2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct) | 979-8-3503-2891-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/ISMAR-Adjunct60411.2023.00100

The UUXR-Framework: A Draft Classification for Using Extended Reality in Usability and User Experience Research

Danny Schott
University of Magdeburg

Julia Moritz
USE-Ing. GmbH

Christian Hansen*
University of Magdeburg

Fabian Joeres†
University of Magdeburg
USE-Ing. GmbH

ABSTRACT

Conducting human-centered evaluations in extended reality (XR) environments is a growing trend in user research and usability engineering. However, there has been little to no systematic investigation of the emerging methods in this field published to date. The motivation behind our work is to explore and classify strategies and methods for utilizing XR technologies in the context of usability and user experience (UUX) activities. This paper proposes a draft classification framework for the use of XR technologies in UUX activities, combining an informal exploration of relevant literature with established UUX methods. Within this framework, we propose 12 dimensions that we consider potentially relevant for determining whether and how the use of XR technologies can benefit product development and user research. To evaluate the structure and phrasing of our proposed dimensions, we conducted an initial evaluation with UUX professionals (N = 11). We believe that our dimensions form an early-stage foundation for future guidelines aimed at UUX researchers. The framework serves as a tool for assessing different levels of virtualization in UUX work and facilitating knowledge transfer between academia and industry.

Index Terms: Human-centered computing—HCI theory, concepts and models; Human-centered computing—Mixed / augmented reality; Human-centered computing—Virtual reality

1 INTRODUCTION

With the increasing growth and improvements in Augmented Reality (AR), Virtual Reality (VR), and Mixed Reality (MR) technologies, promising UUX studies can be conducted in immersive environments for human-computer interfaces research. Compared to real-world conditions, conducting empirical, human-centered studies with XR technologies offers many advantages, as time and costs can be saved, especially in product development. Thus, studies in virtual replicas enable earlier and more efficient evaluation of real systems and can, in some applications, reduce recruitment efforts while increasing subject diversity through remote execution [20]. Empirical research using XR technology covers a wide range of use cases in science and industry [26]. These include the evaluation of physical interfaces in the automotive industry [10, 12, 23], ergonomic assessments of industrial workplaces and maintenance activities [2, 7, 9], analysis of assembly and manufacturing tasks [1, 25, 30], early insights into use problems in risk-related scenarios [24], studies of locomotion behavior in public spaces [18], immersive simulations in aerospace [4, 11], and the study of consumer products in XR environments [6].

This breadth of application areas and product industries demonstrates the value that XR will bring to the UUX sector over the coming years. The current technological barriers to immersive technologies are shrinking and the differences to the real world for the user are becoming smaller and smaller. This has an effect on the use of XR technologies in user research [17].

A classification approach based on a systematic literature review of how MR can be used to support prototyping is presented by Kent et al. [14]. The result is the characterization of five evidence-based value dimensions that show implications for science and industry. In the area of designing immersive virtual environments (VE) for industrial purposes, Hoecherl et al. [13] compile factors that influence overall quality and user experience, as well as tools to quantify this quality. The methodological approach includes analysis of use cases and selection criteria for technology in the design of VE. Saeghe et al. [27] identified design dimensions for AR Television by analyzing existing prototypes to understand the impact of design decisions on viewer experiences. An evaluation framework for Building Information Modeling-based VR applications in the architecture, engineering, and construction industry was presented by Kim et al. [15]. There are attempts to categorize and shed light on the importance of user-centered evaluation and the design decisions to be made in the context of virtual environments [19, 29, 31]. In this context, Barricelli et al. [3] pursue a promising approach by introducing a framework for the classification of VR applications for testing usability and UX in the VR field.

Considering the works described above, it can be observed that there is currently a lack of a systematic approach for the use of XR technologies in UUX studies in general. With this work, we propose a draft classification framework that builds towards a systematic foundation of XR technology use in UUX work (UUXR). We strive for a design that links findings from the scientific literature with the expertise of UUX in order to promote a mutual knowledge transfer between academia and industry in the long term [8].

We aim for a systematic development process that includes three key steps: In the first step, presented in this paper, we conceive the basic design of our classification framework. This was an exploratory phase in which we formulated the initial taxonomy in an internal workshop between three of the authors (DS, JM, FJ), based on our knowledge of the field, established UUX methods, and the relevant literature. The draft framework was subjected to an initial evaluation by a group of UUX professionals to assess consistency and comprehensibility of the framework wording. The second step is to foster collaboration with the broader scientific community. By sharing our research, we aim to engage other researchers and practitioners and stimulate discussion, insights, and exchange of ideas. This interaction serves as an iterative process in which feedback and discussion contribute to the progressive refinement of the nuances of our framework. Building on the foundation laid in Steps 1 and 2, we set our sights on a systematic literature review in the subsequent step, conducted in concert with the evolving taxonomy.

The development journey ends with the synthesis of our efforts into a systematically constructed taxonomy that provides a holistic and organized explanation of strategies and methodologies that underlie the role of XR technology in UUX activities. This taxonomy can serve as a navigational aid for researchers, practitioners, and academics, guiding them to make informed decisions and effectively implement XR technologies in their respective areas of activity.

*e-mail: hansen@isg.cs.uni-magdeburg.de
†e-mail: fabian.joeres@use-ing.de

DOI 10.1109/ISMAR-Adjunct60411.2023.00100

2 FRAMEWORK DEVELOPMENT

The initial motivation driving this work was to explore strategies that enable the effective utilization of XR technologies in human-centered evaluations conducted within a usability lab setting at our own university. To achieve this objective, we attempted to conduct a review of existing literature, aiming to gain an understanding of user research evaluation methodologies utilizing XR technologies. The first and last author independently conducted an informal literature review of the current state of scientific research and published industrial applications, resulting in the identification of 64 relevant publications. We started analyzing these publications, focusing on the employed use cases, technologies, and evaluation methods. We saw the need for a systematic framework to classify the existing literature and applications, and aimed to develop a systematic framework that can serve as a valuable resource to researchers and practitioners in the field of UUX investigations using XR technologies. Based on the case studies identified in the literature, and our own expertise in usability engineering, UX design, and product development, we have formulated the initial version of the framework presented in this article. We envision HCI experts with UUX knowledge who are engaged in evaluation planning as the intended user group of our framework. It aims to serve as a tool for evaluating UUX aspects of projects during product development, offering guidance in classifying the object of study, defining research goals, and providing recommendations for appropriate virtualization. We propose a framework with 12 dimensions of UUX projects and believe that these 12 dimensions are potentially relevant in the decision about whether and how the use of XR technology can be beneficial. These dimensions concern three main categories, including: the UUX project circumstances, the user task under investigation, and the interface under investigation. As our framework aims to serve as a future aid to decision making as much as a foundation for an upcoming literature review on the matter, we have added a fourth category covering the resulting virtualization solutions (i.e., the use of XR technology in existing projects). This fourth category will eventually aid framework users in formalizing the extent of XR technology utilization in their UUX work. For some dimensions, a project can be assigned to exactly one rating option, whereas, for other dimensions, multiple options may apply to the same project, which is highlighted in the corresponding tables listed below. The four categories are defined and explained as follows:

1. **Project goals:** *Dimensions that describe the project context in which the UUX research is conducted and the resulting research objectives.* The product maturity (in product development) and the overall research goals (in generalized, non-product specific research) determine the objectives and resources that shape a given UUX activity. These, in turn, determine the appropriate methodology. Since the use of XR technologies represents a series of methodological decisions, we believe that the research objectives and their determining factors should be considered in the framework. (See Table 1)

2. **Task:** *Dimensions describing the task that is represented or simulated in the UUX study; i.e., they do not describe the study task that subjects were given, but the real task, for which the interface/device is intended.* Understanding the task that is to be performed with the intended product or in the investigated scenario is necessary to assess whether, how, and which aspects of it can be simulated in different XR environments. Therefore, the *Task* dimensions review the complexity, predictability, and environmental dependence of the task. (See table 2)

3. **Interface:** *Dimensions describing the task-relevant user interface under development; i.e., they do not describe the prototypical interface used in the UUX study, but the envisioned interface for the real task.* Reviewing the user interface is intended to help framework users assessing the simulation effort and potential benefit of simulating the interface in an XR environment. The dimensions aim to help understand the size, location and interaction means of the interactive system. This is intended to aid framework users in reflecting on which interaction aspects can and should be beneficially simulated in XR environments. (See table 4)

4. **Virtualization:** *Methods of device, interface, and task virtualization in the UUX study.* This category serves the classification of existing XR solutions for UUX work and of framework users' own ideas, concepts, and study setups. (See table 5)

3 EVALUATION

This study intended to quantitatively evaluate the structure and phrasing of the framework dimensions we have proposed. Primarily, we wanted to investigate if the dimensions are understood consistently by different members of the UUX community. This common interpretation is an important foundation for further work (including the planned literature review) that builds on the framework. We approached this objective by selecting three published UUX research projects that used some degree of XR technology. Subsequently, these publications were rated by a sample of 11 domain experts in a within-subject design. Krippendorff's Alpha [16] was calculated to measure the inter-rater reliability for each dimension.

3.1 Literature selection

The publications were selected using six criteria: (1) Product orientation: we only selected publications that reported UUX research for a given product or prototype, rather than methodological research papers that aimed to investigate UUX methods as such (somewhat limiting the *research intention* dimension in the *project methods* category); (2) Recency: To capture recent advances and trends in the rapidly evolving XR field, publications older than five years at the time of the study were excluded; (3) Length: the length of the article should be proportionate to the processing time of our study participants and still sufficiently illuminate the content of the respective evaluation; (4) MR continuum: the selected articles should represent different stages of the MR continuum [21], if possible; (5) Variation: The studies should represent a variation of products/projects so that different industries were covered; (6) Accessibility: Because we needed to distribute the articles to our participants, we selected only publications that were either open access or for which we could obtain a pre-print with the authors' permission of distribution. Following these criteria, we selected the following three publications: Cheiran et al. [5] investigate a comparison between a physical desktop workstation and an exact copy of the real interface in the form of a VR mockup. The task involves a remote control station for an industrial robot operating out of direct view, where the user should monitor and navigate the robot. A concept for integrating GUI prototypes into an AR visualization was presented by Morozova et al. [22]. A 3D rendering of an automated coffee machine was displayed using Microsoft HoloLens around a mobile device that was on a tripod. Schrom-Feiertag et al. [28] tested mobile applications in AR environments by evaluating a ride-sharing application on a cell phone in a disrupted traffic scenario.

3.2 Sample design

We recruited UUX professionals with the inclusion criterion of having self-reported first-hand experience with planning and/or conducting UUX research. Participants volunteered their time without payment or other direct compensation. Due to the considerable time effort required of the participants (approximately three hours) and the associated recruitment challenge, we did not set a fixed sample size. The recruitment was conducted via the authors' professional

Table 1: Overview of the category *"Project goals"* including respective dimensions and detailed descriptions. * = Multiple rating option may apply.

Dimension	Description
Research intention	Primary intention of the reported research activities.
Method comparison	The study / work compares different methods of virtualization (e.g., mixed reality vs reality baseline).
Product development	The study reports only one virtualisation method and focuses on the product under development.
Prototype state	Which stage / phase is the tested prototype in?
Requirement research / no prototype yet	Investigating the user requirements at the early design or development stage.
Look and feel prototype	Prototype in early (design) stage to gain visual and/or haptic impressions.
Functional prototype	Prototype with some functionality, but not yet final design.
Final design prototype	Final design of fully functional prototype as intended for release.
Objectives*	Research goal(s) of the project / study.
Context research	Investigating the context of use (e.g. use environment, user groups, etc.).
Issue detection	Identification of interaction errors and difficulties, pain points.
AB test	Comparison of design variants.
Performance prediction	Assessment of task-specific performance parameters of the tested device (e.g. efficacy, efficiency, joy of use).
UX aspects*	UX dimension(s) under investigation.
Joy of use	Elicitation of hedonic user experience.
Cognitive ergonomics	Elicitation of cognitive ease of use and use obstacles.
Physical ergonomics	Elicitation of physical ease of use and use obstacles.

networks through social media posts and direct contact. Twenty-four (24) persons agreed to receive the study materials. Due to the required time effort, 13 persons were unable to complete the questionnaires, leaving 11 data sets for analysis.

3.3 Questionnaire design

The framework was formatted as a digital PDF questionnaire to enable the participants to rate each publication. The order of the categories was altered in comparison to the above listing order: the *virtualization* category was listed at the top of the questionnaire. The overall PDF questionnaire included three copies of the framework: one for each publication. Participants were instructed to rate each dimension. The instruction included information on whether a given dimension was a single choice or a multiple choice dimension. Each category, dimension, and rating level were separately explained in the questionnaire.

A "not reported" (NR) option was added to each dimension for cases in which participants could not find the information they would need to rate a dimension in the respective publication. Finally, a "not applicable" (NA) option was added in case participants deemed a dimension inappropriate for classifying a given UUX study.

In addition, the questionnaire included a section for demographic information, in which participants were asked to provide their age, gender, professional background, and their experience in the UUX sector, as well as their familiarity with XR technology in general and its application in HCI investigations. Participants were instructed to maintain anonymity and confidentiality in their replies.

3.4 Procedure

After providing electronic written consent to participate, participants were sent a digital package including the publications (or pre-prints thereof) and the questionnaire file. They were asked to read and rate the publications in their own time and return the filled-in questionnaire after a defined period. The exact duration that was available to each participant varied, but all participants had at least one week to complete the questionnaire.

3.5 Data analysis

The ratings for the eight single-choice dimensions were directly transcribed into nominal scores. The multiple-choice ratings were converted to nominal values encoding the specific combination of options that each participant had ticked. In order to quantify inter-rater reliability, Krippendorff's Alpha was calculated separately for each of the 14 dimensions, using the *irrCAC* R library.

This was calculated using bespoke difference weight matrices. For the single-choice dimensions, each framework category level was treated as equidistant (weight = 1.0). The NR and NA ratings were treated as lying outside the regular classification rating spectrum and were, therefore, weighted with a distance of 0.5 to all other scores. The multiple-choice weighting matrix was based on the overlap between two rating scores. The NR and NA options were treated equivalently to the single-choice dimensions and weighted with a distance of 0.5 to all other ratings. Dimensions on which participants did not provide any rating were excluded from the analysis.

3.6 Results

Eleven (11) respondents participated in the questionnaire study (six female and five male). Participant ages ranged between 25 and 43 years (median = 33 years). Self-reported professional experience in the UUX sector ranged between 1.5 and 13 years (median = 6 years) and participants reported to have worked on between 0 and 30 usability/UX studies (median = 5). The fields of activity included both hardware and software development, design, science, consulting, and management, spanning diverse sectors such as the automotive industry, biomedical engineering, healthcare, education, art, digital products, and enterprise solutions. The reported experience in their current professions ranged from 0 to 10 years (median = 4.5). Eight participants reported sporadic private usage of XR technology (such as AR filters, VR gaming, etc.) and/or professional utilization (AR/VR research, software development), while the remainder indicated a lack of practical experience. Five of the participants used XR technology in the context of user research, with a focus on evaluating the XR application itself. Three data points were excluded because participants did not provide any rating; one data point each for the *objectives*, *MR continuum*, and *virtualized content* dimensions. The resulting values for Krippendorff's Alpha are listed in Table 3. Due to the small sample size and low inter-rater reliability, we only report the descriptive results here.

4 Discussion & Limitations

In our evaluation, we made an exploratory attempt to validate our classifications in terms of structure and phrasing in a quantitative

Table 2: Overview of the category *"Task"* including respective dimensions and detailed descriptions. * = Multiple rating option may apply.

Dimension	**Description**
Task timeline	The type of timeframe in which the task is executed.
Continuous	Performing a continuous task without interruption.
Interrupted	The task is interrupted once or multiple times.
Sporadic	The task requires sporadic attention/execution between longer periods of downtime.
Task trigger	Stimuli that trigger the user to execute/continue the task.
Workflow based	The task is part of a set workflow.
Supervision	The task is triggered by set states / changes in a supervised system.
Trigger based	An external signal / stimulus triggers the task execution.
Intrinsic	The user initiates the task when they intrinsically feel like it.
Environment relevance*	Influence that the use environment has on the task.
Environment not relevant	User´s surroundings do not affect task completion.
Environment information	Information from the environment is necessary or helpful to carry out the task.
Environment conditions	Completion of the task is influenced by conditions of the environment (e.g. lighting, noise, crowdness, . . .).
User independence	Dependence / Independence of the user from other (human or technical) agents in the environment.
Solo task	The user can complete the task alone.
Information exchange	The user depends on exchanging information with other agents.
Object exchange	The user depends on exchanging physical objects with other agents.
Collaboration	The user closely collaborates with other agents (info & object exchange).
Solution space complexity	Number and length / complexity of solution paths for completing the task.
Few paths, Short paths	The user has few different options (solution paths) on how to complete the task and all options are simple and short. For example, inflating your bicycle tire is a task that consists of few steps and has little space for process variation.
Few paths, Long paths	The user has few different options (solution paths) on how to complete the task and all options are complex or long. For example, getting an airplane ready for take-off is a task that includes many sub-tasks, steps, and even multiple agents, but it is highly checklist-driven and allows for limited process variation.
Many paths, Short paths	The user has many different options (solution paths) on how to complete the task and all options are simple and short. For example, arranging a meeting with your colleague is a task that can be achieved in multiple ways (e-mail, phone call, messenger app(s), walking over to their office, etc.) but each solution path is completed within a few steps.
Many paths, Long paths	The user has many different options (solution paths) on how to complete the task and all options are complex or long. For example, getting from your home in Munich to a meeting in Berlin is a task that can be achieved in multiple ways and each of these ways requires many sub-tasks (planning, booking, travelling), steps, and potentially multiple agents.

Table 3: Krippendorff's Alpha results for each of the 14 classification dimensions. RI: research intention; PS: prototype state; Ob: objectives; UXA: UX aspects; TTL: task timeline; TTR: task trigger; ER: environment relevance; UI: user independence; SSC: solution space complexity; DS: device size; EL: effect location; ID: interface digitalization; MRC: MR continuum; VC: virtualized content.

	Project goals				Task					Interface			Virtualization	
Dimensions	RI	PS	Ob	UXA	TTL	TTR	ER	UI	SSC	DS	EL	ID	MRC	VC
Krippendorff's α	0.47	-0.2	0.16	0.10	0.06	0.13	0.04	-0.03	0.13	0.32	0.15	0.13	0.47	0.05

approach. Eleven UUX experts rated a given set of three studies accoring to the proposed framework with 12 dimensions of UUX projects. Our results show that in-depth qualitative research is needed to understand how users interpret the dimensions and their level. Krippendorff recommends Alpha values of at least 0.8 for a rating system to be viewed as reliable [16]. The values achieved in our evaluation study, while only being descriptive approximations, lie considerably and consistently below this threshold. We see a number of potential reasons for these results.

The first possible reason may be that the classification criteria seem to be not clearly defined or poorly expressed, causing a validly inconsistent interpretation by UUX domain experts. While this may well be the case for more abstract concepts like the *solution space complexity*, it is somewhat surprising that the respondents disagreed even on the more tangible concepts, such as *device size*. This requires further investigation. For the dimension *user independence*, we identified an additional reason: the small number of rated studies led to a very low diversity between the studies. While there was high agreement between raters on the *solo task* rating, this rating was also consistent across all three studies. This very low diversity between the rated objects (i.e. the studies) reduces the diversity in observed rating levels and is likely to have caused the low Krippendorff's Alpha coefficient for that dimension. However, in other dimensions, the variation was larger, reducing this statistical artifact to the dimension *user independence*. Beyond valid interpretation discrepancies, there are potential methodological reasons for the high levels of disagreement. The evaluation of only three selected publications represents a limitation. Our decision was influenced by the pragmatic consideration of balancing research depth with participant time constraints.It is important to note that the three studies we examined were not reported in a uniform or standardized manner. This discrepancy made obtaining relevant information difficult. The effects of this difficulty appeared to vary among respondents. In addition, the project reports in these publications were tailored to a scientific audience and focused on specific research objectives rather than conforming to the format of traditional product development documentation. This distinction is critical because it affects the context and type of information presented.

The alternative approach of synthesizing project documents for classification, however, would heavily bias the study as the documents would be written with the classification answers in mind. The third option would be the use of authentic product development doc-

Table 4: Overview of the category *"Interface"* including respective dimensions and detailed descriptions. * = Multiple rating option may apply.

Dimension	Description
Device size	Physical dimensions of the tested device.
Handheld	Portable, small, lightweight and can be held with one hand (e.g. smartphone).
Tabletop	Portable, but mostly stationary (e.g. printer, washing machine).
Large-scale	Non-portable large device (e.g. vehicle in maintenance and repair).
Environment	User(s) can move around inside the tested system (e.g. Industrial plant, vehicle for the driver/pilot).
Effect location*	Location where user actions / manipulations take effect.
Peripersonal	Area of space that immediately surrounds the user.
Line of sight	Within direct sight of the user, but outside peripersonal space.
Remote	The effect happens remotely, the user receives feedback through sensors and technical means (e.g. telemanipulators, UAV pilots).
Digital	The user's actions do not have direct physical, only digital effects (e.g. database manipulation, e-mail sending, document editing).
Interaction digitalization*	Degree of interface digitalization.
Mechanical manipulation	Direct mechanical steering (e.g. mechanical lever, moving an object from a to b).
Physical/tangible UI	Tangible interface of physical elements (e.g. buttons, switches, dials).
Touchless UI	Physical contact is not necessary for control (e.g. gestural input, voice commands).
Graphical UI	Graphical, mostly 2-dimensional interface (e.g. touchscreen).
App on universal device	Dedicated application on a universal device (smartphone, laptop).

Table 5: Overview of the category *"Virtualization"* including respective dimensions and detailed descriptions. * = Multiple rating option may apply.

Dimension	Description
MR continuum	Mixed reality solution according to the MR continuum by Milgram & Kishino [21].
Reality	Environment consists solely of physical objects.
Augmented reality	Real world is augmented with digital elements.
Augmented virtuality	Virtual world is augmented by the inclusion of real or physical objects.
Virtual reality	Environment consists solely of digital elements/objects.
Virtualised content*	Content that is virtualized in the MR environment.
Interface	Interface or components of the device that is being tested.
Device	Entire device that is being tested.
Environment components	Elements of the use environment (e.g. surrounding objects, furniture, buildings).
Agents	Human or technical agents that are associated with the task.
Full environment	The entire use environment is a virtual scene.

umentation that commonly underlies confidentiality. We believe that the chosen route was suitable for an initial exploration of the framework. In-depth follow-up investigations of respondents' framework interpretation is required to identify shortcomings in the conceptual definition and phrasing.

Another possible cause could be the small sample size of our evaluation. However, we do not yet see any evidence that increasing the number of participants has an impact on the results. In addition, due to the study design, we were only able to provide three scientific publications, although a larger number and broader sample could have an impact on the interpretation of the dimensions.

The results imply that the current framework phrasing leaves some room for improvement to aid consistent understanding. Although we supplemented the descriptions of each dimension for complex and potentially unknown concepts such as *solution space complexity* with examples, there was some ambiguity or misinterpretation among the participating UUX experts. This may be due to the fact that some dimensions are very abstract and, therefore, need clearer phrasing or a taxonomy that is more closely aligned with conventional UUX standards. This demonstrates that further work is required to turn the proposed framework draft into a tool that can benefit UUX experts with different professional backgrounds such as engineers, designers, and researchers in industries as divers as aviation and consumer product development, that were identified in section 1.

5 Conclusion and Ongoing Work

This paper introduces a draft of a classification framework for the use of XR technology in human-centered evaluations. In doing so, we combined findings from the emerging scientific research field for the use of XR as a tool for human-centered investigations with usability and UX concepts, proposing twelve dimensions. As an immediate next step, we want to conduct qualitative interviews with UUX professionals to improve the dimension phrasing and descriptions, as well as domain experts' perception of the dimensions' relevance. Moving forward, we plan a comprehensive literature review to determine the state of the art in conducting UUX investigations using XR technologies and to create a literature database. This process will be carried out in conjunction with the framework as a means of categorization and decision support, with the verification of the exclusivity and orthogonality of the individual dimensions being part of it. Our goal is to develop a framework that allows HCI researchers and practitioners to locate their individual project proposals in the landscape of existing studies. This can be used to derive recommendations on the degree of virtualization and to identify similar reference projects from the literature. Despite inconsistency in ratings, we believe that the twelve dimensions can serve as a starting point for developing a framework to guide any type of UUX investigation - even outside the XR context - and thus help UUX researchers make decisions when analyzing their project proposals and conducting studies. We welcome broad feedback from the XR community to help us turn our vision of virtuality into reality.

ACKNOWLEDGMENTS

This project received funding from the Federal Ministry of Education and Research as part of the Magdeburg Research Campus *STIMULATE* program (grant no. 13GW0473A), and from the German Federal Ministry for Economic Affairs and Climate Action (grant no. 16KN093943).

REFERENCES

[1] P. Agethen, V. S. Sekar, F. Gaisbauer, T. Pfeiffer, M. Otto, and E. Rukzio. Behavior analysis of human locomotion in the real world and virtual reality for the manufacturing industry. *ACM Transactions on Applied Perception (TAP)*, 15(3):1–19, 2018.

[2] S. Aromaa, V. Goriachev, and T. Kymäläinen. Virtual prototyping in the design of see-through features in mobile machinery. *Virtual Reality*, 24(1):23–37, 2020.

[3] B. R. Barricelli, A. De Bonis, S. Di Gaetano, S. Valtolina, et al. Semiotic framework for virtual reality usability and ux evaluation: a pilot study. In *GHItaly@ AVI*, 2018.

[4] F. Bernard, M. Zare, J.-C. Sagot, and R. Paquin. Using digital and physical simulation to focus on human factors and ergonomics in aviation maintainability. *Human factors*, 62(1):37–54, 2020.

[5] J. F. P. Cheiran, L. A. Torres, A. A. S. da Silva, G. A. de Souza, L. P. Nedel, A. Maciel, and D. A. C. Barone. Comparing physical and immersive vr prototypes for evaluation of an industrial system user interface. In N. Magnenat-Thalmann, C. Stephanidis, E. Wu, D. Thalmann, B. Sheng, J. Kim, G. Papagiannakis, and M. Gavrilova, eds., *Advances in Computer Graphics*, pp. 3–15. Springer International Publishing, Cham, 2020.

[6] Y. M. Choi, S. Mittal, et al. Exploring benefits of using augmented reality for usability testing. In *DS 80-4 Proceedings of the 20th International Conference on Engineering Design (ICED 15) Vol 4: Design for X, Design to X, Milan, Italy, 27-30.07. 15*, pp. 101–110, 2015.

[7] A. Colley, J. Väyrynen, and J. Häkkilä. Exploring the use of virtual environments in an industrial site design process. In *IFIP Conference on Human-Computer Interaction*, pp. 363–380. Springer, 2015.

[8] E. De Wit-de Vries, W. A. Dolfsma, H. J. van der Windt, and M. P. Gerkema. Knowledge transfer in university–industry research partnerships: a review. *The Journal of Technology Transfer*, 44(4):1236–1255, 2019.

[9] R. Dias Barkokebas and X. Li. Use of virtual reality to assess the ergonomic risk of industrialized construction tasks. *Journal of Construction Engineering and Management*, 147(3):04020183, 2021.

[10] F. Freitas, H. Oliveira, I. Winkler, and M. Gomes. Virtual reality on product usability testing: A systematic literature review. In *2020 22nd Symposium on Virtual and Augmented Reality (SVR)*, pp. 67–73, 2020. doi: 10.1109/SVR51698.2020.00025

[11] M. Guida and P. Leoncini. Regional aircraft interiors evaluation in a real time ray-traced immersive virtual environment. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*, pp. 483–498. Springer, 2021.

[12] M. Hoggenmüller, M. Tomitsch, L. Hespanhol, T. T. M. Tran, S. Worrall, and E. Nebot. Context-based interface prototyping: Understanding the effect of prototype representation on user feedback. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3411764.3445159

[13] J. Höcherl, A. Adam, T. Schlegl, and B. Wrede. Human-robot assembly: Methodical design and assessment of an immersive virtual environment for real-world significance. In *2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, vol. 1, pp. 549–556, 2020.

[14] L. Kent, C. Snider, J. Gopsill, and B. Hicks. Mixed reality in design prototyping: A systematic review. *Design Studies*, 77:101046, 2021. doi: 10.1016/j.destud.2021.101046

[15] J. I. Kim, S. Li, X. Chen, C. Keung, M. Suh, and T. W. Kim. Evaluation framework for bim-based vr applications in design phase. *Journal of Computational Design and Engineering*, 8(3):910–922, 2021.

[16] K. Krippendorff. *Content analysis: An introduction to its methodology. Thousand Oaks, California: Sage, 2004 978-0-7619-1545-4*. 2004.

[17] J. Lacoche, E. Villain, and A. Foulonneau. Evaluating usability and user experience of ar applications in vr simulation. *Frontiers in Virtual Reality*, 3, 2022. doi: 10.3389/frvir.2022.881318

[18] V. Mäkelä, R. Radiah, S. Alsherif, M. Khamis, C. Xiao, L. Borchert, A. Schmidt, and F. Alt. Virtual field studies: Conducting studies on public displays in virtual reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–15. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3313831.3376796

[19] D. Martens. Virtually usable: A review of virtual reality usability evaluation methods. *Recuperado de: https://danamartensmfadt. files. wordpress. com/2016/08/virtuallyusable. pdf*, 2016.

[20] F. Mathis, K. Vaniea, and M. Khamis. Replicueauth: Validating the use of a lab-based virtual reality setup for evaluating authentication systems. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3411764.3445478

[21] P. Milgram and F. Kishino. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12):1321–1329, 1994.

[22] A. Morozova, V. Rheinstädter, and D. Wallach. Mixedux: A mixed prototyping framework for usability testing in augmented reality. In *Companion Publication of the 2019 on Designing Interactive Systems Conference 2019 Companion*, DIS '19 Companion, p. 41–44. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3301019.3325146

[23] I. Pettersson, M. Karlsson, and F. T. Ghiurau. Virtually the same experience? learning from user experience evaluation of in-vehicle systems in vr and in the field. In *Proceedings of the 2019 on Designing Interactive Systems Conference*, DIS '19, p. 463–473. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3322276.3322288

[24] E. F. Picka, A. Vogel, M.-S. Roder, J. Birkle, D. Schrenk, J. Linnemann, J. Moritz, and S. Pfeffer. Virtual usability testing (virtuse) - development of a methodical approach for usability testing in vr. In C. Stephanidis, M. Antona, and S. Ntoa, eds., *HCI International 2022 Posters*, pp. 109–116. Springer International Publishing, Cham, 2022.

[25] C. Pontonnier, G. Dumont, A. Samani, P. Madeleine, and M. Badawi. Designing and evaluating a workstation in real and virtual environment: toward virtual reality based ergonomic design sessions. *Journal on Multimodal User Interfaces*, 8(2):199–208, 2014.

[26] P. A. Rauschnabel, R. Felix, C. Hinsch, H. Shahab, and F. Alt. What is xr? towards a framework for augmented and virtual reality. *Computers in Human Behavior*, 133:107289, 2022. doi: 10.1016/j.chb.2022.107289

[27] P. Saeghe, M. McGill, B. Weir, S. Clinch, and R. Stevens. Evaluating and updating a design space for augmented reality television. In *ACM International Conference on Interactive Media Experiences*, pp. 79–94, 2022.

[28] H. Schrom-Feiertag, G. Regal, J. Puthenkalam, and S. Suette. Immersive experience prototyping: Using mixed reality to integrate real devices in virtual simulated contexts to prototype experiences with mobile apps. In *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 75–81, 2021. doi: 10.1109/ISMAR-Adjunct54149.2021.00025

[29] V. Souza, A. Maciel, L. Nedel, and R. Kopper. Measuring presence in virtual environments: A survey. *ACM Computing Surveys (CSUR)*, 54(8):1–37, 2021.

[30] M. Weiß, K. Angerbauer, A. Voit, M. Schwarzl, M. Sedlmair, and S. Mayer. Revisited: Comparison of empirical methods to evaluate visualizations supporting crafting and assembly purposes. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1204–1213, 2020.

[31] T. Zhang, R. Booth, R. Jean-Louis, R. Chan, A. Yeung, D. Gratzer, G. Strudwick, et al. A primer on usability assessment approaches for health-related applications of virtual reality. *JMIR serious games*, 8(4):e18153, 2020.