Input Device Adequacy for Multimodal and Bimanual Object Manipulation in Virtual Environments

Martin Fischbach, Maximilian Neff, Immanuel Pelzer,
Jean-Luc Lugrin, Marc Erich Latoschik

Human-Computer Interaction
Universität Würzburg
Tel.: +49 (0)931 / 31 86314
E-Mail: martin.fischbach@uni-wuerzburg.de

Abstract: This article describes a benchmark for evaluating adequacy of input devices for bimanual direct interaction techniques typically found in VR/AR applications. The benchmark implements a puzzle-like scenario inspired from shape and color sorting games, in which the user manipulates the position, rotation and scale of 3D spheres. The continuous interactions are combined with multimodal state-change actions using either a button or speech interface. A follow-up usability study utilizes the benchmark to evaluate the performance of one professional and one consumer-grade tracking system for both state-changing interfaces. The results reveal similar adequacy for both tracking systems under both multimodal conditions.

Keywords: Multimodal, Interaction, Evaluation, VR, AR

1 Introduction

Interacting with 3D objects is a mandatory requirement for all VR/AR applications. Some objects represent functional application aspects: They are either virtual replicas of real world objects of a given application domain (e.g. engine parts in a virtual construction scenario), or they represent parts from the 3D user interface domain (e.g. 3D menus, buttons, or widgets). The typical interaction procedure with these objects can be separated into distinct functional steps:

I_1 Selection of the target object(s).

I_2 Selection of the target action(s).

I_3 Specification of the required parameters for the action(s).

I_4 Exit from the interaction.

A specific benefit of interactive systems is on-line modification. For VR/AR systems, I_3 will be coupled into the interaction loop. Here, actions are performed dynamically and the outcome of these actions are constantly reflected as a change of the system state. Such continuous 3D interactions [Lat05] enable users to gradually control and adjust parameters for a wide variety of potential application scenarios based on various 3D interaction techniques, e.g., the virtual hand (cf. [PIWB98]), Go-Go [PBWI96], World-in-Miniature [SCP95], or HOMER [BH97]. For an elaborate overview of 3D interaction techniques see [BKLP05].

Interaction techniques like these rely on two input types: A tracking system to register 3D movements and an event source to register state-changing actions. The quality of the employed tracking system and the event source is considerably influencing the overall usability of such interaction techniques. Following the idea of [LWSL13], we therefore propose to evaluate the adequacy of input devices for a specific interaction technique by benchmarking usability. ISO's definition of usability [ISO98], incorporating effectiveness, efficiency and satisfaction, is adopted as a measure for adequacy. In contrast to the common definition of a *benchmark*, as a fully automated test scenario, we consequently refer to the concept of a *usability benchmark*. However, in order to guarantee correctness and comparability, a formalized and standardized procedure is required.

In this article, we describe one benchmark and present the results of an associated evaluation as one step towards this long term goal. The presented benchmark is designed for the evaluation of input device adequacy for bimanual direct [HHN85] object manipulation. Its application is illustrated by the comparison of two tracking systems and of two event sources, i.e. a handheld joystick and speech commands.

The rest of this article ist structured as follows: Section 2 describes previous work before the presentation of the utilized benchmark in section 3. Section 4 then provides a detailed description of the conducted evaluation. The results are depicted and discussed in section 5, followed by a conclusion and a perspective on feasible future work in section 6.

2 Related Work

Research has been done to compare different object selection and manipulation techniques [BH97, PIWB98, MGG$^+$06] or to evaluate specific styles, like bimanual interaction [BM86, ZFS97]. Furthermore general testbeds provide guidance for the selection of an adequate interaction technique for a specific task [PWBI97, BJH99]. However these approaches only target the interaction technique, not the utilized input devices.

The negative impacts of latency and jitter of tracking systems on user performances, satisfaction and discomfort are widely recognized [SSMM98, MAEH04, Ste08, LBBH10]. Reliable tracking of user movements in real-time appear thus as a core requirement of 3D interaction techniques. Meanwhile, there exist nowadays a myriad of tracking solutions offering various latency and accuracy performances, from high-priced professional-grade systems to low-cost consumer-grade hardware like the Microsoft Kinect. However, raw performances are often insufficient to value the influence of a specific system on the user experience [LWSL13, Ste08, PMK11]. In some cases these properties are not even available. The choice of a tracking system for a particular set of 3D interaction techniques is thus problematic, where guidance would be greatly beneficial.

The high-level rating of the adequacy of input devices has been conducted by utilizing task- or game-based benchmarks to obtain objective measurements, like task completion time or achieved precisions [BHB99, RPB05, MST11]. Other approaches take this idea one step further by measuring the user's experience. Conducted studies range from general VE investigations targeting tracking device issues [SLH00, MAEH04] to evaluations on specific interaction tasks, such as the comparison of different modalities (e.g. [KKHF07]) or the determination of adequacy of novel input devices (e.g. [HSK$^+$05]). Overall, there exist numerous usability evaluation methods, as summarized by [BGH02]. Yet, there exist no standardized comparison or rating procedures for usability-based measurements of input device adequacy for 3D interaction techniques. A recent continuative approach proposes the establishment of a set of canonical usability benchmarks for a task-specific comparison of tracking systems [LWSL13]. Two simple game-based benchmarks, focussing on tracking system quality for head tracking and one-handed unimodal interaction, are presented. In line with the stated need for additional benchmarks dedicated to other typical interaction techniques, we present one benchmark for a multimodal and bimanual direct object manipulation.

3 Design of the Benchmark

The proposed benchmark is a puzzle-like *skill game* (see figure 1). Users have to create, select, translate, rotate, scale, and finally place two-colored spheres (one color per half-sphere) into matching holes punched inside a board with a similar color-coding (turquoise, yellow, purple, orange). A correct match requires three conditions for each sphere: (1) Rotation to match the color coding, (2) scaling to fit the holes diameter, and (3) translation into the center of the matching hole (see figure 1). An ϵ-neighborhood is used during the condition checks in order to deal with expected imprecisions (e.g., human hand shivering, tracker's latency, jitter, position noise, and float point approximation): Rotation $\pm 15°$ (angle between spheres orientation and exact goal orientation), scale ± 0.02 and translation $\pm 5\,\text{cm}$ (point-to-point distance). The ϵ-neighborhoods are the same for all experiments. A text notification informs the user about properties satisfying the ϵ-neighborhood criteria. The creation and correct placement of three spheres wins the game. The sequence of color codings and the board are the same for each session. Altogether the benchmark depends on a 3DOF (degrees of freedom) tracking system to track the position of the user's two hands, an event source triggered by the user, and one 6DOF tracking device to support stereo as well as motion parallax necessary for a correct depth perception during the 3D interaction.

3.1 Interaction Use Cases

Two event sources provide necessary state changes. New spheres are spawned either by pressing a button or by using a voice command ("create ball"). Object selection is indicated by touching objects using two "virtual hands" (green spheres) as proxies of the user's hands.

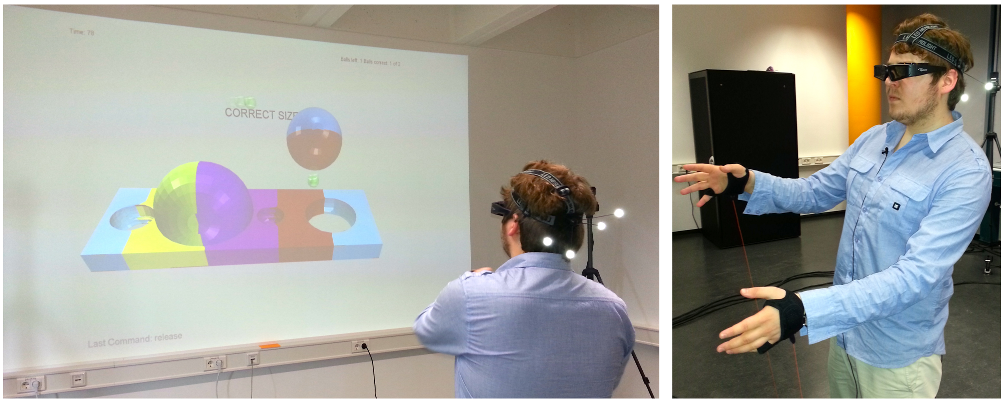

Figure 1: The puzzle-like skill game: A typical interaction (left), and a user wearing *Gametrak* gloves, an ioTracker head target, a clip-on microphone and shutter glasses (right).

A second button or the voice-command (“take”) confirms the selection and starts the direct manipulation. Incoming updates of the users hand position(s) are then continuously mapped to the objects transformation, depending on the number of virtual hands touching it at confirmation time: A one-handed touch initiates displacement relative to the respective hand’s position. A two-handed touch initiates displacement relative to the virtual center between the hands, scaling relative to the distance between the hands, and rotation relative to the vector between the hands. The continuous interaction modes are exited using either a third button or the voice-command (“release”). Note that this interaction description maps the functional steps of typical 3D interactions I_{1-4}. Because the purpose of the interaction (direct manipulation) is predefined for this benchmark, I_2 permits no choice but serves only as a confirmation for an indicated selection (see figure 2).

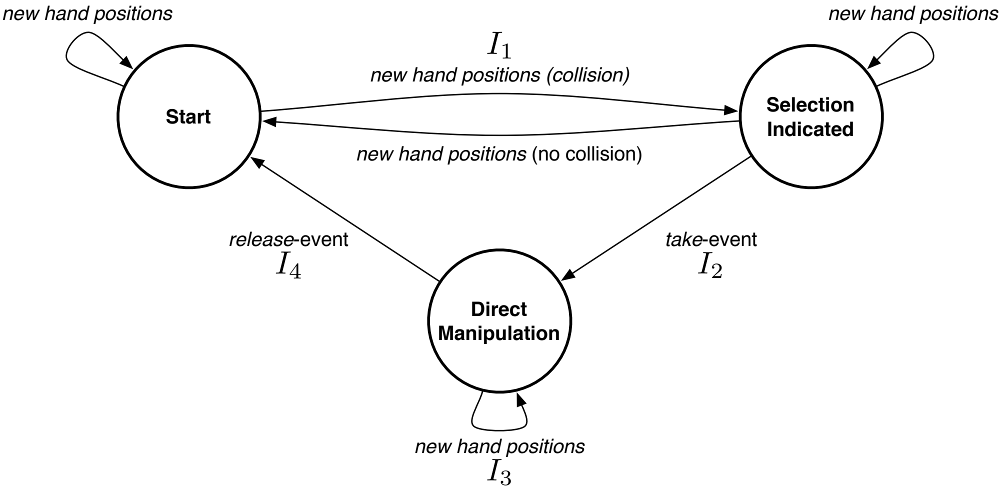

Figure 2: Specification of the two-handed direct manipulation interaction.

Table 1: Mapping to ISO's usability definition.

		Measurements
Usability	Effectivity	Benchmarks winning conditions (implicit), TLX (performance)
	Satisfaction	TLX (frustration), QUESI
	Efficiency	TLX (other), task completion time

3.2 Measurements

The usability is measured by (1) NASA's Task Load indeX (TLX) [HS88] questionnaire and (2) the questionnaire for the subjective consequences of intuitive use (QUESI) [NH10] (representing subjective measurements). Additionally, (3) the time users need to complete the game is derived as an objective measurement. The (4) Simulation Sickness Questionnaire (SSQ) [KLBL93] is conducted to cross check any perceptual drawbacks of the experimental setup. Table 1 maps the conducted questionnaires and the derived measurements to ISO's usability criteria:
Efficiency is measured by the task completion time and the TLX questionnaire. The TLX is designed to access subjective task workload. It measures the perceived mental, physical, and temporal load while carrying out a specific task as well as the effort to accomplish it.
Satisfaction is measured by QUESI and the frustration level measurement of the TLX.
Effectivity, as the capability of achieving a desired result, is implicitly considered by the winning conditions of the games. Effectivity, as a ratio between desired and achieved goal, is only subjectively considered by TLX's performance measurement.

4 Evaluation

The evaluation applies the proposed benchmark and derives four measurements for two tracking systems: ioTracker's *IOTRACKER/8* and In2Games's *Gametrak* [IN204]. Both tracking systems were evaluated for two distinct multimodal conditions dedicated to the execution of the discrete state changes:

C_1 Button presses via Nintendo's *WiiMote*.

C_2 Speech input via Sennheiser's *EW 100 G3* clip-on microphone in combination with the *Sphinx-4* speech recognition tool [WLK$^+$04].

The specific tracking systems were selected for the following reasons. Both use different physical principles (optical vs mechanical). IOTRACKER/8 is a professional-grade tracking system already assumed adequate for VE interactions. It is supposed to be sufficiently precise and fast, and supposed to provide a reasonable low latency (see Table 2). The adequacy of the consumer-grade Gametrak system, however, is unclear. The relevant properties are not

Table 2: Comparison of some technical properties: IOTRACKER/8 vs Gametrak.

	IOTRACKER/8	Gametrak
Type	optical	mechanical
Refresh Rate	∼60 Hz	n/a
Latency	18–40 ms	n/a
Number of cameras	8	-

available for the Gametrak system, but it may provide an affordable input device for VR applications, specifically for public installations without constant monitoring.

The main objective of our usability benchmark is to provide a comparison tool going beyond low-level technical specification comparisons. Since specifications may not or may only partly be available. They might be measured arbitrarily (using different measurement methods or platforms) and their impact on the user experience is not explicit and straight forward.

The two multimodal conditions were selected for the following reasons. First, button-press interfaces are a central (additional) component of many VR/AR applications for decades now. The input devices are mature and reliable, the associated interaction metaphors are well-known. Hence, C_1 provides a baseline condition reflecting the current state-of-the-art. Second, speech (or speech/gesture) interfaces provide an alternative interaction metaphor (cf. [Lat05]) potentially usable in several contexts, e.g, in hands-free scenarios, in complex command scenarios historically provided by terminal access, or in scenarios establishing human-like interactions with artificial humanoid agents.

In any case, such multimodal scenarios rely on the robustness of the speech recognition as a central component. Hence, in the context of the two tracking systems compared, C_2 evaluates the open source Sphinx system as an alternative to a button interface, also matching a low-cost approach as motivated by the Gametrak. In line with these selection criteria and for bimanual direct manipulation tasks it is hypothesized that

H_1 the Gametrak system is adequate for the proposed continuous interaction and

H_2 voice commands via Sphinx are an adequate alternative to a simple button interface.

4.1 Procedure

The conducted user study included 31 participants (15 male, 16 female). For each participant the experiment was divided into three phases: introduction, four skill game sessions and completion of the post-SSQ. In the course of the first phase the participants were introduced, general information (like gender, age, familiarity with VR systems) was collected, and the participants answered the pre-SSQ. Thereafter the four benchmark conditions were tested in random order to avoid *order effects*, such as the practice effect. Each session included the hardware rigging (see figure 1), the execution of the skill game, and the completion of the TLX and QUESI. The total duration for one participant was ∼1 h.

4.2 Setup

The benchmark was implemented using the VR/AR software framework *Simulator X* [LT11]. Execution of the benchmarks was performed by a host system consisting of an NVIDIA Quadro K4000 graphics card, an Intel Xeon e3 1234V2 processor (3.4 GHz), and 8 GB DDR3 memory. Voice input was recorded by a Sennheiser EW 100 G3 microphone and processed by the open-source speech recognition software SphinX-4. Button presses where captured by a Nintendo WiiMote. The front projection was done in stereoscopic 3D with a BenQ LW 61 in combination with Optoma BG-ZF2100GLS active shutter glasses. The resulting working space was located 2.5 m in front of the projection screen, respectively 0.5 m behind the table where the projector was placed. Tracking data was communicated via the Virtual Reality Peripheral Network (VRPN) for both tracking systems. Head-tracking required for motion parallax was provided by IOTRACKER/8 in all four conditions.

5 Results

Figure 3(a) displays the mean times it took to complete the skill game. The two-way repeated-measures ANOVA identifies no significant differences, neither in the main effect of the tracking system ($Wilks' \ Lambda = .987$, $F_{1,30} = .388$, and $p > .05$) nor in the main effect of the event source ($Wilks' \ Lambda = .970$, $F_{1,30} = .924$, and $p > .05$) nor in the interaction effect ($Wilks' \ Lambda = .950$, $F_{1,30} = 1.570$, and $p > .05$). This supports H_{1+2} in terms of efficiency.

Figure 3(b) displays the means for the QUESI. The two-way repeated-measures ANOVA identifies no significant differences, neither in the main effect of the tracking system ($Wilks' \ Lambda = .974$, $F_{1,30} = .795$, and $p > .05$) nor in the main effect of the event source ($Wilks' \ Lambda = .993$, $F_{1,30} = .205$, and $p > .05$) nor in the interaction effect ($Wilks' \ Lambda = .982$, $F_{1,30} = .562$, and $p > .05$). This supports H_{1+2} in terms of satisfaction.

Figure 3(c) displays the means for the TLX. The two-way repeated-measures ANOVA identifies no significant differences, neither in the main effect of the tracking system ($Wilks' \ Lambda = .989$, $F_{1,30} = .343$, and $p > .05$) nor in the main effect of the event source ($Wilks' \ Lambda = .937$, $F_{1,30} = 2.012$, and $p > .05$) nor in the interaction effect ($Wilks' \ Lambda = .994$, $F_{1,30} = .191$, and $p > .05$). This supports H_{1+2} in terms of satisfaction, efficiency and effectiveness.

All findings strongly support H_1 and H_2: The comparison of the Gametrak system and the IOTRACKER/8 system yields no significant difference in the subjective task workload, the satisfaction during use, and the task completion time given the tested interactions (H_1). Same holds true for the comparison of the utilized event sources (H_2).

The supplementary conducted SSQ does not reveal a negative impact of the 3D environment on all the participants. This crosschecks the obtained results with respect to potential unwanted simulation issues and thus improves the overall validity.

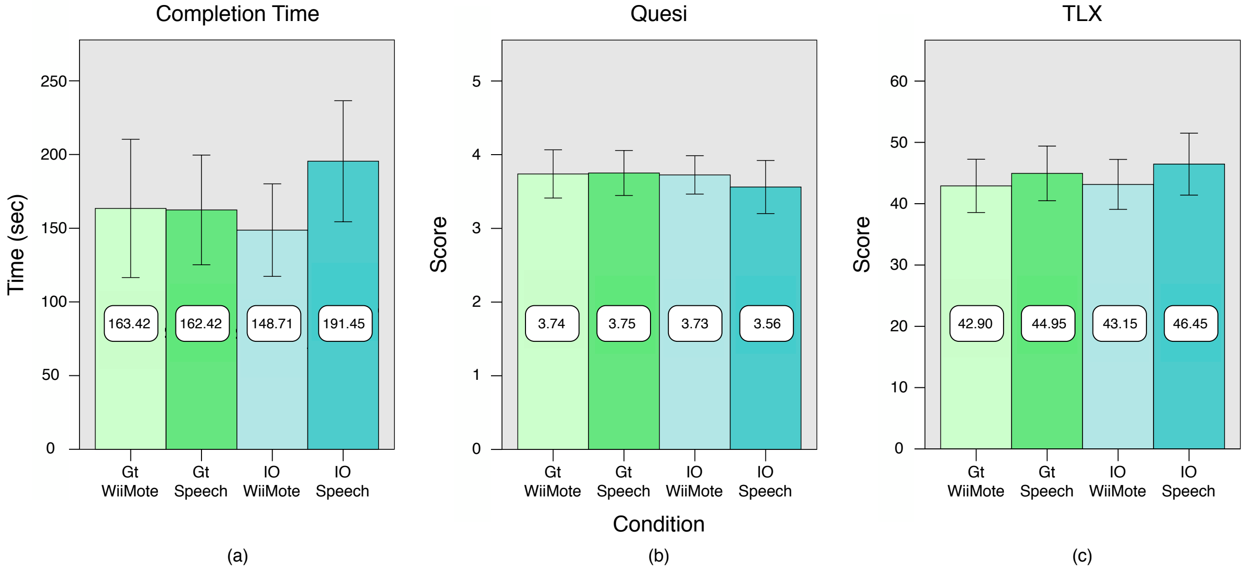

Figure 3: Mean (a) task completion times, (b) QUESI scores, and (c) TLX scores for each of the four conditions (error bars indicate the respective confidence intervals).

5.1 Discussion

The results confirmed the adequacy of the Gametrak system and the equality of the event sources for the given interactions. However, this holds true only for the exact benchmark conditions. In order to get a more detailed comparison, capable of reflecting the assumable superiority of the professional IOTRACKER/8 system, supplementary controlled variables would be feasible (e.g. decreasing ϵ-neighborhoods to provoke errors).

The utilized speech interface is in fact simple, but already sufficient for various fundamental and universal interaction techniques. In accordance with the overall approach, the speech recognition component is used as a black box. Relevant technical properties, like the recognition rate, are not always available and normally require an auxiliary evaluation.

In addition to the benchmark's results, we could make some interesting observations. Speech commands seemed uncommon for many participants as opposed to the handheld button device, frequently resulting in initial hesitant and low voice utterances. We consider this to be caused by the familiarity of the well-known button interfaces and the unfamiliarity of speech input, the negative effect diminished during the ongoing interaction. Furthermore, the haptic feedback provided by a button-press was perceived beneficial compared to the visual feedback of a correctly recognized speech-command. This has to be investigated to find an appropriate feedback mechanism.

The Gametrak system is adequate for selection and manipulation tasks, however it is limited in tracking space ($\sim 3\,\mathrm{m}^3$, no twisting of ropes). In contrast to the optical IO-

TRACKER/8 though, no fragile targets have to be worn and occlusion issues are not existent. In addition, a mechanical principle can provide haptic feedback.

Both tracking systems caused limitations for two-handed manipulations. IOTRACKER/8 had major identification problems for the two hand targets when users moved their hands near to each other, bringing the infrared-reflecting spheres in close proximity (i.e. when trying to rotate by 180° in one movement, resulting in hands crossed with the backs of the hands facing each other). The Gametrak revealed similar limitations caused by the ropes attached to the gloves the user wore. The ropes connected the gloves with the Gametrak's chassis in front of the user, restricting similar movements physically. As a result, most participants divided big rotations into a sequence of several smaller steps. In fact, participants starting with the Gametrak condition seemed to have a steeper learning curve in that regard.

Alternative questionnaires are feasible. For example the pragmatic property measured by AttrakDiff 2 [HBK03] would cover some of ISO's criteria. At the same time the hedonistic property measured by AttrakDiff 2 would actually target the design aspects of a system as a whole and, hence, might blur results targeting the usability of the input devices.

6 Conclusion and Future Work

This article proposed a usability-oriented benchmark to evaluate input device adequacy for multimodal two-handed direct object manipulation tasks. In contrast to selection strategies based purely on technical specifications, this approach is immune to comparison problems caused by unclear measurement conditions or missing values of devices' raw performances (such as latency, precision, or accuracy).

The benchmark has been applied to compare a consumer-grade tracking system (Gametrak) to a professional-grade tracking system (IOTRACKER/8) and to compare two multimodal event interfaces (speech and button press).

It could be shown that the Gametrak system is adequate for the proposed continuous interaction and that the compared modalities are equivalent for this kind of task. Both findings are beneficial: The Gametrak system is an affordable, easy to setup input device for VE applications. Speech-commands keep the users' hands free (for other devices), they allow an extended system control in scenarios where additional devices are undesirable, and they are a basis for human-like speech/gesture interfaces.

Further benchmarks will evaluate additional multimodal interaction techniques, e.g., incorporating gesture recognition and complex speech combinations. In the long term, this will hopefully lead to a set of formalized and standardized methods, that also provide rating matrices and reference values, allowing easy task specific categorization and comparison of input devices and techniques. Measuring the user experience in multimodal scenarios may provide insight into a certain input technique and proof more appropriate than looking at system specifications and performances. Following [LWSL13], we also believe that VR community will benefit from usability-based benchmarks, in order to compare current and future input devices adequacy for 3D interaction techniques.

References

[BGH02] Doug A. Bowman, Joseph L. Gabbard, and Deborah Hix. A survey of usability evaluation in virtual environments: classification and comparison of methods. *Presence: Teleoperators and Virtual Environments*, 11(4):404–424, 2002.

[BH97] Doug A. Bowman and Larry F. Hodges. An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In *Proceedings of the Symposium on Interactive 3D Graphics*, pages 35–38. ACM, 1997.

[BHB99] Woodrow Barfield, Claudia Hendrix, and Karl-Erik Bystrom. Effects of stereopsis and head tracking on performance using desktop virtual environment displays. *Presence: Teleoperators and Virtual Environments*, 8(2):237–240, 1999.

[BJH99] Doug A. Bowman, Donald B. Johnson, and Larry F. Hodges. Testbed evaluation of virtual environment interaction techniques. In *Proceedings of the Symposium on Virtual Reality Software and Technology*, pages 26–33. ACM, 1999.

[BKLP05] Doug A. Bowman, Ernst Kruijff, Joseph J. Jr. LaViola, and Ivan Poupyrev. *3D User Interfaces – Theory and Practice.* Addison-Wesley, 2005.

[BM86] William Buxton and Brad A. Myers. A study in two-handed input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 321–326. ACM, 1986.

[HBK03] Marc Hassenzahl, Michael Burmester, and Franz Koller. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In *Mensch & Computer 2003*, pages 187–196. Springer, 2003.

[HHN85] Edwin L. Hutchins, James D. Hollan, and Donald A. Norman. Direct manipulation interfaces. *Human–Computer Interaction*, 1(4):311–338, 1985.

[HS88] Sandra G. Hart and Lowell E. Staveland. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human Mental Workload*, 1(3):139–183, 1988.

[HSK^{+}05] Anke Huckauf, Alexander Speed, André Kunert, Jan Hochstrate, and Bernd Fröhlich. Evaluation of 12-DOF input devices for navigation and manipulation in virtual environments. In *Human-Computer Interaction*, INTERACT, pages 601–614. Springer, 2005.

[IN204] IN2Games. Gametrak. `http://en.wikipedia.org/wiki/Gametrak`, 2004.

[ISO98] ISO. 9241-11: Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs). *Part 11: Guidance on Usability*, 1998.

[KKHF07] André Kunert, Alexander Kulik, Anke Huckauf, and Bernd Fröhlich. A comparison of tracking- and controller-based input for complex bimanual interaction in virtual environments. In *Proceedings of the Eurographics Conference on Virtual Environments*, pages 43–52. Eurographics Association, 2007.

[KLBL93] Robert S. Kennedy, Norman E. Lane, Kevin S. Berbaum, and Michael G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, 3(3):203–220, 1993.

[Lat05] Marc E. Latoschik. A user interface framework for multimodal VR interactions. In *Proceedings of the International Conference on Multimodal Interfaces*, pages 76–83. ACM, 2005.

[LBBH10] Cha Lee, Scott Bonebrake, Doug A. Bowman, and Tobias Höllerer. The role of latency in the validity of AR simulation. In *Virtual Reality Conference*, pages 11–18. IEEE, 2010.

[LT11] Marc .E. Latoschik and Henrik Tramberend. Simulator X: A scalable and concurrent architecture for intelligent realtime interactive systems. In *Virtual Reality Conference*, pages 171–174. IEEE, 2011.

[LWSL13] Jean-Luc Lugrin, Dennis Wiebusch, Alexander Strehler, and Marc E. Latoschik. Usability Benchmarks for Motion Tracking Systems. *To appear in Proceedings of Symposium on Virtual Reality Software and Technology*, ACM, 2013.

[MAEH04] Katerina Mania, Bernard D. Adelstein, Stephen R. Ellis, and Michael I. Hill. Perceptual sensitivity to head tracking latency in virtual environments with varying degrees of scene complexity. In *Proceedings of the Symposium on Applied Perception in Graphics and Visualization*, pages 39–47. ACM, 2004.

[MGG$^+$06] Ryan P. McMahan, Doug Gorton, Joe Gresock, Will McConnell, and Doug A. Bowman. Separating the effects of level of immersion and 3D interaction techniques. In *Proceedings of the Symposium on Virtual Reality Software and Technology*, pages 108–111. ACM, 2006.

[MST11] Elke Mattheiss, Johann Schrammel, and Manfred Tscheligi. Navigating the space: evaluating a 3D-input device in placement and docking tasks. In *Proceedings of the Conference on Human-Computer Interaction*, pages 309–314. British Computer Society, 2011.

[NH10] Anja Naumann and Jörn Hurtienne. Benchmarks for intuitive interaction with mobile devices. In *Proceedings of the International Conference on Human Computer Interaction with Mobile Devices and Services*, pages 401–402. ACM, 2010.

[PBWI96] Ivan Poupyrev, Mark Billinghurst, Suzanne Weghorst, and Tadao Ichikawa. The Go-Go interaction technique: Non-linear mapping for direct manipulation in VR. In *Proceedings of the Symposium on User Interface Software and Technology*, pages 79–80. ACM, 1996.

[PIWB98] Ivan Poupyrev, T. Ichikawa, S. Weghorst, and M. Billinghurst. Egocentric object manipulation in virtual environments: Empirical evaluation of interaction techniques. *Computer Graphics Forum*, 17(3):41–52, 1998.

[PMK11] Giorgos Papadakis, Katerina Mania, and Eftichios Koutroulis. A system to measure, control and minimize end-to-end head tracking latency in immersive simulations. In *Proceedings of the International Conference on Virtual Reality Continuum and Its Applications in Industry*, pages 581–584. ACM, 2011.

[PWBI97] Ivan Poupyrev, Suzanne Weghorst, Mark Billinghurst, and Tadao Ichikawa. A framework and testbed for studying manipulation techniques for immersive VR. In *Proceedings of the Symposium on Virtual Reality Software and Technology*, pages 21–28. ACM, 1997.

[RPB05] M. Roerdink, C.E. Peper, and P.J. Beek. Effects of correct and transformed visual feedback on rhythmic visuo-motor tracking: Tracking performance and visual search behavior. *Human Movement Science*, 24(3):379–402, 2005.

[SCP95] Richard Stoakley, Matthew J. Conway, and Randy Pausch. Virtual reality on a WIM: Interactive worlds in miniature. In *Proceedings of the Conference on Human Factors in Computing Systems*, pages 265–272. ACM, 1995.

[SLH00] Richard H.Y. So, W.T. Lo, and Andy T.K. Ho. Scene movement: An importance cause of cybersickness. *Virtual Reality: Select Tools and Applications*, 2000.

[SSMM98] Mel Slater, Anthony Steed, John McCarthy, and Francesco Maringelli. The influence of body movement on subjective presence in virtual environments. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 40(3):469–477, 1998.

[Ste08] Anthony Steed. A simple method for estimating the latency of interactive, real-time graphics simulations. In *Proceedings of the Symposium on Virtual Reality Software and Technology*, pages 123–129. ACM, 2008.

[WLK$^+$04] Willie Walker, Paul Lamere, Philip Kwok, Bhiksha Raj, Rita Singh, Evandro Gouvea, Peter Wolf, and Joe Woelfel. Sphinx-4: A flexible open source framework for speech recognition. Technical report, 2004.

[ZFS97] Robert C. Zeleznik, Andrew S. Forsberg, and Paul S. Strauss. Two pointer input for 3D interaction. In *Proceedings of the Symposium on Interactive 3D Graphics*, pages 115–120. ACM, 1997.