

The Performance of Un-Instrumented In-Air Pointing

Michelle A. Brown¹

York University, Toronto, Canada

Wolfgang Stuerzlinger²

York University, Toronto, Canada

E. J. Mendonça Filho³

Federal University of Bahia, Brazil

ABSTRACT

We present an analysis of in-air finger and hand controlled object pointing and selection. The study used a tracking system that required no instrumentation on the user. We compared the performance of the two pointing methods with and without elbow stabilization and found that the method that yielded the best performance varied for each participant, such that there was no method that performed significantly better than all others. We also directly compared user performance between un-instrumented in-air pointing and the mouse. We found that the un-instrumented in-air pointing performed significantly worse, at less than 75% of mouse throughput. Yet, the larger range of applications for un-instrumented 3D hand tracking makes this technology still an attractive option for user interfaces.

Keywords: Human-computer interaction, Fitts' law, pointing tasks.

Index Terms: H.5.2. [Information interfaces and presentation]: User Interfaces — Input devices and strategies (e.g., mouse, touchscreen)

1 INTRODUCTION

In the everyday world, pointing at objects to reference them is a fundamental task that spans across 2D and 3D selection. When people are using computers with others they will often point to objects visible on screen to indicate them to another person. This suggests that exploring the benefits and trade-offs of in-air pointing as an interaction modality is a worthwhile endeavour.

In-air pointing, especially un-instrumented pointing, promises to be less intrusive and possibly more convenient. We identify characteristics of this technology, measure throughput based on the ISO 9241-9 standard relative to the mouse, identify possible technological improvements, and identify situations where this type of selection method is beneficial.

1.1 Motivation

A number of hand tracking devices have recently appeared on the market and many more have been announced. This includes the Leap Motion, the DUO, the CamBoard pico, the touchless control system by Elliptic Labs, and 3 Gear Systems' hand tracking camera setup. These new devices promise revolutionary and "natural" control of your computer. However, the performance of such in-air un-instrumented hand tracking devices relative to other pointing devices has yet to have been evaluated.

While it is widely assumed that 3D un-instrumented tracking does not perform as well as the mouse, there is no scientific study that quantitatively shows this. Furthermore, the reasons for a

possibly lower throughput have not been explored. Currently it is unclear whether these systems have a lower throughput due to various human postures, differences in latency, poor click detection methods, sub-optimal tracking algorithms, or potentially even human limitations. As such, it is unclear what developers could do to improve the performance of such systems.

Un-instrumented in-air pointing technologies are ideal in situations when users are concerned with sterility, situations where there is no mouse such as with a laptop or tablet, and situations where smudges are a concern, such as when using a mobile device while cooking. This last scenario occurs when people today consult the Internet to find new recipes for cooking. Instead of printing the recipe, users then bring their laptops, tablets, or phones into the kitchen while cooking. Yet, currently all interaction with the device requires users to first wash their hands or risk getting their device dirty. Un-instrumented in-air pointing avoids this problem and enables simple interaction with scrolling, unit conversions, cooking timers, or the music player. In order for this dream to become a reality, the performance of such systems needs to be evaluated and the benefits and shortcomings clearly understood.

We performed several pilot studies to identify a reasonable in-air tracking system. The first system we considered was 3 Gear Systems' hand tracker version 0.9.22. This Kinect-based system had a very high end-to-end latency of about 170 ms and was found to yield very noisy 3D data. In a pilot study, throughput was estimated to be at best 1.5 bps, with selection through the space bar. We also experimented with 3D tracking systems, but decided not to use these, as they required the user to put on extra equipment to use the system.

In the end, we decided to use the Leap Motion device. It easily affords pixel-accurate pointing, due to its' low noise level. We identified 85 ms of end-to-end latency in our system with this device. As this is reasonably close to the 75 ms reported for a instrumented 3D tracking system in a study of effects of latency on 3D interaction [1], we decided to use this device in our study. In another pilot study we found that only about half of all performed "click" motions, a short up and down of a finger, registered with the version 0.8.0 of the Leap Motion system as a successful "click". This success rate was deemed much too low to be practically competitive as a method to indicate selection. Thus, we did not consider this further in our study and used an alternate selection approach, which relies on the space bar operated with the non-dominant hand.

Before investigating how well un-instrumented 3D pointing works, we first devised user study 1 to determine the best operational method for un-instrumented pointing. This study also explores two possible explanations for a possibly lower throughput: pointing method (finger vs. hand) and the effect of elbow stabilization. Past work has shown that pointing with the finger alone (with the arm immobilized) affords significantly lower throughput compared to the whole hand. Elbow stabilization has the potential to improve performance through better accuracy. We then designed user study 2 to directly compare un-instrumented 3D tracking to the mouse by using the same selection technique for both devices. This second study thus also explores the effect different selection methods have on user performance. The reason we decided to directly compare un-

¹ brown@cse.yorku.ca

² wolfgang@cse.yorku.ca

³ euclidesmendonca.f@gmail.com

instrumented pointing to the mouse was to enable calibration against other pointing studies, including MacKenzie and Jusoh's work [2].

1.2 Contributions

Our contributions are:

- The first accurate performance measurement for in-air un-instrumented pointing with ISO 9241-9.
- A comparison of two in-air target acquisition methods: finger pointing and whole hand movement.
- An analysis of the effect of elbow stabilization on selection performance.
- An evaluation of the performance benefit of using a ring button vs. selection with the other hand.

1.3 Related Work

Ray pointing is a method for pointing at objects. In this method, the user will move a tracked object, such as a pen or laser pointer, or a tracked arm or finger and orient it in the direction she or he wishes to point to, such that the tracked entity forms a ray toward the desired pointing location. The first object along that ray is then traditionally highlighted and selected when the user indicates selection, e.g., through a button click. Ray pointing remains a popular selection method for large screen and virtual reality systems. Many studies have investigated this technique, including [2]–[14]. All these comparisons used various 3D tracking systems to implement ray pointing. However, users cannot simply walk up to these systems and start using them; they must first put on the proper tracking equipment or grab an instrumented device.

Ray pointing uses 3D input to afford control over a 2D cursor. Effectively users rotate the wrist (or finger) to move the cursor. Balakrishnan and MacKenzie quantified the relative bandwidth of the fingers, wrist, and forearm and found that the wrist and forearm afford bandwidth of about 4.1 bits/s, while a finger had only 3.0 bits/sec [15]. Another option for 2D cursor control is to directly map 3D motion to 2D motions by dropping the third degree of freedom recorded by the 3D tracking system. With this input method, the user then has to move her or his whole hand. Moving the finger or the whole hand to control the cursor both afford efficient pointing [16]. However, tracking very small hand rotations with 3D tracking systems with sufficient accuracy is difficult, as any amount of tracking noise is effectively magnified increasingly along the ray. This is the most likely explanation why ray pointing has been identified as inferior to other pointing methods in *desktop* environments, e.g., [17].

Un-instrumented pointing has been studied in the past. Here, the system tracks the hand or finger(s) of the user without instrumentation on the user. Typically this is done with some form of cameras. This body of work also explores gestures as a method of interaction. Yet, ray pointing with the finger or arm is typically used as the main pointing method even in gesture-based systems. Work by Gallo et al. [18] explored the benefits of an un-instrumented hand tracking device in a medical context, where sterility is a major concern. Here contact-less technologies offer clear benefits. In a paper by Kolarić et al., finger pointing was enabled through a two camera stereo setup [19]. Matikainen et al. tracked multiple users and their pointing gestures with a camera system [20]. In another work [21] the authors look at the requirements of an un-instrumented tracking system and present an implementation similar to previous work. The authors developed three sample applications to demonstrate its use: FingerMouse, FreeHandPresent, and Brainstorm. Most relevant to our context is the FingerMouse application, which tracks a user's index finger to position a mouse pointer on screen. In this

application, a one second dwell time is used for selection [21]. Song et al. also used finger pointing to select and move virtual objects [22]. None of the above work evaluates the performance of un-instrumented in-air pointing with the ISO standard.

Last, but not least, there has been research into the design and use of wearable digital jewellery, specifically rings. This research looks at creating a device that not only functions well, but is also comfortable and attractive to the user. Many of these rings are made of conventional jewellery materials [23], [24], but some are made of more unconventional materials such as elastic [25] and Velcro [26]. These unconventional materials permit users with different finger sizes to use the same ring. Such rings can be used for 3D selection, in particular for indicating which object to select.

1.4 Fitts' Law and Pointing

Fitts' Law is an empirical model that describes the speed accuracy tradeoff in pointing tasks [27]. It can be used both as a predictive model and as a way to calculate throughput. The model is $MT = a + b \times \log_2(D / W + 1)$. In this model, MT is movement time, D is the target distance, W is the target size, and a and b are derived from linear regression. The log term in this model is referred to as the index of difficulty (ID). The ID describes the difficulty of selecting a particular target by combining the distance and the size of the target into a single value. Fitts' law implies that the further away or the smaller a target is, the harder it will be for a user to select. Although Fitts' law was originally developed for one-dimensional pointing tasks, it has been successfully adapted to 2D pointing tasks and can describe even some 3D pointing tasks.

Building on decades of Fitts' law based studies, the ISO 9241-9 standard [28] has been developed to standardize experimental methodologies and to improve the quality of Fitts' law data. In this standard, throughput is the primary measure of performance. Throughput is calculated as $TP = \log_2(D_e / W_e + 1) / MT$. In this equation, D_e is the effective distance and W_e the effective width. These effective values measure the task that the user actually performed, not the one that she or he was presented with [27]. This reduces variability in identical conditions, which also facilitates comparisons between different Fitts' law studies.

1.5 Interaction Techniques

We studied two types of interaction techniques in this work (see Figure 1). The first method is the *Whole Hand* method. With this method the user moves her or his dominant hand in space to indicate the area on the screen the cursor should move to. For example, if the user held her or his hand at the bottom left area of the screen the cursor would appear there, if she or he then began moving her or his hand up then the cursor would gradually move with the hand towards the upper left area of the screen. The second is the *Pointing* method, based on ray casting. With this method the user uses her or his index finger of the dominant hand to point to the relative location on the screen that she or he wishes the cursor to move to.

2 METHODOLOGY

For the remainder of this paper in-air will always refer to in-air un-instrumented interaction.

In this section, we describe the two user studies we performed to evaluate in-air performance. In the first study, two in-air methods and the effect of elbow placement were evaluated in terms of throughput. In the second study, in-air is directly compared to the mouse using either a button or the spacebar for target selection.

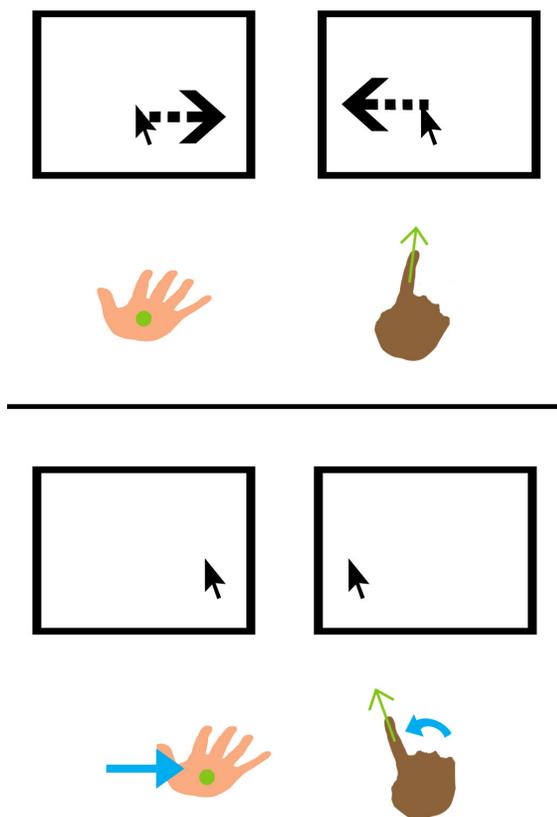


Figure 1: The two interaction methods. The *Whole Hand* method is pictured on the left and the *Pointing* method on the right. The top and bottom frame indicate the cursor position before and after the movement. The green dot and green arrow illustrate what is being tracked by the system.

3 USER STUDY 1

The objective of this study was to determine the interaction method that yields the best throughput for in-air operation. The two compared interaction methods were the *Pointing* interaction method and the *Whole Hand* method (see the section on Interaction Techniques for details). These two interaction styles were tested both with the user's elbow resting on the table and with the user's elbow in the air. When using a mouse, one's hand benefits from the natural stabilization afforded by resting the arm on the desk. We wanted to see if there would be a performance benefit if users grounded their elbows on the table thereby stabilizing their hands and fingers so that they would be less prone to natural hand tremor. On the other hand, grounding the elbow has the potential for restricting the movement over users and thus may also decrease performance from a participant's arm being more constricted in movement by being confined to the surface of the desk. This work is similar to the work by Cockburn et al., except we look at these two interaction methods in a desktop environment [29]. Our participants were seated while using these two methods and the range of motion required was much smaller than in previous work. The range of motion in our experiment was at a comparable level to the range of motion required by a mouse.

3.1 Participants

We recruited 16 participants for this study (mean age 23 years, *SD* 8.5). Seven were female and one was left-handed. None had used in-air devices before to interact with a computer. All users kept

their hands in an open relaxed position when using the *Whole Hand* interaction method.

3.2 Setup

The Leap Motion sensor was placed directly in front of computer display so that it was centered with the middle of the monitor. The sensor was then calibrated to the screen with the default calibration process, which uses a wooden chopstick for a more precise calibration. The Leap Motion device driver and hardware used for this first study was version 0.8.0 respectively v.05. The software used for this Fitts' Law study was FittsStudy [30]. We added support to read data from the LeapMotion to this package.

3.3 Input Conditions

For this user study there were four input conditions for selecting targets. These were the *Whole Hand* method with the participant's elbow supported by the table or by a stack of books (depending on what was more comfortable), the *Whole Hand* method with the participant's elbow raised above the table, i.e., in the air and not supported, the *Pointing* method with the elbow supported, and the *Pointing* method with the elbow unsupported. After targets had been acquired using one of these four methods, targets were selected using the spacebar on a keyboard. The spacebar was operated by the non-dominant hand of the participant and was placed in a comfortable operating position so that the dominant hand used for object acquisition was not obstructed. Figure 2 illustrates the setup. In all these conditions the distance between the participant's hand and the computer was relatively consistent, but an exact distance was not enforced to prevent unnatural poses.

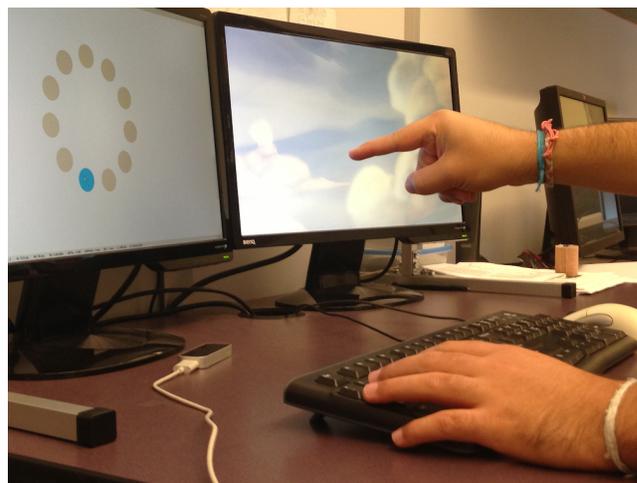


Figure 2: Photo of experimental setup in the condition where users were pointing with the finger and where the elbow was off the table. This user operates the space bar with the left thumb.

3.4 Procedure

First, each participant was given a brief questionnaire about her or his background. The questionnaire recorded gender, age, and handedness. Then, the participant was instructed in the use of one of the four input conditions and was encouraged to practice with this input method until she or he felt comfortable. Once the participant was comfortable with the input method, she or he completed a series of Fitts' law selection tasks using her or his dominant hand to move the cursor to the desired location and operating the spacebar with the other hand for selection. The participant was instructed to select these targets as quickly and accurately as possible. She or he was also instructed that breaks should be taken between circle groups if her or his arm was

getting tired. Each such block consisted of 9 Fitts' law "circles", with 13 trials per circle. The 9 circles used all combinations of target widths of 32, 64, and 96 and amplitudes of 256, 384, and 512 pixels, respectively. This smaller range of ID values was chosen, as target widths below 32 pixels would often be missed, not due to participant error, but due to jitter in the tracking system. We wanted to get a good measurement of the achievable performance with a well-operating system and so did not include these targets. Participants were free to move their arm around while using the *Pointing* method and, as such, movement angles and distances would not be consistent across all participants. Consequently, we report ID values in pixels. The participant would then be presented with another "block" for the next input method and the above process would be repeated until all four input conditions had been completed using a Latin square design across all participants. Overall, the study took about one hour per participant.

3.5 Results

Data was first filtered for participant errors, such as hitting the spacebar twice on the same target or pausing in the middle of a circle to focus her or his attention elsewhere. Removing these errors and outliers, i.e., results more than three standard deviations from the mean, amounted to a 3% loss of total data collected.

3.5.1 Throughput

The data were not normally distributed. Also, Levene's test for homogeneity of variance showed no significance for elbow placement ($F_{1,15} = 0.4332$, ns), but movement type and the interaction between movement type and elbow placement were statically significant ($F_{1,15} = 57.708$, $p < .001$ and $F_{3,13} = 27.724$, $p < .001$). This invalidates the assumption of similar differences between groups variances needed for parametric repeated measures ANOVA. To address these concerns, we used the *Aligned Rank Transform (ART)* for nonparametric factorial data analysis in Human-Computer Interaction treatment [31].

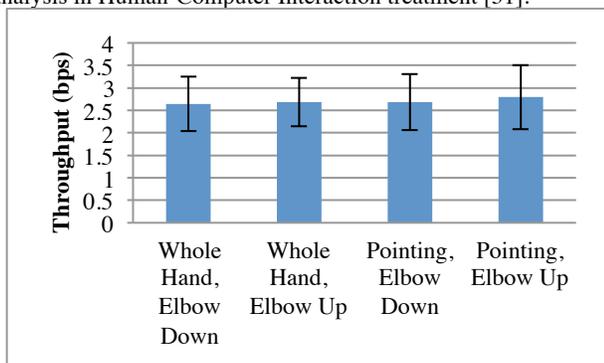


Figure 3: Graph depicting the average throughput (bps) for each condition with error bars. Difference between methods was not significant. Error bars (in this and subsequent figures) show standard deviation.

Overall, there was no significant effect for movement type ($F_{1,15} = 0.52$, ns) or for elbow placement ($F_{1,15} = 0.67$, ns) on (effective) throughput (see Figure 3). There was also no significant effect from the interaction of movement type and elbow placement ($F_{1,15} = 0.44$, ns).

3.5.2 Movement Time

The data were not normally distributed and also failed Levene's test for homogeneity. ART was used again to address this concern.

There was no significant effect on movement time for both movement type ($F_{1,15} = 1.42$, $p > .05$) and elbow placement ($F_{1,15}$

$= 0.92$, ns). There was also no significant effect for the interaction between movement type and elbow placement ($F_{1,15} = 0.04$, ns). See Figure 4 for mean movement times.

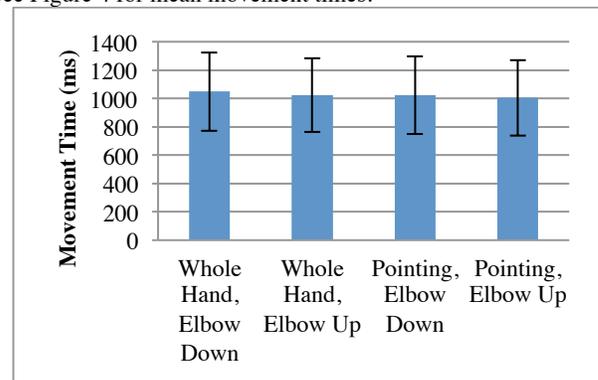


Figure 4: Graph depicting the average effective movement time (bps) for each condition with error bars. Difference between methods was not significant.

3.5.3 Error Rate

The data were not normally distributed and also failed Levene's test for homogeneity. ART was used again to address this concern.

The placement of a participant's elbow had no significant effect on error rate ($F_{1,15} = 0.29$, ns), but the movement type did identify a significant difference ($F_{1,15} = 4.93$, $p < .05$) with the *Whole Hand* method producing fewer errors. Although technically significant, the statistical power of this result is rather low at 0.55, so we cannot claim this to be a strong result.

3.5.4 Index of Difficulty (ID)

There was no significant effect on the interaction between ID and movement type ($F_{1,15} = 1.22$, $p > .05$) or ID and elbow placement ($F_{1,15} = 0.23$, ns) on movement time. See Figure 5 for the data for all conditions. Values for linear trendlines are as follows: *Whole Hand* elbow down: $y = 302.36x + 151.97$, $R^2 = 0.986$, *Whole Hand* elbow up: $y = 295.87x + 156$, $R^2 = 0.9918$, *Pointing* elbow up: $y = 273.15x + 198.42$, $R^2 = 0.9976$, *Pointing* elbow down: $y = 283.81x + 190.5$, $R^2 = 0.9742$.

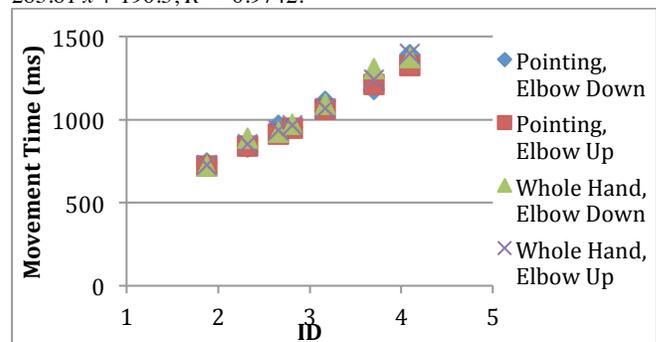


Figure 5: Graph depicting Fitts' law model for input conditions in user study 1.

3.5.5 Learning

As there was only one block per input condition for this experiment it is unclear how strong learning was. Over all participants, no significant learning effects could be detected, but this does not mean that users did not learn. For more information about learning effects with in-air operation see user study 2.

3.5.6 Other Results

We also analyzed in which condition each *individual* participant showed the highest throughput. We classified throughput as being equal if the two throughput values were within 5%, otherwise as different. The results of this analysis identify that the similar results seen for throughput between each method are not due to individual participants performing equally well with each method. Rather, there seems to be a half-half split among the population, both for the question as to which input method is better and if it is better if the elbow is placed on the table or not.

3.6 Discussion

The overall conclusion from this study is that there seems to be no particular method of using in-air devices that works best for all people. As such, we believe that the differences between these methods are not a major contributing factor to lower in-air throughput. Statistical significance between movement types was only found for error rate, but even this significance had low statistical power. These results are in keeping with the results found by Cockburn et al. and show that this trend holds even with modern desktop sensors. Performance for each method was heavily dependent on the individual participant. Some participants exhibited no difference between the methods, while others found that a particular method yielded substantially better results. As such, there seems to be no universal solution that exhibits uniformly high performance. This indicated to us that in-air systems should be configurable to account for different input methods.

The throughput observed for in-air interaction in this study was at best about 2.8 bps, which is lower than the usually observed range of 3.7–4.9 bps for the mouse in Fitts' law studies [32]. With 85 ms end-to-end latency in whole the system, the finger tracking system ranks comparable to other 3D tracking systems in terms of delay (though there are better ones). Even so, 85 ms of latency is much higher than the measured 28 ms of end-to-end latency for the mouse in our setup. While this difference of 57 ms in latency can be expected to have an impact on the performance of in-air interaction, it should decrease throughput by only about 0.3 bps according to previous work on the effects of latency [33]. If this latency were eliminated it would bring the Leap Motion's throughput to about 3.1 bps. This is still not even close to levels that can be easily achieved with a mouse.

4 USER STUDY 2

The purpose of this study was to directly compare in-air interaction with the mouse. The two input methods were compared using the ISO 9241-9 test procedure using both the spacebar and a button click for selection. Both spacebar and button were used in order to determine the effect different click detection methods have on performance. Specifically, we wanted to identify if clicking the left button on the mouse affords an advantage in terms of throughput. The purpose of this study was not to suggest in-air interaction as a mouse replacement, but rather to provide a comparative measurement to the mouse that could be used to calibrate the performance of this system against other systems. This enables comparisons to other hand tracking devices and other input methods.

4.1 Participants

We recruited 16 *different* participants for this within subjects study (mean age 27 years, *SD* 9.5). Six were female and one was left-handed. That participant still preferred to operate the mouse and to point with the right hand. None had used an in-air system before. All users kept their hands in an open relaxed position when using the *Whole Hand* interaction method.

4.2 Setup

The sensor was placed directly in front of computer display so that it was centered at the middle of the monitor and carefully calibrated. The Leap Motion software and hardware used was version 1.0.5+7357, respectively LM-010 (a more up-to-date version relative to user study 1). The measured end-to-end latency of the overall system was 63 ms. The mouse was a Microsoft IntelliMouse Optical set to the default pointer speed in the Windows 7 operating system. The system had an end-to-end latency of 28 ms with the mouse.

We developed a ring button as an alternative method to indicate object selection in the in-air condition. We wanted this button to mimic the left click on the mouse as close as reasonably possible. The ring button consisted of a button glued to a *Hook and Loop* strip. We used a *Hook and Loop* strip to accommodate participants with diverse finger sizes, much as Harrison and Hudson did with their Velcro ring [26]. The button, an Omron B3F-1020 tactile switch, is a 6×6 mm square box, 5 mm tall with a 3.5 mm cylindrical tip. The operating force is specified at 0.98 N and the tip travel during a click is about 0.5 mm. The *Hook and Loop* strip was 20 mm wide and 100 mm long and was glued to the button with a hot-melt adhesive. The button was wired in parallel to the left-button of a desktop mouse. The thin wire connecting the ring to the mouse was held on participants' arms with two 50 mm bands of 3M™ Coban™ Self-Adherent tape in order to prevent the wire from interfering with the arm motions and from possibly confounding the sensor's view. Figure 6 illustrates the ring button, its placement as well as the wire placement. The software used for conducting the Fitts' law tasks was again FittsStudy [30].



Figure 6: Photo of finger ring with the button visible below the users thumb (raised more than normal for illustration). Self adhesive tape was used to keep the cable from interfering with the motions.

4.3 Input Conditions

For this user study there were four input conditions that the participants selected targets with. These were in-air with spacebar selection (using whatever in-air technique the participant preferred), in-air with ring button selection (using the same in-air technique chosen), the mouse with spacebar selection, and the mouse with left mouse button selection. For selection with the spacebar, the keyboard was operated and placed as described in user study 1.

4.4 Procedure

First, each participant was given a brief questionnaire about her or his background. The questionnaire recorded gender, age, and handedness. Then, the participant was introduced to the two different interaction styles for in-air operation, the *Whole Hand* method and the *Pointing* method, as described above under Interaction Techniques. After trying both methods, participants were asked to choose which method they wanted to use. That method was then used for the rest of the experiment. We permitted participants to choose their own in-air method based on

the results of our first user study, where different people performed better with different interaction methods. Keeping with the work by Sparrow and Newell [34], we assumed that participants would choose the in-air style that would yield the best throughput through self-optimization. For the same reason, we also did not control for user elbow placement in this study.

Once a particular input method was chosen, participants completed 2 blocks of 9 Fitts' law circles with 11 trials per circle for practice with this method. This practice period accounted for the inexperience of participants with in-air interaction. Our study thus indicates the performance possible in the early adaptation stages of in-air interaction. The performance in these early stages is frequently just as important as potential top performance: if performance is much lower than an alternative method, users will often just give up and use that alternative method even if their performance could ultimately be better by adapting the higher performing method. One well-known example is QWERTY vs. Metropolis for touchscreen keyboards [35].

After the practice period, the participant started with one of the four conditions, and then experienced the others. The presentation order to participants was determined with a Latin square design.

With each given condition the participant completed 3 blocks of 9 Fitts' law circles with 11 trials per circle. Target widths of 32, 64, and 96 and amplitudes of 256, 384, and 512 pixels were used (see user study 1 for an explanation as to why this range of ID values was chosen and why the values were presented this way). Between blocks, participants were encouraged to rest for about a minute before starting the next one. Participants were also instructed that breaks should be taken between "circles" if they experienced fatigue. At the end of all four conditions, participants were given a brief questionnaire about discomfort they might have experienced while using in-air interaction.

4.5 Results

Data were first filtered for participant errors, such as hitting the spacebar twice on the same target or pausing in the middle of a circle. Removing these errors amounted to less than .005% loss of total data collected. There were no outliers.

4.5.1 Throughput

The data for throughput were not normally distributed. Also, the data failed Levene's test for homogeneity. Consequently, we again used the *Aligned Rank Transform (ART)* for nonparametric factorial data analysis [31] and performed a repeated measures parametric ANOVA on the transformed data.

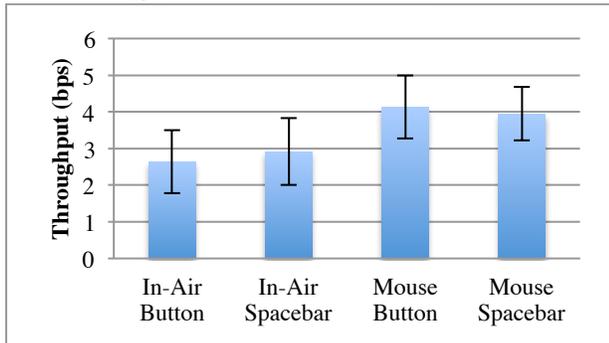


Figure 7: Graph depicting the average throughput (bps) for each condition in the second study with error bars. The difference between methods was significant.

For throughput there was a significant effect of type of device of ($F_{1,15} = 90.76, p < 0.0001$), with effect size (η^2) of 0.3 and power ($1-\beta$) of 0.99. There was no significant effect of selection method on throughput ($F_{1,15} = 0.47, ns$). There was a significant interaction of selection method and type of device ($F_{1,15} = 17.10, p$

< 0.001), with an effect size (η^2) of only 0.03 and very weak power ($1-\beta$) of 0.07. The mouse exhibited throughput around 4 bps, while the in-air conditions showed less than 3 bps, see Figure 7.

4.5.2 Movement Time

The data for movement time were not normally distributed and also failed Levene's tests for homogeneity. Consequently, we used *ART* again.

For Type of device there was a significant effect on MT_e ($F_{1,15} = 49.66, p < 0.0001$), with effect size (η^2) of 0.23, and power ($1-\beta$) = 0.96. There was no significant effect of selection method on movement time ($F_{1,15} = 0.8, ns$). Similar to the results for throughput, there was a significant effect for the interaction of Selection Method and Type of device on MT_e ($F_{1,15} = 14.37, p < 0.05$), with an effect size (η^2) of only 0.03, and with very weak power ($1-\beta$) = 0.065. See Figure 8 for mean movement times.

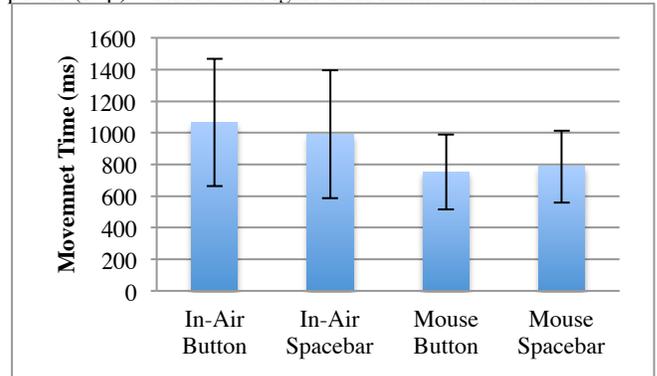


Figure 8: Graph depicting the average effective movement time (milliseconds) for each condition in the second study with error bars. The difference between methods was significant.

4.5.3 Error Rate

We did not find any significant effects in the data for error rates.

4.5.4 Learning

There was no statistically significant difference between blocks on throughput ($F_{1,15} = 0.49, ns$). There was also no statistically significant difference for the interaction of block and selection method on throughput ($F_{1,15} = 0.6, ns$).

4.5.5 Index of Difficulty (ID)

There was no significant effect on the interaction between *ID* and selection method ($F_{1,15} = 2.67, p > .05$). See Figure 9 for a depiction of the Fitts' law data for this study.

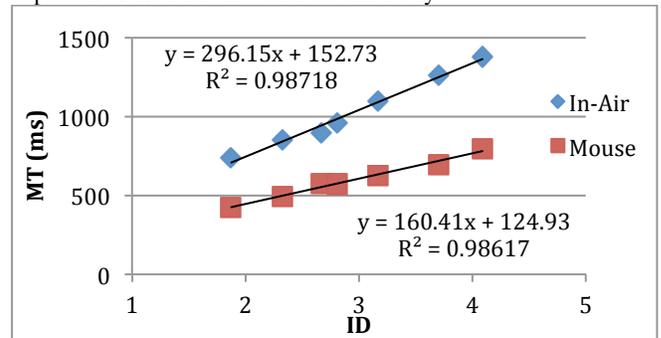


Figure 9: Graph depicting Fitts' law model for in-air interaction in user study 2 and a linear regression

4.6 Discussion

The results of the second user study show that the mouse performs significantly better than in-air interaction, regardless of the target selection method used. The performance of the mouse is around 4 bps, which is in line with other work. The performance of in-air conditions is slightly less than 3 bps, which corresponds (approximately) to our first study and to the results of other recent evaluations of ray pointing techniques [8] with traditional 3D tracking systems. A caveat is that these results are not fully comparable, as our work did not use a stereo display.

One result of this study is that even in a condition designed to be as optimal for in-air interaction as possible, the mouse still outperforms this method by a substantial margin. Moreover, and in all conditions, the selection mechanism (spacebar or ring button) was 100% reliable. Consequently, we are fairly confident in stating that using finger gestures, such as “up-down” click motions, would not improve throughput. The reasoning here is that any inaccuracies in gesture detection will very likely have a negative effect on error rates, as un- or mis-recognized button presses will result either in a “miss” or correction movement. Both alternatives will decrease throughput. Moreover, and even assuming that there is a 100% reliable gesture recognizer, we point out that to make an “up-down” click gesture in free air easily recognizable, the user needs to move the finger a distance that is sufficiently large to be detectable. This distance is likely larger than the small motion needed to operate a mouse or ring button. Consequently, such a motion likely takes longer, which again can only decrease throughput. Other gestures, such as a pinch, making a fist, and closing the gap between the thumb and the side of the hand, suffer from the same problem and are also not likely to exhibit increased throughput.

Moreover, all participants in the study reported some level of fatigue while using the in-air system regardless of whether they were using the *Pointing* or the *Whole Hand* method. All participants reported either ‘mild discomfort’ or ‘discomfort’ (values 2 and 3 respectively) on a five-point scale from ‘none’ to ‘pain’ with an average reported value of 2.5. There was no correlation between interaction method used and level of fatigue.

To get a better idea of the performance potential of in-air interaction, one would need to run a longitudinal study, where participants get several hours of practice.

5 OVERALL DISCUSSION

In these two studies we largely rule out three potential factors (arm stabilization, cursor control method, and selection method) as potential explanations for the lower throughput observed with un-instrumented tracking systems. This considerably narrows the search space for future research and brings the community closer to tracking down the factor or factors that contribute to the reduced throughput for such systems. We also presented an accurate throughput measurement for in-air interaction using the ISO 9241-9 standard. This enables our research to be easily compared to other future work.

In general, we can state that modern instantiations of in-air interaction can achieve throughput comparable to that of state-of-the-art 3D tracking systems. We are particularly excited about the low level of noise in the system we used. Yet, end-to-end system latency affects throughput and reductions of lag may improve throughput levels even further. Based on previous results around latency, such as [33], the potential for improvement is (only) in the range of 15–20%. Nevertheless, it is good to see that the developers of in-air interaction systems have recognized the issue and are continuously working on improving the latency of their devices. They are now even reaching out to developers to explain the issue and give active advice on how to reduce the latency in a given system [36].

Some of our experiments can be considered a replication of the work by Balakrishnan and MacKenzie [15]. However, our work considers in-air interaction, while their work investigated only interaction on a surface. Yet, the throughput results that we see in our experiments are very similar to the results seen in this previous study. There, the throughput for the finger was reported to be approximately 3.0 bits/s and the throughput of the wrist 4.1 bits/s. Taking into account that the experimental setup in this previous work had lower latency than our system, our measured throughput of approximately 2.8 bits/s is right within the expected range from this study (as our participants were free to use both their fingers and wrists to control the system). This can be interpreted both as a validation of these previous results and of our current user study design.

Based on our results, we cannot recommend un-instrumented in-air hand tracking for user interfaces that require the best possible pointing performance or that need to compete with mice or touch screens. However, un-instrumented hand tracking is attractive for many other applications, such as user interfaces that require the use of multiple fingers at the same time, casual games, bi-manual interaction, gesture-based systems, and applications where the user cannot touch or hold a device.

5.1 Limitations

In order to study in-air device performance it was necessary to choose an un-instrumented system to do this with. In the case of this paper, the Leap Motion was selected. While this is a necessary step, it potentially causes particulars of the selected system to be studied, rather than the class of devices it is meant to represent. While the authors have tried to make this research as device-independent as possible, it is possible that particular aspects that are Leap Motion-specific have exhibited themselves in the results.

6 CONCLUSION

We conducted two studies to explore the performance of un-instrumented 3D tracking for controlling a cursor on a screen. We show that elbow support is not uniformly advantageous for in-air interaction. Likewise, the best method of cursor control, either *Pointing* or *Whole Hand*, varies from one individual to the next. This suggests that one method of operation is not universally beneficial and that multiple methods should be permitted to accommodate different users. This result also indicates that neither arm stabilization nor the cursor control method are likely causes for the decreased throughput of in-air interaction. This finding permits future studies to ignore these two factors when designing experiments.

Our second study provided evidence that un-instrumented, in-air tracking is unlikely to reach levels of throughput equivalent to the mouse, even if the technical implementation had minimum latency. Consequently, we believe that some other factor accounts for the decrease in throughput. We also show that there is no statistically significant benefit to using a button for selection with the same hand that is being used to move the cursor. As such, this is another factor that does not need to be controlled for in future experiments. Furthermore, we found evidence that the left-click on the mouse does not contribute to its higher throughput and that there must be some still untested factor in un-instrumented interaction that is contributing to its lower throughput. Consequently, we suggest using un-instrumented 3D tracking only in situations where interaction would benefit from using a hands-free, sterile, multi-hand, or multi-finger device. In other situations, it is likely that the mouse is still better suited to the task.

6.1 Future Work

We plan to further explore potential causes of the gap between pointing and mouse throughput, such as latency, learning, and sub-optimal tracking algorithms. Ultimately, we hope to explore new realms of computer operation that extend beyond what the mouse is capable of.

6.2 Acknowledgements

We would like to thank the GRAND NCE and York University for supporting this work. CNPq, Brazil supported E. Mendonça.

REFERENCES

- [1] R. J. Teather, A. Pavlovych, W. Stuerzlinger, and I. S. MacKenzie, "Effects of tracking technology, latency, and spatial jitter on object movement," in *2009 IEEE 3DUI*, 2009, pp. 43–50.
- [2] I. MacKenzie and S. Jusoh, "An evaluation of two input devices for remote pointing," *Eng. Human-Computer Interact.*, vol. 2254, pp. 235–250, 2001.
- [3] A. Banerjee and J. Burstyn, "Pointable: an in-air pointing technique to manipulate out-of-reach targets on tabletops," in *ACM Tabletops and Surfaces*, 2011, pp. 11–20.
- [4] K. Das and C. W. Borst, "An evaluation of menu properties and pointing techniques in a projection-based VR environment," *2010 IEEE Symp. 3D User Interfaces*, pp. 47–50, Mar. 2010.
- [5] R. Jota, M. Nacenta, and J. Jorge, "A comparison of ray pointing techniques for very large displays," in *GI*, 2010, pp. 269–276.
- [6] A. Kunert, A. Kulik, C. Lux, and B. Fröhlich, "Facilitating system control in ray-based interaction tasks," in *ACM VRST*, 2009, vol. 1, no. 212, pp. 183–186.
- [7] C. Wingrave and D. Bowman, "Baseline factors for raycasting selection," in *Proc. of HCI International*, 2005.
- [8] R. J. Teather and W. Stuerzlinger, "Pointing at 3d target projections with one-eyed and stereo cursors," in *ACM CHI*, 2013, p. 159.
- [9] D. Vogel and R. Balakrishnan, "Distant freehand pointing and clicking on very large, high resolution displays," in *ACM UIST*, 2005, p. 33.
- [10] J. Zigelbaum, A. Browning, D. Leithinger, O. Bau, and H. Ishii, "G-stalt: a chirocentric, spatiotemporal, and telekinetic gestural interface," in *International Conference on Tangible, Embedded, and Embodied Interaction*, 2010, pp. 261–264.
- [11] M. Gokturk and J. L. Sibert, "An analysis of the index finger as a pointing device," in *Extended Abstracts on Human Factors in Computing Systems*, 1999, no. May, p. 286.
- [12] I. Oakley, J. Sunwoo, and I.-Y. Cho, "Pointing with fingers, hands and arms for wearable computing," in *Extended Abstracts on Human Factors in Computing Systems*, 2008, no. 3255, pp. 3255–3260.
- [13] J. Oh and W. Stuerzlinger, "Laser pointers as collaborative pointing devices," *GI* vol. 2002, pp. 141–149, 2002.
- [14] T. Grossman and R. Balakrishnan, "Pointing at trivariate targets in 3D environments," in *ACM CHI*, 2004, vol. 6, no. 1, pp. 447–454.
- [15] R. Balakrishnan and I. S. MacKenzie, "Performance differences in the fingers, wrist, and forearm in computer input control," in *ACM CHI*, 1997, pp. 303–310.
- [16] R. Balakrishnan, T. Baude, G. Kurtenbach, and G. Fitzmaurice, "The Rockin' Mouse : Integral 3D Manipulation on a Plane," in *ACM CHI*, 1997, pp. 311–318.
- [17] R. J. Teather and W. Stuerzlinger, "Pointing at 3D targets in a stereo head-tracked virtual environment," in *2011 IEEE 3DUI*, 2011, no. 1, pp. 87–94.
- [18] L. Gallo, A. Placitelli, and M. Ciampi, "Controller-free exploration of medical image data: Experiencing the Kinect," in *Computer-Based Medical Systems*, 2011, pp. 1–6.
- [19] S. Kolaric, A. Raposo, and M. Gattass, "Direct 3D manipulation using vision-based recognition of uninstrumented hands," in *X Symposium on Virtual and Augmented Reality*, 2008, pp. 212–220.
- [20] P. Matikainen, P. Pillai, L. Mummert, R. Sukthankar, and M. Hebert, "Prop-free pointing detection in dynamic cluttered environments," *Face Gesture*, pp. 374–381, Mar. 2011.
- [21] C. von Hardenberg and F. Bérard, "Bare-hand human-computer interaction," *Work. Perceptive user interfaces*, 2001.
- [22] P. Song, H. Yu, and S. Winkler, "Vision-based 3D finger interactions for mixed reality games with physics simulation," in *ACM VRCAI*, 2008, vol. 1, no. 212.
- [23] C. Miner, D. Chan, and C. Campbell, "Digital jewelry: wearable technology for everyday life," in *Extended Abstracts on Human Factors in Computing Systems*, 2001, pp. 45–46.
- [24] D. Ashbrook, P. Baudisch, and S. White, "Nenya: subtle and eyes-free mobile input with a magnetically-tracked finger ring," in *ACM CHI*, 2011, pp. 2043–2046.
- [25] S. Rhee, B. Yang, and H. Asada, "Artifact-resistant power-efficient design of finger-ring plethysmographic sensors," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 7, pp. 795–805, 2001.
- [26] C. Harrison and S. E. Hudson, "Abracadabra : Wireless , High-Precision , and Unpowered Finger Input for Very Small Mobile Devices," in *ACM UIST*, 2009, pp. 121–124.
- [27] I. S. MacKenzie, "Fitts' Law as a Research and Design Tool in Human-Computer Interaction," *Human-Computer Interact.*, vol. 7, no. 1, pp. 91–139, Mar. 1992.
- [28] ISO, "9241-9 Ergonomic requirements for office work with visual display terminals (VDTs)-Part 9: Requirements for non-keyboard input devices (FDIS-Final Draft International Standard)," *Int. Organ. Stand.*, 2000.
- [29] A. Cockburn, P. Quinn, C. Gutwin, G. Ramos, and J. Looser, "Air pointing: Design and evaluation of spatial target acquisition with and without visual feedback," *Int. J. Hum. Comput. Stud.*, vol. 69, no. 6, pp. 401–414, Jun. 2011.
- [30] J. O. Wobbrock, K. Shinohara, and A. Jansen, "The Effects of Task Dimensionality , Endpoint Deviation , Throughput Calculation , and Experiment Design on Pointing Measures and Models," in *ACM CHI*, 2011, pp. 1639–1648.
- [31] J. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins, "The aligned rank transform for nonparametric factorial analyses using only anova procedures," in *ACM CHI*, 2011, pp. 143–146.
- [32] R. W. Soukoreff and I. S. MacKenzie, "Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI," *IJHCS*, vol. 61, no. 6, pp. 751–789, Dec. 2004.
- [33] A. Pavlovych and W. Stuerzlinger, "The tradeoff between spatial jitter and latency in pointing tasks," in *ACM EICS*, 2009, pp. 187–196.
- [34] W. Sparrow and K. Newell, "Metabolic energy expenditure and the regulation of movement economy," *Psychon. Bull. Rev.*, vol. 5, no. 2, pp. 173–196, 1998.
- [35] S. Zhai, M. Hunter, and B. a. Smith, "The metropolis keyboard - an exploration of quantitative techniques for virtual keyboard design," in *ACM UIST*, 2000, vol. 2, pp. 119–128.
- [36] R. Bedikian, "Understanding Latency," *Leap Motion Developer Labs*, 2013. [Online]. Available: <http://labs.leapmotion.com/post/55354675113/understanding-latency-part-1>.