Wirtschaftsinformatik 2022 Proceedings

Track 16: Human Computer Interaction & Digital Assistance Systems

Jan 17th, 12:00 AM

Typing the Future: Designing Multimodal AR Keyboards

Marius Schenkluhn
Karlsruhe Institute of Technology, Germany, marius.schenkluhn@kit.edu

Christian Peukert
Karlsruhe Institute of Technology, Germany, christian.peukert@kit.edu

Christof Weinhardt
Karlsruhe Institute of Technology, Germany, christof.weinhardt@kit.edu

Recommended Citation

Schenkluhn, Marius; Peukert, Christian; and Weinhardt, Christof, "Typing the Future: Designing Multimodal AR Keyboards" (2022). *Wirtschaftsinformatik 2022 Proceedings*. 11.
https://aisel.aisnet.org/wi2022/hci/hci/11

Typing the Future: Designing Multimodal AR Keyboards

Marius Schenkluhn[1], Christian Peukert[1], and Christof Weinhardt[1]

[1] Karlsruhe Institute of Technology (KIT), IISM, Karlsruhe, Germany
{marius.schenkluhn, christian.peukert, christof.weinhardt}@kit.edu

Abstract. Recent demonstrations of AR showcase engaging spatial features while avoiding text input. However, this is not due to descending relevance but rather because no satisfactory solution to text input in a comprehensive AR system is available yet. Any novel technological device requires rethinking the way we interact with it, including text input. With its variety of sensors, AR devices offer numerous possibilities for uni- and multimodal interaction. However, it is essential to evaluate the actual problem space before suggesting solutions. In our design science research project, we aim to create design knowledge about the learnability and performance of AR keyboards. Based on transfer of learning theory and HCI literature on virtual keyboards, we propose meta requirements and initial design principles that serve as basis for developing a multimodal AR keyboard prototype.

Keywords: Augmented Reality, Multimodal Interaction, Keyboard Input, Transfer of Learning, Design Science Research

1 Introduction

In future, Augmented Reality (AR) devices will be ubiquitous in our everyday life assisting users in various use-cases [1–3]. While the industry is waiting for lightweight, powerful, and unobtrusive AR glasses to emerge, several aspects of the next generation devices already ask for new ideas and improvements today [4]. One important aspect is text input [5]. With every new device category, researchers were exploring adequate ways for users to enter text into the computer (e.g., smartphones [6], smartwatches [7, 8], Smart-TVs [9], or smart speakers [10]) as previous methods often did not perform sufficiently. This pattern holds true for AR and poses a major user experience (UX) design challenge for the already complex transition from traditional systems to this next generation platform [11]. Text input will prevail because speech recognition is not suitable in many use cases, like in noisy environments or when entering confidential information [4, 12, 13]. Especially for expert users, it is highly important that a system facilitates a performant, learnable, portable, non-fatiguing, and unobtrusive way of text input [11]. Yet, the main representative of state-of-the-art AR headsets, Microsoft's HoloLens 2, does not provide a fast, reliable, and user-friendly keyboard. The gesture-based mid-air keyboard lacks haptic feedback, touch-typing capabilities, and visually blocks most of the field of view. However, equipped with various sensors, AR devices

17th International Conference on Wirtschaftsinformatik,
February 2022, Nürnberg, Germany

open up a plentitude of input modalities that application designers may leverage when developing user-centered AR systems, e.g., gaze-, gesture-, or contextual input.

Previous work has made attempts to create text entry techniques with the goal of finding a tailored solution for the AR and VR context [11, 14–21]. However, new approaches often struggle to both perform well and be learnt quickly [11]. Typing speed, accuracy, and learning rate are common metrics for measuring the successful application of new text input techniques and are the foundation of user acceptance [11]. Moreover, several approaches rely on external hardware, e.g., trackers, controllers, or keyboards, limiting the mobility which is essential for ubiquitous AR [19, 22, 23]. Thus, we argue that there is a need for the IS and HCI community to address and research this issue in order for user-centered AR to succeed. As prior approaches often struggled, the underlying design issues and requirements must be analyzed before suggesting novel modes of text entry. Accordingly, this research endeavor pursues the overall objective of investigating how AR keyboards need to be designed. In particular, we examine the following research questions (RQs): *How to design a mobile virtual keyboard for AR systems to increase text entry performance?* (RQ1) and *How to design a mobile virtual keyboard for AR systems to increase learnability?* (RQ2)

To address the RQs, we commenced a design science research (DSR) project to thoroughly examine the theoretical knowledge base and practical challenge, instantiate and evaluate a design artifact, and, eventually, produce design knowledge [24]. Our research is grounded in transfer of learning (ToL) theory and informed by prior HCI research on virtual keyboards. In this paper, we focus on the first three steps of the first design cycle to derive meta requirements and design principles from relevant issues and present a first version of the artifact featuring touch-typing and multimodal input.

2 Theoretical Background

Virtual Keyboards for Augmented Reality. Although consumer-ready AR headsets that are lightweight, small, affordable, and have long-lasting battery life are not yet available, many companies experiment with AR devices such as intermediate smartphone-based solutions or more capable headsets like the Microsoft HoloLens 2 to develop future use cases [25, 26]. In their review, Dube et al. provide a comprehensive overview of text entry techniques in VR [11]. Their suggested input categories and most of the accompanying issues, such as haptic feedback, new layout acceptance, low performance frustration, and physical demand, also apply to AR. The review separates physical from virtual techniques and the regular qwerty keyboard layout (according to the first row of characters on the English keyboard) from other approaches outlining that non-qwerty layouts tend to perform worse and require longer training periods [11]. This issue is attributed to a network effect, as most users are familiar with qwerty layouts [27]. Overall, they conclude that, next to speed and accuracy, a well-designed keyboard needs to pay attention to haptic feedback, comfort, physical and cognitive demand, and potential frustration due to low performance [11].

Transfer of Learning. Depending on the prior knowledge, the teaching method, and the learning target, existent knowledge can have a positive or negative impact on learning [28]. Hajian summarizes four theories in the field of the transfer of learning [29]. There are several aspects that increase the likelihood of successful learning transfer from one context to another. For instance, transfer is more likely to be successful if the learning target and context are similar to the knowledge origin [28, 29]. The theory of low and high road transfer describes two related mechanisms of how transfer can occur [28]: Comparable to the two systems of thinking, *low* road transfer triggers intuitive responses of a well-known concept in a slightly different context [28, 30]. In contrast, *high* road transfer requires "mindful abstraction from the context of learning or application and a deliberate search for connections" [28, p. 8]. Low and high road transfer can be exploited by the concepts of *hugging* and *bridging* [28]. By applying hugging, the prior skill should be well-trained and tightly linked to the learning target. Bridging encourages the learner to actively abstract knowledge from the first context to apply it in the latter. Overall, these insights impact design decisions for the development, teaching, and evaluation of the artifact as a leading theory.

Multimodal Interaction. Multimodal interaction is natural to humans [13]. When we give directions to a foreigner for example, we use spoken language and articulate by using our hands. Research distinguishes parallel and sequential multimodality, depending on the simultaneous or successive application of at least two modes of interaction [13]. In general, multimodality has several advantages e.g., regarding user preference, flexibility, and reliability [13]. Furthermore, multimodal interaction was already applied in AR to improve user experience [31–33]. However, the area of combining multiple non-voice interaction modes is rather unexplored to date [13, 33].

3 Method & First Activities in Design Science Research Cycle 1

To tackle the proposed RQs, we initialized a DSR project following the framework of Kuechler and Vaishnavi [34]. By means of creating a virtual keyboard artifact specifically for AR systems, we aim for knowledge gain to inform both research and practice. DSR offers the adequate research paradigm by providing structured, comprehensive, and iterative frameworks for the construction and observation of a previously non-existent artifact. Within this article, we will present the results of the first three activities of the first design cycle. Based on reviewing relevant literature from the HCI domain (particularly research on virtual keyboards for VR and AR systems), we identified issues (I) (**awareness of problem**). Next, taking the issues, virtual keyboard design knowledge, and ToL theories into account, we have derived meta requirements (MR) and proposed initial design principles (DP) as depicted in Fig. 1 (**suggestion**). The MRs and DPs are then used to implement a situated software artifact (**development**) for evaluation [35]. Overall, we plan to employ two full design cycles. In the following, we describe the already conducted activities in more detail:

Problem Awareness & Suggestion. Across literature on virtual keyboards, several issues were already pointed out that need to be taken into consideration [11]. Depending on the keyboard design, directly mapping more than the 26 letters of the English alphabet to 10 fingers or few buttons on a controller is a challenge (**I1**). Hence, previous research with direct mappings was limited to digits [18], finger combinations (“chords”) [36], or overloading fingers with multiple characters [20]. Yet, solving this issue by capturing multiple touch regions on each finger might lead to complex and ambiguous gesture recognition (**I2**) [18]. Therefore, we suggest a multimodal approach (**MR1**). Using two adequate input modalities results in enough combinations to capture all letters without the necessity to assign multiple touch regions per finger or choosing low performing chorded keyboards [11]. More specifically, *parallel* multimodal interaction can increase entry performance as input combinations can occur simultaneously (**MR2**). Interacting via two simple modalities may also require less cognitive effort than one complex mode [13]. This motivates the first suggested DP: *Provide the Augmented Reality Keyboard (ARKB) with parallel multimodal input in order to quickly access the full alphabet while ensuring mobility.* (**DP1)**

Establishing non-qwerty keyboard layouts comes with further issues. Complex and new techniques can lead to a higher mental load (**I3**) while the training poses an increased entry barrier (**I4**) [11]. This decreased learnability can be ascribed to the dissimilarity between traditional text input and the new technique which complicates transfer of learning [29]. While there might be layouts that could be easier to learn and master for beginners, most users have prior typing experience with the qwerty layout and alternatives show low performance (**I5**) [11]. Therefore, it is imperative to reuse and build on prior knowledge as much as possible. On the one hand, the goal is to exploit low road transfer (i.e., hugging) with a similar design and by addressing internalized intuitive knowledge (**MR3**). On the other hand, high road transfer is exploited (i.e., bridging) by actively pointing out the differences and how to foster them to abstract knowledge (**MR4**). New non-qwerty layouts could even imply effects of negative transfer [28]. Consequently, we suggest the following DP: *Provide the ARKB with interactions based on transferable prior knowledge to increase learnability.* (**DP2**)

Some entry techniques like gaze-based interaction have an inherent performance cap resulting from the required dwelling time that separates intended fixation from unintentional triggers during the search for characters (so-called Midas Touch effect) (**I6**) [37]. Having to wait for the system can lead to user frustration [11]. Therefore, the system’s recognition rate should be faster than users’ entry speeds (**MR5**). Moreover, the event triggers for each character should be time independent (**MR6**), i.e., not requiring two subsequent actions or waiting times. Non-haptic techniques inherit the same issue of a typically lower input performance compared to haptic techniques (**I7**), thus, the system should provide haptic feedback (**MR7**) [5, 11, 36]. Especially for

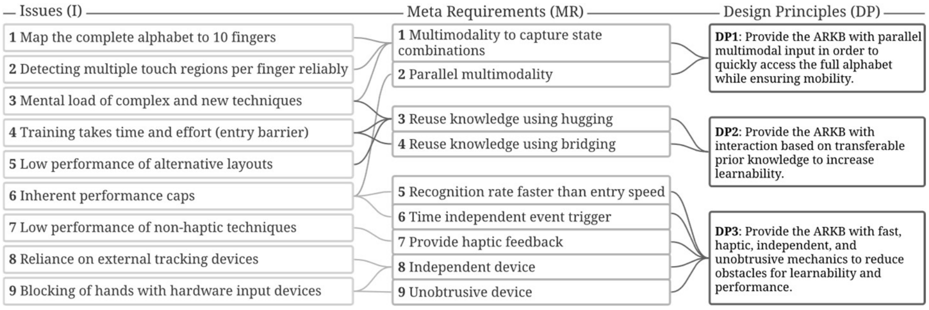

Figure 1. Derived Issues, suggested MRs and DPs for the ARKB artifact development

independent AR glasses, stationary tracking devices hinder mobile usability (**I8**) [22]. The same issue arises for hardware input devices such as controllers [15], wrist-cameras [18], or gloves [20] that block or limit the users' hands, need to be picked up, and stored (**I9**). Hence, the AR device should also be independent from external trackers or input devices (**MR8**). Finally, the device should be unobtrusive to keep the hands free when no text entry is performed (**MR9**). Thus, we suggest the third DP: *Provide the ARKB with fast, haptic, independent, and unobtrusive mechanics to reduce obstacles for learnability and performance.* **(DP3)** In conclusion, DP1 ensures the feasibility, DP2 the learnability, and DP3 the (final) performance of the approach.

Development. For the instantiation of the three DPs, we suggest a gaze- and gesture-based virtual ARKB artifact. The layout should be qwerty to be in line with DP2 and the similarity required by hugging. Moreover, each finger is responsible for the same character set like in regular touch typing. For instance, the left pinky is assigned to q, a, and z and the left middle finger to e, d, and c. The respective key is "pressed" by pinching thumb to finger. To account for the characters t or g, both index and middle fingers are pressed simultaneously. This movement is highly trained [18] and, thus, likely to transfer. In this case, the thumb provides a form of haptic sensation. Furthermore, the regular qwerty layout for the characters is divided into three layers (qwe, asd, yxc) [20]. The selection of the different layers is handled by gazing at one of three virtual areas projected by the AR device. The artifact is implemented in Unity for deployment on a Microsoft HoloLens 2[1]. The HoloLens has eye-tracking and hand-tracking capabilities without the need for an additional device to comply with DP3. Based on suggestions from Yi et al., we analyze the relative speed between thumb and each finger to detect a "key press" [14]. Then, the area the user is currently gazing at is queried which selects the correct character.

4 Concluding Note & Future Research

In this research-in-progress, we contribute to the knowledge base by deriving MRs and initial DPs from prior research on virtual keyboards for AR and ToL theory to implement a multimodal AR keyboard artifact. Further, the current state of the artifact indicates that it is able to recognize both finger taps and gaze-selection solely based on the integrated sensors of a HoloLens 2 at a sufficient rate to provide fast text input. The final implementation of the artifact will then be evaluated in a lab experiment in which we will measure common features, e.g., typing speed over time (**evaluation**) [11]. Based on the obtained findings, we will then be able to draw conclusions regarding the feasibility of the prototype system and applicability of the DPs including the implications of ToL for virtual AR keyboard designs (**conclusion**). In a subsequent design cycle, we want to instantiate the DPs in another artifact for generalization from an artefactual contribution towards a nascent design theory [35]. Additionally, further investigations will be made by integrating predictive text and revision capabilities [11, 38]. Hence, fast and enjoyable typing in AR might just be one gaze and tap away.

---

[1] A preview video of the artifact is available here: https://youtu.be/Aw93rxjk1iU

References

1. Bardeen, L., HBR Analytic Services: Mixed Reality: A new dimension of work (2018)
2. Bain & Company: Augmented or Virtual Reality (2020), https://www.bain.com/insights/customer-experience-tools-augmented-or-virtual-reality/ (Accessed: 12.11.2021)
3. Gurman, M., Webb, A. and Bergen, M.: Apple's Next Big Thing: Augmented Reality (2017), https://www.bloomberg.com/news/articles/2017-03-20/apple-s-next-big-thing (Accessed: 12.11.2021)
4. Masood, T., Egger, J.: Adopting augmented reality in the age of industrial digitalisation. Computers in Industry, vol. 115, 103–112 (2020)
5. Dudley, J., Benko, H., Wigdor, D., Kristensson, P.O.: Performance envelopes of virtual keyboard text input strategies in virtual reality. In: 2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 289–300. IEEE (2019)
6. Balakrishnan, A.: How the first iPhone blew the BlackBerry away with an amazingly accurate on-screen keyboard (2017), https://www.cnbc.com/2017/06/26/creation-of-the-first-iphone-touchscreen-interview-with-scott-forstall-tony-fadell-and-greg-christie.html (Accessed: 12.11.2021)
7. Oney, S., Harrison, C., Ogan, A., Wiese, J.: ZoomBoard: a diminutive qwerty soft keyboard using iterative zooming for ultra-small devices. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2799–2802 (2013)
8. Espósito, F.: Shift Keyboard introduces a new way to write messages on Apple Watch (2020), https://9to5mac.com/2020/01/07/shift-keyboard-introduces-a-new-way-to-write-messages-on-apple-watch/ (Accessed: 12.11.2021)
9. Choi, Y.M., Li, J.: Usability Evaluation of a New Text Input Method for Smart TVs. Journal of Usability Studies, vol. 11 (2016)
10. Peng, T. and Sarazen, M.: Inside the Voice Inside Your Smart Speaker (2017), https://medium.com/syncedreview/inside-the-voice-inside-your-smart-speaker-e8eabb679740 (Accessed: 12.11.2021)
11. Dube, T.J., Arif, A.S.: Text entry in virtual reality: A comprehensive review of the literature. In: International Conference on Human-Computer Interaction, pp. 419–437. Springer (2019)
12. Pyae, A., Joelsson, T.N.: Investigating the usability and user experiences of voice user interface: a case of Google home smart speaker. In: Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct, pp. 127–131 (2018)
13. Turk, M.: Multimodal interaction: A review. Pattern recognition letters, vol. 36, 189–195 (2014)
14. Yi, X., Yu, C., Zhang, M., Gao, S., Sun, K., Shi, Y.: Atk: Enabling ten-finger freehand typing in air based on 3d hand tracking data. In: Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology, pp. 539–548 (2015)
15. Yu, D., Fan, K., Zhang, H., Monteiro, D., Xu, W., Liang, H.-N.: PizzaText: Text entry for virtual reality systems using dual thumbsticks. IEEE transactions on visualization and computer graphics, vol. 24, 2927–2935 (2018)
16. Jiang, H., Weng, D.: HiPad: Text entry for head-mounted displays using circular touchpad. In: 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 692–703. IEEE (2020)

17. Kim, Y.R., Kim, G.J.: Hovr-type: Smartphone as a typing interface in vr using hovering. In: 2017 IEEE International Conference on Consumer Electronics (ICCE), pp. 200–203. IEEE (2017)
18. Prätorius, M., Burgbacher, U., Valkov, D., Hinrichs, K.: Sensing thumb-to-finger taps for symbolic input in vr/ar environments. IEEE computer graphics and applications, vol. 35, 42–54 (2015)
19. Jiang, H., Weng, D., Zhang, Z., Bao, Y., Jia, Y., Nie, M.: Hikeyb: High-efficiency mixed reality system for text entry. In: 2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), pp. 132–137. IEEE (2018)
20. Kuester, F., Chen, M., Phair, M.E., Mehring, C.: Towards keyboard independent touch typing in VR. In: Proceedings of the ACM symposium on Virtual reality software and technology, pp. 86–95 (2005)
21. Rosenberg, R., Slater, M.: The chording glove: a glove-based text input device. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 29, 186–191 (1999)
22. Mourouzis, A., Kilintzis, V., Chouvarda, I., Maglaveras, N.: Using the LEAP 3D motion controller for implementing text entry for all. In: Proceedings of the 2014 European, Mediterranean and Middle Eastern Conference on Information Systems (2014)
23. Grubert, J., Witzani, L., Ofek, E., Pahud, M., Kranz, M., Kristensson, P.O.: Text Entry in Immersive Head-Mounted Display-based Virtual Reality using Standard Keyboards. In: IEEE (ed.) 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR) (2018)
24. Hevner, A., March, S.T., Park, J., Ram, S.: Design science research in information systems. MIS Quarterly, vol. 28, 75–105 (2004)
25. Bohn, D.: Microsoft's HoloLens 2: a $3500 mixed reality headset for the factory (2019), https://www.theverge.com/2019/2/24/18235460/microsoft-hololens-2-price-specs-mixed-reality-ar-vr-business-work-features-mwc-2019 (Accessed: 12.11.2021)
26. Apple: Augmented Reality (2020), https://www.apple.com/augmented-reality/ (Accessed: 12.11.2021)
27. Farrell, J., Klemperer, P.: Coordination and lock-in: Competition with switching costs and network effects. Handbook of industrial organization, vol. 3, 1967–2072 (2007)
28. Perkins, D.N., Salomon, G.: Transfer of learning. International encyclopedia of education, vol. 2, 6452–6457 (1992)
29. Hajian, S.: Transfer of Learning and Teaching: A Review of Transfer Theories and Effective Instructional Practices. IAFOR Journal of education, vol. 7, 93–111 (2019)
30. Kahneman, D.: Thinking, fast and slow. Macmillan (2011)
31. Kaiser, E., Olwal, A., McGee, D., Benko, H., Corradini, A., Li, X., Cohen, P., Feiner, S.: Mutual disambiguation of 3D multimodal interaction in augmented and virtual reality. In: Proceedings of the 5th international conference on Multimodal interfaces, pp. 12–19 (2003)
32. Chen, Z., Li, J., Hua, Y., Shen, R., Basu, A.: Multimodal interaction in augmented reality. In: 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 206–209. IEEE (2017)
33. Nizam, S.S.M., Abidin, R.Z., Hashim, N.C., Lam, M.C., Arshad, H., Majid, N.A.: A review of multimodal interaction technique in augmented reality environment. International Journal on Advanced Science, Engineering and Information Technology, vol. 8, 1460–1469 (2018)
34. Kuechler, B., Vaishnavi, V.: On theory development in design science research: anatomy of a research project. European Journal of Information Systems, vol. 17, 489–504 (2008)

35. Gregor, S., Hevner, A.R.: Positioning and Presenting Design Science Research for Maximum Impact. MIS Quarterly, vol. 37, 337–355 (2013)
36. Bowman, D.A., Rhoton, C.J., Pinho, M.S.: Text input techniques for immersive virtual environments: An empirical comparison. In: Proceedings of the human factors and ergonomics society annual meeting, pp. 2154–2158. SAGE Publications Sage CA: Los Angeles, CA (2002)
37. Vrzakova, H., Bednarik, R.: That's not norma (n/l) a detailed analysis of midas touch in gaze-based problem-solving. In: Proceedings of the 2013 CHI Conference on Human Factors in Computing Systems, pp. 85–90 (2013)
38. Li, Y., Sarcar, S., Zheng, Y., Ren, X.: Exploring Text Revision with Backspace and Caret in Virtual Reality. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, pp. 1–12 (2021)