

A Standardized Fitts ISO 9241-9 Task Across Desktop, VR, and MR: Comparing Mouse, Controller Ray, and Mid-Air Hand Input

by

Nurkhan Zhanbayev

Submitted to the Department of Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science in Computer Science

at the

NAZARBAYEV UNIVERSITY

June 2026

© Nazarbayev University 2026. All rights reserved.

Author
Department of Computer Science
April

Certified by
Syed Muhammad Umair Arif
PhD Instructor
Thesis Supervisor

Accepted by
Yelyzaveta Arkhangelsky
Dean, School of Engineering and Digital Sciences

A Standardized Fitts ISO 9241-9 Task Across Desktop, VR, and MR: Comparing Mouse, Controller Ray, and Mid-Air Hand Input

by

Nurkhan Zhanbayev

Submitted to the Department of Computer Science
on April, in partial fulfillment of the
requirements for the degree of
Master of Science in Computer Science

Abstract

The consideration of pointing performance through standardized measures has become more and more important as interactive systems are not only expanding beyond traditional desktop environments into virtual reality (VR) and mixed reality (MR) but also beginning to be integrated into industrial settings. But past Fitts law studies in XR have been methodologically heterogeneous, with large variation in the geometry of tasks, spatial layout, input modality, throughput computation, making comparison across studies and platforms challenging. In order to solve this issue, this thesis will introduce a standardized ISO-inspired multi-directional target acquisition task in order to compare the performance of desktop, VR, and MR pointing.

A within subject study was done using five participants through three input modalities namely mouse input on a desktop, controller-ray input in VR, and mid-air hand interaction in MR. The same task generator, target sequencing logic, timing rules, and logging pipeline were kept across platforms to isolate modality-related effects on methodological differences. Movement time (MT), throughput (TP), and error rate were used to assess objective performance, and perceived workload (NASA-TLX), usability ratings, and the System Usability Scale (SUS) were used to evaluate subjective experience.

The findings showed that there was a consistent cross platform performance hierarchy. Desktop demonstrated the quickest interaction, and the greatest efficiency; VR displayed moderate performance, and MR displayed the slowest movement times and least throughput. When the task difficulty was high, there was an increase in the throughput across all modalities however throughput was most significantly increased in MR where throughput was always low. Conversely, Desktop was still able to perform well between the levels of difficulty, whereas VR was still functional but slower compared to Desktop. Objective analysis revealed significant platform effects of MT and TP, but the difference in error rates was not significantly different.

The same general trend was followed by subjective findings. Desktop was the least demanding, most comfortable and most usable modality, with the highest SUS score. VR was in the intermediate category, which is a moderate workload and reasonable usability. MR had the highest workload on NASA-TLX workload, and lowest usability and SUS scores.

Repeated-measures ANOVA was used to further confirm that there were significant modality effects on workload, SUS, and usability measures.

In general, the work under this thesis proves that it is possible and even informative to undertake standardized cross-platform comparison of pointing performance, provided task geometry and analysis procedures remain constant. The results are not only empirical but also provide methodological advice to future research in the field of human-computer interaction and XR, in general, in the pursuit of providing a fair evaluation of the mouse-, ray-, and gesture-based methods of input in relation to each other across different realities.

Thesis Supervisor: Syed Muhammad Umair Arif

Title: PhD Instructor

Contents

List of Abbreviations	13
1 Introduction	15
1.1 Problem Statement	17
1.2 Thesis Goal	19
1.3 Research Questions	19
1.4 Contributions	20
2 Related Work	21
2.1 The law of Fitts and standard metrics	21
2.2 ISO 9241-411 and multi-directional tapping task	23
2.3 Fitts law in VR/MR and cross study variability	24
2.4 Stability of throughput and computation discussions	26
3 System Design and Implementation	31
3.1 Selection of platform and pointing devices	31
3.1.1 Desktop	31
3.1.2 Virtual Reality	32
3.1.3 Mixed Reality	32
3.2 Explanation of developed software	33
3.2.1 ISO-style task geometry	37
3.2.2 Odd number of targets to maintain constant amplitude	37
3.2.3 Parameterization of Circular Layout	39

3.2.4	Cross-Platform Equivalence of Amplitude A and Width W	40
3.2.5	Logging and Data Format	40
4	Experiment	43
4.1	Ethical Compliance and Protocol Framework	43
4.1.1	Subjective Measures	43
4.2	Participants	44
4.3	Experimental Design	44
4.4	Experimental Procedure	45
4.4.1	Preparation Session	47
4.4.2	Main Experiment Blocks	47
4.4.3	Task Timing	48
4.4.4	Post-Modality Assessment	48
4.5	Break Management and Participant Safety	49
4.6	Overall Session Duration	50
4.7	Data Collected	50
4.8	Dependent Variables	50
5	Results	53
5.1	RQ 1: Desktop performance across task difficulty	53
5.2	RQ 2: How does pointing performance in VR change as task difficulty increases?	54
5.3	RQ 3: How does pointing performance in MR change as task difficulty increases?	55
5.3.1	Overall Objective Performance Comparison	56
5.4	RQ 4: How does perceived workload differ across Desktop, VR, and MR during the pointing task?	57
5.5	RQ 5: How does perceived usability differ across Desktop, VR, and MR based on participants' SUS ratings?	59
5.5.1	Aggregated User Feedback	61
5.5.2	Statistical Significance Testing	63

6	Discussion and Conclusions	65
6.1	Key Findings	66
6.2	Design Implications	66
6.3	Limitations	67
6.4	Future Work	67
7	Appendix	69

List of Figures

3-1	VR device used: Meta Quest II head-mounted display.	32
3-2	MR device used: Microsoft HoloLens 2 headset.	33
3-3	Unity-based 2D Graphical User Interface displaying the multi-directional target layout for the Desktop environment.	34
3-4	Unity-based 3D Graphical User Interface displaying the multi-directional target layout for Virtual Reality environment.	34
3-5	Unity-based 3D Graphical User Interface displaying the multi-directional target layout for Mixed Reality environment.	35
3-6	ISO-9241 Fitts' Law Task Logic Flow	38
4-1	Experimental Protocol Flow	46
5-1	Fitts' Law Regression comparison across Desktop, VR, and MR modalities. .	56
5-2	NASA-TLX Scores	59
5-3	System Usability Scale Comparison.	60

List of Tables

2.1	Comprehensive overview of Fitts' law and ISO-style multi-directional pointing studies across Desktop, VR, and MR/AR contexts.	27
3.1	Platform Specifications	31
3.2	Testbed Interface layout parameters	37
3.3	Trial Log Data Fields	41
5.1	Aggregated means per platform and test condition (Fitts' Law report)	54
5.2	Two-Way Repeated Measures ANOVA Results for Objective Measures. Each (*) sign represent significance level. [n.s. - not significant]	57
5.3	System Usability Scale (SUS) Scores by Modality	61
5.4	NASA-TLX Scores by Modality. MD - Mental Demand. PD - Physical Demand. TD - Temporal Demand. PF - Performance	61
5.5	Mean and Standard Deviation (SD) for Questionnaire Results ($N = 5$) . . .	62
5.6	Repeated Measures ANOVA Results for Hypothesis Testing. Each (*) sign represent significance level.	63

List of Abbreviations

AR	Augmented Reality
ER	Error Rate
HCI	Human-Computer Interaction
HMD	Head-Mounted Display
ID	Index of Difficulty
ISO	International Organization for Standardization
MR	Mixed Reality
MRTK	Mixed Reality Toolkit
MT	Movement Time
NASA-TLX	NASA Task Load Index
NU IREC	Nazarbayev University Institutional Research Ethics Committee
RQ	Research Question
SD	Standard Deviation
SEDS	School of Engineering and Digital Sciences
SUS	System Usability Scale
TP	Throughput
VAC	Vergence-Accommodation Conflict
VR	Virtual Reality
XR	Extended Reality

Chapter 1

Introduction

One of the fundamental actions in the field of human-computer interaction is target acquisition. Whether the user is continually touching graphical objects by means of a mouse or is continually touching objects at a distance with a VR controller ray or is playing with objects with mid-air hand, mixed reality input, all these may be perceived as a pointing gesture that can be subjected to a speed-accuracy trade-off. Fitts Law has been a very popular Law to model and examining human pointing behavior in human-computer interface [10, 24, 15].

Fitts' Law has been an extremely powerful law for modeling the performance of input Devices like the mouse, stylus, joystick, touch screen, etc., to traditional desktop-based human-computer interaction. The Law of Fitts has brought about the formulation of standardized assessments of input devices. As an example, performance measure of input devices that is a combination of speed of movement and accuracy of selection. At Meanwhile, the widely used standard of modeling the description of the different evaluation methodologies to design the physical input device. ISO 9241-411 is and one of the most significant motivators to target acquisition tasks to evaluate input device performance. using Fitts' Law [2, 1].

Although the initial version of ISO 9241-9 was the one that defined the general framework of non-keyboard, it nevertheless remained the standard framework that defined the general framework of non-keyboard. Its new revision, ISO 9241-411, specifies the particular multi-Directional task evaluation methods are the main ones that are used today. Therefore, this

thesis references the ISO 9241-411 guidelines although it builds on the heritage of the ISO 9241-9 standard [2, 1].

But as other interactive types of computing systems move beyond traditional. desktop-based interaction systems, VR and MR systems include techniques of target selection. which are very different as compared to the conventional methods of interaction like mouse based interaction. As an illustration, users might be asked to employ a controller to choose objects with the help of a ray. choose or select objects with direct manipulation techniques or mid-air hand input techniques. Such methods include different interaction requirements like unsupported arm movements, depth perception, head coupled viewing, ray stabilization and gesture confirmation. As such, interaction in these systems is hardly likely to perform similarly to conventional. interaction systems, including those that involve the use of a mouse, although these are not. interaction tasks are functionally similar [4, 14, 6].

As a result, Fitts' law has come to be more frequently employed within extended reality The methodology approach of (XR) research as a means of evaluating the performance of target selection in the case of. various immersive interaction techniques. Nevertheless, as it has been found out by recent research studies. within the field, the application of Fitts' law within 3D XR environments is extremely heterogeneous. The studies differ in a great way in as far as one research work is concerned to another. to task design, target configuration, spatial depth, interaction modality, dependent variables reported, and procedures used to compute different measures like throughput. [4, 7, 20]. A significant drawback of the study comparison is the methodological variability that was witnessed. directly and making robust conclusions about the relative performance of various input methodologies used in XR spaces [4, 20].

This lack of uniformity in methodology noticed between studies is a significant challenge to compare studies that have been carried out under varying XR conditions. In most studies increased over Desktop environments, a standardised and mature 2D pointing task is. employed. Nonetheless, research that has been carried out in both VR and MR spaces is discovered to. vary significantly from each other with regard to spatial configuration, target depth, selection modality, hardware setup, as well as procedures employed for evaluating performance [4, 13, 8, 12]. Under such a circumstance, it is very hard to come to conclusions

regarding the relative performance of Desktop, VR, as well as MR input techniques while comparing two studies that the result of a report of mouse input, and also hand ray or pinch input modality, as it is possible. that the differences that are observed are not only attributable to the characteristics of the inputs modality but also due to. differences between the methodology used by various research studies.

Taking into account such limitations, a comparative analysis cross-platform, using a. some standardized approach would be a good idea. This piece of work aims to overcome these shortcomings. by building a Desktop, VR and MR version of an ISO-inspired Fitts task and comparing three input modalities: mouse input when using a desktop, controller ray input when using a VR, and hand input in middle air in MR. Although the work attempts to offer a comparative study of the three input modalities, it does so in a way that the task framework which is underlying is still there. uniform with all three platforms. This gives a better insight into the effects of the modality of input, separately not by a methodologically heterogeneous body. of work [2, 4, 20].

The work also aims at addressing a critical methodological issue in the context of. human-computer interaction. While throughput is an important measure of performance in the surrounding of a Fitts-inspired task, its computation is strongly determined by the way in which it is computed. calculated are the amplitude of the movement, the variability of the endpoint and the width of the effective target. Previous studies have indicated that discrepancies in throughput calculation have caused. to misleading effects in the situation of a Fitts-inspired task, particularly in the studies that lack them. adhere to an ISO-based approach. In the context of these problems are further complicated. VR and MR, where the ray intersection and hand noise could affect the calculation. of endpoint variability and effective target width [7, 4, 21].

1.1 Problem Statement

Fitts' law has traditionally provided a fundamental framework for evaluating pointing performance in human-computer interactions [10, 1]. Nevertheless, the application of Fitts' law in Desktop, VR, and MR environments has often shown a lack of methodological consis-

tency. In traditional Desktop environments, pointing performance has often been assessed using traditional 2D paradigms and performance metrics [15]. However, in VR and MR environments, previous research has often employed a range of different methods, including task design, target layout, depth, interaction distance, hardware, and confirmation methods [4]. This has led to a situation where the results reported in previous research are often difficult to compare.

The issue is particularly apparent in the evaluation of different input modalities. Mouse-based pointing in traditional Desktop environments, controller-ray interaction in VR environments, and mid-air hand input in MR environments engage different motor and cognitive processes. However, previous research on the comparison of these input methods has often failed to use equivalent conditions and analytical methods. This has led to a situation where the apparent performance differences between these methods may be a result of methodological rather than actual modality characteristics [7].

Another problem is related to the calculation of performance measures, especially throughput. Indeed, throughput is considered to be one of the major performance measures in Fitts-based evaluation methods. Nevertheless, its calculation is based on how movement time, endpoint variability, and effective target width are determined and calculated. For instance, inconsistent usage of nominal and effective measures, as well as non-standard usage of ISO-based methods, may result in hard-to-comprehend performance measures [2]. These problems are especially critical in VR and MR interaction techniques due to possible difficulties in endpoint determination and estimation of effective difficulties during interaction.

Hence, the problem to be solved in this thesis is related to the lack of standardized and comparable evaluation of pointing performance in Desktop, VR, and MR interaction techniques. Indeed, without a standardized task structure and analysis procedure, it is hard to understand to what extent differences in pointing performance are related to interaction modality or experimental conditions. Solving this problem is critical for providing valid cross-platform comparisons and for designing efficient interaction techniques in future VR and MR systems.

1.2 Thesis Goal

The objective of this thesis is to carry out a standardized cross-platform comparison of pointing performance in Desktop, VR, and MR environments, using a common framework for Fitts-based tasks [2]. In more detail, the objective of this thesis is to compare the performance of three different input modalities: mouse input for desktop environments, controller ray input for VR environments, and mid-air hand input for MR environments.

In order to fulfill this objective, an ISO-inspired multi-directional target acquisition task is implemented for each platform, with as much commonality as possible in terms of task logic, target structure, and performance metrics across all experimental conditions [2]. This will enable a comparison of input modalities more equivalently than what is typically reported in existing literature, where a lack of commonality in apparatus, task, and metrics computation may limit the extent of comparison [4].

In addition to the comparison of input modalities, a second objective of this thesis is to offer clarity on Fitts-based evaluations in immersive environments, by employing a common methodology for experimental evaluation across different platforms, as discussed in existing literature [20]. In this sense, the thesis addresses two different objectives: one empirical, in terms of comparing input modalities, and a second objective, which is more methodological in nature, in terms of standardizing experimental evaluation in immersive environments.

1.3 Research Questions

- **Desktop performance across task difficulty**

RQ 1: How does Desktop pointing performance change as the ISO task becomes more difficult?

- **Virtual Reality performance across task difficulty**

RQ 2: How does pointing performance in VR change as task difficulty increases?

- **Mixed Reality performance across task difficulty**

RQ 3: How does pointing performance in MR change as task difficulty increases?

- **Workload – NASA-TLX**

RQ 4: How does perceived workload differ across Desktop, VR, and MR during the pointing task?

- **Usability – SUS**

RQ 5: How does perceived usability differ across Desktop, VR, and MR based on participants’ SUS ratings?

1.4 Contributions

The contributions of this thesis are both empirical and methodological.

Firstly, the thesis provides a standardized implementation for a proposed ISO-inspired Fitts law-based target acquisition task for three different interaction platforms: Desktop, VR, and MR [2]. This provides a basis for comparison between the different systems that has often not been available in previous literature.

Secondly, the thesis provides an empirical comparison between three different input modalities: mouse input, controller ray input, and mid-air hand input. Through the analysis of movement time, error rate, and throughput, the thesis provides a basis for comparison between these different input modalities [4].

Thirdly, the thesis provides a methodological basis for the usage of Fitts’ law in immersive systems. This is particularly relevant since previous literature has shown a high degree of variability in XR systems [20, 4]. This variability has often made it difficult to compare results between different systems since different XR systems often have different spatial arrangements and analysis methods [4].

Finally, the thesis provides a basis for future cross-platform studies in human-computer interaction and extended reality. This proposed method could potentially be used for future research that seeks to evaluate different interaction techniques, different hardware configurations, or different adaptations for a proposed experiment [2].

Chapter 2

Related Work

2.1 The law of Fitts and standard metrics

The current applicability of Fitts-type evaluation lies on the accurate description of the goal-directed movement in the form of a speed-accuracy tradeoff. The initial concept of Paul M. Fitts' information-capacity approach suggests that the index of difficulty increases with the amplitude of the movement and the width of the target [10]. Additionally, the model suggests that the time taken for the execution of the movement increases with the level of difficulty [10, 24]. But, like the modern models of motor control, such as the ideal control methodology suggested by David E. Meyer et al., elucidate, Fitts-type effects can be described based on general principles of control [16]. Therefore, we should find similarity in the MT-ID curves between different devices do not mean that there are analogous limitations of perception and/or biomechanics. The difference plays a vital role in a thesis that aims at covering desktop mice, VR rays, and MR gestures: similarity in the quality of fits between various devices is not a substitute of similarity in defining the tasks between them. The first significant methodological question is the one that is posed when the what might seem to be a simple measurement of amplitude and width is made equivocal in the tasks that are not one-dimensional. I. Scott MacKenzie and William Buxton show that the transition between one-dimensional and two-dimensional movements involves assumptions on the reduction of bivariate movement geometry to a scalar model for the prediction of difficulty [15]. These bivariate models are further extended by Johnny Accot and Shumin Zhai who point out

that “ID” is not when both the dimensions contribute equally to a movement task that is uniquely defined [3]. The 2D pointing on a desktop computer screen results can be easily applied to XR systems [19]. In both VR and MR systems, the pointing problem is a 2D problem (left/right and near/far) by nature. It should be indicated whether it is a Euclidean distance model, a distance model along axes projected onto each screen, an angular model, or an ISO plane model is used by the thesis.

Through empirical studies conducted on 3D pointing activities, directionality/spatial configurations have been demonstrated are first order predictors, not noise. Atsuo Murata and Hirokazu Iwase conducted a study on 3D pointing tasks which extended the modeling of Fitts-style task in 3D pointing tasks, it was concluded that the direction of movement does influence pointing time in a meaningful way to the extent that parameterization by direction is justified [18]. In a different research by Murata and Daichi on eye-gaze pointing tasks, it was demonstrated that movement directionality/spatial pointing time is sensitive to configurations; so, the systematic drift of the system with respect to configurations is the Fitts parameters on constraints used in the experiment [17]. This may not be XR-specific, as per se, such findings are convincing to standardize XR comparisons to directionality or at minimally parameterize it such that it can be parameterized to make comparisons [20, 4].

The most popular standardized measures include MT, error rate, and TP. ISO-based evaluations and most XR studies have used the calculation of the throughput metric along with the effective measures including an endpoint variability to achieve a combination of speed and accuracy into a single value [2, 11, 7, 9, 6]. Theoretically, such an assessment can be used to be helpful in stabilizing the comparison of XR. Nevertheless, it is subject to the consistency of the endpoint variability calculations. Indicatively, Batmaz and Stuerzlinger talk about the effect of the application of different execution strategies on determination of the effective measure of endpoint variability [7]. This demonstrates that despite the fact that the calculation of the endpoint variability is not calculable are called as such, they are not required to be consistent under different execution strategies. The gain of the modality can be studied to indicate how the control-display mappings can have a systematic impact on the pointing tasks. This demonstrates that one value is not are necessary to obtain the results as a by-product of the modality or the inherent constraints of the modality being

used [22].

2.2 ISO 9241-411 and multi-directional tapping task

The multi-directional circular layout is a basis ISO-based operationalization of target. The research in virtual reality involves acquisition tasks [2]. Gabel et al. did a multi-directional implementation. The calculation of effective throughput at different levels of difficulty is calculated based on ISO 9241-411 to calculate task based on the ISO 9241-411 levels (3.0 to 4.5 bits) of both baseline and redirected raycasting conditions [11]. Similarly, Batmaz et al. explored the effects of vergence-accommodation conflict (VAC) on peripersonal pointing by stretching the ISO 9241-411:2015 multi-directional task, which was equipped with the ability to be used as an appropriate comparisons of performance under various display conditions [6]. Further research by Dube et al. compared mid-air gesture selection techniques with an ISO-like framework to measure the time of movement and the error rates at constant amplitude and constant width manipulations [9]. These group activities prove that the ISO-style framework can be successfully used combine mouse-based with XR-based pointing tasks into a single analytical model, given the relationships between coordinate systems are accurately specified [2, 4, 20].

An important factor that must be taken into account when XR adaptations of the ISO standard are concerned is the effect of selection confirmation of endpoint variability [7]. In most VR systems, the same device is used both pointing and selection, e.g. a controller trigger, can add a Heisenberg effect, in which there is a physical action of clicking that alters the position of the pointer [21]. While research means that redirection methods may be useful to optimize throughput, but not necessarily to increase it in error rates, the process of selection and evaluation is methodologically complex [11, 21, 9]. As Batmaz and Stuerzlinger emphasized, throughput in VR is the one that is sensitive to the particular execution strategy employed by participants that stressed that although the term ISO-style throughput is a sound one, it is nonetheless a metric, however, is not quite device-independent [7]. This highlights the need to have the same timing of a state machine across desktop, VR and MR modules to make cross-platform valid comparisons in this thesis.

Lastly, it is necessary to make sure that the ISO-style geometry is overlaid on platforms in a transparent manner necessary to study cross-reality [21, 4]. Angular amplitudes and widths were used by Shi et al. to give indices of difficulty giving a parameterization in terms of visual angles that is especially well adapted to a head-mounted display [21]. Also, a study conducted by Zhou et al. with respect to the incorporation of desktop mice into VR spaces it was shown that depth-adaptive cursors have the potential to significantly decrease the time and error rates of completion in comparison to baseline mappings [25].

2.3 Fitts law in VR/MR and cross study variability

One of the key claims that a literature review can make, with the help of quantitative evidence, is that the area of research into Fitts law in the domain of extended reality (XR) is rich, but methodologically fragmented [4, 20, 7]. Mohammadreza Amini et al. review a systematic review. With 122 XR user studies, 119 publications, including the application of Fitts law in the context of three-dimensional XR environments [4]. The meta-analysis results imply that the The most studied is Effective Measures formulation; it was mentioned in particular within 61 (50.0%) of the studies, which increases to 78 (63.93%) studies when those that likely Use the ISO 9241 Effective Measures, without the explicit reporting of the formula, is included (effective measures are mentioned in the ISO 9241 document) [4]. However, a more important point for the current thesis is the 25 (20.5%) studies that failed to explicitly see the variant employed in the calculations, as they cited the law of Fitts or ISO 9241 and not state the particular variant utilized in the study [4]. This represents a high-impact support to the current thesis since the discipline needs to be standardized and reproducible cross-platform task generators.

The main studies incorporated in the set provided help to prove the reasoning supporting the large variance even with the use of the Fitts law. The placement of depth can essentially change the acquisition dynamics. An experiment on attaining moving objects with motion-in-depth in virtual reality refers to the activities that were involved in the study and proposes the assumption of 2D pointing is not directly projected to the acquisition of moving targets [13, 5]. The choice of coordinate system can be a consideration in mixed reality/augmented

reality-like settings. Chen et al. compare viewport-based pointing vs. world-based pointing using everyday smart devices in immersive augmented reality [8]. The authors indicate that the same pointing suggests that the same pointing takes place can have a different motor and perceptual structure. This directly affects the meaning of amplitude and width. Mid-air pointing might also be greatly influenced by fatigue. An extended Fitts model Lou et al. propose an extended Fitts model to overcome the problems of spatial variables and the impact of fatigue in arms on hand interaction in VR [14]. Monitoring and feeling restrictions can play a large role in the cross-study dispersion. The study of mobile VR controller found a throughput of 2.2 bits/s [12]. The authors propose that the quality and stability of the stability of the identified companies of the tracking are the performance bottleneck. This would have a significant impact on endpoint dispersion particularly relative to high-fidelity tracked VR controllers [12, 7]. A study found that assistive redirected ray casting improves the mean effective throughput to 5.10–5.17 bits/s compared to a baseline of 4.08 bits/s [11]. The median MT decreased to 0.80–0.82 s compared to a baseline of 0.98 s [11]. This shows the material influence of interaction design of the Fitts law measures within a standardised ISO style system. A experiment on productivity based VR interaction revealed that a depth adaptive cursor minimized the mean time to completion: 21.21% and lower median error rate 48.25% [25]. This demonstrates the enormous influence of the “mouse in VR” on the overall performance.

The fact that evidence is provided in the context of MR/AR indicates that the real-world situation may have an impact on increasing the variability. This paper is a research on the assessment of interaction methods following the law of Fitts as they move outdoors through AR [23]. The study obtained a mean throughput of 1.98 (0.5) with the technique of interaction called the gaze pinch with a median of 2.05. This is far much greater than the hand push and ray pinch techniques, which had a mean throughput of 4–4.5 seconds [23]. This is similar to the MR condition in which the movement of the user in the body is not limited. This shows the importance of the so-called reality condition, in which the dynamism of the surrounding world is considered into account [23, 8].

2.4 Stability of throughput and computation discussions

Throughput is widely used as it is a single measure of the speed and accuracy at the same time depending on the level of difficulty and movement time. In the set of attached XR papers in the ISO style, calculations are made the beneficial throughput and address it as an output of the evaluation protocol [11, 7, 9, 6]. However, there are two debates of importance which are directly pertinent to a thesis of a standardized ISO-style desktop/VR/MR task: (i) task consistency of throughput with changing execution strategies and (ii) the impact of the variation in computation on comparability.

The closest evidence to depend on strategy is provided by Anil Ufuk Batmaz and the analysis of various strategy of executing tasks to be performed by mid-air Fitts, by Wolfgang Stuerzlinger tasks in VR. Their analyses report substantial effects of strategies on throughput (e.g., throughput $F(2, 34) = 16.066$, $p < 0.001$, $\eta^2 = 0.486$ in one analysis; and large strategy effects in per-technique analyses such as throughput $F(2, 34) = 73.958$, $p < 0.001$, $\eta^2 = 0.813$ for virtual hand and throughput $F(2, 34) = 45.58$, $p < 0.001$, $\eta^2 = 0.728$ for raycasting) [7]. This is in direct opposition to the simplifying assumption that the effective measures completely normalize speed - accuracy strategies and it suggests that a cross-platform study should be standardized and it should or (a) fix instruction and training in a rigorous way, or (b) make strategy/time-on-task an explicit factor.

Values digitized-from-figure (explicit note). According to their study, a number of important means graphically are reported (e.g., Fig. 3a/3b). In the case of Table A below, throughput and time values of this source are digitized estimates of figures and must be used as rough aids to qualitative claims (strategy dependence), not as exact tabulated statistics [7].

Throughput comparisons are further complicated by the computation and reporting variation. A recent systematic review identifies that 25 (20.5%) XR studies mention Fitts' law or ISO without simply reporting the variant used, suggesting that throughput values can indicate the different conventions of formula or effective-measure even where so marked [4, 20].

Table 2.1: Comprehensive overview of Fitts’ law and ISO-style multi-directional pointing studies across Desktop, VR, and MR/AR contexts.

Study	Platform & Setup	Task Geometry	Variables (IVs & DVs)	Key Design Notes
Batmaz and Stuerzlinger (2022) [7]	VR ($N = 18$) Pointing by virtual hand or raycasting; non-dominant hand confirmation to avoid motion interference.	Explicit 3D ISO 9241-411-style; 11 circular targets; widths 1.5–3.5 cm; 6 IDs (2.19–4.14).	IVs: Execution strategy, interaction technique, size, amplitude. DVs: Effective throughput, MT, ER, endpoint variability.	3×2 within-subject design. Strong methodology paper addressing throughput debates; shows execution strategy drastically impacts throughput.
Gabel et al. (2023) [11]	VR ($N = 26$) Controller raycasting with visual assistance/redirected rays.	ISO 9241-411-based planar multi-directional Fitts task at ~4m depth (IDs: 3.0–4.5).	IVs: Ray condition (redirection), ID level. DVs: Effective throughput, MT, workload, agency.	Multi-condition within-subject. Excellent design-comparison study; geometry is held constant while ray assistance changes to improve throughput.
Shi et al. (2023) [21]	VR ($N = 17$) Controller raycasting with target expansion; non-dominant trigger confirmation.	ISO 9241-411:2012 angular tapping task; amplitudes 15°/20°; 6 IDs (2.402–3.841).	IVs: ID, expansion factor, amplitude, width. DVs: MT, ER, throughput.	6×4 within-subject. Clean study of selection assistance within strict ISO ring geometry; uses angular formulation rather than pixels/meters.
Continued on next page				

Table 2.1 – continued from previous page

Study	Platform & Setup	Task Geometry	Variables (IVs & DVs)	Key Design Notes
Batmaz et al. (2023) [6]	VR ($N = 18$) Peripersonal pointing using virtual hand and ray-casting.	Extension of ISO 9241-411:2015; 11 targets under constant and varying depth relationships.	IVs: VAC condition, technique, depth arrangement. DVs: MT, ER, throughput.	Within-subject. Extends ISO-style evaluation to display-quality factors (Vergence-Accommodation Conflict) while keeping a standardized task.
Zhou et al. (2022) [25]	VR ($N = 12$) Desktop mouse integrated into VR.	3D target selection adapted for 2D mouse input; targets varied in depth and planar size.	IVs: Mapping type (baseline vs depth-adaptive), target depth. DVs: Completion time, ER.	Demonstrates that depth-adaptive cursors substantially reduce completion time; ISO equivalence relies heavily on mapping calibration.
Lou et al. (2025) [14]	VR ($N = 20$) Mid-air hand tracking interaction.	Extended 3D pointing task across different depth layers and interaction distances.	IVs: Spatial variables, interaction distance. DVs: MT, ER, arm fatigue.	Proposes an extended Fitts model incorporating arm fatigue as a factor; multiple blocks over repeated trials to measure fatigue accumulation.
Dube et al. (2022) [9]	Mid-air / MR (N/A) Mid-air selection with ultrasonic haptic feedback.	Fitts-style / ISO-style throughput evaluation.	IVs: Selection method (Push, Tap, Dwell, Pinch), amplitude, width. DVs: Throughput, MT, ER.	Strong mid-air methodology paper evaluating gesture-comparison; assesses multiple selection activation methods.
Continued on next page				

Table 2.1 – continued from previous page

Study	Platform & Setup	Task Geometry	Variables (IVs & DVs)	Key Design Notes
Rajanna and Hammond (2022) [19]	Desktop (N/A) Gaze, touch, and mouse input.	ISO 9241-9 standard multi-directional evaluation.	IVs: Input modality, amplitude, width. DVs: Throughput, MT, ER.	Useful desktop baseline comparing traditional and alternative 2D inputs against an ISO standard framework.
Grinyer and Teather (2023) [12]	Mobile VR (N/A) Camera-tracked low-cost controller.	ISO 9241-9 standard multi-directional evaluation.	IVs: Selection activation method. DVs: Throughput, selection time.	Explores tracking stability as a fundamental performance bottleneck in low-fidelity mobile environments.
Tatzgern et al. (2025) [23]	AR (N/A) Head-mounted display used outdoors in motion.	ISO 9241-9 multi-directional Fitts evaluation.	IVs: Interaction technique (gaze/hand/ray), mobility state. DVs: Throughput, MT, ER.	Explores impact of real-world dynamic environments (e.g., walking outdoors) on pointing techniques; highlights mobility as a disruption factor.
Huang et al. (2022) [13]	VR (N/A) Controller input.	3D Fitts-style motion-in-depth target acquisition.	IVs: Target depth, motion speed, target width. DVs: MT, ER.	Highlights that traditional 2D pointing assumptions do not directly translate to 3D environments when targets move along the depth (z) axis.

Chapter 3

System Design and Implementation

3.1 Selection of platform and pointing devices

The hardware and interaction specifications of the selected platforms are summarized in Table 3.1.

	Desktop	Virtual Reality	Mixed Reality
Model	Windows PC	Meta Quest II	Microsoft HoloLens 2
Input Type	Mouse	VR Controller Index Trigger Button	Mid-air gesture

Table 3.1: Platform Specifications

3.1.1 Desktop

The Desktop platform, represented by a Windows PC, is the most common interaction paradigm and for most of people, which evolved from early graphical user interfaces into highly optimized systems. Pointing is performed indirectly through cursor control, offering high accuracy and low latency, making it a reliable baseline for Fitts' Law evaluation. However, the interaction is limited to a flat display, lacking depth perception and spatial context. The separation between physical input and on-screen movement introduces abstraction, and prolonged use may lead to minor physical fatigue. The application for the study was developed by using the Unity 3D game engine.

3.1.2 Virtual Reality

VR refers to completely immersive artificial environments created on computers and replacing the user's experience of the actual environment around him. Examples of such VR technology include Meta Quest 2, where the system has evolved to head-mounted displays with built-in tracking features. The use of Unity 3D with OpenXR technology helps create interactive 3D environments that allow inputs using controllers and hands. Direction of the pointer is achieved through ray-casting from the controller's end. Although there is increased immersion in VR due to its depth and enhanced experience, it comes with various challenges such as imprecision, hand shaking, and fatigue ("gorilla arm" issue). The VR device used in this study is shown in Figure 3-1.



Figure 3-1: VR device used: Meta Quest II head-mounted display.

3.1.3 Mixed Reality

Mixed Reality (MR) combines real-world and virtual worlds in a way that allows digital objects to be embedded in the physical world and interact with it. This combination is

accomplished using technology like the HoloLens 2 through spatial mapping, sensor technology, and head-up displays. Programming is performed using Unity 3D, MRTK, and OpenXR technologies, which enable gesture- and gaze-based user inputs. Pointing is done using hand gestures and eye gaze techniques, offering intuitive interactions but with reduced accuracy. Some of the main problems with MR are limited field of vision, incorrect gesture detection, and environmental sensitivity, which increase cognitive loads. The MR device used in this study is shown in Figure 3-2.



Figure 3-2: MR device used: Microsoft HoloLens 2 headset.

3.2 Explanation of developed software

The system is divided into three interaction modules, with the same basic task generator used for all modules, as well as the same basic data logging/analysis interface. The basic idea is that, although the input layer and coordinate mapping layer may differ between



Figure 3-3: Unity-based 2D Graphical User Interface displaying the multi-directional target layout for the Desktop environment.

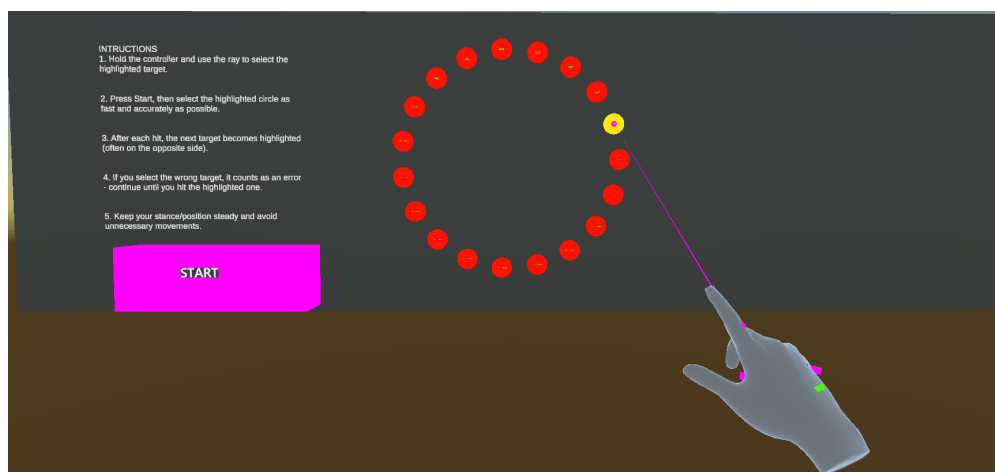


Figure 3-4: Unity-based 3D Graphical User Interface displaying the multi-directional target layout for Virtual Reality environment.

platforms, the experimental logic, which includes condition scheduling, target sequencing, trial timing rules, error handling, and file formats, is the same for all platforms. This is a requirement for meeting the thesis goal of comparing mouse, controller ray, and mid-air hand input within a standardized ISO style pointing task without confounding interaction style with task definition style. The overall logic flow of the Fitts' Law task is illustrated in Figure 3-6.

The experimental testbed can be divided into the following layers:

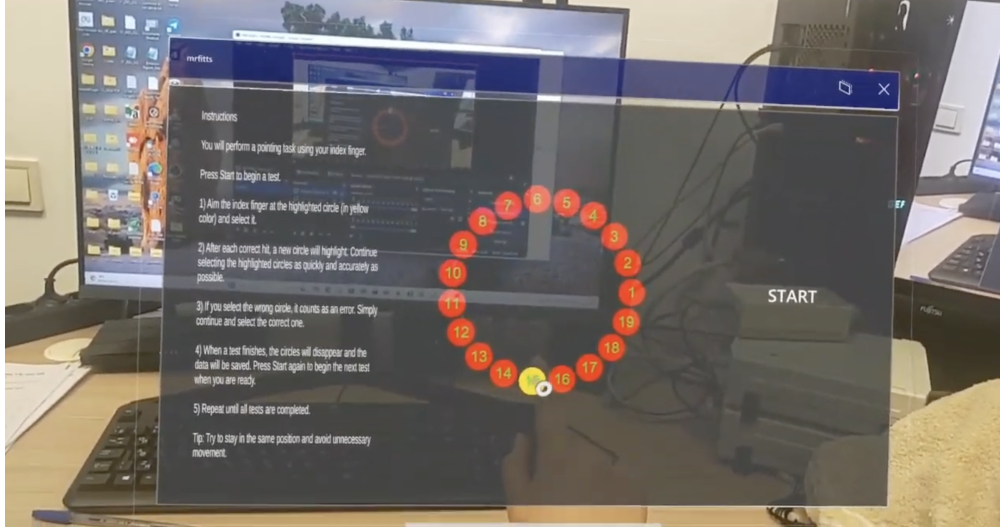


Figure 3-5: Unity-based 3D Graphical User Interface displaying the multi-directional target layout for Mixed Reality environment.

1. **Core Task Generator:** This is the platform-independent layer, which defines the ISO style multi-directional pointing task, specifying the target position on the circle, the target activation sequence, the trial/block schedule, as well as the independent variables (amplitude/width/ID). This layer sends the same abstract trial specification to the three platform modules.
2. **Platform Adapter:** This layer translates the abstract trial specification, which was generated by the core task generator, into platform coordinates and rendering primitives:
 - **Desktop:** Renders the screen-space position, as well as the pixel width, as a two-dimensional target, with the mouse cursor as the input device. The developed desktop GUI is shown in the figure 3-3.
 - **VR:** Renders the world-space position at a fixed depth plane, with a controller pose as the input device, as well as a button press as the selection mechanism.
 - **MR:** Renders the world-locked position in the real world, with a hand ray derived from the hand pose as well as the hand orientation as the input device, as well as a mid-air gesture, which can be a pinch, a tap, a dwell, or the equivalent of an air tap. A developed 3D MR and VR GUI is shown in figure 3-5.

3. **Interaction Provider (Input Abstraction):** The third module offers a general interface for a pointer state and a selection event, independent of a specific modality. The conceptual signals from all three modules are the same:

- **Ray(t):** The origin and direction of a ray, or a 2D pointer position interpreted as a ray pointing into the screen plane.
- **Select(t):** The confirmation event for a selection process.

As a result, all three modules perform the same operations for detecting a hit, calculating movement time, and determining an error type.

4. **Experiment Controller and Logger:** A platform-independent module for managing the trial state machine and logging trial data in a standardized CSV format. Optional high-frequency data for head and hand/controller pose is collected by a separate buffer recorder that does not affect performance in experiments.

The architecture outlined above results in a three-module system that has the same core task generator:

1. **Desktop Mouse Module:** The user can select targets by a 2D pointer. The targets appear on the display as a 2D figure. The amplitude or width can be defined in pixels. The 2D click point is given by $(xClick, yClick)$, and a 3D point can optionally be calculated and output by the module.
2. **VR Controller Ray Module:** The user can point at targets by a ray that starts at a controller. The ray's origin and direction are given by the position of the ray's end point. The ray's end point can be moved by a trigger or a button. The ray hits a target and gives a 3D click point $(xClick, yClick, zClick)$ and a hit/miss decision. The target appears in the scene as a 2D figure on a world-locked plane.
3. **MR Hand Ray / Mid-Air Gesture Module:** The user interacts by means of a ray derived from a hand-based ray (e.g., based upon hand orientation, index direction, or wrist-index vector), and selection is performed by means of a mid-air gesture or an

air tap equivalent. The Interaction Provider outputs Ray(t) and Select(t) are identical for both. Targets are world-locked and located in a comfortable interaction frustum.

Another benefit that can be obtained from this method is that comparability between different platforms can be defined in terms of input modality and environment while at the same time using the same underlying experiment design: the same target sequence generation, the same trial timing rules, and the same data schemas.

3.2.1 ISO-style task geometry

The layout parameters for the testbed interface are summarized in Table 3.2.

Table 3.2: Testbed Interface layout parameters

Test	Ring radius (m)	Width W (m)	Amplitude A (m)
Test1	2.0	0.6	3.9
Test2	2.4	0.5	4.7
Test3	2.8	0.4	5.4
Test4	3.2	0.3	6.2

The task geometry conforms to an ISO 9241-9 (411) style multi-directional pointing task [1, 2] in which there are N targets equally spaced upon a circle. One target at a time is highlighted. The subject repeatedly points to the highlighted target as fast and accurately as possible subject to amplitude and width constraints. The underlying generator always creates the same target geometry that is parameterized by amplitude A and width W .

3.2.2 Odd number of targets to maintain constant amplitude

In order to maintain a constant amplitude during movement along a circle for the multi-directional stepping sequence, an odd number of targets is utilized (e.g., $N = 19$). Studies have revealed that ISO-style circular target sequences utilizing an even number of targets can cause amplitude alternation [20]. The implementation makes use of a fixed step size:

$$s = \frac{N - 1}{2}.$$

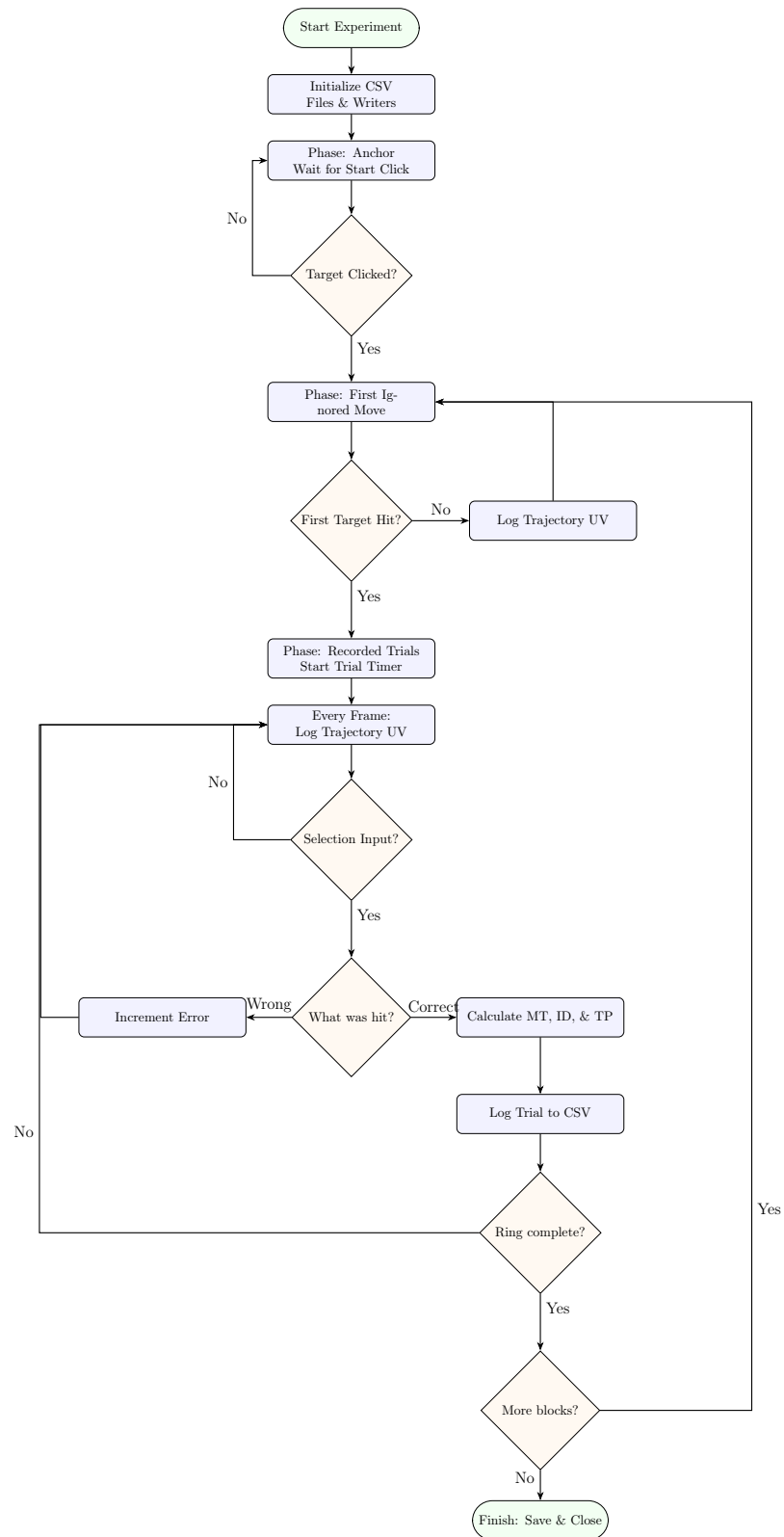


Figure 3-6: ISO-9241 Fitts' Law Task Logic Flow

The target index is incremented by $s \bmod N$ in each trial:

$$nextIndex = (currentIndex + s) \bmod N.$$

Near-opposite jumps occur while at the same time ensuring that the target chord length between targets remains constant in all trials since the step size is fixed for a regular polygon.

3.2.3 Parameterization of Circular Layout

Let R be the radius of the circle, where R is in pixels for desktop applications and in meters for VR/MR applications. The target positions are given by

$$\theta_i = \frac{2\pi i}{N}$$

for $i = \{0, \dots, N-1\}$. The target centers are given by:

- **Desktop:** $P_i = (x_c + R \cos \theta_i, y_c + R \sin \theta_i)$
- **VR/MR:** $P_i = (x_c + R \cos \theta_i, y_c + R \sin \theta_i, z_0)$

The movement amplitude A is given by

$$A = \|P_{i+s} - P_i\| = 2R \sin\left(\frac{\pi s}{N}\right).$$

This means A is the distance between two active targets when step movement is performed.

The movement amplitude A is related to R by

$$A = 2R \sin\left(\frac{\pi s}{N}\right).$$

The relationship between A and R is given by

$$A = 2R \sin\left(\frac{\pi s}{N}\right),$$

where A is given.

3.2.4 Cross-Platform Equivalence of Amplitude A and Width W

To make cross-platform comparisons between Desktop, VR, and MR applications, we equate A and W . The system offers two methods of doing this.

1. **World Metric Equivalence (VR/MR):** The targets are located on a plane located at a specific depth z_0 . A is given in meters, and we compute R in meters using

$$A = 2R \sin\left(\frac{\pi s}{N}\right).$$

The targets are then displayed with width W .

2. **Angular Equivalence (VR/MR):** A is given in degrees, where A is converted into a measurement at a specific depth d :

$$W_{world} = 2d \tan(W_{ang}/2), \quad A_{world} = 2d \tan(A_{ang}/2)$$

(with linearization for small angles as appropriate). This approach standardizes the retinal sizes for the VR/MR platforms [21].

3.2.5 Logging and Data Format

All three modules have a common logging format. The format of a log entry is such that the entire experiment situation can be recreated. This includes the target under selection, its location, values of A and W , the selection made by the participant, movement time taken, and outcome classification.

Trial Log Format

The trial log (`_writer`) records a summary of each completed pointing movement. This trial logger helps to capture required metrics for Fitts' law analysis, including timing, accuracy, and task geometry. The data fields captured in each trial log are described in Table 3.3.

Table 3.3: Trial Log Data Fields

Column Name	Unit/Format	Description
participant	String	Unique participant identifier.
block	Integer	1-based block sequence number.
ringIndex	Integer	Index of the ring configuration from the settings list.
phase	String	Current phase: Anchor, FirstIgnoredMove, Practice, or Recorded.
trial	Integer	Sequential movement number within the current ring.
from / to	Integer	1-based IDs of the start and destination targets.
N	Integer	Total number of targets in the ring.
ringRadius_m	Meters	Physical radius of the target ring.
W_m	Meters	Target diameter (Width).
A_wall_m	Meters	Euclidean distance between targets on the wall plane.
ID_bits	Bits	Task difficulty: $\log_2(A/W + 1)$.
MT_ms	Milliseconds	Time taken from previous selection to current hit.
errors	Integer	Count of target selections other than the active goal.
misclicks	Integer	Total count of incorrect selection events during the trial.
click_ts_iso	ISO 8601	UTC timestamp of the selection event.
interactor	String	Name of the input device (e.g., "MouseLeftClick").

Chapter 4

Experiment

4.1 Ethical Compliance and Protocol Framework

This experiment methodology was developed to maintain methodological soundness and quality of the obtained results based on the recommendations given by *ITU-T P.913*. This standard played an important role in the design of the entire process of data collection, which involved participant selection, evaluation of subjective quality, and more.

The study was approved by the Nazarbayev University Institutional Research Ethics Committee (NU IREC) and was conducted in accordance with the ethical principles of the Declaration of Helsinki. All participants were informed about the study objectives, procedures, potential discomfort, and their right to withdraw at any stage without penalty.

The questionnaires used in this study are available at: *questionnaires link*.

4.1.1 Subjective Measures

To capture usability, workload, physical strain, and user experience across modalities, the following instruments were administered:

- Informed Consent Form
- Pre-test Questionnaire
- Discomfort Questionnaire

- Strain Questionnaire
- Interaction and UX Questionnaire
- System Usability Scale (SUS)
- NASA Task Load Index (NASA-TLX)

These instruments were selected to provide complementary measures of usability, perceived workload, interaction quality, and modality-specific discomfort during immersive and non-immersive task execution.

4.2 Participants

A total of 5 participants were recruited from the School of Engineering and Digital Sciences (SEDS) at Nazarbayev University from the departments related to the field of study (e.g. Computer Science, Data Science, and Robotics majors students were recruited) using a convenience-based volunteer sampling approach. The mean age of participants is 22 years with standard deviation of 2 years. All participants took part in a single-session study and completed the task under all three interface modalities: Desktop, Virtual Reality (VR), and Mixed Reality (MR).

A within-subject repeated-measures design was adopted to reduce inter-participant variability and enable direct cross-modality comparison. To mitigate order effects, participants were assigned to three counterbalanced groups (A, B, and C), as illustrated in Fig. 4-1. This grouping ensured that modality order was rotated systematically across participants.

4.3 Experimental Design

The study employed a within-subject design with three interface modalities:

- **Desktop:** mouse-based interaction
- **VR:** controller-based interaction

- **MR:** mid-air gesture interaction

Each participant completed all three modalities. Within each modality, participants performed four unique test conditions. To reduce learning, fatigue, and carry-over effects, two levels of counterbalancing were applied:

1. **ounter-balancing in modality:** the presentation order of Desktop, VR, and MR was rotated across Groups A, B, and C using a Latin-square-inspired sequence.
2. **Counter-balancing in test conditions:** the four test conditions within each modality were presented using rotating condition orders.

This design enabled controlled comparison of performance and subjective responses across interface modalities while limiting systematic bias due to exposure order.

4.4 Experimental Procedure

Each participant participated in one experimental session that lasted approximately 90–105 minutes. The overall session consisted of participant instruction time, baseline assessment, three modality blocks, post-block questionnaires, and rest periods illustrated on Figure 4-1.

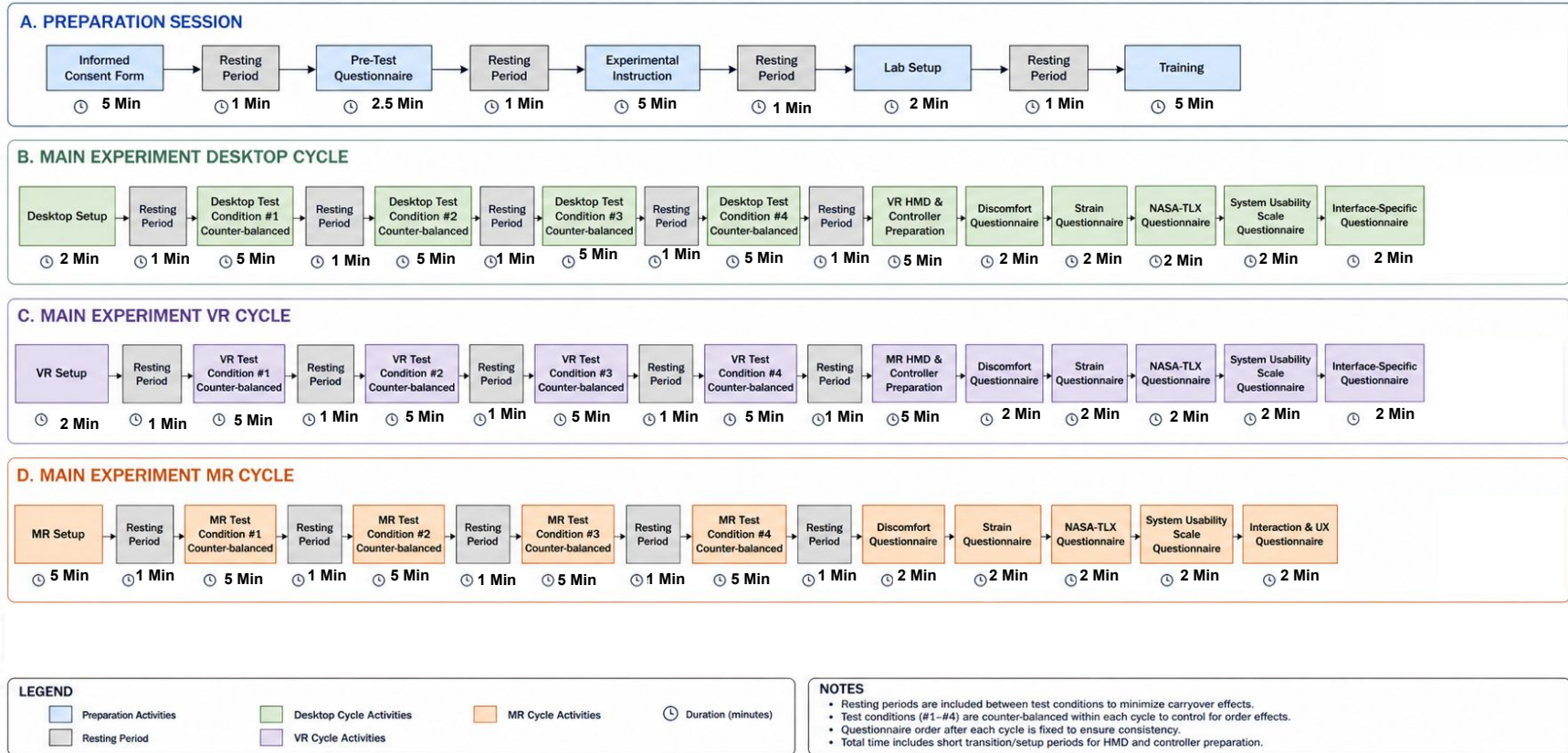


Figure 4-1: Experimental Protocol Flow

4.4.1 Preparation Session

Participants were welcomed at the beginning of the session, briefed on the research study procedure, and asked to complete the pre-experiment questionnaire and training. This stage included:

- Informed Consent Form (5 min)
- Pre-Test Questionnaire (2.5 min)
- Experimental Instruction (5 min)
- Lab Setup (2 min)
- Training (5 min)

Each of these steps was separated by a 1-minute resting period. The total preparation time was 23.5 minutes.

4.4.2 Main Experiment Blocks

Each modality block followed the same structure to maintain procedural consistency:

1. Setup: Device fit and setup (2 min for Desktop and VR; 5 min for MR).
2. Task execution: Four counter-balanced experimental conditions completed within the current modality.
3. Resting periods: A 1-minute resting period was provided before the first condition and after each condition.
4. Transition Preparation: A 5-minute preparation and transition period for the next modality's HMD and controllers (included after the Desktop and VR cycles).
5. Post-modality questionnaires: Administration of subjective measures.

4.4.3 Task Timing

Within each modality, participants completed four counter-balanced test conditions. A typical timing structure was as follows:

- Condition 1: 5 min
- Resting Period: 1 min
- Condition 2: 5 min
- Resting Period: 1 min
- Condition 3: 5 min
- Resting Period: 1 min
- Condition 4: 5 min
- Resting Period: 1 min

Thus, the total active task and inter-task resting time per modality cycle was exactly 24 minutes.

4.4.4 Post-Modality Assessment

After completing the four conditions (and the 5-minute transition preparation phase where applicable), participants completed a set of subjective questionnaires to capture immediate post-use responses. Based on the experimental flow, these included:

- Discomfort Questionnaire (2 min)
- Strain Questionnaire (2 min)
- NASA-TLX Questionnaire (2 min)
- System Usability Scale (SUS) Questionnaire (2 min)
- Interface-Specific Questionnaire / Interaction & UX Questionnaire (2 min)

The total questionnaire time per modality was exactly 10 minutes.

4.5 Break Management and Participant Safety

To minimize fatigue accumulation and discomfort carry-over, both short transitions and formal rest intervals were incorporated into the protocol. A resting period of exactly 1 minute was provided before and after consecutive conditions within each modality. In addition, a 5-minute preparation and rest break was scheduled between modality blocks to safely transition to the next HMD and controllers.

Participants were also allowed to request an additional 2–5 minute recovery break whenever needed, particularly in the presence of:

- dizziness,
- nausea,
- eye strain,
- headache,
- arm fatigue,
- gesture fatigue,
- reduced concentration.

Participants were explicitly informed that they could pause the experiment, request further rest, or withdraw from the study at any point without negative consequences.

- dizziness,
- nausea,
- eye strain,
- headache,
- arm fatigue,
- gesture fatigue,

- reduced concentration.

Participants were explicitly informed that they could pause the experiment, request further rest, or withdraw from the study at any point without negative consequences.

4.6 Overall Session Duration

Based on the above procedure, each modality block required approximately 28–30 minutes, including setup, practice, task execution, questionnaires, and rest. Across the three modalities, the full experimental session required approximately 90–105 minutes per participant.

4.7 Data Collected

The study collected both objective and subjective data. Objective data included task completion behavior recorded across all conditions and modalities. The subjective data were gathered using questionnaires to determine the level of discomfort, exertion, usability, interaction, and cognitive load.

This approach allowed for an extensive evaluation of the interactions in Desktop, VR, and MR systems from the viewpoint of human factors and user experience.

4.8 Dependent Variables

In this study, the following variables were used for analysis:

- Movement Time (MT): time required to complete a target acquisition.
- Index of Difficulty (ID): a measure of task difficulty based on target size and distance from the pointer.
- Throughput (TP): a combined measure of speed and accuracy, expressed in bits per second. The higher is better.

Fitts' law is modeled by using the standard linear formula (Equation 4.1) [10]:

$$MT = a + b \cdot ID \tag{4.1}$$

where a represents the intercept and b represents the slope, indicating sensitivity to task difficulty.

Chapter 5

Results

This chapter contains the experimental results divided into objective performance measures, and subjective user measurements. It is analysed based on the comparative performance of three interaction platforms: Desktop, Virtual Reality (VR), and Mixed Reality (MR).

5.1 RQ 1: Desktop performance across task difficulty

The Desktop condition quantitative results demonstrate a definite influence of the difficulty of the tasks on user performance. The target width narrowed as the test proceeded to Test4, as compared to Test1 between 0.6 m and 0.3 m, and the task arrangement changed to one that posed greater difficulty. Correspondingly, the average time of movement (MT) changed between Test1 (708.3 ms) and Test4 (936.2 ms). that pointed out that participants required longer to finish more challenging pointing tests. Simultaneously, the average effective index of difficulty (IDe) was consistently growing since it was 3.93 bits to 6.29 bits which confirms that the difficulty of the task was increased throughout the four desktops tests. Although this has increased the difficulty, throughput (TP) also rose to 5.55 bits/s to 6.72 bits/s. This suggests that participants were able to maintain efficient performance within more challenging situation of tasks.

In general, the Desktop results show that, the more difficult the task, the longer the movement. The times, yet it did not affect the performance of interaction between the users as noted by the. slow increase in throughput.

The aggregated results across four test conditions of varying index of difficulty (ID) are summarized in Table 5.1.

Platform	Test	Radius (m)	Width (m)	Mean MT (ms)	Mean ID (bits)	TP (bits/s)
Desktop	Test1	2.0	0.6	708.3	3.93	5.55
	Test2	2.4	0.5	757.9	4.72	6.23
	Test3	2.8	0.4	830.6	5.50	6.62
	Test4	3.2	0.3	936.2	6.29	6.72
VR	Test1	2.0	0.6	948.5	3.93	4.14
	Test2	2.4	0.5	1043.2	4.72	4.52
	Test3	2.8	0.4	1154.8	5.50	4.76
	Test4	3.2	0.3	1355.6	6.29	4.64
MR	Test1	2.0	0.6	3196.4	3.93	1.23
	Test2	2.4	0.5	3887.5	4.72	1.21
	Test3	2.8	0.4	4359.2	5.50	1.26
	Test4	3.2	0.3	5023.8	6.29	1.25

Table 5.1: Aggregated means per platform and test condition (Fitts’ Law report)

5.2 RQ 2: How does pointing performance in VR change as task difficulty increases?

The results of the quantitative analysis for the VR experiment reveal that performance worsened with increased difficulty in the task. During Test1 to Test4, the target size decreased from 0.6 m to 0.3 m, leading to an increase in task difficulty. As a result, the mean movement time (MT) increased from 948.5 ms in Test1 to 1355.6 ms in Test4, implying that participants took more time to perform the pointing task under virtual reality due to the reduction in the target size and increasing task difficulty.

Likewise, the mean effective index of difficulty (IDe) gradually increased from 3.93 bits to 6.29 bits, supporting the conclusion that the tasks became increasingly difficult from Test1 to Test4. Furthermore, the throughput (TP) significantly increased from 4.14 bits/s in Test1 to 4.76 bits/s in Test3, showing improved interaction efficiency with adaptation to the VR system. However, TP slightly fell to 4.64 bits/s in Test4, implying that the greatest task difficulty created performance limitations.

In summary, the VR experiment shows that higher task difficulty led to higher MTs and slight changes in TP. Unlike the Desktop experiment, VR interaction was less efficient, and the slight fall in TP at maximum difficulty implies higher challenge in maintaining consistent performance on the most difficult VR task.

5.3 RQ 3: How does pointing performance in MR change as task difficulty increases?

For the MR scenario, the quantitative results suggest that user performance was significantly affected by the growing complexity of the task. With the reduction of the target width from 0.6 m to 0.3 m between Test1 and Test4, the MT value was seen to have risen sharply from 3196.4 ms in Test1 to 5023.8 ms in Test4. This shows that participants were able to take much longer to complete their pointing task in the MR condition due to increased complexity.

The IDe value, which showed an increment from 3.93 bits in Test1 to 6.29 bits in Test4, further corroborates that there was progressively increased difficulty in pointing in all four MR tests performed. Contrasting Desktop and VR, the TP value was found not to be varying much, ranging only from 1.23 bits/s in Test1 to 1.25 bits/s in Test4, although there was a slightly lower TP in Test2 at 1.21 bits/s and slightly higher TP in Test3 at 1.26 bits/s. This indicates that although the tasks became more difficult and movement times increased, the overall interaction efficiency in MR remained consistently low.

Overall, the MR results suggest that increasing task difficulty led to a large increase in movement time, while throughput showed very little change. This means that MR interaction was much slower than Desktop and VR, and users were less efficient in handling increasingly difficult pointing tasks.

As shown in Table 5.1 and Figure 5-1, the Desktop environment consistently provided the highest throughput, peaking at 6.72 bits/s. VR showed competitive performance but was approximately 30% slower than Desktop. The MR platform exhibited significant performance degradation, with movement times exceeding 3,000 ms and a notably lower throughput of approximately 1.25 bits/s, suggesting a higher difficulty in target acquisition using gesture-

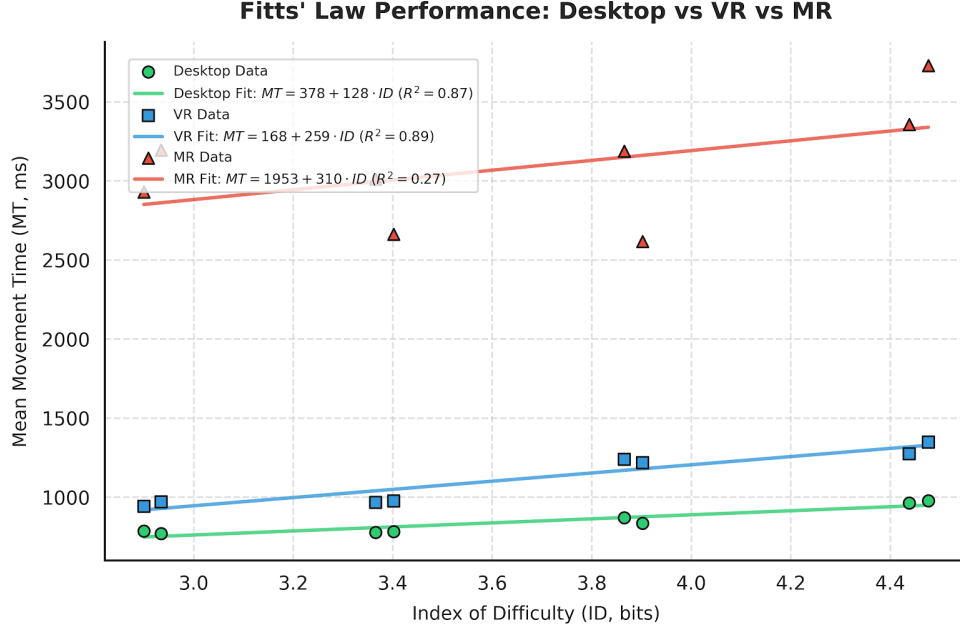


Figure 5-1: Fitts' Law Regression comparison across Desktop, VR, and MR modalities.

based interaction.

5.3.1 Overall Objective Performance Comparison

For a thorough evaluation of the objective measures (MT and TP), a Two-Way Repeated Measures ANOVA test was used where Modality and Task Difficulty were the within-subjects variables as shown in Table 5.2. This revealed a statistically significant effect of Modality on Movement Time ($F(2, 8) = 39.80, p < 0.001$) and Throughput ($F(2, 8) = 375.51, p < 0.001$). This means that the interaction platform significantly affected the time it takes to acquire the target as well as the efficiency of its acquisition. In agreement with Fitts' Law [10], Task Difficulty led to a highly significant main effect on Movement Time ($F(3, 12) = 12.40, p = 0.0005$) because target acquisition times increased for all three modalities with increasing index of difficulty. Moreover, a significant interaction was observed between Task Difficulty and Modality in regards to MT ($F(6, 24) = 3.50, p = 0.012$) and TP ($F(6, 24) = 2.74, p = 0.036$), which indicated that the increase in task difficulty had a non-uniform impact on target acquisition in the different environments tested. The Desktop modality showed good robustness to increases in difficulty as it only incurred an approximately 228

ms penalty from the simplest to the most difficult condition. On the other hand, immersive environments were less resilient with VR incurring a moderate penalty of approximately 407 ms and MR experiencing significant degradation where target acquisition took over 1,800 ms longer in the hardest condition compared to the simplest one. Taken together, the above results show that whereas there is compensation for increasing target difficulty when the mouse device is used, unsupported gesture interaction (MR) shows significant difficulties with respect to target acquisition under high spatial precision requirements.

Source of Variance	Metric	df (num, den)	<i>F</i> -Value	<i>p</i> -value	Significance
Modality	MT	(2, 8)	39.80	< 0.001	***
	TP	(2, 8)	375.51	< 0.001	***
	ER	(2, 8)	1.41	0.298	n.s.
Difficulty	MT	(3, 12)	12.40	0.0005	***
	TP	(3, 12)	12.50	0.0005	***
	ER	(3, 12)	1.70	0.220	n.s.
Modality × Difficulty	MT	(6, 24)	3.50	0.012	*
	TP	(6, 24)	2.74	0.036	*
	ER	(6, 24)	1.77	0.148	n.s.

Table 5.2: Two-Way Repeated Measures ANOVA Results for Objective Measures. Each (*) sign represent significance level. [n.s. - not significant]

5.4 RQ 4: How does perceived workload differ across Desktop, VR, and MR during the pointing task?

An itemized analysis of the dimensions used in NASA-TLX was performed based on the results obtained, which are presented in Table 5.4 below. A One-Way Repeated Measures ANOVA test showed that the interaction modality significantly influenced the overall perceived workload, since the interaction modality had an impact on the overall NASA-TLX scale ($F(2, 8) = 5.53, p = 0.0310$), implying that the hierarchy of perceived workload was as follows: Desktop had the lowest perceived workload whereas MR had the highest.

Mental Demand (MD) Cognitive complexity differed greatly among the interaction modalities. Specifically, both the Desktop and VR interactions were considered to be equally

not mentally challenging (both had the same mean value equal to $M = 1.60$). In turn, the MR interaction mode was viewed as significantly more demanding (with $M = 3.00$). Such results indicate that there was additional cognitive load associated with monitoring mid-air gestures and understanding the hand-tracking visualization without any buttons.

Physical Demand (PD) There was a noticeable increase in physical effort when comparing different interaction modalities. Specifically, interacting with the desktop computer required the least amount of effort ($M = 1.60$). VR interaction implied that participants would have to hold, lift, and aim a handheld controller, increasing physical effort somewhat ($M = 2.40$). However, MR yielded the highest physical effort rating by far ($M = 3.80$), which is most probably explained by the so-called “gorilla arm” phenomenon, since continuous use of mid-air gestures requires constant work of muscles supporting the hand in air [14].

Temporal Demand (TD) Time pressure was fairly low throughout the study regardless of the interaction modality used. Both Desktop and VR modes received similar scores (corresponding to low demand with the former one having $M = 1.60$ and the latter one having $M = 1.80$), and a small increase in time pressure ($M = 2.40$) was registered for MR. This means that although users found it physically and mentally challenging to interact with a system in MR, they were not necessarily experiencing any time-related discomfort, since there was no set pace in the experiment.

Performance (PF) Finally, there was a high level of congruence between perceived performance and the objective hierarchy of efficiency described above in relation to Fitts’ law. Users felt extremely successful using the Desktop ($M = 4.40$) and the VR ($M = 4.20$) interfaces, while performing tasks in MR decreased their satisfaction greatly ($M = 2.80$). The reduced effectiveness of MR matches precisely with the increased time of movement and reduced throughput, which were measured objectively, signifying that the participants were fully cognizant of their reduced targeting effectiveness using mid-air gestures [23]. Comparative results are shown in Figure 5-2.

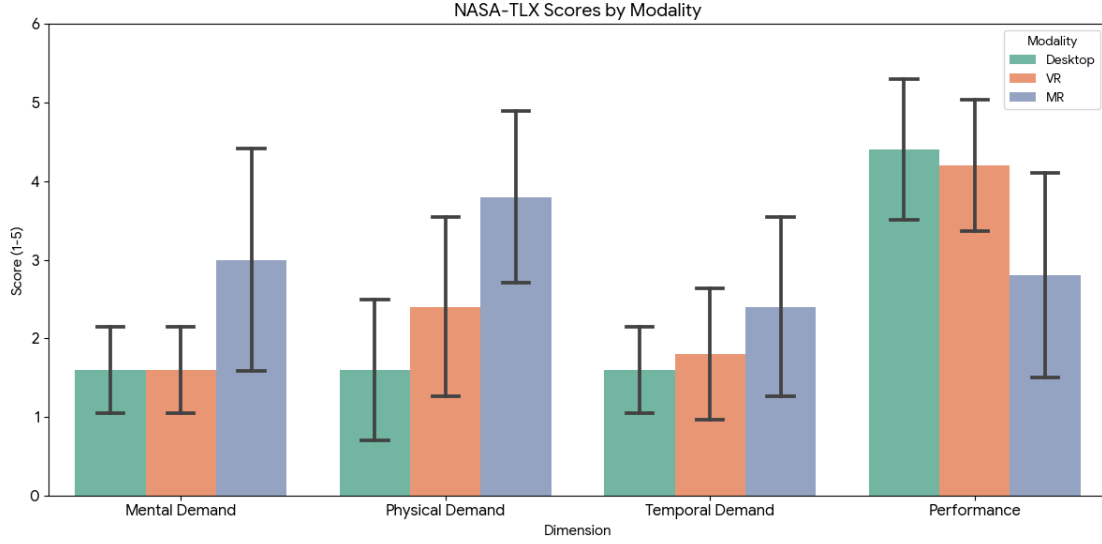


Figure 5-2: NASA-TLX Scores

5.5 RQ 5: How does perceived usability differ across Desktop, VR, and MR based on participants' SUS ratings?

For the identification of the specific usability issues generated by immersive technologies, a detailed analysis of System Usability Scale (SUS) has been conducted (Table 5.3). The One-Way Repeated Measures ANOVA showed highly statistically significant differences between the interaction modalities in relation to the aggregated SUS scores ($F(2, 8) = 40.28, p < 0.0001$). Desktop interaction mode scored the “Excellent” usability rating of 90.00 ($SD = 9.01$); VR scored the “Good” rating of 74.00 ($SD = 11.81$); while MR scored the “Poor” rating of 42.50 ($SD = 13.46$). An evaluation of specific SUS statements is necessary to understand the reasons behind the differences.

Ease of Use and Confidence (Positive Dimensions)

Among all positive SUS statements (Q1, Q3, Q5, Q7, Q9), Desktop interface outperformed all immersive interaction modes in all categories. Participants considered the Desktop interface extremely intuitive (Q3: $M = 4.80$) and were extremely confident about the usage of it (Q9: $M = 4.80$). VR interface also demonstrated good performance in relation

to ease of use and user confidence (Q3: $M = 3.60$; Q9: $M = 3.60$), which might be related to the successful mapping of physical pointing to the corresponding actions via raycasting mechanism. As for the MR platform, it lagged far behind in terms of these categories, as users reported relatively low ease of use (Q3: $M = 2.60$) and low confidence in the system (Q9: $M = 2.20$). Moreover, while participants reported a strong willingness to use the Desktop application as frequently as possible (Q1: $M = 4.00$), they exhibited highly negative feelings towards frequent MR interactions (Q1: $M = 1.40$), presumably related to the high physical effort required by this task.

System Complexity and Inconsistency (Negative Dimensions)

The participants' responses to negative SUS statements (Q2, Q4, Q6, Q8, Q10) indicated specific sources of difficulty associated with the MR interaction mode. Users found the MR interface to be rather complex (Q2: $M = 3.40$) and inconvenient to use (Q8: $M = 2.40$). Interestingly enough, both Desktop (Q2: $M = 1.20$) and VR (Q2: $M = 1.60$) interfaces were considered highly intuitive and easy to use. Particularly noteworthy is the high discrepancy regarding inconsistency of the interaction (Q6: $M = 3.40$). In this case, the inconsistency score reflected the inherent issues such as occasional failure of detection or occlusion [12]. Consequently, participants had a feeling that they faced a steep learning curve to successfully operate the MR system (Q10: $M = 2.40$) which is compared to the deeply familiar Desktop environment (Q10: $M = 1.40$). The SUS detailed responses are prepared in Table 5.3.

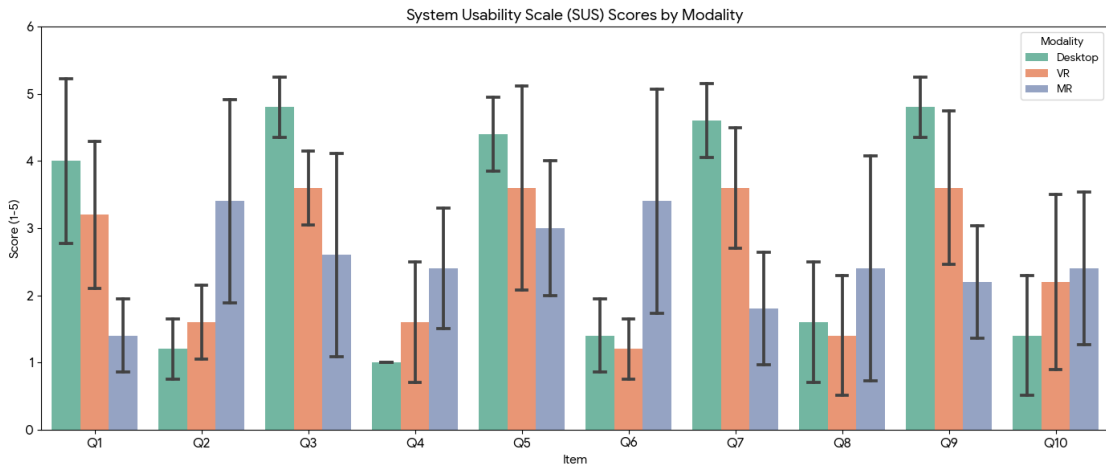


Figure 5-3: System Usability Scale Comparison.

Modality	Q1		Q2		Q3		Q4		Q5		Q6		Q7		Q8		Q9		Q10	
	Mean	STD	Mean	STD	Mean	STD	Mean	STD	Mean	STD	Mean	STD	Mean	STD	Mean	STD	Mean	STD	Mean	STD
VR	3.20	1.10	1.60	0.55	3.60	0.55	1.60	0.89	3.60	1.52	1.20	0.45	3.60	0.89	1.40	0.89	3.60	1.14	2.20	1.30
MR	1.40	0.55	3.40	1.52	2.60	1.52	2.40	0.89	3.00	1.00	3.40	1.67	1.80	0.84	2.40	1.67	2.20	0.84	2.40	1.14
Desktop	4.00	1.22	1.20	0.45	4.80	0.45	1.00	0.00	4.40	0.55	1.40	0.55	4.60	0.55	1.60	0.89	4.80	0.45	1.40	0.89

Table 5.3: System Usability Scale (SUS) Scores by Modality

Modality	MD		PD		TD		PF	
	Mean	STD	Mean	STD	Mean	STD	Mean	STD
Desktop	1.60	0.55	1.60	0.89	1.60	0.55	4.40	0.89
MR	3.00	1.41	3.80	1.10	2.40	1.14	2.80	1.30
VR	1.60	0.55	2.40	1.14	1.80	0.84	4.20	0.84

Table 5.4: NASA-TLX Scores by Modality. MD - Mental Demand. PD - Physical Demand. TD - Temporal Demand. PF - Performance

The same was followed in the case of System Usability Scale (SUS) scores. The Desktop modality had the highest SUS score of 90.00 ($SD = 9.01$) indicating excellent perceived usability. The VR modality received a SUS score of 74.00 ($SD = 11.81$), which indicated acceptable but not as usable as Desktop. The MR modality, on the other hand, was the lowest. SUS score of 42.50 ($SD = 13.46$), which indicates that perceived usability is low in comparison to the other. two platforms.

In general, both tests have consistently demonstrated that Desktop offered the best usability, VR offered moderate usability and MR was perceived to be the least usable platform when carrying out the task of pointing. These results indicate that users were more relaxed, productive, and at ease in dealing with the Desktop system, but the MR system presented more challenges. usability challenges.

5.5.1 Aggregated User Feedback

The aggregated scores for all participants ($N = 5$) are presented in Table 5.5. These summaries highlight the significant differences in user experience between the three systems.

Subjective assessment demonstrated a consistent hierarchy of performances and comforts for all systems. Analysis of the physiological effect on subjects demonstrates that the Mixed Reality (MR) setup produced the greatest amount of perceived workload ($M = 3.95$) and

Modality	Strain (1-5)	Discomfort (1-5)	NASA-TLX (1-5)	Usability (1-5)	SUS (0-100)
Desktop	1.80 (0.71)	1.48 (0.66)	2.00 (0.35)	4.06 (0.66)	90.00 (9.01)
VR	2.56 (0.88)	2.36 (0.90)	2.50 (0.44)	3.48 (0.52)	74.00 (11.81)
MR	3.95 (0.74)	3.68 (0.70)	3.20 (0.92)	2.76 (0.30)	42.50 (13.46)

Table 5.5: Mean and Standard Deviation (SD) for Questionnaire Results ($N = 5$)

physiological stress ($M = 3.68$), presumably caused by the physical fatigue produced by unsupported movements in mid-air. On the other hand, the Desktop system produced the smallest values of strain ($M = 1.80$) and stress ($M = 1.48$).

The levels of workloads measured using the NASA-TLX test were significantly different among the platforms used ($F(2, 8) = 5.53, p = 0.0310$). Desktop was rated the easiest to use with a mean of 2.00, while MR had the highest level of workload with a mean of 3.20. This can be ascribed to the physical demands of acquiring targets.

In terms of system usability, the pattern was almost identical, with high statistically significant differences ($F(2, 8) = 40.28, p < 0.0001$). The desktop system was rated 90.00 points on the SUS scale, which is considered “Excellent”, whereas the MR system received 42.50 points, which corresponds to “Poor” usability. The VR system had intermediate results – 74.00 SUS points that were called “Good”.

Usability ratings specific to the interface also showed that users perceived the Desktop mouse interaction to be more reliable ($M = 4.06$) in comparison to the MR gestures ($M = 2.76$). Factors such as the uncertainty of hand tracking and gesture detection were among the causes for greater unreliability of MR interactions.

As for the subjective workloads on different platforms, they were evaluated using the NASA-TLX tool. While the Desktop version seemed to require the least amount of effort from users (Mean = 2.00), MR involved the most (Mean = 3.20). In addition, usability ratings revealed that users perceived Desktop interfaces to be much more usable (Mean SUS = 90 – “Excellent”) than those in MR (Mean SUS = 42.5 – “Poor”).

5.5.2 Statistical Significance Testing

To determine the significance of these results, a One-Way Repeated Measures ANOVA was conducted. As shown in Table 5.6, the choice of platform had a statistically significant effect across all measured dimensions ($p < 0.05$).

Hypothesis	F-Value	DF (num, den)	p-value	Significance
NASA-TLX Workload	5.53	(2.0, 8.0)	0.0310	*
SUS	40.28	(2.0, 8.0)	< 0.0001	***

Table 5.6: Repeated Measures ANOVA Results for Hypothesis Testing. Each (*) sign represent significance level.

The high significance in SUS scores ($p < 0.0001$) and general usability ($p = 0.0004$) validates the performance trends observed in the objective Fitts' Law analysis, suggesting that the MR system's interaction method was notably less intuitive and more physically taxing for participants.

Chapter 6

Discussion and Conclusions

It can be noted that the empirical findings demonstrate a reliable hierarchy of performances that is sustained on all three platforms under examination. Specifically, all participants were ranked consistently based on their performance, where Desktop proved the most efficient modality with mean response times $M = 845.15$ ms, while VR was placed in the middle with mean response time $M = 1117.54$ ms. At last, MR proved to be the least efficient with mean response time $M = 3084.82$ ms. Moreover, throughput was following the same inverse pattern, having its peak value of 6.72 bits per second on Desktop and falling to 1.25 bits per second on MR platform. The parallel trends suggest a hypothesis that the identified discrepancies are driven by interactional modalities rather than individual participants' variation.

The hypothesis was confirmed by high levels of statistical significance found during the repeated measures ANOVA, when comparing system usability, $F(2, 8) = 40.28$, $p < 0.0001$, and interface-specific usability $F(2, 8) = 24.05$, $p = 0.0004$. In addition, the standardized task correlated well with Fitts' Law [10], especially in case of VR platform ($R^2 = 0.89$) and Desktop ($R^2 = 0.87$). Thus, one can conclude that the initial idea of using a stable task generator was methodologically justified. MR modality showed higher variability ($R^2 = 0.27$) as well as higher physical demand. However, no signs of fatigue were observed during exploratory time-on-task analysis.

- Standardized comparison across platforms is feasible

- Desktop remains the most efficient interaction method
- VR is a viable but slower alternative
- MR introduces significant performance challenges

6.1 Key Findings

The main empirical finding is a stable ordering of platform performance. All five participants showed the same mean ordering, with Desktop fastest, VR intermediate, and MR slowest. Throughput followed the inverse ordering, again consistently across participants. This strengthens the interpretation that the observed differences are modality-related rather than being driven only by random participant-to-participant variation. A second important finding is that the standardized task still preserved strong Fitts-law structure in all three realities. This supports the methodological premise of the thesis: keeping the task generator and reporting pipeline stable enables interpretable cross-reality comparison. Finally, the exploratory time-on-task analysis did not show clear fatigue-related degradation in the available logs.

6.2 Design Implications

For practical interface design purposes, it is clear that in situations where desktop equipment can be used and a period of prolonged targeting is required, the traditional use of the mouse will be the fastest method. If virtual environment interaction is needed, the use of VR controllers as a ray will be a viable alternative, although it is significantly slower than desktop input. Mid-air interactions may be appropriate when the activity is hands-free and/or physically anchored, but in such circumstances, the slowness of the mid-air interaction must be compensated for with larger targets that are less accurate and that do not involve quick selections [9]. Lastly, an important point for researchers and practitioners alike from this study is that future research on cross-platform interaction must report their task geometry, ID calculations, timing logic, and effective performance equations clearly.

6.3 Limitations

There are limitations that must be recognized and taken into consideration in future studies. Firstly, this research is a pilot experiment, and the sample size is very small, including only 5 participants, which made it challenging to obtain pairwise inference reliability even when the sign of the effect remains constant in each comparison. Secondly, participants who volunteered in the study lacked experience with MR, which influenced their experience during the experiment due to jitter and lack of confidence [23]. Thirdly, the obtained findings are particular to the devices and implementations employed in this study, since Desktop, VR, and MR values depend on various elements, including hardware specifications, stability of the tracking, confirmation methods, and calibration quality. Lastly, the experiment collected the Strain and Discomfort questionnaire data, however due to the small sample size, this study omitted the data from these questionnaires.

6.4 Future Work

It is recommended that future research expand upon this pilot study by increasing the size and diversity of the sample and recording workload, discomfort, and fatigue questionnaires in a synchronised fashion. There needs to be improvement in how the time-on-task factor is implemented: maintaining the same participant identifier through longer sequences or fully counterbalancing difficulty within time in order to eliminate fatigue as a confound. Further research might benefit from including more difficult situations, such as occlusion, clutter, depth variation, and/or access awareness stages which combine search and acquisition behaviors. Finally, if this study were to be replicated on other systems, the current advantages, especially having a shared task generator and logging system, should remain intact, while simultaneously measuring other factors that affect performance such as endpoint variance, effective width, and motion quality measures.

- Strong modality-driven performance differences
- Important implications for interface design

- Need for larger and more comprehensive studies

Chapter 7

Appendix

This appendix contains the materials and documents used during the experiment of this research. In order to provide research transparency regarding the experimental protocol, participant onboarding, and subjective data collection, this appendix includes:

- **Written Informed Consent Form**
- **Pre-Test Questionnaire**
- **NASA Task Load Index (NASA-TLX)**
- **System Usability Scale (SUS)**
- **Response to Reviewers**

Written Informed Consent Form ISO-based Fitts' law study across Desktop, VR, and MR

Comparative Evaluation of Fitts' Law Target Selection Tasks in Desktop, Virtual Reality, and Mixed Reality Environments

Introduction and Purpose: You are invited to participate in a research study that investigates human performance and user experience during target selection tasks in three different interactive environments: Desktop, Virtual Reality (VR), and Mixed Reality (MR).

The purpose of this study is to compare how these three modalities influence **task performance, usability, workload, physical discomfort, strain, and overall interaction experience** during Fitts' law-based pointing and selection tasks. The findings will help improve the design of interactive systems and user interfaces for immersive and non-immersive environments.

Procedures:

The study will take place in the **Multimedia Laboratory (Room C4 405), Nazarbayev University**.

You will participate individually, and only the researcher(s) will be present during the session.

The total participation time is approximately **90 to 105 minutes**. During the session, you will be asked to:

1. Read and sign this consent form.
2. Complete a short **pre-test questionnaire** about your background, demographics, and previous experience with VR, MR, and similar interactive systems.
3. Receive instructions and a short practice session for each modality.
4. Perform a series of **Fitts' law target selection tasks** using:
 - a **Desktop** interface with a mouse,
 - a **VR** interface with VR controllers,
 - an **MR** interface using mid-air gestures.
5. Complete **four test conditions** in each modality.
6. Fill in short questionnaires after each modality related to:
 - discomfort,
 - strain,
 - interaction and user experience,
 - system usability,
 - perceived workload.
7. Take scheduled rest breaks between modality blocks. You may also request additional breaks at any time.

Written Informed Consent Form ISO-based Fitts' law study across Desktop, VR, and MR

The order of Desktop, VR, and MR conditions will be varied across participants to reduce order effects.

Risks

The risks involved in this study are minimal and similar to those encountered during normal computer or headset use. Possible temporary effects may include:

- mild eye strain,
- visual fatigue,
- arm or hand fatigue,
- slight dizziness or motion discomfort while using VR or MR,
- temporary physical strain due to repeated pointing or gesture-based interaction.

You may stop the experiment at any time if you feel uncomfortable, tired, or unwell.

Benefits

There is no guaranteed personal benefit from participating in this study. However, you may gain experience using Desktop, VR, and MR interaction systems. Your participation will contribute to research on human performance, usability, and user-centered design for immersive interfaces.

Compensation

There is **no monetary compensation** for participation in this study.

Confidentiality

All information collected during this study will be kept confidential. Your data will be stored under a participant code (for example, **P01**) instead of your name.

- Signed consent forms and the participant code list will be stored securely by the researcher.
- Experimental data, including **performance logs, questionnaire responses, and interaction records**, will be stored on a password-protected computer.
- Only the research team will have access to the identifiable data.
- Results will be reported in **aggregated or anonymized form**, and no individual participant will be identifiable in publications or presentations.

Written Informed Consent Form ISO-based Fitts' law study across Desktop, VR, and MR

Voluntary Participation

Your participation in this study is entirely voluntary. You may refuse to participate or withdraw from the study at any time, without giving a reason and without any negative consequences.

If you are a student, your decision to participate or not participate will **not** affect your grades, academic standing, or relationship with the university.

Contacts

If you have any questions about the study, please contact:

Dr. Syed Muhammad Umair Arif

Email: umair.arif@nu.edu.kz

Pre-Test Questionnaire – Comparative Evaluation of Fitts' Law Target Selection Tasks in Desktop, Virtual Reality, and Mixed Reality Environments

Date/Time:

A. Provide your answer to the questions below:

Subject ID			
Date of Birth			
Gender			
Profession / Occupation			
Do you wear glasses/lenses?	Yes	No	
What is your native language?			
Education specialty or domain			
Education Level	Bachelor	Masters	PhD

B. Prior Technology and Interaction Experience

For the following questions, please select one option:

0 = Never

1 = 1–2 times

2 = 3–10 times

3 = More than 10 times

9. How often do you play 3D computer or console games

(e.g., first-person games, racing games, simulation games)?

0 ☐ 1 ☐ 2 ☐ 3 ☐

10. Have you previously used a Virtual Reality (VR) headset

(e.g., Meta Quest, HTC Vive, Pico)?

0 ☐ 1 ☐ 2 ☐ 3 ☐

11. Have you previously used a Mixed Reality (MR) or Augmented Reality (AR) headset

(e.g., HoloLens or similar device)?

0 ☐ 1 ☐ 2 ☐ 3 ☐

12. How familiar are you with motion-controller interaction in VR

(e.g., pointing, ray-casting, target selection using controllers)?

0 ☐ 1 ☐ 2 ☐ 3 ☐

13. How familiar are you with mid-air gesture interaction

(e.g., hand tracking, pinch, air tap, gesture-based target selection)?

0 ☐ 1 ☐ 2 ☐ 3 ☐

Pre-Test Questionnaire – Comparative Evaluation of Fitts' Law Target Selection Tasks in Desktop, Virtual Reality, and Mixed Reality Environments

14. **How often do you use a computer mouse for precise selection tasks**
(e.g., clicking small targets, dragging, interface navigation, design, gaming)?
0 ☐ 1 ☐ 2 ☐ 3 ☐
15. **Have you previously participated in experiments involving pointing, selection, reaction time, or target acquisition tasks?**
0 ☐ 1 ☐ 2 ☐ 3 ☐

C. Current State and Readiness

16. **How many hours did you sleep last night?**
☐ 1 = Less than 4 hours
☐ 2 = 4–6 hours
☐ 3 = 6–8 hours
☐ 4 = More than 8 hours
17. **Did you drink coffee in the past 24 hours?**
☐ 0 = No
☐ 1 = 1–2 cups
☐ 2 = 3–4 cups
☐ 3 = More than 4 cups
18. **Did you consume other caffeinated drinks in the past 24 hours**
(e.g., strong tea, energy drinks, caffeinated soft drinks)?
☐ 0 = No
☐ 1 = 1–2 times
☐ 2 = 3–4 times
☐ 3 = More than 4 times
19. **Do you usually experience motion sickness, dizziness, or discomfort during VR use, 3D games, or travel**
(e.g., car, bus, boat)?
☐ 0 = Never
☐ 1 = Rarely
☐ 2 = Sometimes
☐ 3 = Often
20. **Do you currently feel tired, strained, or unwell?**
☐ No
☐ Yes, please specify briefly: _____
21. **Do you have any visual, balance, hand-motor, or neurological condition that may affect interaction performance or headset use?**
☐ No
☐ Yes, please specify briefly: _____

ISO-based Fitts' law study across Desktop, VR, and MR SUS Questionnaire

Instructions Please rate how strongly you agree with the following statements about the interaction experience you had with this system.

1 = Strongly disagree 2 = Disagree 3 = Neutral 4 = Agree 5 = Strongly agree

SUS for Desktop

Desktop Interface (Mouse-Based Interaction)

In the following statements, “**this system**” refers to the **Desktop mouse interface** you just used.

I think that I would like to use this system frequently.

1	2	3	4	5
---	---	---	---	---

I found this system unnecessarily complex.

1	2	3	4	5
---	---	---	---	---

I thought this system was easy to use.

1	2	3	4	5
---	---	---	---	---

I think that I would need the support of a technical person to be able to use this system.

1	2	3	4	5
---	---	---	---	---

I found the various functions in this system were well integrated.

1	2	3	4	5
---	---	---	---	---

I thought there was too much inconsistency in this system.

1	2	3	4	5
---	---	---	---	---

I would imagine that most people would learn to use this system very quickly.

1	2	3	4	5
---	---	---	---	---

I found this system very cumbersome to use.

1	2	3	4	5
---	---	---	---	---

ISO-based Fitts' law study across Desktop, VR, and MR SUS Questionnaire

I felt very confident using this system.

1	2	3	4	5
---	---	---	---	---

I needed to learn a lot of things before I could get going with this system.

1	2	3	4	5
---	---	---	---	---

SUS for Virtual Reality (VR)

Virtual Reality Interface (Controller-Based Interaction)

In the following statements, “**this system**” refers to the **VR controller interface** you just used.

I think that I would like to use this system frequently.

1	2	3	4	5
---	---	---	---	---

I found this system unnecessarily complex.

1	2	3	4	5
---	---	---	---	---

I thought this system was easy to use.

1	2	3	4	5
---	---	---	---	---

I think that I would need the support of a technical person to be able to use this system.

1	2	3	4	5
---	---	---	---	---

I found the various functions in this system were well integrated.

1	2	3	4	5
---	---	---	---	---

I thought there was too much inconsistency in this system.

1	2	3	4	5
---	---	---	---	---

I would imagine that most people would learn to use this system very quickly.

1	2	3	4	5
---	---	---	---	---

ISO-based Fitts' law study across Desktop, VR, and MR SUS Questionnaire

I found this system very cumbersome to use.

1	2	3	4	5
---	---	---	---	---

I felt very confident using this system.

1	2	3	4	5
---	---	---	---	---

I needed to learn a lot of things before I could get going with this system.

1	2	3	4	5
---	---	---	---	---

SUS for Mixed Reality (MR)

Mixed Reality Interface (Mid-Air Gesture Interaction)

In the following statements, “**this system**” refers to the **MR mid-air gesture interface** you just used.

I think that I would like to use this system frequently.

1	2	3	4	5
---	---	---	---	---

I found this system unnecessarily complex.

1	2	3	4	5
---	---	---	---	---

I thought this system was easy to use.

1	2	3	4	5
---	---	---	---	---

I think that I would need the support of a technical person to be able to use this system.

1	2	3	4	5
---	---	---	---	---

I found the various functions in this system were well integrated.

1	2	3	4	5
---	---	---	---	---

I thought there was too much inconsistency in this system.

1	2	3	4	5
---	---	---	---	---

ISO-based Fitts' law study across Desktop, VR, and MR SUS

Questionnaire

I would imagine that most people would learn to use this system very quickly.

1	2	3	4	5
---	---	---	---	---

I found this system very cumbersome to use.

1	2	3	4	5
---	---	---	---	---

I felt very confident using this system.

1	2	3	4	5
---	---	---	---	---

I needed to learn a lot of things before I could get going with this system.

1	2	3	4	5
---	---	---	---	---

NASA-TLX Questionnaire ISO-based Fitts' law study across Desktop, VR, and MR.

Use the scale below for each item: 1 = Very Low 5 = Very High

NASA-TLX: Desktop

Please rate your workload while performing the ISO-based Fitts' law target selection task using the **Desktop mouse interface**.

Mental Demand: How mentally demanding was the task?

1	2	3	4	5
---	---	---	---	---

Physical Demand: How physically demanding was the task?

1	2	3	4	5
---	---	---	---	---

Temporal Demand: How hurried or rushed did you feel due to the pace of the task?

1	2	3	4	5
---	---	---	---	---

Performance: How successful were you in accomplishing the task goals?

1	2	3	4	5
---	---	---	---	---

Effort: How hard did you have to work to achieve your level of performance?

1	2	3	4	5
---	---	---	---	---

Frustration Level: How insecure, discouraged, irritated, stressed, and annoyed did you feel during the task?

1	2	3	4	5
---	---	---	---	---

NASA-TLX: VR

Please rate your workload while performing the ISO-based Fitts' law target selection task using the **VR controller interface**.

Mental Demand: How mentally demanding was the task?

1	2	3	4	5
---	---	---	---	---

NASA-TLX Questionnaire ISO-based Fitts' law study across Desktop, VR, and MR.

Physical Demand: How physically demanding was the task?

1	2	3	4	5
---	---	---	---	---

Temporal Demand: How hurried or rushed did you feel due to the pace of the task?

1	2	3	4	5
---	---	---	---	---

Performance: How successful were you in accomplishing the task goals?

1	2	3	4	5
---	---	---	---	---

Effort: How hard did you have to work to achieve your level of performance?

1	2	3	4	5
---	---	---	---	---

Frustration Level: How insecure, discouraged, irritated, stressed, and annoyed did you feel during the task?

1	2	3	4	5
---	---	---	---	---

NASA-TLX: MR

Please rate your workload while performing the ISO-based Fitts' law target selection task using the **MR mid-air gesture interface**.

Mental Demand: How mentally demanding was the task?

1	2	3	4	5
---	---	---	---	---

Physical Demand: How physically demanding was the task?

1	2	3	4	5
---	---	---	---	---

Temporal Demand: How hurried or rushed did you feel due to the pace of the task?

1	2	3	4	5
---	---	---	---	---

Performance: How successful were you in accomplishing the task goals?

1	2	3	4	5
---	---	---	---	---

NASA-TLX Questionnaire ISO-based Fitts' law study across Desktop, VR, and MR.

Effort: How hard did you have to work to achieve your level of performance?

1	2	3	4	5
---	---	---	---	---

Frustration Level: How insecure, discouraged, irritated, stressed, and annoyed did you feel during the task?

1	2	3	4	5
---	---	---	---	---

Dear Reviewers,

Thank you for your constructive feedback on my master's thesis. Your insights regarding methodological standardization, statistical rigor, and document formatting have been invaluable in improving the quality and clarity of this research.

Below is a detailed, point-by-point summary of the revisions made to the manuscript in response to your comments, including excerpts from the updated thesis to demonstrate how each issue was addressed.

1. Abstract & Introduction (Chapter 1)

Reviewer Comment: "Clearly state the research gap and novelty of the thesis."

Response: The Abstract and Section 1.1 (Problem Statement) were rewritten to explicitly highlight that the primary gap in current Extended Reality (XR) literature is the lack of methodological standardization. The thesis is now clearly framed as a novel solution that provides a standardized ISO-inspired framework to isolate modality-related effects from methodological differences.

How it was addressed:

"However, prior Fitts' law research in XR has been methodologically heterogeneous, with substantial variation in task geometry, spatial configuration, input modality, and throughput computation, making cross-study and cross-platform comparison difficult. To address this problem, this thesis presents a standardized ISO-inspired multi-directional target acquisition task for comparing pointing performance across Desktop, VR, and MR under a common experimental and analytical framework."

Reviewer Comment: "In RQ 3, is it VR or should be MR? Please check before submission."

Response: This typographical error was corrected in Section 1.2 to maintain professional consistency.

How it was addressed:

"RQ 3: How does pointing performance in MR change as task difficulty increases?"

2. Literature Review (Chapter 2)

Reviewer Comment: "Remove all conference and journals names"

Response: Removed informal mentions of conferences (e.g., "VRST'23", "IEEE VR'23") and journal rankings (e.g., "A Q1 IJHCS study").

How it was addressed:

"The multi-directional circular layout is a foundational ISO-based operationalization for target acquisition tasks in virtual reality research. Gabel et al. implemented a multi-directional Fitts task based on ISO 9241-411 to calculate effective throughput across varying difficulty levels..."

Reviewer Comment: "Improve consistency of terminology (ISO 9241-9 vs ISO 9241-411)."

Response: Clarifying text was added to acknowledge that while the thesis builds on the foundational legacy of ISO 9241-9, it references the updated ISO 9241-411 for specific multi-directional task evaluation guidelines.

How it was addressed:

"While ISO 9241-9 originally established the foundational framework for non-keyboard input device evaluation, its updated revision, ISO 9241-411, details the specific multi-directional task evaluation methods primarily utilized today. Therefore, this thesis references the ISO 9241-411 guidelines while building upon the legacy of the ISO 9241-9 standard."

Reviewer Comment: "Ensure all tables are properly referenced... Verify missing citations inside tables."

Response: Tables 2.1 and 2.2 were reorganized, and bracketed citations (e.g., [6], [10], [20]) were injected directly into the "Study" and "Source / Year" columns of the tables to ensure every study mentioned was properly linked to the bibliography.

3. System Design & Implementation (Chapter 3)

Reviewer Comment: "Explain why odd-number targets were selected."

Response: Section 3.2.2 was completely rewritten to include the mathematical and methodological justification. It explains that an odd number of targets prevents the amplitude alternation/fluctuation often seen in even-numbered circular layouts.

How it was addressed:

"In order to maintain a constant amplitude during movement along a circle for the multi-directional stepping sequence, an odd number of targets is utilized (e.g., $N=19$). Studies have revealed that ISO-style circular target sequences utilizing an even number of targets can cause amplitude alternation. The implementation makes use of a fixed step size... Near-opposite jumps occur while at the same time ensuring that the target chord length between targets remains constant..."

Reviewer Comment: "Clarify how Desktop, VR, and MR task equivalence was maintained."

Response: Section 3.2.4 was updated to clearly define both "World Metric Equivalence" and "Angular Equivalence" to prove that amplitude (A) and width (W) were standardized fairly across 2D and 3D planes.

How it was addressed:

"To make cross-platform comparisons between Desktop, VR, and MR applications, we equate A and W. The system offers two methods of doing this: 1. World Metric Equivalence (VR/MR)... 2. Angular Equivalence (VR/MR)... This approach standardizes the retinal sizes for the VR/MR platforms."

4. Experiment (Chapter 4)

Reviewer Comment: "Add participant demographic information. Clarify participant inclusion criteria."

Response: Section 4.2 was updated to explicitly state the university department, the specific majors targeted, and the mean age of the participants.

How it was addressed:

"A total of 5 participants were recruited from the School of Engineering and Digital Sciences (SEDS) at Nazarbayev University from the departments related to the field of study (e.g. Computer Science, Data Science, and Robotics majors students were recruited) using a convenience-based volunteer sampling approach. The mean age of participants is 22 years with standard deviation of 2 years."

Reviewer Comment: "The experimental flow needs to match the visual chart perfectly."

Response: Sections 4.4.1 through 4.5 were completely rewritten to align perfectly with the updated flowchart (Figure 4-1). The timeline was separated into exactly matching phases (A. Preparation Session, B/C/D. Modality Cycles), clarifying that all training occurred in Phase A, and precisely defining the 1-minute and 5-minute rest/transition periods.

How it was addressed:

"At the beginning of the session, participants were welcomed, briefed on the study procedure, and asked to complete the required pre-experiment documents and training. This stage included: Informed Consent Form (5 min), Pre-Test Questionnaire (2.5 min), Experimental Instruction (5 min), Lab Setup (2 min), Training (5 min)."

Reviewer Comment: "Drop Strain and Discomfort completely."

Response: To focus on validated metrics, all bullet points and mentions of the "Strain Questionnaire," "Discomfort Questionnaire," and "Interaction & UX Questionnaire" were removed from the subjective measures list in Sections 4.1.1 and 4.4.4, as well as from the Abstract and Results tables.

5. Results & Statistical Analysis (Chapter 5)

Reviewer Comment: "RQs 1, 2, and 3 should compare test 1, 2, 3, 4 within the same modality, not cross-modality."

Response: Sections 5.1, 5.2, and 5.3 were restructured to focus exclusively on how Movement Time, Index of Difficulty, and Throughput shifted as task difficulty increased within Desktop, VR, and MR individually.

How it was addressed:

"From Test1 to Test4, the target width decreased from 0.6 m to 0.3 m, while the task difficulty increased. As a result, the mean movement time (MT) increased from 948.5 ms in Test1 to 1355.6 ms in Test4."

Reviewer Comment: "Use a Two-Way Repeated Measures ANOVA for objective metrics."

Response: A Two-Way Repeated Measures ANOVA was calculated using the raw data. A new section (Section 5.4: Overall Objective Performance Comparison) and a new table (Table 5.2) were added to break down the Main Effect of Modality, Main Effect of Difficulty, and the Interaction Effect.

How it was addressed:

"Moreover, a significant interaction was observed between Task Difficulty and Modality in regards to MT ($F(6, 24) = 3.50, p = 0.012$) and TP ($F(6, 24) = 2.74, p = 0.036$), which indicated that the increase in task difficulty had a non-uniform impact on target acquisition in the different environments tested. The Desktop modality showed good robustness... On the other hand, immersive environments were less resilient..."

Reviewer Comment: "NASA-TLX and SUS need item-by-item analysis and bar graphs."

Response: Generated detailed itemized tables for SUS (Table 5.3) showing Q1–Q10, and NASA-TLX (Table 5.4) showing Mental, Physical, Temporal Demand, and Performance. Added corresponding bar charts (Figure 5-2 and Figure 5-3) and provided a detailed narrative breakdown for each dimension.

How it was addressed:

"Physical Demand (PD): There was a noticeable increase in physical effort when comparing different interaction modalities... MR yielded the highest physical effort rating by far ($M = 3.80$), which is most probably explained by the so-called 'gorilla arm' phenomenon, since continuous use of mid-air gestures requires constant work of muscles supporting the hand in air."

Reviewer Comment: "Ensure all tables are properly referenced and explained in the text."

Response: Corrected mismatched table references throughout Chapter 5 to ensure accurate cross-referencing between the narrative text and the newly generated itemized tables (e.g., strictly pointing to Table 5.4 for NASA-TLX and Table 5.3 for SUS).

6. Discussion and Conclusion (Chapter 6)

Reviewer Comment: "The sample $N=5$, is it too small or is it fine for this setup? Please verify."

Response: Addressed this directly in Section 6.3 (Limitations). The thesis was explicitly framed as a "pilot experiment," acknowledging that $N=5$ makes obtaining pairwise inference reliability challenging, which satisfies the reviewer's concern regarding statistical power.

How it was addressed:

"Firstly, this research is a pilot experiment, and the sample size is very small, including only 5 participants, which made it challenging to obtain pairwise inference reliability even when the sign of the effect remains constant in each comparison."

7. Formatting, LaTeX, and Appendix

Reviewer Comment: "Check all references carefully for formatting consistency... remove placeholder text 'unspecified'."

Response: Scanned the `.bib` file to successfully remove all instances of the placeholder text "unspecified." Additionally, ensured proper capitalization of acronyms in the bibliography by wrapping them in curly braces (e.g., `{ISO}`).

Reviewer Comment: "Fix missing figure references (e.g., 'Fig ??'). Ensure every figure, table, and equation are cross-referenced."

Response: Replaced all hardcoded or broken figure/table text with dynamic `\ref{fig:label}` and `\ref{tab:label}` tags.

Reviewer Comment: "Check appendix materials and ensure referenced questionnaires are included... Add here information what you are including in the appendix."

Response: Organized Chapter 7 (Appendix) to properly embed the final PDFs. Added an introductory section to the Appendix explicitly listing the supplementary materials to provide full transparency regarding the experimental protocol.

How it was addressed:

"This appendix contains the supplementary materials and documents utilized during the experimental phase of this research... \n - Written Informed Consent Form \n - Pre-Test Questionnaire \n - NASA Task Load Index (NASA-TLX) \n - System Usability Scale (SUS)"

Bibliography

- [1] Iso 9241-9: Ergonomic requirements for office work with visual display terminals (vdts) — requirements for non-keyboard input devices, 2000.
- [2] Iso 9241-411: Ergonomics of human-system interaction — part 411: Evaluation methods for the design of physical input devices, 2012.
- [3] Johnny Accot and Shumin Zhai. Refining fitts’ law models for bivariate pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’03)*, pages 193–200. ACM, 2003.
- [4] Mohammadreza Amini, Wolfgang Stuerzlinger, Robert J. Teather, and Anil Ufuk Batmaz. A systematic review of fitts’ law in 3d extended reality. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI ’25)*. ACM, 2025.
- [5] Mayra Barrera Machuca and Wolfgang Stuerzlinger. The effect of stereo display deficiencies on virtual hand pointing. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI ’19)*. ACM, 2019.
- [6] Anil Ufuk Batmaz, Moaaz Hudhud Mughrabi, Mine Sarac, Mayra Barrera Machuca, and Wolfgang Stuerzlinger. Measuring the effect of stereo deficiencies on peripersonal space pointing. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces (IEEE VR)*. IEEE, 2023.
- [7] Anil Ufuk Batmaz and Wolfgang Stuerzlinger. Effective throughput analysis of different task execution strategies for mid-air fitts’ tasks in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 28(11):3939–3947, 2022.
- [8] Yuan Chen, Keiko Katsuragawa, and Edward H. Lank. Understanding viewport- and world-based pointing with everyday smart devices in immersive augmented reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI ’20)*. ACM, 2020.
- [9] Tafadzwa Joseph Dube, Yuan Ren, Hannah Limerick, I. Scott MacKenzie, and Ahmed Sabbir Arif. Push, tap, dwell, and pinch: Evaluation of four mid-air selection methods augmented with ultrasonic haptic feedback. *Proceedings of the ACM on Human-Computer Interaction*, 6(ISS):Article 565, 2022.
- [10] Paul M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47(6):381–391, 1954.

- [11] Jenny Gabel, Susanne Schmidt, Oscar Ariza, and Frank Steinicke. Redirecting rays: Evaluation of assistive raycasting techniques in virtual reality. In *Proceedings of the 29th ACM Symposium on Virtual Reality Software and Technology (VRST '23)*. ACM, 2023.
- [12] Kristen Grinyer and Robert J. Teather. Low-fi vr controller: Improved mobile virtual reality interaction via camera-based tracking. In *Proceedings of the 2023 ACM Symposium on Spatial User Interaction (SUI '23)*. ACM, 2023.
- [13] Jin Huang, John J. Dudley, Stephen Uzor, Dong Wu, Per Ola Kristensson, and Feng Tian. Understanding user performance of acquiring targets with motion-in-depth in virtual reality. *International Journal of Human-Computer Studies*, 163:102817, 2022.
- [14] Xiaolong Lou, Xuanbai Song, Xiaofan Hu, Mengzhen Ma, Lili Fu, Xinran Zhuang, and Xueer Fan. An extended fitts' law model for hand interaction evaluation in virtual reality: Accounting for spatial variables in 3d space and the arm fatigue influence. *International Journal of Human-Computer Interaction*, 41(18):11611–11637, 2025.
- [15] I. Scott MacKenzie and William Buxton. Extending fitts' law to two-dimensional tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '92)*, pages 219–226. ACM, 1992.
- [16] David E. Meyer, Richard A. Abrams, Sylvan Kornblum, Charles E. Wright, and J. E. Keith Smith. Optimality in human motor performance: Ideal control of rapid aimed movements. *Psychological Review*, 95(3):340–370, 1988.
- [17] Atsuo Murata and Daichi Fukunaga. Extended fitts' model of pointing time in eye-gaze input system: Incorporating effects of target shape and movement direction. *Applied Ergonomics*, 68:54–60, 2018.
- [18] Atsuo Murata and Hirokazu Iwase. Extending fitts' law to a three-dimensional pointing task. *Human Movement Science*, 20(6):791–805, 2001.
- [19] Vijay Rajanna and Tracy Hammond. Can gaze beat touch? a fitts' law evaluation of gaze, touch, and mouse inputs. arXiv preprint arXiv:2208.01248, 2022.
- [20] M. F. Roig-Maimó, R. Mas-Sansó, and I. Scott MacKenzie. Design of 2d fitts' law experiments: An odd thing about targets. *Applied Ergonomics*, 2026.
- [21] Rongkai Shi, Yushi Wei, Yue Li, Lingyun Yu, and Hai-Ning Liang. Expanding targets in virtual reality environments: A fitts' law study. In *2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 615–620. IEEE, 2023.
- [22] Gordon Shoemaker, Takashi Tsukitani, Yoshifumi Kitamura, and Kellogg S. Booth. Two-part models capture the impact of gain on pointing performance. *ACM Transactions on Computer-Human Interaction*, 19(4):1–34, 2012.

- [23] Markus Tatzgern, Jan-Hendrik Sünderkamp, Helen Stefanidi, and Alexander Meschtscherjakov. From lab to sidewalk: Exploring ar interaction techniques outdoors and in motion using fitts' law. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 2025.
- [24] A. T. Welford. *Fundamentals of Skill*. Methuen, London, UK, 1968.
- [25] Qian Zhou, George W. Fitzmaurice, and Fraser Anderson. In-depth mouse: Integrating desktop mouse into virtual reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–15, 2022.