AUGMENTED REALITY DESIGN HEURISTICS: DESIGNING FOR DYNAMIC INTERACTIONS

Tristan C. Endsley, Kelly A. Sprehn, Ryan M. Brill, Kimberly J. Ryan, Emily C. Vincent, James M. Martin Draper
Cambridge, MA

Augmented Reality (AR) has emerged as a rapidly developing technology, capable of a wide scope of applications across a variety of domains. AR technologies allow for a virtual experience to be overlaid on top of a physical environment, creating a hybrid experience in which virtual objects become a part of the user's perceptual and physical environment. Rapid progression of the AR field requires that effective and validated methods of design evaluation be developed. Failure to consider the usability of AR applications during the design process will result in an increase in user errors and accidents, limiting user trust of the technology and undermining user perceptions of the technology, for both AR and Virtual Reality (VR) technologies (Nordrum, 2016). Through a robust and iterative process, a set of Design Heuristics for AR were developed for multidimensional augmented environments with the aim of advancing AR design methods for human factors, ergonomics, and user experience practitioners within the expanding AR community.

INTRODUCTION

Augmented Reality (AR) technologies project virtual information onto real world environments (Krevelen & Poelman, 2010). While Virtual Reality (VR) typically creates a fully immersive virtual experience, AR augments the real world by overlapping virtual information onto the physical environment (Ko, et al., 2013). Beyond AR, this emerging field also uses the term "Mixed Reality" (MR) to indicate the capabilities of the technology to appropriately map and place objects in a user's environment (deSouza et al., 2009). For the purpose of this paper, we will use the term AR, primarily referring to the more advanced technologies. While many AR capabilities have emerged as mobile applications, there has been a growth of new technologies that promote a more immersive experience in the physical environment or context of operation (e.g. Moore, 2006). Wearable devices such as the Microsoft HoloLens, Magic Leap, Meta, Epson Moverio Smart Glasses, and DAQRI allow for a more integrated virtual experience in a multidimensional space. This capability has spawned a large number of new applications, which are pushing the boundaries of current practices. Many AR applications focus on augmenting vision, with some instances of augmentation of other senses. While the main thrusts of research have been in visual dimensions, the extension of current lessons learned to auditory, haptic, and other human sensory modalities is feasible. However, with these more sophisticated capabilities come specific nuances and challenges that need to be addressed.

Much focus of augmented reality is on the augmentation of human cognition (Schmorrow, 2005; Webb et al., 2016). While the benefits of this could be boundless (a multidimensional understanding of complex data, faster navigation in convoluted spaces, more interaction opportunities with yet-to-be-imagined objects), the cost of failure of inappropriately designed AR applications could be significant, resulting in a loss of situation awareness, creation of cognitive overload, interruption of work flows, leading to both poor performance and higher levels of human error (Carmigniani, 2011).

The three-dimensional nature of the AR experience makes evaluation of design challenging. Testing any hypothesis requires a certain amount of accuracy in measurement instruments, data collection devices, and the task under scrutiny. The resulting hypothesis testing model inherently contains error associated with the effects of this accuracy. Reporting the error describes the trust and confidence that we can have in the results being generalizable. A number of assessment and prediction methods are used throughout the development of a system. Considering the ideation, development, and refinement cycles that many products go through, reducing errors and increasing the accuracy of predicted success or failure as the product advances is important. However, there is a significant tradeoff between cost and accuracy of measurement, which leads designers to rely on low-cost, lower-effort methods in the beginning of the design process and refine their designs with higher-cost, higher accuracy methods at the end of the cycle. Both Neilsen (1994) and John & Marks (1996) discuss cost and accuracy tradeoffs within certain usability evaluation methods.

Design Heuristics are guidelines that can be used to evaluate the usability of a system and guide its design. Heuristics have a broad range of use. In psychology, "heuristic" refers to cognitive shortcuts that help (or harm) humans in making decisions (Groner, Groner, and Bischof, 1983). In computer science and algorithm development, a heuristic is a function that approximates a solution, which may not be optimal, but trades off speed for optimality in returning a solution (Apter, 1970). In user interface design, the term is almost synonymous with guidelines, and can be used by both designers and evaluators. Certainly, the most famous design heuristics are the ten usability heuristics from Molich and Nielsen (1990). Like psychological heuristics, usability heuristics are often used as shortcuts when data from a formal user study is not available. Heuristics can be used across the
stages of the design process, guiding small and large decisions and predicting success or failure of the usability in prototypes with varying levels of fidelity.

Evaluation methods using heuristics as measurement instruments have been formalized in the design and user experience contexts, enabling rapid and low cost usability evaluation (Mack & Nielsen, 1994). The method involves a flexible number of expert evaluators that are familiar with the heuristics. The evaluators review the experience, noting where the heuristics are violated. In more formal methods, the evaluators develop Action Reports for each violation that describe where and how, and which heuristic is not being met (Mack & Nielsen, 1994). These Action Reports can number from 300 to as few as two and are useful in both early design reviews to higher fidelity systems (Mack & Nielsen, 1994). Additionally, Action Reports can provide formative feedback on what actions to take in order to effectively address identified issues. In less formal instances of the method, evaluators will provide a score, or relative performance metric of the system along each of the heuristics, offering designers and developers insight into the predicted success or failure of the product usability at a high level (Mack & Nielsen, 1994).

With the advantages of audience, speed, cost, and applicability, heuristics provide a powerful development and evaluation tool for user experiences. In more traditional interaction environments, such as desktop and mobile applications, Neilsen’s heuristics have been used to effectively inform design. This approach, however, does not require practitioners to consider user needs within more dynamic environments. Through our AR design process, it was found that Neilsen’s design heuristics do not fully address the challenges of designing for an integrated digital/physical user experience. The development of a set of usable design heuristics for augmented reality experiences is needed. In this paper, we aim to provide practitioners with a usable, tractable set of Augmented Reality heuristics. This set will result in a quick, low cost method able to influence and reflect the design quality of augmented reality experiences. A multi-dimensional creation and evaluation approach was utilized in order to develop the current set of heuristics, presented below.

**PRACTICE INNOVATION**

The current set of AR design heuristics were developed by leveraging affinity diagramming, expert evaluations, feedback from active AR designers, and statistical analyses. From the method of Contextual Inquiry, affinity diagramming offers a systematic method for organic hierarchies to emerge from a large set of concepts (Beyer and Holtzblatt, 1997). This was helpful in organizing many different heuristics into an initial set. Because designers, developers, and evaluators utilize heuristic evaluations, we sought initial feedback from experts from these areas. The set of heuristics was updated based on this feedback. We then recruited another set of expert evaluators and designers. Qualitative feedback indicated usability and semantic changes that were necessary. Quantitative ratings were analyzed for inter-rater reliability and internal item consistency. The analyses then provided insights upon which to iteratively update the set again.

**Affinity Diagramming**

Affinity diagramming is an exercise used to identify themes from individual thoughts, notes, concepts, and/or pieces of information (Beyer and Holtzblatt, 2003). As a group process, participants then organize the individual concepts on a board one at a time, reading each aloud as they do so. As new notes are placed on the board, they are grouped with others based on thematic similarity. Groups can split and change as new notes are added to the board, and as new themes become apparent. In this case, we used literature review to populate each note with a singular heuristic concept. We did this for each AR oriented concept, guideline, and heuristic identified in the literature. In this way, the heuristics formed a thematic structure through a bottom-up approach.

A review of the literature revealed few validated, generalized AR heuristics lists in use. Many of the AR heuristics identified focused on specific applications of AR technologies, namely mobile or smartphone AR devices (Ko et al., 2013), or on more general principles or design features of the AR design space (Dünser et al., 2007). While these provided some valuable insight into distinct aspects of AR design evaluation, they lacked parsimonious application to current ongoing design applications. Table 1 summarizes the sources of the heuristics lists that were leveraged.

| Reference                  | Heuristics Contribution Summary                                      |
|----------------------------|-----------------------------------------------------------------------|
| Atkinson et al., 2007      | 12 design and usability heuristics                                   |
| Dünser et al., 2004        | 9 general principles for AR design                                   |
| Franklin et al., 2014      | 9 AR heuristics for collaborative and distributed AR systems         |
| Furmanski et al., 2002     | 14 cognitive principles for design features                         |
| Gong and Tarasewich, 2004  | 15 guidelines for handheld mobile device interface design            |
| Kalalahti, 2015            | 6 design heuristics, plus Neilsen’s 10                               |
| Kim et al., 2007           | 25 guidelines for tangible user interfaces                           |
| Ko et al., 2013            | 5 categories of 22 AR heuristics for smartphones                     |
| Kourouthanassis et al., 2013| 5 principles for mobile augmented reality                           |
| Molich and Nielsen, 1990   | 10 commonly accepted usability heuristics                           |
| Pinelle, et al., 2008      | 10 video game heuristics                                             |

Our affinity diagramming process sought to leverage the existing heuristics in the AR space, include related fields (such as video game design), and generate themes for immersive AR interactions. Findings from these sources were broken into 97
notes. These notes were organized using the affinity diagramming process by two evaluation experts and a design expert. Each note was organized by theme. Effort was made to separate notes from video game design heuristics that were not generalizable to AR applications. Several additional heuristics grounded in design explorations with AR technologies were added to this diagram once gaps were realized. The diagramming process delivered an initial set of heuristics, which were then used in a formal heuristics evaluation.

**Formative Heuristic Evaluations**

Much of the power of the heuristic evaluation method comes from the number of users that can take advantage of the information. Designers and evaluators leverage the guidelines defined by the set of heuristics. With this in mind, we sought feedback and usability from both groups.

In the first iteration, we recruited three experts familiar with the process of heuristic evaluations. Two evaluators had used Nielsen’s 10 usability heuristics to assess designs. The third evaluator was relatively new to the practice, but had learned the method and completed a small assessment. The first set of heuristics included 18 guidelines. Each was titled with a description. Evaluators were asked to complete a task aided by an augmented reality experience and then score the experience along each of the 18 heuristics. The evaluators scored each heuristic along a 10-point Likert scale, which was then converted into a binomial representation. Doing this produced a Cohen's kappa ($\kappa$) of 0.45, or "moderate" inter-rater reliability (IRR) according to Fleiss (1981), as shown in Table 2. They were also asked to evaluate the usability of the heuristic and its ability to guide a good AR experience. The evaluators started to exhibit signs of survey fatigue at the point of usability evaluations. Verbal feedback was much more meaningful than the incomplete quantitative responses. The three expert evaluators found the heuristics lacking consistency of voice: some were directive while some were lightly suggestive. While the set of 18 covered all of the aspects of the AR experience, there were duplicates that caused some confusion. These comments led to clarification and shortening of the heuristic descriptions, so as to construct larger distinctions between heuristics. This set of heuristics were collapsed into 15 separate topics, each supported by a new description.

**Usability Evaluations**

In the second iteration we sought feedback from five expert evaluators, three of whom were experts in AR design as well. Following an experience with an AR design, evaluators filled out a heuristic evaluation for the design, delivered via an online survey. In addition to understanding the usability of the heuristics, we looked at objective measures of internal consistency and IRR. A Cronbach’s Alpha indicating internal consistency ($\alpha$) = 0.65, while "questionable" according to Kline (2000), can be considered a good improvement and starting point for the continued refinement of the heuristic set. The measures of IRR suggest that the raters converged much more effectively than the first iteration. Converting the scoring methods to binomial codes, we found Cohen’s kappa ($\kappa$) = 0.65, which is rated as "substantial" according to the Fleiss scale (1981). Feedback was collected from a post-evaluation group interview method.

Overall, the heuristics were well received, although some points of clarification were again necessary. Evaluators commented on the constraints of the survey method preventing them from specifying particular cases of heuristics violations. It was also thought that inconsistencies in language content might make the heuristics difficult to use when forming the AR designs. From this feedback, we iterated around the clarity of language and discussed the possibility of visual examples for evaluators to be able to identify individual heuristic issues in the evaluation process, as a part of the Action Report development. Emerging from this set of evaluations, a new set of 14 design heuristics were generated. This initial set was then put into practice within a more complex design use case.

The designers used this set of 14 heuristics while designing two fully functional, complex AR applications, which were evaluated in an experimental setting in order to test the validity of the heuristics. These AR experiences were designed to assess correlations between user performances with varying degrees of designer compliance with the heuristics. Through the design process of these two applications, it became clear that the heuristics had irregularities, namely overlap in concepts represented, and inconsistent designer interpretations of the heuristics and their applications. Intermediary statistics were not run on the set of 14 heuristics as the identified lack of usability of the 14 design heuristics suggested an additional iteration of the heuristics was required.

A set of 9 heuristics (presented in Findings below) was created through an additional affinity diagramming process with statistical evaluation. The content of this affinity diagramming exercise, however, was generated through a set of 84 deconstructed items and descriptions of elements, “heuristic principles,” from the initial set of 14 heuristics, as well as Nielsen’s usability heuristics. Six members of the evaluation team utilized these 84 heuristic principles to create six individual sets of affinitized heuristics. These six sets were then consolidated and integrated into a set of 12 separable, AR-specific heuristics through an additional affinitizing process. Three additional heuristics centered on good “multimodality” and “direct manipulation” of virtual objects.

Issues and discrepancies were sorted through and a final set of nine separable, AR-specific heuristics were created. From the 12, several proposed heuristics were removed due to lack of specificity to AR-only contexts (i.e. good multimodal design practices), and general usability principles, which overlap with established design sets (i.e. Nielsen’s heuristics).

For this fourth iteration of heuristics, the evaluation process assessed inter-item consistency and inter-rater reliability of 5 reviewers across a finalized AR application ($N=5$). Inter-rater reliability, Cohen’s Kappa ($\kappa$) = 0.53, fell as compared to the previous set of heuristics, but internal consistency, Cronbach’s Alpha ($\alpha$) = 0.81, rose to a “Good” level, along the convention described by Kline (2000). These results indicate that the set of nine heuristics has high internal
consistency and inter-item separability. However, the low inter-rater reliability indicated that our raters did not reach agreement about the quality of the application that they evaluated, suggesting further steps to support better consensus are necessary.

Table 2. Description of Iterative Heuristic Set Statistics. Some values are omitted as described in the text.

| Iteration | Number of Heuristics | Interrater Reliability (Cohen's \(k\)) | Internal Item Consistency (Chronbach's \(\alpha\)) |
|-----------|----------------------|----------------------------------------|-----------------------------------------------|
| 1         | 18                   | 0.45                                   | --                                            |
| 2         | 15                   | 0.65                                   | 0.65                                          |
| 3         | 14                   | --                                     | --                                            |
| 4         | 9                    | 0.53                                   | 0.81                                          |

**FINDINGS**

The following AR heuristics frame the multiple aspects of a design that must be considered by designers of AR experiences particularly given the integrated context of the experience.

1. **Fit with user environment and task.** AR experiences should use visualizations and metaphors that have meaning within the physical and task environment in which they are presented. The choice of visualizations & metaphors should match the mental models that the user will have based on their physical environment and task.

2. **Form communicates function.** The form of a virtual element should rely on existing metaphors that the user will know in order to communicate affordances and capabilities.

3. **Minimize distraction and overload.** AR experiences can easily become visually overwhelming. Designs should work to minimize accidental distraction due to designs that are overly cluttered, busy, and/or movement filled.

4. **Adaptation to user position and motion.** The system should adapt such that virtual elements are useful and usable from the variety of viewing angles, distances, and movements that will be taken by the user.

5. **Alignment of physical and virtual worlds.** Placement of virtual elements should make sense in the physical environment. If virtual elements are aligned with physical objects, this alignment should be continuous over time and viewing perspectives.

6. **Fit with user’s physical abilities.** Interaction with AR experiences should not require the user to perform actions that are physically challenging, dangerous, or that require excess amounts of coordination. All physical motion required should be easy.

7. **Fit with user’s perceptual abilities.** AR experiences should not present information in ways that fall outside of an intended user's perceptual thresholds. Designers should consider size, color, motion, distance, and resolution when designing for AR.

8. **Accessibility of off screen objects.** Interfaces that require direct manipulation (for example, AR & touch screens) should make it easy for users to find or recall the items they need to manipulate when those items are outside the field of view.

9. **Accounting for hardware capabilities.** AR experiences should be designed to accommodate for the capabilities & limitations of the hardware platform.

**DISCUSSION**

**Future Work**

Our results indicated that a process of iteration through evaluation is effective for the creation of robust sets of heuristics for designers and evaluators. Expert evaluations of the heuristics within a real design context provide useful insight into the design process of dynamic augmented experiences. Further iterations of the current list of heuristics will be formulated through similar testing methodologies, including testing the inter-rater reliability and measures of internal consistency. Evaluation across multiple AR applications, intended to correlate user task performance to heuristic scores, will provide the empirical evidence to the power of the final set. This process will also serve to further evaluate and hone the current set of heuristics into meaningful, usable design guidelines for multidimensional augmented reality experiences.

**Impact to the Field**

AR has a specific and unique set of challenges designers need to overcome in order for this technology and its applications to be integrated into workflows, tasks, and experiences more seamlessly. The medium of AR introduces many new considerations into the design process. For example, AR applications are implemented within and reliant on the contexts of physical environments. Whereas applications in traditional interaction environments can rely on consistent and designer-defined context, designers of AR applications must take dynamic physical environments into account. The surrounding physical environment frames the designer’s visual assets and invites tangible, physics-based metaphors for interaction. Virtual objects move within a threedimensional space and can exist in a scene while being outside of a user’s field of view. As a result, designers must consider positioning and attention direction in a more dynamic manner. Inputs based on body movement, gesture (Merrill & Maes, 2007), position and collaborative interaction (Franklin et al., 2014) open up new possibilities and new challenges.

These features, among others, demand the expansion of our current understanding of how to design and evaluate these experiences. Luckily, there are a many diverse fields (including art, design, architecture, and film) from which we can pull elements of effective immersive design. The design of virtual objects needs to fit within the physical world in such a way that allows users to make sense of it. Building appropriate mental models is critical to user experience (Qian et al., 2011). The use of storytelling and metaphors to build user understanding of an interface must also be adapted to the open ended nature of the experience. Additionally, designers must
factor in limitations to a user’s physical capacity and ensure that their users can safely navigate their environment. These and other significant differences from the desktop, mobile, and VR environments bring new complexity to the design process, demanding an extension of Nielsen’s heuristics. The growth and expansion of new AR capabilities needs to be informed by good design. We need to evolve our design methods to incorporate this expansion.

REFERENCES

Atkinson, B. F. W., Bennett, T. O., Bahr, G. S., & Nelson, M. M. W. (2007, July). Development of a multiple heuristics evaluation table (MHET) to support software development and usability analysis. In International Conference on Universal Access in Human-Computer Interaction (pp. 563-572). Springer Berlin Heidelberg.

Apter, M. J. (1970). The Computer Simulation of Behaviour. London: Hutchinson & Co. p. 83.

Beyer, H., & Holtzblatt, K. (1997). Contextual design: defining customer-centered systems. Elsevier.

carmigniani, J., Furfet, B., Anissetti, M., Ceravolo, P., Damiani, E., & Iivkovic, M. (2011). Augmented reality technologies, systems and applications. Multimedia Tools and Applications, 51(1), 341-377

de Souza e Silva, A.; Sutko, D. M. (2009). Digital Cityscapes: merging digital and urban playspaces. New York: Peter Lang Publishing, Inc.

Dörsä, A., Grasset, R., Seichter, H., & Billinghurst, M. (2004). Applying HCI principles to AR systems design. HIT Lab NZ, University of Canterbury, New Zealand.

Fleiss, J. L. (1981). Statistical methods for rates and proportions (2nd ed.). New York: John Wiley. ISBN 0-471-26370-2.

Franklin, F., Breyer, F., & Kelner, J. (2014, May). Usability heuristics for collaborative augmented reality remote systems. In Virtual and Augmented Reality (SVR), 2014 XVI Symposium on (pp. 53-62). IEEE.

Furmanski, C., Azuma, R., & Daily, M. (2002). Augmented-reality visualizations guided by cognition: Perceptual heuristics for combining visible and obscured information. In Mixed and Augmented Reality, 2002. ISMAR 2002. Proceedings. International Symposium on (pp. 215-320). IEEE.

Gabbard, J. L., & Swan II, J. E. (2008). Usability engineering for augmented reality: Employing user-based studies to inform design. IEEE Transactions on visualization and computer graphics, 14(3), 513-525

Gigerenzer G. and Engel, C. eds. (2007). Heuristics and the Law, Cambridge, The MIT Press, ISBN 978-0-262-07275-5

Gong, J., & Tarasewich, P. (2004, November). Guidelines for handheld mobile device interface design. In Proceedings of DSI 2004 Annual Meeting (pp. 3751-3756).

Groner, R.Groner M., & Bischof, W. F. (1983). Methods of heuristics. Hillsdale N.J., Lawrence Erlbaum.

John, B. E., & Marks, S. J. (1997). Tracking the effectiveness of usability evaluation methods. Behaviour & Information Technology, 16(4-5), 188-202.

Kalalahti, J. (2015). Developing usability heuristic for augmented reality applications. Masters Thesis.

Kim, H. J., Kim, M. H., Chio, J. K., & Ji, Y. G. (2008). A study of evaluation framework for tangible user interface. Proceedings of the 2nd International Conference on Applied Human Factors and Ergonomics AHFE2008. July 14–17, 2008, Las Vegas, USA.

Kline, P. (2000). The handbook of psychological testing (2nd ed.). London: Routledge, page 13

Ko, S. M., Chang, W. S., & Ji, Y. G. (2013). Usability principles for augmented reality applications in a smartphone environment. International Journal of Human-Computer Interaction, 29(8), 501-515.

Kourouthanassi, P. E., Boletis, C., & Lekakos, G. (2015). Demystifying the design of mobile augmented reality applications. Multimedia Tools and Applications, 74(3), 1045-1066.

Krevelen D.W.F, Poelman R (2010) A survey of augmented reality technologies, applications and limitations. Int J Virtual Reality 9(2):1–20

Mack, R. L., & Nielsen, J. (Eds.). (1994). Usability inspection methods (pp. 1-414). New York, NY: Wiley & Sons.

Mölich, R., & Nielsen, J. (1990). Improving a human-computer dialogue. Communications of the ACM, 33(3), 338-348.

Moore, A. (2006). A tangible augmented reality interface to tiled street maps and its usability testing (pp. 511-528). Springer Berlin Heidelberg.

Nielsen, J. (2004, April). Enhancing the explanatory power of usability heuristics. In Proceedings of the SIGCHI conference on Human Factors in Computing Systems (pp. 152-158). ACM.

Nordam, A. (2016, Nov 15) The fuzzy future of virtual reality and augmented reality. http://spectrum.ieee.org/tech-talk/consumer-electronics/gadgets/can-you-see-it-the-future-of-virtual-and-augmented-reality. Retrieved 2017 March 9.

Pinelle, D., Wong, N., & Stach, T. (2008, April). Heuristic evaluation for games: usability principles for video game design. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1453-1462). ACM.

Schmorrow, D. (Ed.). (2005). Foundations of augmented cognition. Hillsdale, NJ: Lawrence Erlbaum.

Qian, X., Yang, Y., & Gong, Y. (2011, December). The art of metaphor: a method for interface design based on mental models. In Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry (pp. 171-178). ACM.

Webb, A. K., Vincent, E. C., Patnaik, P., & Schwartz, J. L. (2016, July). A Systems Approach for Augmented Reality Design. In International Conference on Augmented Cognition (pp. 382-389). Springer International Publishing.