viable alternatives, though they were significantly slower and more taxing than the familiar mouse. In order to make effective use of these devices, researchers, designers, and developers must find or create control schemes which take advantage of the alternative devices' benefits while curtailing the drawbacks.

**Keywords:** Gaze · Gesture · Mouse · Leap motion · Eyetribe · Fitts' law

## 1 Introduction

The ever-present computer mouse has been a staple of computing for many years. This remains true even as the number devices grows to unprecedented levels. Cisco (2014) reports that 526 million mobile devices and connections were added in 2013, with smartphones accounting for 77 % of the growth. They predict that by the end of 2014, the number of mobile connected devices will start to exceed the number of people, and by 2018 there will be 1.4 mobile devices per capita. As more devices are added and integrated into work and play, more data is being generated as is the need to access it. Alternatives to the mouse have been created and suggest the possibility of more intuitive interfaces to provide better access for a variety of populations. Other devices solve more specific problems, such as allowing access to data in sterile environments, overcoming physical handicaps, and exploring big data [1]. Sterile environments exist for users in surgical settings, where accessing a keyboard and mouse would require rescrubbing, a process that can take significant time and is critical to patient safety [2]. Alternate input devices such as vision-based interfaces, have the potential to create more accessible devices for users living with disabilities [3].

The success of a new input device depends on many different factors. The longevity of the computer mouse has led to the development of user interfaces which are largely mouse-driven (e.g., a considerable amount of human-computer interaction can be described as simply pointing and clicking). The success of a new device may depend on its ability to transition users from these point-and-click interfaces to new control schemes, and part of that transition will require retrofitting new devices to point-and-click interfaces. This study examines the viability of new input devices using a point-and-click Fitts' Law task [4].

Performance is not the only factor contributing to the success or failure of a new device. Wachs et al. describe several requirements for gesture-based input devices - some of which can be generalized to all input devices [1]. As is often the case, cost is a factor, as the adoption of a new device is only plausible if the masses can afford it. Intuitiveness is another important factor to consider in a new input device. While the mouse has enjoyed many years as the preferred device without being the most intuitive, new input devices must now compete with users' experience, expertise, and preference for the computer mouse. Devices lacking intuitive control schemes are unlikely to compete. Requiring users to modify their appearance, wear special equipment, or alter their environment in order to use a new device will likely be seen as cumbersome and hinder a device's adoption.

The Leap Motion Controller is one alternative input device that relies on hand gestures as a means of input. The controller is sold as a stand-alone device and has been integrated into some laptop designs [5]. It is intuitive for users to physically manipulate tangible objects, and many of these physical manipulations have or approach one-to-one translations to a gesture-controlled interface. Depending on the sensing technique, gesture-based devices have the potential for a much larger bandwidth than a standard computer mouse [6]. The Leap Motion controller uses a depth-sensing camera with a high frame rate and level of detail. The depth-sensing method allows the device to detect gestures without the need for special markers or contrasting backgrounds. Because these gestures are performed in the air, the device is touchless, making it ideal for sterile environments.

The Eye Tribe gaze-tracking device is another alternative input device that is capable of approximating the location of a user's gaze using a high-power LED light source and high-resolution camera. The device is sold as a development kit for a price comparable to some computer mice [7]. Gaze-tracking has the potential to improve human-to-computer data transmission by opening an entirely new channel of communication. A person's gaze naturally indicates the area of their attention, and gaze-tracking equipment can help systems realize this data in meaningful ways [8]. One minor disadvantage with the Eye Tribe device is that it requires calibration for individual users. However, the calibration process is usually brief, and there are no other requirements for the user to begin using the device. Lastly, gaze-tracking is naturally touchless.

Gaze-tracking and gesture-control input devices have been studied by various researchers in the past. Zhai used gaze-tracking to provide context for actions [8]. More specifically, gaze information was combined with statistical analysis of the user interface to place the cursor in a predicted area of the user's intention. A study of this

system found that, with just ten minutes of training, performance was similar to standard mouse usage. Wachs et al. developed a gesture-controlled interface through which surgeons could browse radiological images in a sterile environment while avoiding many of the typical drawbacks (rescrubbing, change in focus of attention, etc.) [1]. Their usability tests indicated that users found the system to be "easy to use, with fast response and quick training times" [1, 323].

The aim of this study was to determine how well gesture-control and gaze-tracking devices could be used for point-and-click tasks using Fitts's Law performance measures when compared to the traditional computer mouse.

## 2 Method

### 2.1 Participants

Twenty-two college students were recruited for this study. Of those, 3 were used for pilot data to modify and optimize parameters for gesture control, gaze control, and testing procedure. Of the remaining 19 participants, 4 could not be properly calibrated to the gaze-tracking equipment, and their data for all devices was discarded. The remaining 15 participants (5 female, 10 male) aged 18–56 (M = 25.6, SD = 10.8) reported that they owned either a laptop or desktop computer and used a computer regularly throughout the day. Participants were compensated with course credit for their participation.

### 2.2 Materials

The primary hardware components of this study included a standard computer mouse, a Leap Motion Controller for gesture control, and an Eye Tribe for gaze control. An ASUS Zenbook Prime connected to an LCD monitor (1280x1024) was also used. Participants sat at a desk with an adjustable chair, and the monitor was raised to eye-level, approximately 2 feet away from the participant.

Software was created to control the mouse cursor using the gesture and gaze-tracking devices. For gesture control, participants moved the on-screen cursor by pointing a finger and moving it about in the space in front of them, above the desk, and parallel to the monitor. X-Y air space was translated to X-Y screen coordinates absolutely, so that a pointing finger positioned on the left of the desk would relocate the mouse to the left side of the screen. The angle of the pointing finger relative to the palm was also tracked and used as an adjustment to the final cursor location in all directions. A pilot test was performed to determine the best gesture for selection between poking, tapping, and looping. Drawing a loop was found to be the easiest gesture to perform consistently, but accuracy was a challenge. To overcome this, the system was modified so that when the participants were required to first hover over the location where a click was desired for 50 ms within a 3 pixel radius, followed by gesturing a loop or circle. Upon detecting a hover event, a small red circle was displayed to provide feedback to

the user, with the center marking the location where the selection would occur if the hover event was followed by a loop gesture.

Gaze location was sampled at approximately 30 Hz, and the cursor was positioned to the location of the participant's gaze. A selection event (mouse-click) was generated whenever the gaze was fixated on a 25 pixel area (approximately 0.59 visual degrees) for 500 ms, with the selection occurring at the average location of the fixation.

## 2.3     Procedure

Each participant was informed that the study was designed to evaluate the viability of alternative input devices to a traditional computer mouse. Prior to performing the experiment, participants completed a consent form and brief survey about their experience with computers and pointing devices.

The experiment itself consisted of 100 trials for each of the three devices (mouse, gesture, gaze). The order of devices was randomized for each participant. The trials were separated into 10 blocks with 10 trials in each. At the start of a trial, a circular target (the start target) was presented to the user with a 40 pixel radius in a random location on the screen. Once the subject selected the target, it disappeared and a new target (the stop target) appeared elsewhere on the screen for the user to select. Between trials there was a 1.5 s delay, and between blocks the experiment paused until the participant selected an area of the screen to indicate that they were ready to resume.

Within each block, the stop targets were one of ten combinations of sizes and distances. The size of the stop target was either large (120 pixel radius) or small (40 pixel radius), and the distance between the start and stop target was either 100, 250, 375, 500, or 750 pixels. The order of these 10 combinations was randomized within each block, and the angle between the targets was generated randomly. After completing the 100 trials for a device, participants completed the NASA Task Load Index survey to collect workload measures.

The index of performance, as defined by Fitts' Law [4], was the focal data for analysis. For an individual trial, the index of performance was calculated as a ratio of the difficulty of a trial to the amount of time required to complete it. More specifically, the time was measured between selecting the start target and selecting the stop target. Having a shorter movement time or a higher difficulty raised the index of performance, while a longer movement time or lower difficulty decreased the index of performance:

$$performance \; = \; difficulty/movementTime$$

The difficulty of a trial is a function of the stop target's size and distance from the start target. Targets which are further apart are more difficult than targets closer together. Likewise, smaller targets are more difficult to hit than larger targets:

$$difficulty \; = \log_2(2 * distance/size)$$

# 3   Results

## 3.1   Performance

A one-way repeated-measures ANOVA test was performed to compare the index of performance across all three devices. There was a significant difference in performance between all devices, $F(2, 28) = 196.49, p < 0.01$. Participants performed best with the mouse (M = 4.56, SD = 0.792), followed by the gaze-control (M = 2.321, SD = 0.352) and then the gesture-control (M = 1.56, SD = 0.20), and all of these differences were significant at $p < 0.05$ (See Fig. 1).

A 2-way (3x10) repeated-measures ANOVA was performed to determine whether participants improved with experience across the three input methods. There was a significant main effect of input method, $F(2, 28) = 199.41, p < 0.01$, experience, $F(9, 126) = 3.79, p < 0.01$, and a device x time interaction, $F(18, 252) = 1.86, p < 0.05$. Post hoc comparisons of the interaction showed that the mouse had a significant improvement in performance from the first block to the last, $F(1, 14) = 6.788, p < 0.05$, as did the gesture controlled device, $F(1, 14) = 7.196, p < 0.05$. The gaze-tracking device showed no significant change in performance amongst any of the blocks in this experiment.

Two types of errors were tracked during experimentation, and repeated-measures ANOVA tests were performed for these as well. A miss-click was counted anytime a participant performed a selection with the cursor outside of a target. The difference in miss-clicks was significant between all devices, $F(2, 28) = 41.238, p < 0.01$. The mouse had the fewest miss-clicks (M = 4.800, SD = 4.507), followed by gesture-control (M = 9.867, SD = 6.379), and then gaze-tracking (M = 69.533, SD = 35.377). Post hoc analysis indicates a significant difference between mouse and gesture, $p < 0.05$; mouse and gaze-tracking, $p < 0.01$; and between gesture and gaze-tracking, $p < 0.01$ (See Fig. 2).
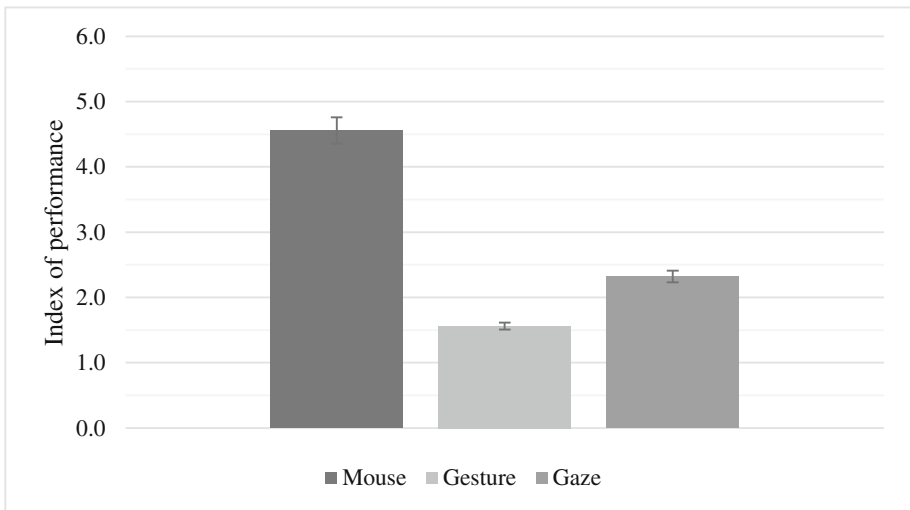


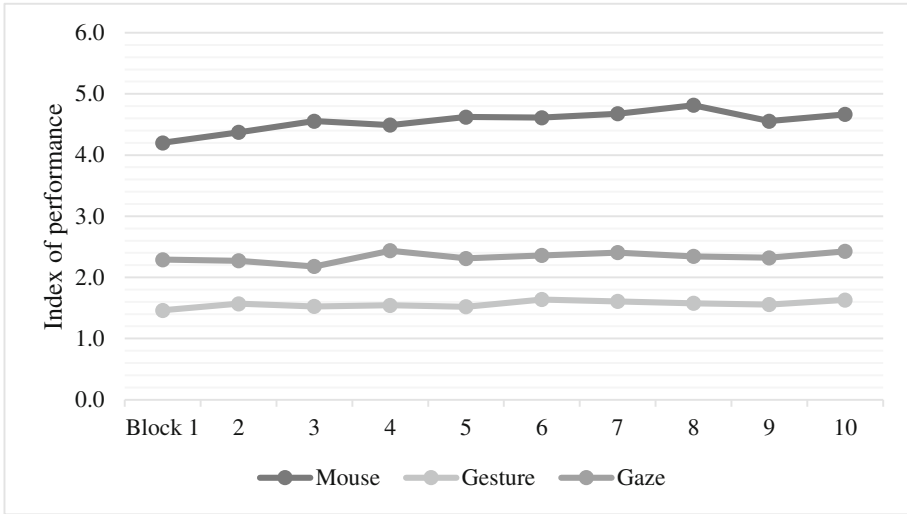**Fig. 1.**  Comparison of performance across input methods (± 1.0 SE)

**Fig. 2.** Performance over time

A failure was counted whenever a participant could not successfully select a target within 10 s. The difference between devices for failures was also significant, $F(2, 28) = 8.966$, $p < 0.01$. No participant recorded any failures with the mouse. Gesture-control did yield some failures (M = 1.133, SD = 1.598), but was again eclipsed by gaze-tracking (M = 4.267, SD = 4.803). Post hoc analysis indicates a significant difference between mouse and gesture, $p < 0.05$; mouse and gaze-tracking, $p < 0.01$, and between gesture and gaze-tracking, $p < 0.05$ (See Fig. 3).

### 3.2   Target and Distance Size

A 3x5 repeated-measures ANOVA was performed to determine if there was any significant difference in movement time which varied with target distance. A main effect of distance was found, $F(4, 56) = 23.285$, $p < 0.01$, as well as a device x distance interaction, $F(8, 112) = 2.350$, $p < 0.05$. Post hoc analysis was performed to determine specific differences in distances. For the mouse, all pairwise comparisons were significant at $p < 0.01$. For the gesture-control device, participants had significantly slower movement time for the furthest targets (750 px) when compared to the nearest three (100, 250, 375px) at $p < 0.01$. The gaze-tracking device had several significant pairwise comparisons at $p < 0.05$. The most interesting point is the middle-range target (375 px), which had the slowest movement time (M = 1.551, SD = 0.091), and was significantly slower than targets at 100, 500, and 750 px (See Fig. 4).

A main effect of target size was also found, $F(1, 14) = 72.209$, $p < 0.01$, as well as the device x target size interaction, $F(2, 28) = 6.183$, $p < 0.01$. Post hoc analysis shows a significant difference in movement time for all three devices when comparing small and large targets at $p < 0.01$ (See Fig. 5).
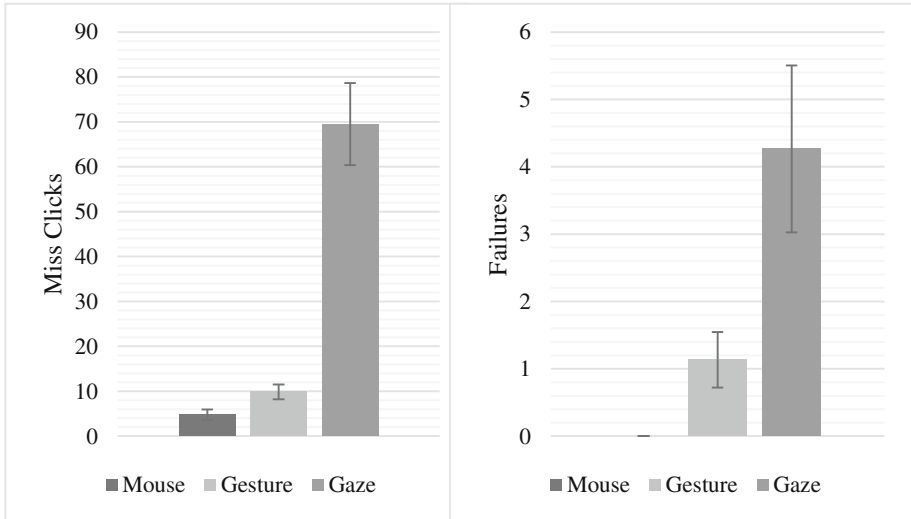
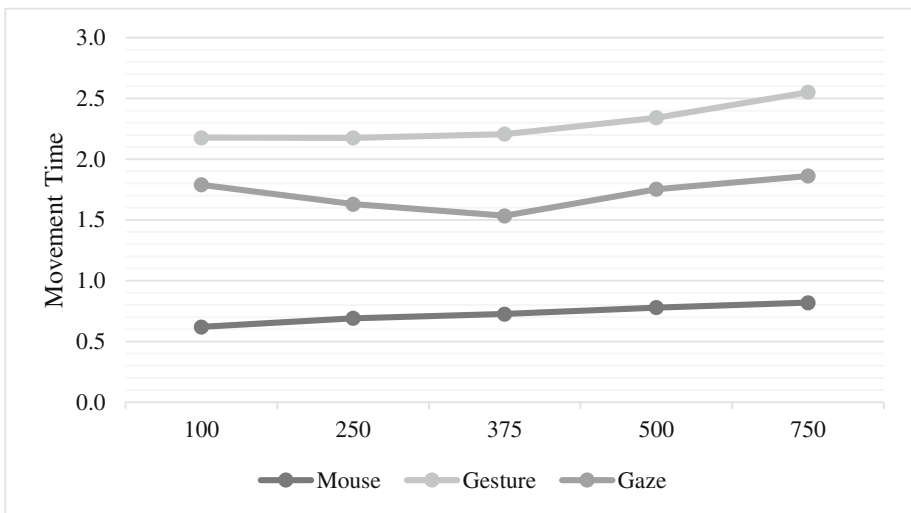**Fig. 3.** Errors encountered for each input method (± 1.0 SE)



**Fig. 4.** Effect of target distance on movement time

## 3.3 Workload

A repeated-measures ANOVA was performed to compare the perceived workload using the weighted NASA Task Load Index across input methods. There was a significant difference between devices, $F(2, 28) = 10.02, p < 0.01$, with the mouse having lower perceived workload (M = 48.88, SD = 9.64), than both gaze-tracking (M = 60.95, SD = 15.37), $p < 0.05$, and gesture-control (M = 62.36, SD = 10.23),
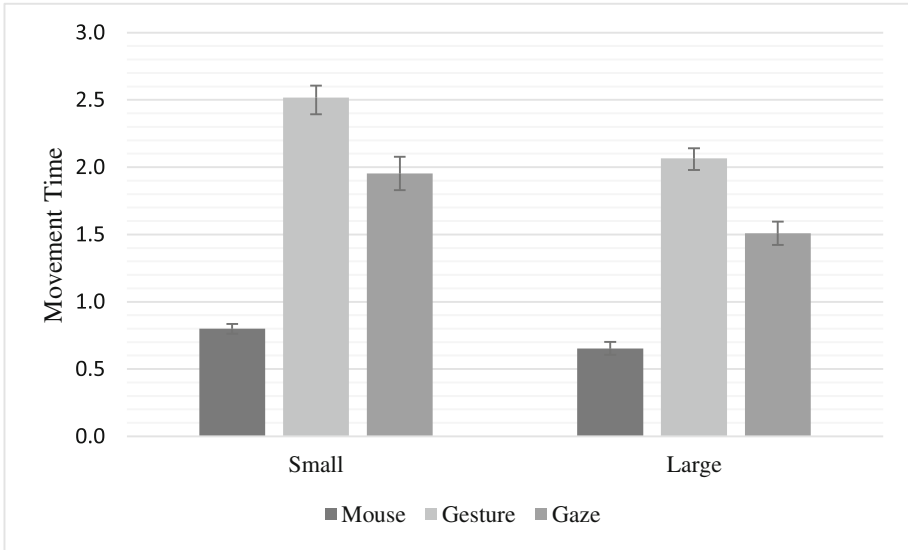
**Fig. 5.** Effect of target size on movement time (± 1.0 SE)

$p < 0.05$. There was no significant difference between the gaze and gesture-controlled devices (See Figs. 6 and 7).

## 4   Discussion

The data from this study indicate that participants were able to complete the point-and-click tasks with the gesture-control and gaze-tracking devices, though their performance was significantly inferior to the traditional computer mouse. Of the two alternative devices, the gaze-tracking device yielded slightly better performance than the gesture-controlled system, but also resulted in more invalid and failed selections.

It is important to note that this study examined first-time usage of the alternative devices. Performance with the input devices [6] may improve with more practice - a longer study is necessary to determine this.

The continued development and improvement of the gaze-tracking device may help improve its accuracy, and with that, the fixation detection parameters can be narrowed. This will likely improve performance and accuracy while reducing errors. Some participants reported difficulty when adjusting for the inaccuracies with the gaze-tracking equipment. The targets were presented on an otherwise blank screen, and participants struggled to fixate on a blank region offset from the target to adjust for inaccurate tracking.

There are many configurable parameters and different control schemes for a gesture-controlled interface, and the best settings will likely vary between individuals. The best implementation may require an adaptive control scheme or training period, or take into account hand velocity and acceleration to create a non-linear mapping of physical-space to screen space (like mouse-acceleration).
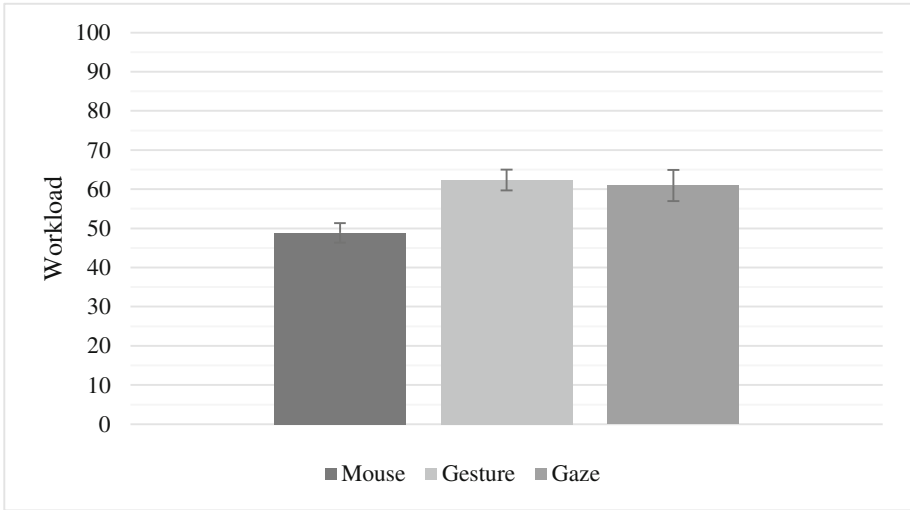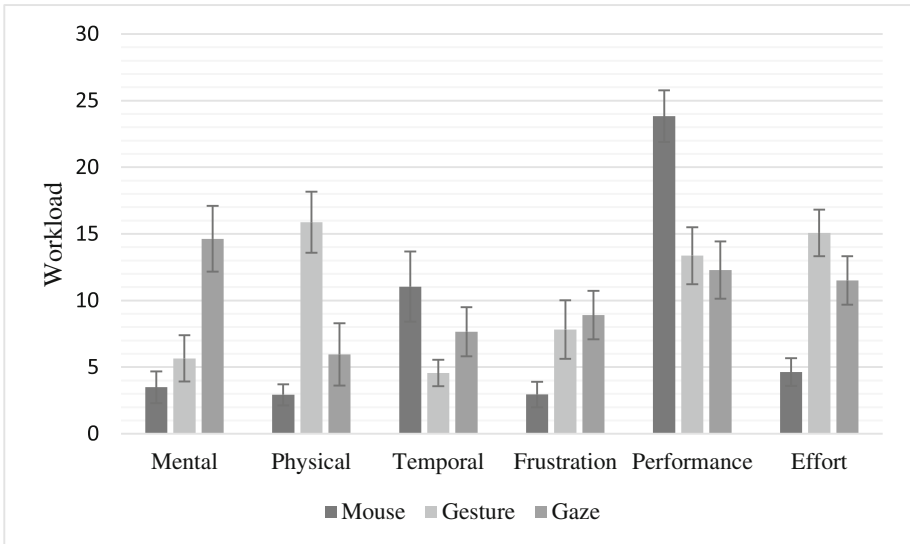
**Fig. 6.** NASA task load index (± 1.0 SE)



**Fig. 7.** Perceived workload factors by input method (± 1.0 SE)

While the results of this study are not in favor of the quick and easy adoption of these alternative input devices, they do not preclude the need for these devices or their potential. Rather, the results of this study should serve as motivation for further research exploring new user interfaces and control schemes which capitalize on the strengths of these devices while minimizing their drawbacks. This may involve multimodal input, combining data from multiple input devices [6]. For example, Zhai was

able to improve task performance by capturing gaze data as an implicit input source [8]. With knowledge of the user's gaze and the UI topology, his MAGIC Pointing system predictively positions the mouse cursor, from which the user only needs to make minor adjustments. The combination of gesture and gaze-tracking devices is particularly interesting. One possibility is a control scheme which uses gestures to provide the action to be performed and gaze-tracking to provide context for the action or a target on which the action can be performed. These alternative devices in their current state solve some existing problems, but more research and experimentation is necessary to use them effectively and achieve their full potential.

# References

1. Wachs, J.P., Stern, H., Edan, Y., Gillam, M., Handler, J., Feied, C., Smith, M.: A gesture-based tool for sterile browsing of radiology images. J. Am. Med. Inform. Assoc. **15**(3), 321–323 (2008)
2. O'hara, K., Harper, R., Mentis, H., Sellen, A., Taylor, A.: On the naturalness of touchless putting the "interaction" back into NUI. ACM Trans. Comput.-Hum. Interact. **20**(1), 5 (2013)
3. Manresa-Yee, C., Ponsa, P., Varona, J., Perales, F.: User experience to improve the usability of a vision-based interface. Interact. Comput. **22**, 594–605 (2010)
4. MacKenzie, S.: Fitts' law as a research and design tool in human-computer interaction. Hum.-Comput. Interact. **7**, 91–139 (1992)
5. Leap Motion Controller LM-010, Best Buy, [Online]. Available: http://www.bestbuy.com/site/leap-motion-controller/9526603.p/. [Accessed 9 May 2014]
6. Savage, N.: More than a mouse. Commun. ACM **56**(11), 15–16 (2013)
7. The Eye Tribe, [Online]. Available: http://theeyetribe.com. [Accessed 9 May 2014]
8. Zhai, S.: What's in the eyes for attentive input. Commun. ACM **46**(3), 34–39 (2003)
9. Cisco, Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013–2018 (2014)
10. NASA, Nasa Task Load Index (TLX) v. 1.0 Manual (1986)