Comparing Mixed Reality Hand Gestures to Artificial Instruction Means for Small Target Objects

Lukas Walker, Joy Gisler, Kordian Caplazi, Valentin Holzwarth, Christian Hirt, and Andreas Kunz

Abstract Hand gestures are a valuable means for the instruction of complex handling processes. They are used and perceived in an intuitive way and outperform artificial representations such as arrows or symbols. On the other hand, referring finger gestures require a certain object size to avoid ambiguities, and often they are replaced by artificial means. However, this comes to the cost of reduced intuition due to the change of a hand gesture to an artificial gesture, which consequently makes it more difficult to learn long instruction sequences and keep them in mind. This paper thus introduces study results showing that hand pointing gestures perform well even for small objects, so that unnecessary switches to artificial representations can be avoided in the future.

Keywords Augmented reality · Gestures · Visualization

1 Introduction

When explaining complex manual operation tasks, gestures play an important role together with speech. In this context, hand gestures are considered particularly important [3]. Hand gestures for handling objects are unique since they constrain and predefine the gesture [2]. Aigner et al. [2] classify occurring gestures into pointing, semaphoric, pantomimic, iconic, and manipulative gestures, from which only the pointing gesture is not constrained by geometry.

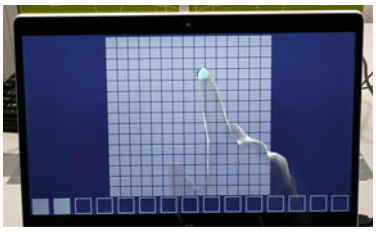
L. Walker · J. Gisler · C. Hirt · A. Kunz (✉)
ETH Zurich, 8092 Zurich, Switzerland
e-mail: kunz@iwf.mavt.ethz.ch

J. Gisler
e-mail: lukas.walker@inf.ethz.ch
URL: http://www.ethz.ch

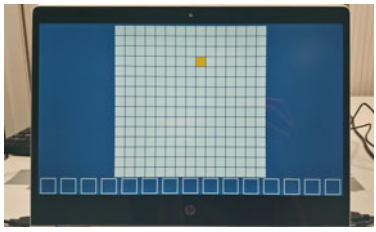
K. Caplazi
Rimon Technologies GmbH, 8092 Zurich, Switzerland

V. Holzwarth
RhySearch, 9471 Buchs, Switzerland

X.-S. Yang et al. (eds.), *Proceedings of Eighth International Congress on Information and Communication Technology*, Lecture Notes in Networks and Systems 693,
https://doi.org/10.1007/978-981-99-3243-6_62



(a) Fingerpointing overlay displayed by Microsoft HoloLens II



(b) Highlighting the element displayed by laptop screen

Fig. 1 Overview on the two study conditions

If hand gestures are used for an instruction on how real objects should be handled, Mixed Reality (MR) can be used. MR technologies, especially head mounted displays (HMD), receive a lot of attention since a few years. Their success lies in enabling users to interact with digital data seamlessly, which is a great support in both training and in productive settings. The field of MR is subject to many research projects, both in academia and industry, being driven by the question how we can best profit from the new ways to interact with the digital world. Typically, MR superimposes virtual instructions on real objects, as shown in an early work by Thomas and David [25]. This study was continued three decades later by Hoover et al. [9]. The study had different groups performing the task using a tablet with a 2D guide, a tablet with an MR guide, a desktop computer and a HoloLens 1, respectively. The study showed that participants who used the Hololens guide made fewer errors and had faster assembly times by 15%. The study concluded that HMDs like the HoloLens can be a better alternative for state-of-the-art approaches. Also the work from Scurati et al. [23] uses abstract virtual instructions such as arrows or symbols to inform the worker about the next step. This is also confirmed by a systematic review by Palmarini et al. [18], who only listed MR systems that use abstract virtual augmentation. Consequently, it was stated by LaViola et al. [13] that for presenting positions in the real environment, pointing arrows are superior to other virtual overlays. This might be due to the fact that hand tracking was computationally expensive in the past and just recently became available in devices such as the Microsoft HoloLens (I+II) for MR and Oculus Quest (I+II) for Virtual Reality (VR).

In fact, using hand gestures as an instruction means initially required a larger technical effort, as shown in *SEMarbeta* [6], or in *Augmented 3D hands* [10]. Just recently, with the advance of HoloLens, hand gestures could be easily used as an instructive overlay, as shown in [17], who used pointing gestures to emphasize certain objects during the support by a remote expert. Using hand gestures as an overlay for augmented work instructions is a promising approach since humans have an easier access to images, such as gestures rather than to abstract information, such as icons or pictograms. The latter need to be interpreted by a user in order for him or her to understand and perform a task. A hand animation on the other hand shows the task to be performed in real-time and in three dimensions, allowing the user to imitate

the movements and view them from different sides. In other words, hand animations show to the user immediately how to complete a task instead of explaining it to him or her, which increases the level of immersion and eventually supports memorization.

Showing hand gestures as an overlay on a manual assembly task will lead to the user mimicking these gestures and thus learning the correct behavior in an intuitive way. However, little is known about the efficiency of such hand gesture overlays in contrast to iconographic instructions particularly when it comes to smaller object sizes. It seems that the accuracy and clearness of hand gestures is limited, since [24] state that hand gestures could also be replaced by virtual pointing rays. Based on literature, there is no clear evidence on the accuracy of hand gesture overlays when it comes to pointing to smaller parts in the real environment. Further, existing MR applications using MR hand gesture overlays focus mainly on a remote support and do not give information on memorization effects which could become relevant in training scenarios. This could help in particular for memorizing longer sequences of numbers, since not the numbers are kept in mind, but the complete fingerpointing gesture which is then related to the underlying matrix for recalling the numbers.

We hypothesize that there is a certain size limit for small neighboring target objects, where a user cannot clearly distinguish anymore, where the fingerpointing gesture is directing them at. Further, little is known on the memorization of sequences that are indicated by pointing gestures in comparison to a regular highlighting using virtual objects. Since Liu et al. [16] describe the positive effect of pointing gestures on spatial memorization, and Aldugom et al. [4] describe gestures' positive effect in learning mathematics, we secondly hypothesize that pointing will positively effect the memorization of work sequences.

This paper focuses on one of the most important gestures—the pointing gesture. We compare this pointing gesture to a common highlighting of relevant positions. The overall goal is to completely use gesture-based instructions, without switching back and forth between gesture overlays and iconographic overlays. This would also reduce development efforts of MR and VR applications that direct a user to certain targets, since hand gestures can be recorded by the device, whereas iconographic overlays have to be manually implemented.

After an overview on pointing possibilities in MR, the paper introduces the user study, which evaluates the limits in accuracy for a pointing gesture, as well as the impact of natural pointing gestures on memorization. This is followed by an evaluation of the achieved data regarding accuracy and memorization. The remainder of the paper will give a brief summary and an outlook on future work.

2 Related Work

Pointing gestures in MR or VR are mainly done using a supporting ray that can be controlled by the hand and the index finger, e.g., for selecting objects as shown by Yusof et al. [27], or for controlling a highlighter as described by Lin et al. [15]. The latter inspired us for our user study, in which also certain fields of a matrix should be

Table 1 Previous works' target sizes

Literature	Target size
Tsang et al. [26]	20
Park et al. [19]	10
Gao and Sun [7]	15.9 × 9
Komine and Nakanishi [12]	7
Leitão and Silva [14]	14
Schedlbauer [22]	15

selected. Highlighters or controlled rays only allow for a remote interaction and thus contain a certain amount of unnaturalness when selecting objects. We therefore use the "direct touch" as being used e.g., by Kervegaut et al. [11] as an inherent HoloLens functionality. For providing feedback when touching virtual objects in MR, handheld devices such as smartphones or tablets are used, as described by Prilla et al. [20], who compared this also to hands-free interaction.

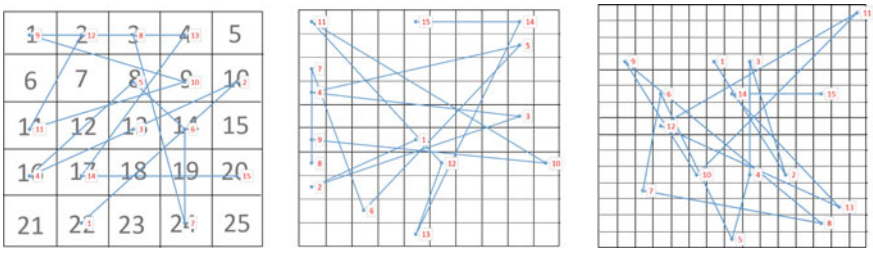
3 Study Design and System Setup

The study described in this paragraph has the following purpose:

- Determine the minimum target size that can be unequivocally detected using a pointing instruction as opposed to a highlighting of objects.
- Effect of a pointing gesture's naturalness on memorizing a sequence of numbers.

To answer these questions, a study was designed where a user was instructed to touch buttons of a matrix on a tablet screen. The target sizes were based on existing manifold works listed in Table 1.

The study consists of three matrices with 5 × 5, 10 × 10, and 15 × 15 elements. Given the resolution of the laptop screen, this results in button sizes of 29 × 29 mm, 14 × 14 mm, and 9 × 9 mm. For all matrices, there was a 0.5 mm spacing between the buttons. This spacing can be seen as irrelevant for the results [22]. For each matrix, the user was instructed which button to press either by a fingerpointing overlay or by highlighting the elements (Fig. 1). For each matrix size, the user had to press 15 buttons. In both cases—the HoloLens instruction and the highlighting—the user's input was detected by the touch sensitive overlay of the laptop computer. The system waits until the user performed the instructed pointing gestures and measures the time for the gesture. As soon as the user touched a button on the screen, the next instruction follows. The instructions where such that the user had to traverse the matrices in seemingly random order (Fig. 2). To avoid any biasing effects due to memorization, every trial of the user study used a different sequence of numbers,



(a) 5×5 matrix (b) 10×10 matrix (c) 15×15 matrix

Fig. 2 Pointing trajectories on the three different matrices

Fig. 3 User study in front of the laptop, showing a 15 × 15 grid for fingerpointing instruction. The lower bar shows the amount of touch inputs already entered

and all of them were randomly chosen so that the generated trajectory cannot be kept in mind.

In a next part of the study, the user was instructed to memorize a sequence of eight buttons that were either highlighted or shown by a fingerpointing overlay in a 4 × 4 matrix (button size 29 × 29 mm). The highlighting and fingerpointing were automized, showing a new button every two seconds. After having seen the sequence, the user had to reenter the memorized sequence into the system.

The complete study was designed as a within-subject study, i.e., after signing a consent form, filling out initial questionnaires, and becoming acquainted with the setup, each subject had to do both, the highlighting and the fingerpointing instruction. In order to balance the study, all participants were divided into two groups, which differ in the order of the two initial experiments (fingerpointing and highlighting). The whole study took about 25 min. In order to avoid any technical biasing, users had to wear the HoloLens in both trials of the study, although it was not required for the highlighting task. For the fingerpointing overlay, MS HoloLens II was used, while the highlighting was displayed on a laptop screen (HP ProBook x360 435 G8). The touch sensitive screen of the laptop was used to detect the user's input (Fig. 3).

During the study, three questionnaires were filled out. The first questionnaire collected demographic data as well as a computer confidence level. After each element

of the study, the NASA TLX [8] (scale: 1–10) and the SUS [5] (scale = 1–100) questionnaires were filled out by the participants. In addition, the cognitive absorption (CA) [1] was completed (scale: 1–10). Finally, after completing both parts of the user study, a user preference was questioned. 11 participants recruited from the local university staff with an average age of 29 years (SD = 6.05) took part in our user study (2 female, 9 male), from which all had normal or corrected to normal vision.

4 Study Results

4.1 Comparison to Fitts' Law

To make the setup comparable with other works on touch interaction, the three matrices are characterized by measures from the Fitts' law: Index of difficulty (ID), performance index (IP), and motion time (MT). D is the mean length of the paths between the individual touch points in Fig. 2, and w is the size of the buttons in the matrix. The results are summarized in Table 2.

$$\mathrm{ID} = \log_2\left(\frac{2D}{w}\right); \qquad \mathrm{IP} = \left(\frac{\mathrm{ID}}{MT}\right) \tag{1}$$

A comparison of the tables shows that for both—the HoloLens instruction and the highlighting instruction—the index of difficulty (ID) slightly increased with a decreasing button size w, while the traveling distance for a pointing gestures was tried to be kept constant. However, the mean time for performing an action was significantly smaller for the highlighting condition than for the HoloLens, which is also reflected in the performance index IP, which is more than two times higher. The main reason for this reduced performance index when using the Hololens with fingerpointing gestures is that users wait until the fingerpointing gesture is starting to move away from the button, and then start doing the fingerpointing by themselves. Instead of a hand and finger that simply pop up at the correct position (which would be similar to regular highlighting), we consciously chose a moving hand and thus a resulting lower IP, since we believe that a natural movement of a hand is crucial for intuitiveness and more efficient memorization. Consequently, such a "waiting" behavior was not observed for the highlighting condition, and it shows in principle that the pointing gesture overlay is perceived as natural, so that there cannot be "two fingers at the same location".

Table 2 Fitts law measures for holoLens and highlighting

		HoloLens				Highlighting			
Matrix size	w (mm)	D (mm)	ID	IP	MT	D (mm)	ID	IP	MT
5 × 5	29	86.32	2.57	1.07	2.42	84.76	2.55	2.76	0.92
10 × 10	14	93.02	3.73	1.77	2.12	81.21	3.53	3.65	0.97
15 × 15	9	85.04	4.24	159	2.67	57.99	3.69	3.59	1.03

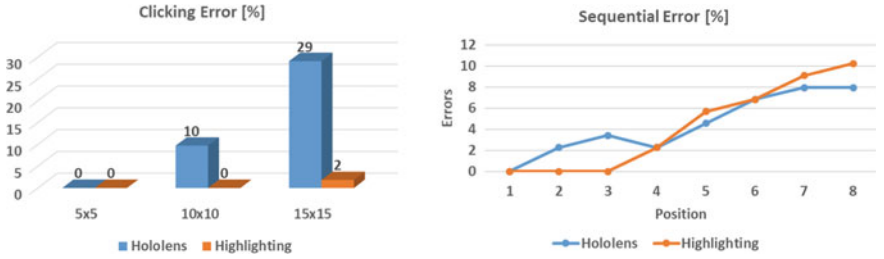


Fig. 4 Cumulative error for HoloLens and highlighting instruction (left); Accumulated errors of the memorization task (right)

4.2 Evaluation of the Clicking Accuracy

For all three matrix sizes, the user was instructed to press the button shown to him either by fingerpointing overlay using the HoloLens or by highlighting directly on the laptop screen. In all cases, the error is defined as the sum of wrongly pressed buttons. It was not possible to press two buttons simultaneously or to press a gap between the buttons. The results are shown in Fig. 4 (left). The results show that both—the HoloLens and the highlighting—perform equally well for the 5×5 matrix, while for the larger matrix sizes the amount of wrongly pressed buttons significantly increases for the fingerpointing instruction compared to the highlighting instruction. This is mainly due to the fact that the fingerpointing gesture could not unequivocally be assigned to a button anymore, since it was completely or partially occluded by the finger. For the 10×10 matrix, occlusion was one of the reasons for the occurring errors, since an accidental mistyping due to button size can be excluded as there are no errors for the highlighting instruction. For the 15 × 15 matrix, errors also occur for the highlighting method, which are probably due to accidental mistyping because of the small button size. However, also here the majority of the errors is likely from the wrongly detected buttons due to occlusion by the fingerpointing overlay. Since the forearm, the hand, and the fingers were semi-transparent (Fig. 1a), it was mainly the opaque shape attached to the fingertip that caused the occlusions. However, this shape was necessary to clearly detect the fingertip.

For the memorization study, a 4 × 4 matrix was used to avoid biasing of the results by erroneous readouts due to the small button size. The user had to keep in

mind 8 numbers in the right order being shown to him by either fingerpointing or highlighting. Once the instruction sequence was finished, the user had to reenter this sequence on the touchscreen. Additionally, the buttons show numbers that should be kept in mind.

Participants showed a similar error rate in memorization for fingerpointing (m = 2.82, SD = 1.66) and highlighting (m = 2.73, SD = 1.62) (Fig. 4 (right)). The results show that short sequences of numbers (<4) can be kept in mind better when they are shown to the user by highlighting. Thus there is evidence that there is no significant tendency toward a positive effect of fingerpointing for memorizing short sequences (<4). This finding is in-line with [21] who found that pointing positively effects the memorization of a final position, but the cognitive absorption based on the hand movement negatively impacts the memorization. However, for sequences >4 there seems to be a tendency in favor of the fingerpointing, since the amount of errors was then equal or smaller than the errors for the highlighting condition (Fig. 4 (right)). Thus, our findings do not clearly support the results from [4], who also described the supportive character of gestures for memorization of mathematical contexts, which were, however, not related to the learning of sequences.

4.3 Evaluation of Questionnaires

The results from the NASA TLX questionnaire showed a slightly higher perceived task load when using the HoloLens (m = 2.55, SD = 1.20) compared to the highlighting procedure (m = 2.08, SD = 0.88), which comes from the fact that in particular for smaller grid sizes the finger pointing was less easy to detect. This might also be due to the limited field of view, which made pointing gestures to appear more suddenly by just seeing the finger but not the forearm in the peripheral field of view. Although the users also had to wear the HoloLens during the highlighting study, this effect of a limited field of view did not occur, since the HoloLens was switched off. The reasons from the above might also be responsible for the outcomes of the SUS questionnaire, where the HoloLens was rated worse (m = 83.41, SD = 11.74) compared to highlighting (m = 91.59, SD = 7.44). Further analysis with a one-tailed paired samples t-test reveals a statistically significant difference between the usability of the HoloLens compared to highlighting, $t(10) = 2.3$, $p = 0.023$. With regard to cognitive absorption, HoloLens had a higher level (m = 5.04, SD = 0.76, $t(10) = 1.5$) than highlighting (m = 4.67, SD = 0.86). This result is intuitive considering the fact that MR is more immersive than a laptop screen and thus increases users' cognitive absorption.

5 Summary and Outlook

We showed that there is no need to switch from explaining hand gestures to artificial symbols when precise pointing on a specific object is needed. As long as the target object is considerably larger (e.g., $> 29 \times 29$ mm), regular pointing gestures with the finger can be unequivocally detected. Thus, using an MR overlay of hand gestures for explaining, e.g., the operation of machines is sufficient and no further need for pointing arrows is required. This allows for a more natural and intuitive explanation and better understanding of complex contexts. The paper further showed that the intuitive nature of hand gestures also allows for better memorization of longer instructional sequences, since human beings have better access to hand gestures than to other more artificial means of pointing.

Future work will focus on reducing the pointing error further which also stems from sources such as gaze point calibration of the HoloLens when recording or replaying the pointing gestures. Moreover, we will also investigate whether a combination of pointing gestures and highlighting will improve both—the pointing accuracy as well as the memorization of longer instructional sequences. Another study will also focus on the memorization of sequences other than numbers, such as objects, colors, or shapes, to which users might have a more intuitive access.

References

1. Agarwal R, Karahanna E (2000) Time flies when you're having fun: cognitive absorption and beliefs about information technology usage. MIS Q 24(4):665–694
2. Aigner R, Wigdor D, Benko H, Haller M, Lindbauer D, Ion A, Zhao S, Koh J (2012) Understanding mid-air hand gestures: a study of human preferences in usage of gesture types for hci. Microsoft Res TechReport MSR-TR-2012-111 2:30 (2012)
3. Alaçam S (2014) The many functions of hand gestures while communicating spatial ideas-an empirical case study. In: 18th Conference of the iberoamerican society of digital graphics, online, CUMINCAD pp 106–109
4. Aldugom M, Fenn K, Cook SW (2020) Gesture during math instruction specifically benefits learners with high visuospatial working memory capacity. Cogn Res Principles Implications 5(1):1–12
5. Brooke J (1996) Sus: A "quick and dirty" usability. Usability Eval Ind 189(3):189–194
6. Chen S, Chen M, Kunz A, Yantac AE, Bergmark M, Sundin A, Fjeld M (2013) Semarbeta: mobile sketch-gesture-video remote support for car drivers. Christian SABAH 4th augmented human international conference. USA, ACM, New York, pp 69–76
7. Gao Q, Sun Q (2015) Examining the usability of touch screen gestures for older and younger adults. Human Factors 57(5):835–863
8. Hart SG, Staveland LE (1988) Development of nasa-tlx (task load index): results of empirical and theoretical research. Adv Psychol 52:139–183
9. Hoover M, Miller J, Gilbert S, Winer E (2020) Measuring the performance impact of using the microsoft hololens 1 to provide guided assembly work instructions. J Comput Inf Sci Eng 20(6):061001
10. Huang W, Alem L, Tecchia F, Duh HBL (2018) Augmented 3d hands: a gesture-based mixed reality system for distributed collaboration. J Mult User Interfaces 12(2):77–89

11. Kervegant C, Raymond F, Graeff D, Castet J (2017) Touch hologram in mid-air. ACM SIGGRAPH 2017 emerging technologies. NY, USA, ACM, New York, pp 1–2
12. Komine S, Nakanishi M (2013) Optimization of gui on touchscreen smartphones based on physiological evaluation–feasibility of small button size and spacing for graphical objects. In: International conference on human interface and the management of information, Springer, pp 80–88
13. Laviola E, Gattullo M, Manghisi VM, Fiorentino M, Uva AE (2022) Minimal AR: visual asset optimization for the authoring of augmented reality work instructions in manufacturing. Int J Adv Manuf Technol 119(3):1769–1784
14. Leitão R, Silva PA (2012) Target and spacing sizes for smartphone user interfaces for older adults: design patterns based on an evaluation with users. In: 19th conference on pattern languages of programs
15. Lin J, Harris-Adamson C, Rempel D (2019) The design of hand gestures for selecting virtual objects. Int J Human-Comput Int 35(18):1729–1735
16. Liu X, Thomas GW, Cook SW (2018) The effect of pointing on spatial working memory in a 3d virtual environment. Appl Cogn Psychol 32(3):383–389
17. Oyama E, Tokoi K, Suzuki R, Nakamura S, Shiroma N, Watanabe N, Agah A, Okada H, Omori T (2021) Augmented reality and mixed reality behavior navigation system for telexistence remote assistance. Adv Robot 35(20):1223–1241
18. Palmarini R, Erkoyuncu JA, Roy R, Torabmostaedi H (2018) A systematic review of augmented reality applications in maintenance. Robot Comput-Interact Manuf 49:215–228
19. Park YS, Han SH, Park J, Cho Y (2008) Touch key design for target selection on a mobile phone. In: Proceedings of the 10th international conference on human computer interaction with mobile devices and services, pp 423–426
20. Prilla M, Janßen M, Kunzendorff T (2019) How to interact with augmented reality head mounted devices in care work? a study comparing handheld touch (hands-on) and gesture (hands-free) interaction. AIS Trans Human-Comput Interact 11(3):157–178
21. Rossi-Arnaud C, Longobardi E, Spataro P (2017) Pointing movements both impair and improve visuospatial working memory depending on serial position. Mem Cogn 45(6):903–915
22. Schedlbauer M (2007) Effects of key size and spacing on the completion time and accuracy of input tasks on soft keypads using trackball and touch input. Proc Human Factors Ergonom Soc Ann Meet 51(5):429–433
23. Scurati GW, Gattullo M, Fiorentino M, Ferrise F, Bordegoni M, Uva AE (2018) Converting maintenance actions into standard symbols for augmented reality applications in industry 4.0. Comput Ind 98:68–79
24. Teo T, Lee GA, Billinghurst M, Adcock M (2018) Hand gestures and visual annotation in live 360 panorama-based mixed reality remote collaboration. In: Proceedings of the 30th Australian conference on computer-human interaction, pp 406–410
25. Thomas P, David W (1992) Augmented reality: an application of heads-up display technology to manual manufacturing processes. Hawaii international conference on system sciences, vol 2. NY, USA, ACM, New York, pp 659–669
26. Tsang S, Chan A, Chen K (2013) A study on touch screen numeric keypads: effects of key size and key layout. In: International Multi-conference of engineers and computer scientists, vol 324
27. Yusof C, Halim N, Nor'a M, Ismail A (2020) Finger-ray interaction using real hand in handheld augmented reality interface. In: IOP conference series: materials science and engineering, vol 979. UK, IOP Publishing, Bristol, p 012009