UBIACTION 2020
Edited by Florian Lang, Pascal Knierim, Jakob Karolus, Fiona Draxler, Ville Mäkelä, Francesco Chiossi, Luke Haliburton, Matthias Hoppe, Albrecht Schmidt

UBIACTION 2020

5th Seminar on Ubiquitous Interaction
January 27, 2020, Munich, Germany

Edited by
Florian Lang
Pascal Knierim
Albrecht Schmidt

UBIACTION 2020 – Vol. 5

Editors

Florian Lang, Pascal Knierim, Jakob Karolus, Fiona Draxler, Ville Mäkelä,
Francesco Chiossi, Luke Haliburton, Matthias Hoppe, Albrecht Schmidt

Human-Centered Ubiquitous Media Group
Institut für Informatik
Ludwig-Maxmilians-Universität München
`hsuc@um.ifi.lmu.de`

ACM Classification 1998 - H.5 INFORMATION INTERFACES AND PRESENTATION

ISBN-13: 979-8635125052

Publication date
27. January 2020

License:

This work is licensed under a Creative Commons Attribution 3.0 Unported license (CC-BY 3.0): `http://creativecommons.org/licenses/by/3.0/legalcode`. In brief, this license authorizes each and everybody to share (to copy, distribute and transmit) the work under the following conditions, without impairing or restricting the authors' moral rights:

- Attribution: The work must be attributed to its authors.

The copyright is retained by the corresponding authors.

Vorwort

In der heutigen Gesellschaft ist die Allgegenwärtigkeit von Computern nahezu unabdingbar und die Digitalisierung hat massive Auswirkungen auf die verschiedenen Seiten unseres Lebens und Arbeitens. So wandelt sich die Interaktion mit unserer Umgebung zunehmend von der Mensch-Mensch-Interaktion zur Mensch-Computer-Interaktion und Kommunikation mit anderen Menschen findet zunehmend durch digitale Technologien statt. Dieser Wandel eröffnet vor Allem in der Forschung die Tore zu neuen Möglichkeiten der Interaktion zwischen Menschen und Computern. Das als „Ubiquitous Computing“ bezeichnete Forschungsgebiet untersucht unter anderem die Zugänglichkeit und Einfachheit der Interaktion mit mobilen und tragbaren Geräten und deren Einfluss auf die Erfahrungen des Menschen. Am Lehrstuhl „Human-Centered Ubiquitous Media“ der Ludwig-Maximilians-Universität München forschen Studierende kontinuierlich in diesem Gebiet. Im Rahmen des Seminars „Menschenzentrierte Interaktion mit ubiquitären Computersystemen“ fertigen die Studierende wissenschaftliche Arbeiten, um einen Einblick in zukünftige Möglichkeiten der Mensch-Computer-Interaktion zu gewähren. Auf der Veranstaltung „Ubiaction“ wurden am 27.01.2020 die Forschungsergebnisse des Seminars in Form von 15-minütigen Vorträgen präsentiert. Zudem wurden auf dieser Veranstaltung die Ergebnisse der Praktika „Web Programmierung (PWP)“ und „Entwicklung von Mediensystemen 2 (PEM2)“ vorgestellt. Die Veranstaltung wurde von Herr Prof. Dr. Albrecht Schmidt eröffnet und geleitet. Die schriftlichen Ausarbeitungen der Studierenden aus dem Wintersemester 2019 wurden in diesem Buch als fünfter Band der Serie „Ubiaction“ zusammengeführt.

Inhaltsverzeichnis

Editors: Florian Lang, Pascal Knierim, Jakob Karolus, Francesco Chiossi, Matthias Hoppe, Fiona Draxler, Luke Haliburton, Ville Mäkelä, Albrecht Schmidt. January 27, 2020. Munich, Germany.

Teilnehmer

Daniel Onumbu

d.onumbu@campus.lmu.de
Mensch-Computer-Interaktion

Simon von der Au

simonanian.Au@campus.lmu.de
Medieninformatik

Lisa Gärttner

lisa.gaerttner@campus.lmu.de
Mensch-Computer-Interaktion

Jan Paolo Holinksi

p.holinski@campus.lmu.de
Mensch-Computer-Interaktion

Vanessa Sarakiotis

v.sarakiotis@campus.lmu.de
Medieninformatik

Anke Pellhammer

anke.pellhammer@campus.lmu.de
Mensch-Computer-Interaktion

Sybil Bast

sybil.bast@campus.lmu.de
Mensch-Computer-Interaktion

Alexander Hiesinger

alexander.hiesinger@campus.lmu.de
Mensch-Computer-Interaktion

Julian Lorenz

julian.lorenz@campus.lmu.de
Mensch-Computer-Interaktion

Julian Preissing

julian.preissing@campus.lmu.de
Mensch-Computer-Interaktion

David Klein

d.klein@campus.lmu.de
Mensch-Computer-Interaktion

Overview on the current techniques to display contextual information with adaptive AR

Daniel Onumbu

Ludwig-Maximilians-Universität München, München, Deutschland
`d.onumbu@campus.lmu.de`

Abstract

This paper provides an overview of the current state of research on contextual information access using adaptive augmented reality. First, a basis for adaptive augmented reality and contextual information access is provided. Afterwards, different research projects are presented that use several techniques such as gaze tracking, speech detection and global positioning system (GPS) tracking to obtain information about the environment. Based on these techniques, the augmented reality content adapts to display relevant contextual information. Finally, the current state of research is determined and research gaps are revealed.

2012 ACM Computing Classification Human-centered computing → Mixed / augmented reality

Keywords and phrases Context-Aware Augmented Reality; Contextual Information Augmented Reality; Adaptive Augmented Reality; Context-Awareness.

1 Introduction

Imagine a tool that can support you in any situation. Something that recognizes what situation you are currently in and what information could help you best at the moment. A tool that shows the way to the products on your shopping list in the supermarket, and can also be a help to build furniture by illustrating the individual steps of the construction process. With the help of Augmented Reality (AR) similar applications have already been developed [28]. Currently, such an application would simply be able to project the assembly instructions, but what if the application could react to influences from the environment? For example, a camera could take a picture of a pile of screws and the application would highlight the ones necessary for the respective step. The user could then easily place the screws in the right place. To do this, the application must understand the context in which it is used and what support is most useful.

© Daniel Onumbu;
licensed under Creative Commons License CC-BY

Cite as: Daniel Onumbu. Overview on the current techniques to display contextual information with adaptive AR. In *5th Seminar on Ubiquitous Interaction (UBIACTION 2020)*. Editors: Florian Lang, Pascal Knierim, Jakob Karolus, Fiona Draxler, Ville Mäkelä, Francesco Chiossi, Luke Haliburton, Matthias Hoppe, Albrecht Schmidt. January 27, 2020. Munich, Germany. pp. 1:1–1:16.

The following is a literature analysis of different techniques for adaptive AR applications, which try to adapt to external influences by means of contextual information access. First, a theoretical basis is provided. Afterwards, this interdisciplinary topic is reviewed, first considering research on adaptive AR applications and possibilities of contextual information access. Then the current state of research on contextual information access with adaptive AR is analysed. Finally, it should become clear which research gaps still exist and in which fields much has already been developed.

2 Theoretical foundation

The following chapter will give a theoretical basis for the further course of the paper. A short overview and definitions of the terms augmented reality (AR) and contextual information will be provided.

2.1 Adaptive AR

To define adaptive AR, first AR and adaptive interfaces are defined. With these definitions as a foundation, adaptive AR will be defined afterwards. Augmented Reality enhances a user's perception of and interaction with the real world. Mekni and Lemieux define AR as systems that have the following characteristics: firstly combines real and virtual, secondly interactive in real time, and lastly registered in 3D. With this definition they aim to not restrict AR to head-mounted-displays (HMDs) as many researchers do and try to allow mobile technologies and many more [25]. Many other and older definitions mainly refer to the use of HMDs. These have become obsolete due to the technological development in AR [38]. These advances and the new mobility due to the use of handheld displays and others make the deployment of adaptive AR possible in the first place [7]. The central technology of this paper will be outlined in the next section. There are also further approaches to AR. Thus Azuma et. al. define AR technology independently. In their view, an AR system must firstly combine real and virtual objects in a real environment, secondly run interactively and in real time and lastly register real and virtual objects with each other. They distance themselves from the fact that AR is limited to the visual sense, as the technology is normally applied, and propose that other senses such as hearing and smelling can also be supported [4]. However, one premise that AR must fulfil is mentioned by all authors. The environment must be real. Otherwise, we no longer speak of AR, but of augmented virtuality (AV) and others (see Figure 1). For the distinction between AR and AV, Klopfer has introduced a spectrum within the AR. A light AR describes an environment in which the user accesses only little virtual information. A heavy AR environment, on the other hand, exposes the user almost permanently to virtual input. This is mainly implemented by HMDs [22].

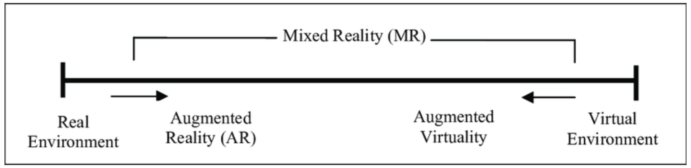


Figure 1 Milgram Reality-Virtuality Continuum [26].

Mekni and Lemieux divide the AR applications into at least twelve categories. These include military, medical and entertainment applications. Examples include displaying real battlefield scenes and augmenting them with information [35], displaying correctly rendered image of the ultrasound of a fetus overlaid on the abdomen of the pregnant woman [5] or increasing the visibility of important game aspects in sports like the line of scrimmage in a football field [25]. In this paper AR will be seen in the way Azuma et. al. defined it combining real and virtual objects in a real environment, running interactively and in real time and registering real and virtual objects with each other [25].

Next, adaptive interfaces should be defined. Langley defines them as a software artifact that improves its ability to interact with an user by constructing a user model based on partial experience with that user [23]. Therefore the interfaces should memorize the interaction of a user and adapt its content accordingly. According to Benyon and Murray an adaptive interface consists of the models, the user-, task- and interaction model. Using these models, the interface is given the ability to adapt to certain situations [6].

After defining AR and adaptive interfaces we can finally take a look at adaptive AR, the combination of AR and adaptive interfaces. For Damala et al. it is essential that the human being is once again at the centre of the development process. Thus they define adaptive AR as an immersive application which is continuously monitoring the visual, acoustic, and affective context within which the experience takes place. Thus every user will have a unique, personalised experience compared to the unpersonalised standard applications. These applications can be implemented with both HMDs and hand-held displays. As Azuma et al. already describe, more than just the sense of sight must be supported. There are possibilities, for example, to augment hearing in an adaptive AR environment [10]. Ghouaiel et. al. describe a concept of plasticity whereby AR scenes adapt to the context and accordingly become adaptive AR. The concept is based on the elasticity of materials, which adapts to different temperatures without breaking the material. In the same way, the AR scene should be able to adapt to different contexts and remain functional at all times. In Human-Computer-Interaction (HCI) Plasticity describes a context shaped by the environment, the platform and the user. For Plasticity in adaptive AR, however, this definition is only broken down to the environment. How context is exactly defined will be described in the next chapter [16].

Basically, adaptive AR is a niche topic of AR, so there are not many definition approaches yet. In this paper adaptive AR describes the adaptation of AR scenes to the context under which the application is influenced. Normally, this is done by evaluating the user's interest and environment and adapting the scenes accordingly. For the former, technologies such as gaze tracking and speech detection are used. The description of contextual information clearly shows that the evaluation of the environment is much more difficult because it is necessary to define which parameters are considered important [16]. Various approaches are described in the literature analysis.

2.2 Contextual information

After describing what adaptive AR is and that the influence of context is of great importance in such applications, it has to be clarified what exactly is meant by context. First, the term context is defined with regard to AR. Then it is described which possibilities there are to generate context, among other things with the help of the Smartphones, which nearly everyone carries with them. Finally, it is described which kinds of context are useful for adaptive AR applications.

Previous definitions of context are often based on synonyms such as situation or environment. These, as well as definitions based on examples, are difficult to implement in this case. They do not contain every piece of information that could be useful for the application [14]. Thus Schilit and Theimer define context as place, identities of people in the environment or objects as well as changes to them [31]. Dey and Abowd, on the other hand, define context as any information that can be used to describe a situation of an entity. An entity in these cases can be a person, a place or an object that is relevant for the interaction between the user and the application. This also includes the user and the application itself. This definition is not overly specific. In short, if the information can be used to characterize the situations of a participant in an interaction, it is context [1]. Theoretically, this information can also be provided by the users. On the one hand this can overload the users and on the other hand they are not necessarily aware which information is relevant in each particular situation. The definition of Dey and Abowd allows both cases. Information that is explicitly entered by the user as well as information that is implicitly recognised by the application. Accordingly, a profile can always be generated to describe the environment [1]. This can be explained as follows as an example. If considering, for example, the running speed of a hobby runner. Whether the runner runs in a red coloured or blue coloured shirt should not influence his speed. Accordingly, this is not context. The weather, on the other hand, has an influence on the runner's performance, so the weather information is context.

Important at this point is another definition from Dey. All adaptive AR applications are at least context-aware. Thus they can adapt their support for the user to the context. The use of context to provide useful information to the user is called context-awareness in this sense. Dey speaks of three categories that a context-aware application should support. First presentation of information, second automatic execution of a service, and finally tagging of context to information to support later retrieval [14]. Due to today's technology, there is a lot of information that can be guaranteed by the smartphone, for example, to describe the context. This includes location, identity, activity, and time. Dey attaches particular importance to these [1]. Nevertheless, the analysis must clarify whether this information is sufficient for adaptive AR applications or whether smartphones can guarantee even more important information or whether even essential information has to be provided by other devices and so forth. On the basis of the definitions presented and other fundamentals, the literature analysis describes previous approaches and techniques for generating contextual information and making it available in adaptive AR applications.

3 Literature analysis

The following presents the current state of the literature regarding adaptive AR and contextual information. This has to be done before finally considering to what extent the interdisciplinary topic contextual information access with adaptive AR has already been explored. Thus it will become clear whether there is a need for further research on one of the basic topics or whether the combination of both disciplines needs to be deepened.

3.1 Adaptive AR

As described earlier, there are several techniques, such as gaze tracking and speech detection, that are used to grant AR an adaptive character. The following two applications serve to illustrate the gaze technologies and as a basis for the subsequent adaptive AR applications. Shahid et al. have developed an assistance system for drivers, which uses gaze tracking to control the driver's attention to the traffic. The approaching route is projected onto the windscreen by AR. Face recognition enables detailed gaze tracking and the projections are adjusted according to the gaze. If the system notices that the driver is, for example, turning away from the road, it alerts him [32]. Here Shahid et al. use neither a handheld device nor an HMD, in comparison Renner and Pfeiffer use smartglasses. They use AR to inform the user about objects that are outside their field of view. Eye-tracking is used to determine the gaze and accordingly it can be evaluated where the object is located relative to the user [27]. Similar types of attention evaluation are used in other projects, e.g. by Toyama et al., to identify what the user is focusing on and to design the AR scenes accordingly. Their system recognizes whether the user is focusing on the AR content or on the natural environment. As shown in Figure 2, it evaluates whether the focus is on the AR content. If so, the content is adjusted according to the user behavior. If not, the display is dimmed [34]. This technology can be used in applications that are already being discussed, such as museum visits [10]. The AR labeling technique described by Sinclair and Martinez could also be useful here, which makes it possible to expand existing AR systems by integrating individual labels into the scenes [33].

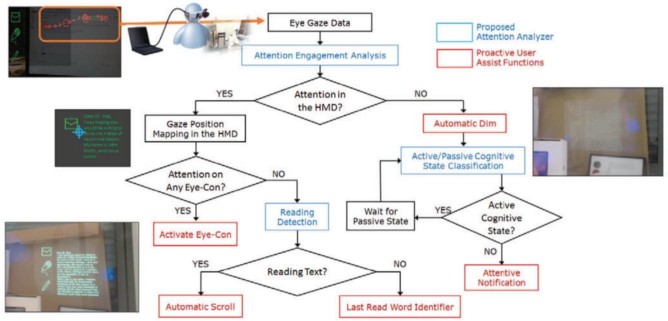


Figure 2 Flowchart of the evalution system of Toyama et al [34].

Besides gaze tracking, the location can also be used as input to make AR applications adaptive. In their Multi-Object Oriented Augmented Reality (MOOAR) system for mobile learning, Chang and Tan use global positioning system (GPS) data from the iPhone and its orientation to find out the exact location of the user. According to the orientation of the smartphone, the AR Scene displays different content for learning [8]. The part on adaptive AR can be completed with the project by Damala et al., in which gaze, speech and location are used as input. With their ARtSENSE project they personalize museum visits by providing visual and acoustic content based on the above mentioned input parameters. Figure 3 shows the model implemented the ARtSENSE project. This illustrates how the current AR scene adapts to the evaluated input, which is composed of the user's interest [11].

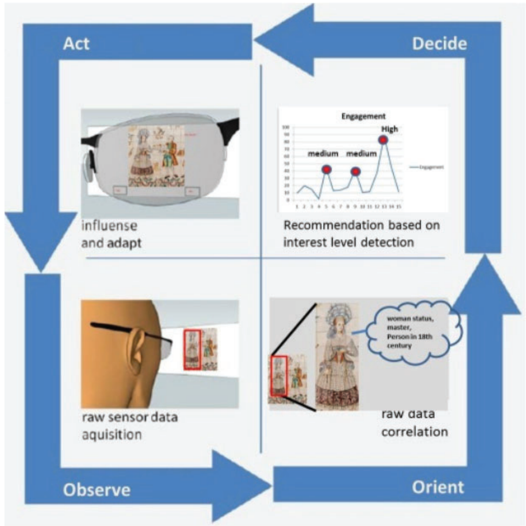


Figure 3 Observe, Orient, Decide, Act cycle of Damala et al. to rerender AR scene [12].

In the Observe phase, the necessary data for the evaluation of the user behaviour are collected. This includes visual data, which reveals the focus of the user, and acoustic data, which represents the sounds from the environment and hand gesture data for interaction with the system. In the Orient phase all these data are evaluated in real time and combined to an attention state. In the Decide phase, the user's interest in the artwork and the AR content is determined on the basis of this attention state and the content of the AR scene is recomposed. The new content is displayed on the glasses in the Act phase. When a visitor shows interest in a work of art by looking at it over a longer period of time, the AR System provides more information and details about the work. In the case of no interest, other works of art are suggested, taking into account content that has already attracted the visitor's interest and other behavioral data [12]. Even though the projects presented, especially the last one, are all already very adaptive, they usually only work under specific conditions. The use of contextual information could lead to more universal applications.

3.2 Contextual information

With the ability to recognize context, adaptive AR applications can adjust their information to best support the user by displaying appropriate information in the corresponding situations [9]. In the following, it is analysed which possibilities there are to perceive context and how it is filtered in order to capture the important elements. In the end it should be clear wether the context information which can be gathered by common AR devices is useful for adaptive AR applications.

White et al. demonstrate that it makes more sense to use multiple contextual sources than to restrict oneself to one type of context. Based on five different context information sources they have created different user interest models to predict the future user interest. These include user interaction, task, social conditions and long-term interaction history. It became clear that the user interaction gives the best results in real time. In the long term, however, the interaction history is the most meaningful, whereby the social context, i.e. the interest of others, also has an influence here [37]. Accordingly, ARtSense project evaluates that the conditions for users are very good, since three sources are used [12]. According to Schilit et al. different places support different behaviours [30]. The PARCTAB system by Want et al. makes use of this and uses location as context source. It is, among other things, a file system which stores files in a local directory. If one is at work, information related to the calendar and Meetings, which have large relevance in the office, are faster to access, than information necessary, for example, in the kitchen. Here the context is filtered based on location and the application adapts accordingly [36].

Huang et al. try to predict the usage behavior of apps on smartphones through contextual information. For this they use information about the time, location and a self-generated user profile. For the influence of time they describe two patterns. On the one hand, a different behaviour is predicted on weekends compared to weekdays. On the other hand, the behavior should be different throughout a day. The location of the users is also dependent on time. In order to determine this as accurately as possible, it is not only tracked via GPS, but also via the WiFi routers to which the users are connected. In this way, the room in a building can be determined precisely. Both information are implicitly opposed by the user profile, which is created explicitly by each user. On the basis of all this context information, the usage behaviour of apps is predicted [19].

The context sources used in the various projects presented can be generated by handheld devices such as smartphones. Accordingly, common devices used in AR applications should be sufficient to generate enough context input for a reasonable adaptation. It is also clear that it makes sense to use multiple context sources at the same time in order to guarantee the most accurate adaptation. However, Want et al. suggest that it makes sense to limit the input of the sources and not to evaluate every detail [36]. A visionary application would of course be one that can use every detail in context in real time. Before starting the development of an application, it should be checked which context sources should have an impact and if they are compatible with AR applications.

3.3 Contextual information access with adaptive AR

This section presents three aspects from the literature. First, which types of context can be detected by common hand-held and head-mounted devices. These include location, time, weather, movement and user interest. Second, which applications have been implemented so far to utilize contextual information in adaptive AR applications. And finally which challenges still exist.

The current state of the technology provides a lot of information about the user. Thus, the input required by Richards et al. such as an interface for the user and recordings of the environment can be accomplished with smartphones [29]. In addition, it is possible to determine the current time or the current location of the user with the help of GPS [19]. The user does not need to explictly express all the relevant input, which makes the adaptation easier [1]. It is also possible to see whether the user is currently moving or remains at a certain location. Further information, which can be helpful for the context of adaptive AR applications, includes other people close to the user. This can be accomplished using speech detection or image recognition [10]. Lastly a mechanism is needed to recognize the user's interest. In most of the presented applications gaze tracking and speech detection are used for this purpose.

AR is used in many ways, there are applications from the fields of technology and entertainment as well as support in everyday life [39]. Liu et al. describe a user interface for smart glasses that adapts to the environment by adjusting color, shape and size of the objects to look more natural and not distract users [24]. Erkoyuncu et al., on the other hand, describe an adaptive system which is used in industry to support maintenance work by means of AR. The developed system proposes a method for maintenance that adapts based on the user's skills and the available tools. Furthermore, it becomes clear which part of the object has to be maintained (see Figure 4) [15].

Henrysson and Ollila describe a concept of Ubiquitous Mobile Augmented Reality (UMAR) in which, depending on the spatial relationship, information relevant to the context, e.g. a map on which a target is entered, is represented either as a 2D map or by AR [17]. Delail et al. describe a Context Aware Visual indoor Augmented Reality (CAViAR) which is intended to improve life on campus using a smartphone. The position of the users is determined

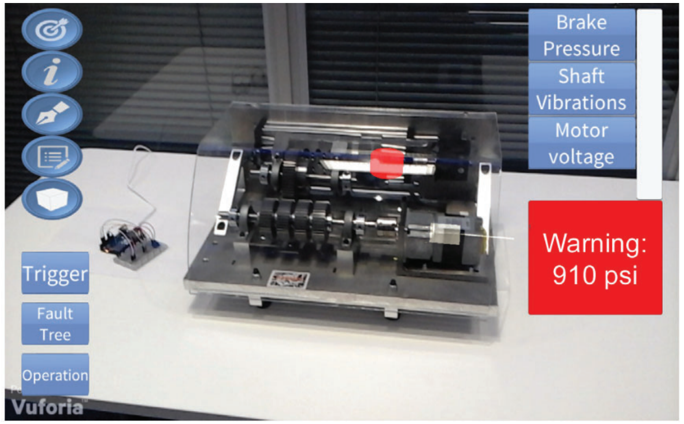


Figure 4 Context-aware maintenance module by Erkoyuncu et al [15].

by object recognition through camera images. In addition, the users can enter their timetable and other data into a database. Finally, a pre-defined database of the campus is used to display information such as the best route from the location to the next classroom or further information on the current location [13].

In order to support a surgeon during surgery, Katic et al. have developed a system that serves as an assistant in the operating room. The application recognizes different phases of the operation and can give the surgeon further information by means of context-related visualizations. As shown in Figure 5, the surgeon can be guided to a tumor or endangered stromal structures are highlighted [21]. A similar application has been developed by Kalkofen et al. They support the visualization of hidden objects. With their framework they reduce the amount of lost information by projecting contextual elements onto the obstacles so the object that is hidden is displayed completely. For example, as shown in Figure 6, the interior of a car can be recognised from the outside without having to be disassembled [20]. Most applications try to generate the contextual information implicitly, as described by Abowd et al [1]. However, Hervás et al. take a different approach: their application provides information that can support the user, such as instructions for washing machines and others, as soon as they scan a QR-code with their smartphone. These are then projected onto the corresponding objects using AR [18]. Ajanki et al. have developed a system similar to that of Damala et al. Their application evaluates gaze, speech and implicit feedback such as user interaction and presents relevant

information about people and objects in the environment [2]. The presentation is constantly re-evaluated by means of the input and accordingly unimportant information is hidden and further information is shown if there is interest by the user. For this they used an HMD and a small portable computing unit [3].

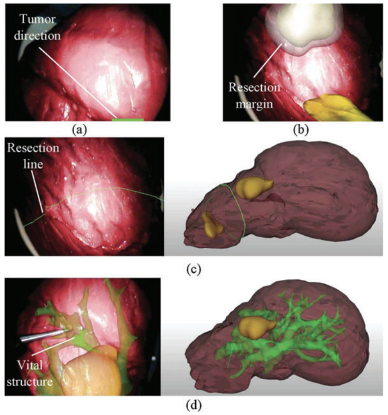


Figure 5 Visualizations for (a) directing the surgeon to the tumor, (b) showing the resection margin and (c) the resection line and (d) displaying vital structures currently at risk by Katic et al [21].

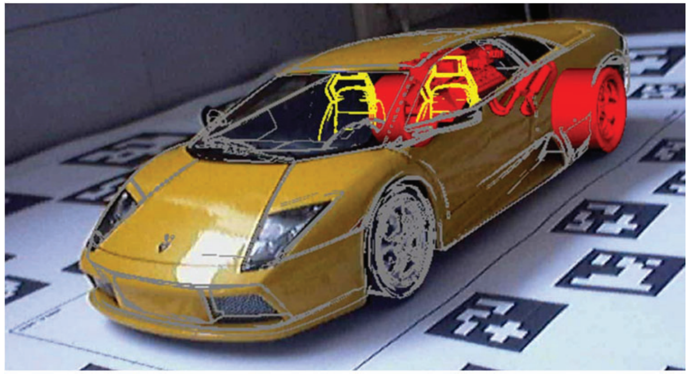


Figure 6 Visualization of occluded parts of car by Kalkofen et al [20].

4 Research gaps

Over the years, much has been done in this field of research. Technical progress has solved many of the hardware problems of the past and many applications can now be realised with smartphones alone. Nevertheless, many of the analysed applications still show weaknesses in the evaluation of in-depth information about the analysed environment. Thus the AR content can be represented partly incorrectly, or important information does not flow into the context evaluation. On the other hand, the means for gaze-tracking have to be further improved for mobile use. With regard to the evaluation of context information, there are various approaches that need to be pursued further. In general, the technology and knowledge gaps have become smaller over the years. Much has been tried and tested regarding to the interaction with the different systems and the presentation of the content and should be tested again with the new possibilities. In the end, however, it can still be concluded that almost all applications have been developed for a specific application only and that there is no concrete elaboration of a system that can be used universally. The work of Ajakni et al. and Damala et al. can probably best be extended for this purpose, since they use many different input sources.

5 Conclusion

This paper gives an overview of the current techniques to display contextual information with adaptive AR. For this purpose, the literature on adaptive AR and contextual information was analysed in order to create a basis for the interdisciplinary analysis. The current state of research in adaptive AR shows that much has been tried and different approaches have been developed. Unfortunately, much of it is still conceptual or prototypical work that will need to be further elaborated in the future.

6 Acknowledgments

I would like to thank Luke Haliburton for the support during the prepartion of this research paper.

References

1 Gregory D Abowd, Anind K Dey, Peter J Brown, Nigel Davies, Mark Smith, and Pete Steggles. Towards a better understanding of context and context-awareness. In *International symposium on handheld and ubiquitous computing*, pages 304–307. Springer, 1999.

2 Antti Ajanki, Mark Billinghurst, Hannes Gamper, Toni Järvenpää, Melih Kandemir, Samuel Kaski, Markus Koskela, Mikko Kurimo, Jorma Laaksonen, Kai Puolamäki, et al. An augmented reality interface to contextual information. *Virtual reality*, 15(2-3):161–173, 2011.

3 Antti Ajanki, Mark Billinghurst, Toni Järvenpää, Melih Kandemir, Samuel Kaski, Markus Koskela, Mikko Kurimo, Jorma Laaksonen, Kai Puolamäki, Teemu Ruokolainen, et al. Contextual information access with augmented reality. In *2010 IEEE International Workshop on Machine Learning for Signal Processing*, pages 95–100. IEEE, 2010.

4 Ronald Azuma, Yohan Baillot, Reinhold Behringer, Steven Feiner, Simon Julier, and Blair MacIntyre. Recent advances in augmented reality. *IEEE computer graphics and applications*, 21(6):34–47, 2001.

5 Michael Bajura, Henry Fuchs, and Ryutarou Ohbuchi. Merging virtual objects with the real world: Seeing ultrasound imagery within the patient. *ACM SIGGRAPH Computer Graphics*, 26(2):203–210, 1992.

6 David Benyon and Dianne Murray. Experience with adaptive interfaces. *The Computer Journal*, 31(5):465–473, 1988.

7 Wolfgang Broll, Irma Lindt, Iris Herbst, Jan Ohlenburg, Anne-Kathrin Braun, and Richard Wetzel. Toward next-gen mobile ar games. *IEEE Computer Graphics and Applications*, 28(4):40–48, 2008.

8 William Chang and Qing Tan. Augmented reality system design and scenario study for location-based adaptive mobile learning. In *2010 13th IEEE International Conference on Computational Science and Engineering*, pages 20–27. IEEE, 2010.

9 Fabio Crestani and Ian Ruthven. Introduction to special issue on contextual information retrieval systems. *Information Retrieval*, 10(2):111–113, 2007.

10 Areti Damala, Tobias Schuchert, Isabel Rodriguez, Jorge Moragues, Kiel Gilleade, and Nenad Stojanovic. Exploring the affective museum visiting experience: adaptive augmented reality (a2r) and cultural heritage. *International Journal of Heritage in the Digital Era*, 2(1):117–142, 2013.

11 Areti Damala and Nenad Stojanovic. Tailoring the adaptive augmented reality (a 2 r) museum visit: Identifying cultural heritage professionals' motivations and needs. In *2012 IEEE international symposium on mixed and augmented reality-arts, media, and humanities (ISMAR-AMH)*, pages 71–80. IEEE, 2012.

12 Areti Damala, Nenad Stojanovic, Tobias Schuchert, Jorge Moragues, Ana Cabrera, and Kiel Gilleade. Adaptive augmented reality for cultural heritage: Artsense project. In *Euro-Mediterranean Conference*, pages 746–755. Springer, 2012.

13 Buti Al Delail, Luis Weruaga, and M Jamal Zemerly. Caviar: Context aware visual indoor augmented reality for a university campus. In *Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web*

Intelligence and Intelligent Agent Technology-Volume 03, pages 286–290. IEEE Computer Society, 2012.

14 Anind K Dey. Understanding and using context. *Personal and ubiquitous computing*, 5(1):4–7, 2001.

15 John Ahmet Erkoyuncu, Inigo Fernandez del Amo, Michela Dalle Mura, Rajkumar Roy, and Gino Dini. Improving efficiency of industrial maintenance with context aware adaptive authoring in augmented reality. *Cirp Annals*, 66(1):465–468, 2017.

16 Nehla Ghouaiel, Jean-Marc Cieutat, and Jean-Pierre Jessel. Adaptive augmented reality: plasticity of augmentations. In *Proceedings of the 2014 Virtual Reality International Conference*, page 10. ACM, 2014.

17 Anders Henrysson and Mark Ollila. Umar: Ubiquitous mobile augmented reality. In *Proceedings of the 3rd international conference on Mobile and ubiquitous multimedia*, pages 41–45. ACM, 2004.

18 Ramón Hervás, José Bravo, Jesús Fontecha, and Vladimir Villarreal. Achieving adaptive augmented reality through ontological context-awareness applied to aal scenarios. 2013.

19 Ke Huang, Chunhui Zhang, Xiaoxiao Ma, and Guanling Chen. Predicting mobile application usage using contextual information. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 1059–1065. ACM, 2012.

20 Denis Kalkofen, Erick Mendez, and Dieter Schmalstieg. Interactive focus and context visualization for augmented reality. In *Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 1–10. IEEE Computer Society, 2007.

21 Darko Katić, Anna-Laura Wekerle, Jochen Görtler, Patrick Spengler, Sebastian Bodenstedt, Sebastian Röhl, Stefan Suwelack, Hannes Götz Kenngott, Martin Wagner, Beat Peter Müller-Stich, et al. Context-aware augmented reality in laparoscopic surgery. *Computerized Medical Imaging and Graphics*, 37(2):174–182, 2013.

22 Eric Klopfer et al. *Augmented learning: Research and design of mobile educational games*. MIT press, 2008.

23 Pat Langley. User modeling in adaptive interface. In *UM99 User Modeling*, pages 357–370. Springer, 1999.

24 James C Liu, Anton O Andrews, Benjamin I Vaught, Craig R Maitlen, Christopher M Novak, and Sheridan Martin Small. Context adaptive user interface for augmented reality display, February 21 2013. US Patent App. 13/212,172.

25 Mehdi Mekni and Andre Lemieux. Augmented reality: Applications, challenges and future trends. *Applied Computational Science*, pages 205–214, 2014.

26 Danakorn Nincarean, Mohamad Bilal Alia, Noor Dayana Abdul Halim, and Mohd Hishamuddin Abdul Rahman. Mobile augmented reality: the potential for education. *Procedia-social and behavioral sciences*, 103:657–664, 2013.

27 Patrick Renner and Thies Pfeiffer. Attention guiding techniques using peripheral vision and eye tracking for feedback in augmented-reality-based assistance systems. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 186–194. IEEE, 2017.

28 Alexandra Rese, Daniel Baier, Andreas Geyer-Schulz, and Stefanie Schreiber. How augmented reality apps are accepted by consumers: A comparative analysis using scales and opinions. *Technological Forecasting and Social Change*, 124:306–319, 2017.

29 Brian Richards, Brent Robert Blum, Timothy Li, Byron John Schmidt, and Amjad-ali Khoja. Context based augmented reality, November 26 2015. US Patent App. 14/653,208.

30 Bill Schilit, Norman Adams, and Roy Want. Context-aware computing applications. In *1994 First Workshop on Mobile Computing Systems and Applications*, pages 85–90. IEEE, 1994.

31 Bill N Schilit and Marvin M Theimer. Disseminating active mop infonncition to mobile hosts. *IEEE network*, 1994.

32 Muhammad Shahid, Tabassam Nawaz, and Hafiz Adnan Habib. Eye-gaze and augmented reality framework for driver assistance. *Life Science Journal*, 10(3), 2013.

33 Patrick Sinclair and Kirk Martinez. Adaptive hypermedia in augmented reality. In *Proceedings of Hypertext*, 2001.

34 Takumi Toyama, Daniel Sonntag, Jason Orlosky, and Kiyoshi Kiyokawa. Attention engagement and cognitive state analysis for augmented reality text display functions. In *Proceedings of the 20th international conference on Intelligent user interfaces*, pages 322–332. ACM, 2015.

35 Ellison C Urban. The information warrior. In *Technology and society*, pages 493–501. Prentice-Hall, Inc., 1999.

36 Roy Want, Bill N Schilit, Norman I Adams, Rich Gold, Karin Petersen, David Goldberg, John R Ellis, and Mark Weiser. The parctab ubiquitous computing experiment. In *Mobile computing*, pages 45–101. Springer, 1996.

37 Ryen W White, Peter Bailey, and Liwei Chen. Predicting user interests from contextual information. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pages 363–370. ACM, 2009.

38 Hsin-Kai Wu, Silvia Wen-Yu Lee, Hsin-Yi Chang, and Jyh-Chong Liang. Current status, opportunities and challenges of augmented reality in education. *Computers & Education*, 62:41 – 49, 2013. URL: `http://www.sciencedirect.com/science/article/pii/S0360131512002527`, `doi:https://doi.org/10.1016/j.compedu.2012.10.024`.

39 Darrell L Young, Gregory R Gondran, Thomas G Ribardo, Mark A Bigham, et al. Systems and methods for context based information delivery using augmented reality, September 19 2017. US Patent 9,767,615.

Lässt sich Mixed Reality für das Lernen nutzen?

Sybil Bast

Ludwig-Maximilians-Universität München, München, Deutschland
`sybil.bast@campus.lmu.de`

Zusammenfassung

Obwohl Mixed Reality in den letzten Jahren immer attraktiver wurde, wird es in der heutigen Zeit noch kaum für das Lernen eingesetzt. In der Forschung und Produktion sowie in der Unterhaltungsbranche ist Mixed Reality keine Seltenheit mehr. Dort wird Mixed Reality beispielsweise für virtuelle 3D-Spiele oder kostengünstige Prototypen genutzt. Trotzdem greifen Schulen und Universitäten auf alte Technologien, wie das Tafelbild oder Übungblätter, zurück, sobald es um die Weiterbildung ihrer Schüler oder Studenten geht. Obwohl Mixed Reality neue Lernmöglichkeiten bietet, stellt es viele Pädagogen, Lehrer und Professoren vor eine neue Herausforderung. Wie kann diese sehr attraktive neue Technik in der Bildung eingesetzt werden, um das Lernen zu erleichtern und für den Lernenden keine Überlastung oder unüberwindbare Herausforderung zu stellen?

Keywords and phrases Augmented Reality, Multimedia, Vokabellernen, Mixed Reality, Lernen, Virtual Reality, Immersives Lernen, 3D, Mobile, Sprachen

1 Einleitung

Was wäre, wenn das Lernen so gestaltet werden könnte, dass es den Kindern und Erwachsenen Spaß macht? Was wäre, wenn diese Personen an fremde Orte reisen könnten um etwas darüber zu lernen? Und was wäre, wenn man die Umgebung jedes Menschen durch Objekte erweitern und somit die Lernerfahrung verbessern könnte? Mit Mixed Reality (MR) ist dies möglich. In den letzten Jahren hat sich Mixed Reality rasant weiterentwickelt und bietet somit unzählige Möglichkeiten für ein effektiveres Lernen.

Betrachtet man das Lernen durch technische Geräte in der Vergangenheit, so erkennt man, dass „die technische Entwicklung exponentiell verläuft“ [22]. Die Technik entwickelt sich rasant weiter und wird auch in den kommenden Jahren immer schneller wachsen. Ebenso gewinnt der Wissenstand eine immer höhere Bedeutung. Kinder bildern sich schon im jungen Alter weiter. Vor allem der technische Wissenstand muss frühzeitig zur Verfügung stehen. „So wird das technische Wissen von heute nur noch 1% des technischen Wissens

© Sybil Bast;
licensed under Creative Commons License CC-BY

Cite as: Sybil Bast. Lässt sich Mixed Reality für das Lernen nutzen?. In *5th Seminar on Ubiquitous Interaction (UBIACTION 2020)*. Editors: Florian Lang, Pascal Knierim, Jakob Karolus, Fiona Draxler, Ville Mäkelä, Francesco Chiossi, Luke Haliburton, Matthias Hoppe, Albrecht Schmidt. January 27, 2020. Munich, Germany. pp. 2:1–2:16.

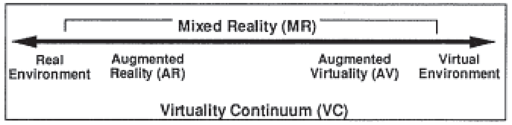


Abbildung 1 Das Mixed-Reality-Kontinuum [20]

im Jahr 2050 ausmachen“ [22]. Einen bedeutenden Aspekt bildet hierbei das virtuelle Lernen. Dieses kann bereits im frühen Kindesalter zur Weiterbildung eingesetzt werden und sich durch alle Lebensphasen des Lernens ziehen [22].

Durch Virtual-Reality-Medien kann der Lernerfolg gesteigert werden. Allein der Einsatz der Mixed-Reality-Technik verspricht jedoch noch keine Lernsteigerung. Ebenso wie beim Offline-lernen führt die richtige Umsetzung zum Ziel [8].

2 Mixed Reality

Mixed Reality kann in Augmented Reality (AR) und Virtual Reality (VR) unterteilt werden. Während Virtual Reality in der heutigen Zeit ein geläufiger Begriff ist, kann Augmented Reality häufig nicht so einfach erklärt werden.

Hierbei werden natürliche und künstliche Sinnesreize kombiniert. Während Virtual Reality eine komplett computergesteuerte Umgebung beschreibt, definiert Augmented Reality eine Überlappung der physischen Welt mit einer computergesteuerten Umgebung [25].

Bei AR und VR wird häufig von einem Kontinuum gesprochen. Das Mixed-Reality-Kontinuum wurde von Paul Milgram erfasst und wird häufig auch als Reality-Virtuality-Kontinuum bezeichnet. Bei Augmented Reality überwiegt der Anteil der Realität, während bei der Virtual Reality nur der virtuelle Inhalt betrachtet wird. In der Praxis können diese zwei Bereiche nicht immer vollkommen voneinander getrennt verstanden werden. In den Punkten 2.1 und 2.2 wird jedoch von der hier definierten Reinform ausgegangen [23]. Der Aufbau von Mixed Reality wird in der Abbildung 1 verdeutlicht.

In dieser Abbildung wird gezeigt, dass sich Mixed Reality in zwei Richtungen bewegen kann. Zum Einen in die Richtung der realen Welt. Solange man sich im Virtuality Continuum (VC) befindet und die reale Umgebung genutzt wird spricht man von Augmented Reality. Wird sich in Richtung der virtuellen Umgebung bewegt spricht Milgram von Augmented Virtuality (AV). Sobald die reale Welt keinen Einfluss mehr hat und sich der Nutzer nur noch in der virtuellen Welt befindet, spricht man von Virtual Reality. Milgram bezeichnet dies ebenfalls als Augmented Virtuality [20]. Im Folgenden wird jedoch der Begriff Virtual Reality verwendet.

2.1 Virtual Reality

In der VR nimmt der Nutzer, wie bereits beschrieben, nur noch die computeranimierte Umgebung wahr. Er erlebt die virtuelle Welt, meist durch eine Brille und Kopfhörer, indem er seine Umgebung sieht, spürt und hört. Es wird auf spezifische Hardware gesetzt, um dem Benutzer die Illusion einer physischen Präsenz innerhalb dieser „Realität“ zu vermitteln. Somit wirkt sie für den Betrachter äußerst real und kann in 3D und 360 Grad erlebt werden. Die virtuelle Welt verhält sich wie die reale Welt. Dies bewirkt, dass der Nutzer intuitiv richtig handeln kann [12].

Die virtuellen Systeme nutzen Ein- und Ausgabegeräte, welche in der realen Welt existieren, um alle Sinne des Nutzers anzusprechen. Die Möglichkeiten, welche Virtual Reality bietet, steigen immer weiter an. Der Nutzer verwendet VR um in eine vollkommen neue Welt zu reisen. Da sich die Entwicklung rasend schnell verbessert, werden VR-Systeme auch immer kostengünstiger. Produkte, welche sich vor wenigen Jahren noch kein Hobbynutzer leisten konnte, werden nun bezahlbar und für jedermann zugänglich [23].

Virtual Reality stößt auch in der Unterhaltungsbranche zunehmend auf Interesse. Hier wird VR für virtuelle 3D-Spiele, wie “Escape-Games” genutzt. Einen vermutlich noch weitreichenderen Verwendungsgrund bietet die Aufnahme in Industrie- und Fertigungsprozesse [3]. Hierbei ermöglicht VR zum Beispiel in der Autoindustrie Prototypen schon weit vor der Produktion realitätsnah in 3D zu betrachten. Somit können Design und Formen ohne Material- und Arbeitsaufwand angepasst werden [24]. Auch die Segmente Gesundheit und Einzelhandel nutzen Virtual Reality [3]. Hier wird VR beispielsweise bei chronischen Schmerzen eingesetzt. Das Spielen von VR-Spielen und das enspannte Leben in der VR-Welt kann die chronischen Schmerzen lindern. Außerdem können Phobiepatienten mit ihrer Phobie konfrontiert werden, ohne sich dieser in der Realität aussetzen zu müssen [6].

Dörner, Broll, Grimm und Jung beschreiben in ihrem Paper „Virtual Reality und Augmented Reality“, dass Menschen ausblenden können, dass sie sich in einer „künstlichen Welt“ befinden. Obwohl es dem Nutzer bewusst ist, kann sein Gedächtnis diesen künstlichen Zustand akzeptieren. Die Autoren vergleichen dies mit der deutschen Synchronisation von Filmen. Dort ist dem Zuschauer ebenfalls bewusst, dass es sich um einen synchronisierten Film handelt, jedoch wird sein Gedächtnis dies mit der Zeit akzeptieren und verdrängen [23].

2.2 Augmented Reality

Augmented Reality hingegen ergänzt die reale Welt des Betrachters durch virtuelle Objekte oder Informationen. Auch bei AR spielt die 3D-Erfassung der Umgebung eine wichtige Rolle. Diese kann jedoch auch über einem Bildschirm, wie der eines Smartphones, Tablets oder Head-Up-Displays geboten werden [14].

Durch die Kameras dieser Geräte werden virtuelle Objekte in Echtzeit in die reale Welt integriert [19]. Die virtuellen Objekte sind im besten Fall weder beim Betrachten noch beim Benutzen von einem realen Objekt zu unterscheiden [23]. Die virtuelle Darstellung verschmilzt mit der realen Darstellung. Dies hat den Vorteil, dass dem Nutzer alle wichtigen Informationen an einer Stelle bereitstehen [19]. Bei herkömmlichen Lernmethoden stehen die wichtigsten Informationen ebenfalls an einer Stelle bereit. Jedoch werden dafür 2D-Bilder oder Bücher genutzt. Eine 3D-Darstellung ist für den Nutzer um einiges leichter zu verstehen. Er kann von verschiedenen Sichtpunkten auf die in 3D dargestellten Objekte blicken, mit dem Objekt interagieren und seine Erkenntnisse teilen [17].

Früher waren die AR-Systeme schwer zu nutzen. Es mussten große und schwere Rucksäcke mit hoher Rechenleistung, jedoch niedriger Akkulaufzeit mitgetragen werden. Heutzutage kann der Nutzer über sein Smartphone oder Tablet AR anwenden. Somit stieg die Nutzung in den vergangenen Jahren rasant an.

Durch GPS und die Kamera lassen sich Positionen, Bewegungsrichtung und die Lage des Betrachters ermitteln [23]. Eine Funktionalität von AR welche häufig genutzt wird ist das Tracking. Hierbei wird die Umgebung erfasst und entsprechend dazu werden virtuelle Objekte sichtbar. Durch das GPS, die Kamera und den Kompass, im Smartphone oder einem anderen ähnlichen Gerät, können die Richtung, reale Objekte und die Position wiedergegeben werden. Somit ergibt sich die Möglichkeit reale Objekte durch virtuelle Zusätze zu ergänzen [19].

Augmented Reality ist im Vergleich zu Virtual Reality und auch anderen Technologien relativ neu [15]. Einsatzgebiete von Augmented Reality sind die Entertainment- und Spielebranche, die Industrie und der Einsatz in der Aus-, Fort- und Weiterbildung von Kindern und Erwachsenen [19]. Dieses Paper befasst sich mit dem zuletzt genannten Aspekt.

3 Lernen und Mixed Reality

Durch Mixed Reality ist es möglich, unsere Realitätswahrnehmung zu verändern und durch virtuelle Inhalte zu erweitern. Somit kann der Nutzer sich durch diese Technik neues Wissen aneignen.

Von dem Philosophen Lao-Tse stammt das Zitat: "If you tell me, I will listen. If you show me, I will see. But if you let me experience, I will learn."

Obwohl dieses Zitat bereits im 5. Jahrhundert v. Chr. verfasst wurde, beschreibt Lao-Tse das Lernen durch Mixed Reality äußerst treffend [13, S.147].

Beim Lernen mit MR wird durch die erlebte Erfahrung gelernt. Dies bringt einige Vorteile mit sich, welche in den folgenden Absätzen beschrieben werden.

Beim Lernen durch Mixed Reality werden im Vergleich zum traditionellen Lernen unterschiedliche Sinne zur selben Zeit angesprochen. Desto mehr Sinne verwendet werden, desto höher ist die Lernwahrscheinlichkeit des Nutzers. Die kommunizierten Inhalte können effektiver erfasst werden, wodurch schneller und intensiver gelernt wird [9].

Weitere Vorteile beim Lernen durch Mixed Reality ergeben sich durch die Einbeziehung der Umgebung.

Es können Objekte in 3D dargestellt werden, was wiederum zu einem kollaborativen und situierten Lernen führt. Dies trägt wiederum zu einer besseren Gruppendynamik bei Gruppenarbeiten oder in Unterrichtseinheiten bei [9].

Außerdem können unsichtbare oder nicht präsente Gegenstände und Objekte aus unterschiedlichen Perspektiven erfasst werden. Dadurch wird die visuelle Wahrnehmung des Lernenden verbessert [9].

Durch Motivation und Interesse wird der Lernvorgang ebenfalls gesteigert, was einen weiteren Vorteil von Mixed Reality birgt. Informationen können interessanter vermittelt werden und somit das Lernerlebnis stärken. Für den Nutzer wird das Lernen einfacher und das Selbstvertrauen gestärkt [16].

Viele Studien haben ergeben, dass Wissen gelernt, verinnerlicht und angewendet werden muss, um im Gedächtnis zu bleiben [7]. Das Lernen und Verinnerlichen von Wissen ist keine zu große Schwierigkeit und wird auch im herkömmliche Schulunterricht genutzt. Das Anwenden jedoch führt häufig zu Problemen. Nicht in jedem Themenbereich hat der Lernende die Möglichkeit durch Anwendung das Gelernte zu testen oder realitätsnah zu erleben. Nun kommt Mixed Reality zum Einsatz [7]. Hierbei kann sowohl durch Virtual Reality, als auch durch Augmented Reality das Gelernte durch Anwendung verinnerlicht werden. Der Fokus dieser Arbeit wird auf das Lernen durch Augmented Reality gelegt. Auch wenn der Schwerpunkt dieser Arbeit auf dem Lernen durch Augmented Reality liegt, kann durch Virtual Reality ebenfalls umfangreich gelernt werden. Diese Art von Lernen wird als immersives Lernen bezeichnet und im folgenden Kapitel näher erläutert.

3.1 Immersives Lernen

Beim Immersiven Lernen hat der Nutzer das Gefühl, sich in einer virtuellen dreidimensionalen Welt zu befinden. Die Wahrnehmung der realen Welt wird vermindert. Dadurch hat der Nutzer das Gefühl direkt in die simulierte Welt einbezogen zu sein [26]. Dieser Effekt wird häufig zum Lernen genutzt. Sowohl in Klassenzimmern als auch bei Fortbildungen und Schulungen werden immersive Technologien bereits angewendet. Der Nutzer wird durch immersive

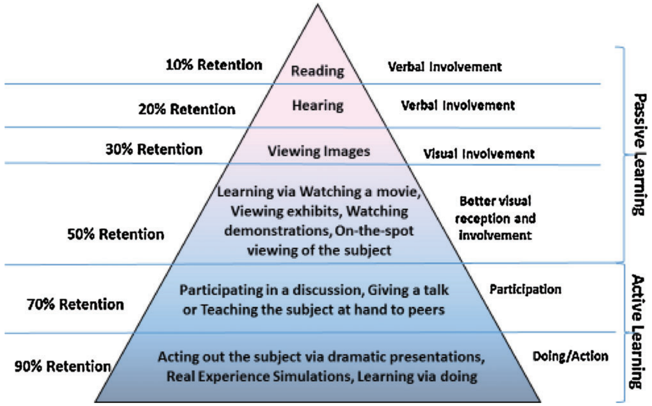


Abbildung 2 Stufen des Lernens [4]

Erfahrungen aktiver und somit zum Lernen ermutigt, denn durch die Interaktion mit der virtuellen Umgebung kann sowohl das Interesse als auch das Durchhaltevermögen beim Lernen gesteigert werden. Durch die in Echtzeit stattfindende Interaktionen in der dreidimensionalen Lernumgebung entsteht beim Nutzer ein sofortiger Lerneffekt [26].

In der Abbildung 2 wird der Anteil dessen, was sich der Lernende merkt im Verhältnis zur Lernmethode aufgezeigt. Hier wird sichtbar, dass das passive Lernen zu keiner hohen Speicherung des Gelernten führt. Sobald der Lernende jedoch aktiv mit dem Lernstoff interagiert, steigt die Speicherung drastisch an. Dieser Effekt wird beim immersiven Lernen genutzt. Dadurch, dass der Lernende Aktionen ausführt und interagiert, kann die Auffassung des Gelernten bis zu 90% des Gesamtstoffes erreichen. [4]

Beim Lernvorgang macht sich der Schüler dies zunutze, indem er sich beim Lesen oder Hören das Gelesene oder Gehörte visuell vorstellt um es sich merken zu können. Dies kann noch weiter verbessert werden, indem die Ereignisse in einer virtuellen Welt erlebt werden [4]. Nun kommt das Lernen durch VR zum Einsatz.

Versuche der letzten Jahre haben gezeigt, dass eine Weiterbildung durch VR-Anwendungen mit immersiven Inhalten einen höheren Lernerfolg bietet als traditionelle Ansätze. Jan Hellriegel und Dino Ćubela beschreiben in ihrem Paper “Das Potenzial von Virtual Reality für den schulischen Unterricht“ einige Potentiale, nach welchen die Einbindung von Virtual Reality im Lernvorgang einen höheren Lernerfolg liefert. Ein Beispiel, worauf dies zurückzuführen ist, wäre, dass Lernende sobald sie mit ihrer Umgebung interagieren, nachhaltige Lernerfolge erzielen [8]. Durch Virtual Reality kann während des Lernens mit der Umgebung interagiert werden. Dies ist beim herkömmlichen Lernen kaum möglich. Aus diesem Grund wird in vielen Fächern und Themenbereich schnell das Interesse verloren. Beim immersiven Lernen, wird dieses durch visuelle Qualität, Audioqualität und intuitive Interaktion länger aufrechterhalten. Vor allem bei Kindern wird durch die spielerische Darstellung und Interaktion das Interesse geweckt und länger gehalten [10].

Da es sich bei Virtual Reality um eine Neuerung im Bezug auf das Lernen handelt, haben die meisten Nutzer noch keine oder wenige Erfahrung damit. Dies führt zur kognitiven Neugierde (Malone 1981). Lernende haben noch keine ausreichende Erfahrung um sich ein mentales Bild von der Verwendung von VR zu machen. Aus diesem Grund überwiegt zu Beginn die Neugierde, welche im Laufe der Zeit abflachen kann [21]. Mit diesem Aspekt befasst sich der Novelty Effect. Dieser besagt, dass bei der Einführung neuer Technologien die Leistung zunächst verbessert wird. Der Grund dafür ist das zunehmende Interesse an der neuen Technologie und nicht die eigentliche Verbesserung. Neuartige Reize bleiben besser in der Erinnerung als bekannte [18]. Lernenden sollte dies beim Einsatz von VR bewusst sein. Der Novelty Effect schwächt mit der Zeit ab und kann dazu führen, dass auch der Lerneffekt wieder abnimmt.

3.2 Lernen in Augmented Reality-erweiterter Umgebung

Mit Hilfe von Augmented Reality kann man virtuelle Objekte in der realen Welt so platzieren, dass sie sich der Umgebung anpassen. Dadurch kann vor allem die Kognition, Interaktivität, Leistung und Konzentration gesteigert werden. Der Nutzer lernt allgegenwärtig. Die Verbindung des virtuellen Objekts mit der realen Umgebung, in der es sich befindet, wird durch diese Darstellung gestärkt [15].

Ähnlich wie beim immersiven Lernen wird durch den spielerischen Aspekt effektiver gelernt. Virtuelle 3D-Darstellungen werden in der realen Umgebung betrachtet oder es wird mit ihnen interagiert. Für den maximalen Lernerfolg werden einfache Systeme oder Dinge in AR nachgebildet [17].

Einige Beispiele dafür sind das 3D-animierte Buch MagicBook, die Darstellung von früheren oder zukünftigen Gebäuden, sowie jeglichen andere Gegenstände. Im folgenden Kapitel werden diese Beispiele weiter ausgearbeitet.

Da die Nutzung von Smartphones und Tablets in den letzten Jahren rasant angestiegen ist und sich die Prozessoren stetig verbessern, bietet sich nun die Möglichkeit der Nutzung von AR für nahezu jeden Besitzer eines solchen Geräts an. Das AR-Erlebnis kann an beliebigen Orten stattfinden. Somit haben Lernende die Möglichkeit, sich innerhalb und außerhalb ihres Klassenzimmers oder ihres Lernraums weiterzubilden. Trotzdem gilt zu beachten, dass AR

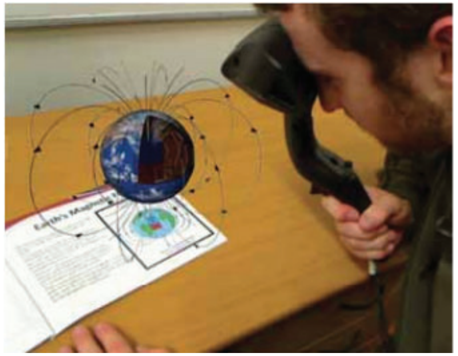


Abbildung 3 Virtueller 3D-Globus [17]

richtig eingesetzt werden muss, damit ein Lerneffekt entsteht. Das Lernen ist am effektivsten, wenn es zur Ergänzung traditioneller Lernmethoden genutzt wird. Häufig kann durch das alleinige Erstellen einer AR-Szene schon gelernt werden. Da man sich vor dem Aufbau der Szene ausgiebig mit dem Thema beschäftigten muss, wird man sich bereits in dieser Phase schon Wissen aneignen. Durch ein Analysetool kann die gewünschte AR-Szene bereits von Kindern auf unkomplizierte Weise und mit breiten Lernerfolg erstellt werden [17].

3.2.1 Darstellung eines virtuellen 3D-Objekts

Im letzten Kapitel wurde erläutert, wie die Umgebung durch Augmented Reality erweitert werden kann. Nun werden die zuvor genannten Beispiele zur Darstellung eines virtuellen 3D-Objekts genauer erfasst.

Zunächst kann die reale Umgebung durch ein oder mehrere 3D-Objekte ergänzt werden. In der Abbildung 3 betrachtet der Nutzer ein erweitertes Buch, in dem ein dreidimensionales Erdmagnetfeld durch ein Augmented-Reality-Gerät erfasst wird. Hier wird eine 2D-Zeichnung dreidimensional dargestellt um die räumliche Komponente zu erhöhen. Umso stärker diese ist, desto besser kann der Nutzer mit der Darstellung interagieren. Bei einem Test konnte herausgefunden werden, dass die Personengruppe, welche das Magnetfeld in Augmented Reality betrachtet hat, die Fragen zu dem Gelernten um 12% besser beantworten konnte, als das Team, welches nur ein 2D-Buch als Lernvorlage gestellt bekam. Auch in dem Folgetest nach 4 Wochen erzielte die AR-Gruppe einen höheren Wert [17]. Bei dem in Abbildung 4 dargestellten erweiterten Buch handelt es sich um eine der ersten Versionen. Modernere 3D-Bücher können nicht nur einzelne Objekte, sondern ganze Szenen darstellen.

Ein weiteres Beispiel eines erweiterten Buches ist das 3D-animierte MagicBook. Dieses Buch wird wie ein normales Buch geöffnet und gelesen. Zusätzlich bekommt der Nutzer eine 3D-Ansicht der Szene oder eines Elements aus der Szene. Somit kann er aus jeder Position und von jeder Richtung den Buchinhalt betrachten. Auf jeder Seite des Buches wird eine andere AR-Szene angezeigt. Eine weitere Besonderheit des MagicBooks ist es, dass der Nutzer mit unterschiedlichen virtuellen Objekten interagieren kann, indem er sie bewegt oder berührt. Verknüpft man diese Augmented Reality-Darstellung mit dem immersiven Lernen, so bietet sich die Möglichkeit, an die im Buch gezeigten virtuellen Orte zu reisen. Außerdem bietet das MagicBook eine Anzeige für zwei oder mehrere Personen. Diese Gruppe kann sowohl miteinander als auch mit dem Buch interagieren [17].

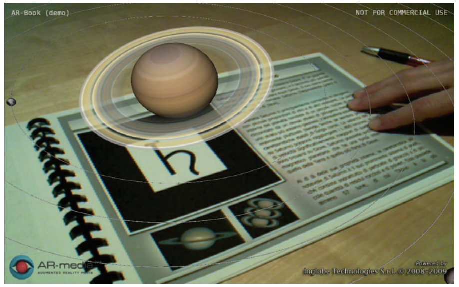


Abbildung 4 Eine der ersten Versionen des 3D-Buches Books [17]

3.2.2 Mobile Anwendung von Augmented Reality

Ein weit verbreitetes Einsatzgebiet von AR bildet die mobile Anwendung. Hierbei verwendet der Nutzer die Kamera sein Smartphone oder Tablet für die AR-Darstellung. Da die mobilen Geräte immer schneller und stärker werden, können sie auch für Augmented Reality eingesetzt werden. Über das Smartphone kann die Umgebung ergänzt werden und neue Informationen werden bereitgestellt. Installiert der Nutzer eine notwendige Bildverarbeitungssoftware, so können reale Bilder von virtuellen Inhalten überlagert werden. Die Wahrnehmung der realen Umgebung geschieht über die Smartphone- oder Tabletkamera. Diese wird um virtuelle Objekte ergänzt und auf dem Bildschirm dargestellt [17].

Ein Beispiel ist in der Abbildung 5 sichtbar. Hier werden für den Nutzer interessante Beschriftungen von Gebäuden oder Gegenständen dargestellt. Somit kann er sich leicht in seiner Umgebung, beispielsweise einer fremden Stadt zurechtfinden. Diese Technik kann ebenfalls für Stadtführungen genutzt werden. Die Teilnehmer erhalten die Möglichkeit den historischen Aufbau von

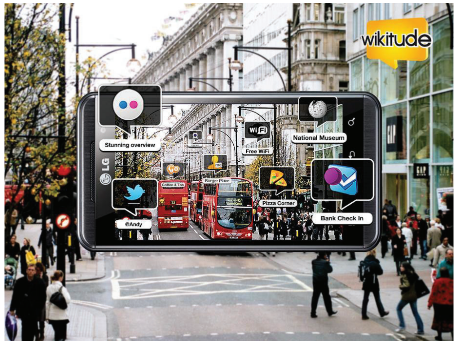


Abbildung 5 Bedeutende Orte sind gekennzeichnet und beschriftet [5]

Gebäuden durch ihr Smartphone zu betrachten. Sie lernen die Geschichte von Gebäuden und haben gleichzeitig den direkten Bezug zur historischen Geschichte. Somit kann diese Technik bestens für den Geschichtsunterricht oder für Stadtführungen einsetzt werden [17].

3.2.2.1 Beschriftung von Objekten

Ein weiteres Beispiel von mobilen Anwendungen bildet die Beschriftung von Objekten der realen Umgebung. Durch das Anzeigen von virtuellen Labels und Symbolen kann sich der Nutzer neues Wissen zu bestimmten Gegenständen aneignen. Vor allem ortsbezogene Informationen wie Gebäudenamen, Entfernungen und Navigationen können dabei unterstützend sein [15]. Augmented Reality kann auch in anderen Bereichen sinnvoll eingesetzt werden um Gegenstände zu beschriften und somit zu lernen. So zum Beispiel in der Medizin, wie in der Abbildung 6 dargestellt wird.

Hier wird durch die Beschriftung von Körperteilen der Aufbau des Körpers gelernt. Der Nutzer betrachtet das Modell eines Körpers über die Kamera seines Tablets. Die darauf abgestimmte Software kann nun einzelne Körperteile beschriften.

Durch Beschriftung kann ebenfalls eine fremde Sprache gelernt werden. In einer Unterrichtsstunde hat nicht jeder Schüler denselben Wissensstand. Der Lehrer kann ein bestimmtes Niveau voraussetzen, doch nur Ehrgeiz und Fleiß führen dazu, sich die Fremdsprache richtig anzueignen. Mit dem Druck, nicht mit anderen Personen mithalten zu können, können viele Menschen nicht umgehen. Sie ziehen sich zurück und üben wenig. Vor allem beim Erlernen einer Fremdsprache kommt es darauf an, sich zu unterhalten und seine Angst beiseite zu schieben. Diese Angst nimmt mit steigenden Alter zu [7].

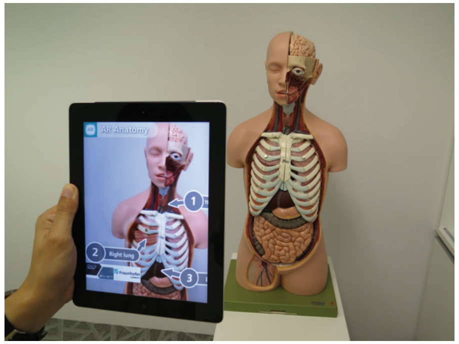


Abbildung 6 Beschriftung von Organen [1]

Der Einsatz von Technologien kann dem entgegenwirken. Sie bieten Unterstützung beim Erlernen einer Fremdsprache und erhöhen die positive Einstellung [7].

Beim Lernen einer neuen Sprache spielt vor allem der Aufbau eines Vokabulars eine große Rolle. Dieses bildet die Grundlage für das Erlernen einer Fremdsprache. Durch Wiederholung können die Wörter vertieft werden und durch den Kontext werden Assoziationen zwischen dem Wort und einem Objekt gebildet. Es besteht die Möglichkeit das Vokabellernen durch Augmented Reality umzusetzen. Hierbei werden Objekte in der Umgebung durch AR beschriftet. Daraus ergibt sich ein situiertes Vokabellernen. Die Vokabeln, welche in der Umgebung relevant sind, können sich besser eingeprägt werden. Der Lernende assoziiert sie mit einem ihm bekannten Objekt. Dadurch wird dieses Wort einfacher zu merken [15]. In der Abbildung 7 werden Alltagsgegenstände in einer fremden Sprache beschriftet. Über die Kamera des Smartphones oder Tablets werden die Gegenstände erfasst und mit der entsprechenden Software können die Beschriftungen in einer ausgewählten Sprache angezeigt werden.

Dies verstärkt den Lerneffekt beim Vokabellernen, da die Vokabel im direkten Zusammenhang mit dem Objekt gesehen wird. Der Nutzer assoziiert die Vokabel mit dem Gegenstand und erleichtert sich somit das Lernen.

4 Verbindung von Lerntheorien mit Mixed Reality

Nachdem in den letzten Kapiteln der Einsatz von Mixed Reality beim Lernen erläutert wurde, wird nun auf die Verbindung von Lerntheorien zu Mixed Reality eingegangen. Mit Hilfe von Lerntheorien können Lernvorgänge beschrieben werden. Es stellt sich jedoch die Frage, ob sich herkömmliche Lerntheorien mit Mixed Reality vereinen lassen. Bekannte Psychologen, unter anderem Wil-

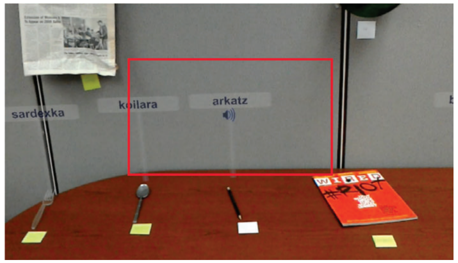


Abbildung 7 Vokabellernen durch Beschriftung [2]

liam Winn, sind der Meinung, dass Mixed Reality und Lerntheorien vereinbar sind [27]. Nach einer Erklärung und Benennung der bekanntesten Lerntheorien wird darauf näher eingegangen. Bei den Lerntheorien handelt es sich um den behavioristischen, den kognitiven, den motivationstheoretischen, den handlungsorientierten und den konstruktivistischen Ansatz. Bei dem behavioristischen Ansatz betrachtet der Psychologe das sichtbare Verhalten des Lernenden. Alle inneren Vorgänge werden nicht beachtet. Im Gegensatz dazu wird bei kognitivistischen Ansätzen davon ausgegangen, dass der Mensch durch Beobachtung lernt. Er orientiert sich an bekannten oder fremden Personen um sich Wissen anzueignen. Eine dritte bekannte Lerntheorie ist der motivationstheoretische Ansatz. Er konzentriert sich auf die Motivation, auf Grund welcher gelernt wird. Weiterhin beschäftigt sich eine vierte Theorie, der handlungsorientierte Ansatz, mit der Handlung des Lernenden, dem Soll-Ist-Vergleich. Zuletzt beschreibt auch der konstruktivistische Ansatz eine Lerntheorie, welche den Lernenden aktiv in den Mittelpunkt stellt [11].

Denkt man nun zurück an das Lernen durch Mixed Reality, so fällt auf, dass dies mit einigen der genannten Ansätze vereinbar ist. Mixed Reality wäre somit als Behandlungsform zum Lernen einsetzbar. William Winn vergleicht die unterschiedlichen Lerntheorien mit den Möglichkeiten, welche diese für das Lernen durch Mixed Reality bieten. Dabei stellt er vor allem den konstruktivistischen Ansatz in den Mittelpunkt. Seiner Meinung nach bietet der Konstruktivismus die beste Theorie, um pädagogische Anwendungen von VR zu entwickeln. Jedoch spielen auch die übrigen Ansätze eine Rolle. So ist der behaviouristische Ansatz für die traditionelle Unterrichtsplanung genutzt worden [27]. Bei Mixed Reality kann der Psychologe ebenfalls das Verhalten des Nutzers beobachten. Sowohl die Interaktion mit dem AR-Gerät als auch die Reaktion auf Situationen in VR können betrachtet und interpretiert werden. Weiterhin ist auch der motivationstheoretische Ansatz in Mixed Reality einsetzbar. Wie bereits erwähnt, ist die Motivation eine neuen Technologie zu nutzen zu Beginn sehr hoch. Aus diesem Grund kann die Motivation der Nutzung von MR-Geräte zu einem erhöhten Lernerfolg führen. William Winn legt seinen Fokus auf den konstruktivistischen Ansatz, da hierbei der Lernende in den Mittelpunkt gestellt wird. Dies ist bei Mixed Reality, vor allem jedoch bei Virtual Reality, der Fall. Hierbei spielt die Immersion die Hauptrolle. Erfahrungen machen einen großen Teil des Lernens aus. Dabei handelt es sich um alle Erlebnisse, welche der Umwelt weitgehend nicht mitgeteilt werden können. Durch Virtual Reality sammelt der Nutzer solche Erfahrungen. Es wird Wissen aus der direkten Erfahrung und nicht aus der Erzählung gezogen [27].

5 Conclusion

Zusammenfassend gilt zu sagen, dass Augmented Reality trotz seines jungen Alters bereits in vielen Bereichen des Lernens Einsatz findet. Da bei AR die Situation angesprochen wird, lernt der Nutzer immer durch das allgegenwärtige Erleben. Es konnte bereits in einigen Veröffentlichungen, unter anderem in dem Paper "Augmented reality as multimedia: the case for situated vocabulary learning", bewiesen werden, dass durch AR das Lernen erleichtert wird. Lernende schnitten bei Lerntest durch AR besser ab, als beim herkömmlichen Lernen. Durch AR wird die kognitive Belastung verringert, die Aufmerksamkeit verbessert und die Zufriedenheit erhöht. Außerdem können Gegenstände oder Beschriftungen angezeigt werden, welche in der realen Welt unmöglich sind. Die Erkundung dieser Bereiche bietet neue Möglichkeiten des Lernens. Neben abstrakten 3D-Darstellungen von Gegenständen können auch historische Gebäude wieder zum Leben erweckt werden. Es besteht die Möglichkeit Dinge der direkten Umgebung beschriften zu lassen. In allen Themengebieten kann dies zu einem erhöhten Lernergebnis führen. Vor allem im Lernen einer neuen Fremdsprache wird durch das Anzeigen von Vokabular der Lerneffekt gesteigert. Im Vergleich zum herkömmlichen Lernen wird bei AR der Fokus hauptsächlich auf den visuellen Kanal gelegt. Jedoch können auch auditive und taktile Sinne genutzt werden. Die in diesem Paper aufgeführten Beispiele sind natürlich nur einige aus einem unglaublich breiten Themengebiet. In nahezu allen Fachbereichen kann Mixed Reality zum effektiven, verteilten und flexiblen Lernen genutzt werden. Dies ist auch der Grund, warum dieser Themenbereich sich großer Beliebtheit erfreut.

Jedoch bietet das Gebiet von Virtual Reality und Augmented Reality noch viel Potential für Weiterentwicklungen. Ein Problem bildet die teilweise örtliche ungenaue Darstellung. Durch GPS-Fehler werden virtuelle Objekte an nicht realitätsgetreuen Orten angezeigt. Es kommt zu leichten Verschiebungen, wodurch dem Nutzer der Realitätsbezug verloren geht.

Eine weitere Problematik bilden die Kosten für die Technologie. Wie in diesem Paper bereits erwähnt, findet AR und VR heutzutage bereits über Smartphones und Tablets Einsatz. Jedoch sind dabei eingeschränkte Interaktionsmöglichkeiten geboten. Stärkere Geräte sind dagegen teurer und unhandlicher zu betätigen. Für den Hobbynutzer sind diese somit eher unrelevant.

Durch den Novelty-Effekt wird das Lernen durch AR und VR begünstigt. Da dieser nach gewisser Zeit verfliegt, sollte die Technologie den Fokus nicht auf die technischen Möglichkeiten sondern auf den Mehrwert fürs Lernen legen.

Abschließend ist zu sagen, dass das Lernen durch Mixed Reality großes Potential mit sich bringt. Die Technik muss jedoch weiter ausgearbeitet werden um einen Lernerfolg garantieren zu können.

Literatur

1 AR Anatomy. Retrieved December 03, 2019 from `https://pxhere.com/de/photo/1201753`.

2 Ibrahim Adam, Huynh Brandon, Downey, Höllerer Tobias, Chun Dorothy, and O'Donovan John. ARbis Pictus: A Study of Language Learning with Augmented Reality. 2017. `doi:10.1109/TVCG.2018.2868568`.

3 Consultancy.uk. Virtual and Augmented Reality market to boom to $170 billion by 2022, July 2018. Retrieved February 25, 2020 from `https://www.consultancy.uk/news/17876/virtual-and-augmented-reality-market-to-boom-to-170-billion-by-2022`.

4 Bhattacharjee Deblina and Paul Anand. An immersive learning model using evolutionary learning. *Computers and Electrical Engineering*, pages 236–249, 2018. `doi:10.1016/j.compeleceng.2017.08.023`.

5 LG Electronics. Retrieved February 25, 2020 from `https://www.flickr.com/photos/lge/5708231997`, License: CC BY 2.0 (`https://creativecommons.org/licenses/by/2.0/`).

6 Inside360 GmbH. Die Mehrwerte von Virtual Reality im Gesundheitswesen. Retrieved February 25, 2020 from `https://inside360.tv/blog/die-mehrwerte-von-virtual-reality-im-gesundheitswesen/`.

7 Ting-Chia Hsu. Learning english with augmented reality: Do learning styles matter? *Computers & Education*, 106:137 – 149, 2017. URL: `http://www.sciencedirect.com/science/article/pii/S0360131516302494`, `doi:https://doi.org/10.1016/j.compedu.2016.12.007`.

8 Hellriegel Jan and Cubela Dino. Das Potenzial von Virtual Reality für den schulischen Unterricht-Eine konstruktivistische Sicht und Lehrende im Lehramt View project. 2018. `doi:10.21240/mpaed/00/2018.12.11.X`.

9 Wu Hsin Kai, Lee Silvia Wen Yu, Chang Hsin Yi, and Liang Jyh Chong. Current status, opportunities and challenges of augmented reality in education. *Computers and Education*, 2013. `doi:10.1016/j.compedu.2012.10.024`.

10 Choi Kanghyun, Yoon Yeo Jin, Song Oh Young, and Choi Soo Mi. Interactive and immersive learning using 360° virtual reality contents on mobile platforms. *Mobile Information Systems*, 2018. `doi:10.1155/2018/2306031`.

11 Simone Kauffeld. *Nachhaltige Weiterbildung*. Springer Verlag, 01 2010. `doi:10.1007/978-3-540-95954-0`.

12 Thorsten Alexander Kern. *Entwicklung Haptischer Geräte: Ein Einstieg für Ingenieure*. 01 2009. `doi:10.1007/978-3-540-87644-1`.

13 D.C. Leonard. *Learning Theories, A to Z*. ABC-Clio ebook. Oryx Press, 2002. URL: `https://books.google.de/books?id=nNcoAO5Za9YC`.

14 Matthias Mangold. Unterschied Virtual Reality (VR) und Augmented Reality (AR). Retrieved November 26, 2019 from `https://magic-holo.com/unterschied-virtual-reality-vr-und-augmented-reality-ar/`.

15 Santos Marc, Ericson C., Lübke Arno in Wolde, Taketomi Takafumi, Yamamoto Goshiro, Rodrigo Ma Mercedes T., Sandor Christian, and Kato Hirokazu. Augmented reality as multimedia: the case for situated vocabulary learning. *Research and Practice in Technology Enhanced Learning*, 12 2016. `doi:10.1186/s41039-016-0028-2`.

16 Stambolieva Marija. Horizont erweitern: Aktives Lernen mit Virtual Reality/Augmented Reality, December 2017. Retrieved December 1, 2019 from `https://hochschulforumdigitalisierung.de/de/blog/horizont-erweitern-aktives-lernen-virtual-augmented-reality`.

17 Billinghurst Mark and Dünser Andreas. Augmented reality in the classroom. *Computer*, pages 56–63, 2012. `doi:10.1109/MC.2012.111`.

18 Rangel-Gomez Mauricio and Janenaite Sigita. Novelty's effect on memory encoding. *Acta Psychologica*, pages 14–21, 2015. `doi:10.1016/j.actpsy.2015.05.004`.

19 A. Mehler-Bicher and L. Steiger. *Augmented Reality: Theorie und Praxis*. De Gruyter, 2014. URL: `https://books.google.de/books?id=TqvoBQAAQBAJ`.

20 Paul Milgram and Fumio Kishino. A taxonomy of mixed reality visual displays. *IEICE Trans. Information Systems*, vol. E77-D, no. 12:1321–1329, 12 1994.

21 Fredriksson Pontus and Rödström Herman. Virtual Reality's Effect on Engagement in Educational Games. 2017.

22 Textor Martin R. Zukunftsentwicklungen. Retrieved November 26, 2019 from `http://www.zukunftsentwicklungen.de/technik.html`.

23 Döner Ralf and Broll Wolfgang. Virtual Reality und Augmented Reality (VR/AR): Auf dem Weg von der Nische zum Massenmarkt. *Informatik-Spektrum*, pages 30–37, 2016. `doi:10.1007/s00287-014-0838-9`.

24 ARTS Holding SE. Augmented Reality und Virtual Reality in der Industrie. Retrieved November 26, 2019 from `https://arts.eu/de_DE/blog/augmented-reality-and-virtual-reality-in-manufacturing-and-industry`.

25 Kind Sonja, Ferdinand Jan-Peter, Jetzke Tobias, Richter Stephan, and Weide Sebastian. Virtual and augmented reality. 2019. URL: `http://www.tab-beim-bundestag.de/de/pdf/publikationen/berichte/TAB-Arbeitsbericht-ab180.pdf`.

26 Höntzsch Susan, Katzky Uwe, Bredl Klaus, Kappe Frank, and Krause Dirk. Simulationen und simulierte Welten. Lernen in immersiven Lernumgebungen. 2013.

27 Winn William. A Conceptual Basis for Educational Applications of Virtual Reality. 1993.

Technology-Supported Context-Based Learning

Simon von der Au

Ludwig-Maximilians-Universität München, München, Deutschland
`simonanian.au@campus.lmu.de`

Abstract

The classroom is not the ideal place to acquire knowledge, however, for a long time, there has been no practical alternative. Prior research in pedagogical science shows that a learner's context has a major impact on gaining new knowledge. Over the last decades, context-based learning seemed impossible. The massive improvement in technologies now offers new ways for learning. This paper provides an overview of technology-supported learning approaches, which take the context of the learner into account. By presenting representative examples, this work shows how context-based learning can be realized. Furthermore, the context, which is used in these approaches, will be analysed to provide starting points for future work.

2012 ACM Computing Classification Human-centered computing → Empirical studies in ubiquitous and mobile computing

Keywords and phrases Context-Based Learning; Mobile Learning; Ubiquitous Learning; XR Learning.

1 Introduction

Over the last century, little has changed in the way we learn at school. Learning is more or less tied to the classroom. However, there has always been the attempt to teach not only at school desks, but outside in the field. Good examples for learning outside the classroom are foreign language exchange programs in secondary schools. A week in an Italian family is sometimes more effective for language learning than a year spent in the classroom. Besides the omnipresence of the foreign language in such a situation, the context the student is exposed to supports the language learning. This phenomenon is well researched in educational science [24, 11]. The context, in which learning takes place, has a huge impact on the gained knowledge. Therefore, learning Italian in Italy and using the language in an authentic manner can be much more effective than learning it in a classroom in a constructed and

© Simon von der Au;
licensed under Creative Commons License CC-BY

Cite as: Simon von der Au. Technology-Supported Context-Based Learning. In *5th Seminar on Ubiquitous Interaction (UBIACTION 2020)*. Editors: Florian Lang, Pascal Knierim, Jakob Karolus, Fiona Draxler, Ville Mäkelä, Francesco Chiossi, Luke Haliburton, Matthias Hoppe, Albrecht Schmidt. January 27, 2020. Munich, Germany. pp. 3:1–3:18.

abstract way. In the article "How Children Learn Words" by Miller and Gildea [14], they mark out that language learning is attached to context. The amount of vocabulary children learn in context is much bigger than what they learn in classes. However, providing learning content right at the time it is needed in context is largely impossible for schools and teachers. Modern technologies, like smartphones, in contrast, are capable of sensing the students' surroundings and therefore reconstruct the context the student is in and provide the corresponding educational material. This paper displays approaches which deal with this challenge.

In the following part, the essential theories for context-based learning and technology-supported learning will be summarized. In Section 3, context-dependent learning systems are presented sorted into mobile learning, Augmented-Reality and Virtual-Reality approaches. In Section 4 the context used in the prior examples is analysed. Research results and limitations will be discussed in Section 5. The last section deals with starting points and research gaps for future research.

2 Learning Theories

This section provides a theoretical background on why context in which learning takes place is important. It will define what exactly is meant by context and how context can be structured. Furthermore, a quick overview of technology-supported learning will be provided.

2.1 Authentic Learning

In their paper "Situated cognition and the culture of learning" Brown et al. [3] explain why the situation and activity in which we learn has an essential impact on gaining knowledge. People learn by extracting knowledge out of the context. As a prominent example, they mention how children learn their mother language. Unfortunately, learning in school is mostly driven by abstract concepts. Brown et al. propose that learning has to take place in the same context in which the knowledge is used later on. Learning in school is inauthentic, so students should gain their knowledge in *Authentic Activities*.

2.1.1 Authentic Activities

Learning in school often tends to be abstract and ripped apart from context of use. Activities of a domain are always related to its culture. Meaning, social groups and conceptual tools of the domain form an authentic environment to gain knowledge. As an example,learning how to repair a car should take place in a repair shop supervised by a car mechanic. In the first step, learners observe the mechanic repair a car, in the second step they work together with the mechanic and in the last step, they repair a car alone, observed and supervised by the mechanic. In this way, students can learn implicitly by observing and adapting the domain's context and explicitly by doing. Therefore, they will develop a deeper understanding of the learned content. Brown et al. suggest that problems should always be solved in the context they occur. In this way, students are encouraged to come up with their own solution path and apply knowledge rather than recall it.

2.1.2 Cognitive Apprenticeship

The higher the level of education, the harder it gets to learn in Authentic Activities. Abstract concepts and theories have to be imparted. Nevertheless, Brown et al. argue that these abstract concepts should be taught in an authentic manner. Through *Cognitive Apprenticeship* methods, students can be enculturated in a domain similar to craft apprenticeship. In order to gain authentic practice, activity and social interaction are important. Learning about an abstract mathematical concept should take place in an environment where this concept appears, surrounded by professionals who work with it. Also, to create ideal conditions for learning, the task should be embedded in situations familiar to students, to provide relevance to them. If tasks or learning contents gain relevance for learners, they are not abstract anymore. Content that is relevant to the learner increases learning success. Therefore, problems should be broken down into smaller sub-tasks and provided to students in a way they can empathise with.

> "Cognitive Apprenticeship supports learning in a domain by enabling students so acquire, develop, and use cognitive tools in authentic domain activity"
>
> — Brown et al. [3]

Learning content adapted to the users' context causes relevance and supports learning. Therefore learning should take place in an authentic context.

2.2 Context in Learning

Schmidt et al. [19] claim that location is commonly used in relation to context. However, context is not restricted to location. In their paper, they outline how context can be understood and used for application design. Schmidt et al. define the concept of context with four characteristics: (1) Context describes a situation and the environment it takes place in. (2) It is defined by a unique name. (3) It has relevant features and (4) each feature has a range of values determined by the context. Furthermore, they build up a framework to classify

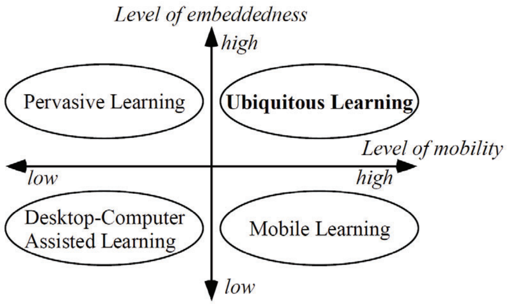


Figure 1 Comparison of learning environments [17]

context. They divide context into *Human Factors* and *Physical Environment*. Both can be subdivided further on. Human Factors can be understood as user information, social environment and the user's task. Physical environment can be considered as location, infrastructure and physical conditions. With the help of the provided framework, context can be built up hierarchically. This will be used to analyze the context of technology-supported context-based learning approaches.

2.3 Technology-Supported Learning

Creating Authentic Activities to support context-based learning is very complicated. In school, environment authenticity is hardly reachable. Providing lessons outside the classroom therefore needs technological support. Through emerging mobility of information technology, different learning environments and learning methods are possible. In the following, environments and concepts of technology-supported learning will be outlined. These concepts are applied in many technology-supported context-based learning approaches.

2.3.1 Learning Environments

Different learning environments can be applied to support learning. Based on the work of Lyytinen et al. [13], Ogata et al. [16, 17] adapted the dimensions of ubiquitous computing, to compare learning environments (see Figure 1).

For this paper, *Mobile Learning* and *Ubiquitous Learning* are relevant. Both terms are often used in a similar way, but in fact there is a difference. Mobile learning defines the capability to move the learning environment around [23]. Learners, equipped with mobile computing devices can learn wherever they like. Connected with the Internet, learners can access information any time they need it. However, mobile learning itself does not support context awareness. Ubiquitous Learning supports a high mobility on the one hand, on the other, it dynamically supports the learner by observing the environment. The system communicates with embedded computers in the surroundings or builds up its own understanding of the environment. Approaches presented in Section 3 are partly mobile learning environments and partly ubiquitous learning environments.

2.3.2 Integrated Micro Learning

Some of the approaches use a method called Micro Learning [8]. Integrated Micro Learning takes advantage of the learners' mobility and their use of technological devices. Learning material is split up into small sections and provided to the learners. Learners will be confronted with material in a more or less subconscious way, in gap times between tasks or by switching devices. Learning, therefore, is implemented in the daily routine of the learner. Combined with contextual understanding, micro learning has a huge potential and a motivational effect.

3 Examples

There are many ways how technology can support learning. Below, examples using mobile devices, Augmented-Reality devices and Virtual-Reality devices are presented. All these approaches support context-dependent learning in various ways and at different levels.

3.1 Context-Aware Mobile Learning

Ubiquitous computing and the availability of mobile devices like the early PDAs (Personal Digital Assistants) or today's smartphones enable learners to carry course material whereever they are. Using these devices combined with sensory data allows to provide context-dependent content. Learning content may be generated to the learners' best needs or to support learning in authentic activities. In the following, context-aware mobile systems are presented.

Ogata et al. [15] used PDAs to enable learning Japanese outside classrooms. They developed LOCH (Language-learning Outside the Classroom with Handhelds system) to assist exchange students to learn in real-life situations. Teachers could assign tasks to students and monitor progress or offer assistance. Students equipped with the PDAs solved tasks like interviewing locals or going shopping. PDAs are used for annotations, taking pictures or contacting the teacher, if they need help. For example, students can use the devices to record

Figure 2 Screenshot of the Live Wallpaper by Dearman and Truong [5]

voice messages to ask the teacher for translation or teachers can provide hints and suggestions. Participants in the study liked the approach and would like to carry on working with the system. Teachers liked the system as well, and found it useful for students to learn Japanese in a real social context. LOCH is a medium of communication between teacher and student, to assist learning in the student's context.

Another system developed by Ogata et al. [17] helps overseas students to learn complex Japanese polite expressions. The application has a model of the campus and detects the students' location with GPS and RFID tags. Furthermore, users have the possibility to enter information about themselves and their conversation partner. Based on the learners' location and the social parameters, polite expressions are recommended. In a study, users rated the system as useful for language learning.

Edge et al. [6] developed a context-aware mobile app, using micro learning principles for language learning. Depending on the users' location, the system displays flashcards related to the context. The location can be detected by GPS or manually set by the users. In their study, Edge et al. compared the context-dependent approach with a frequency approach. Results showed that more vocabulary was learned in the frequency approach, but contextual micro learning led to *situated cognition* and resulted in direct practical usage by the learner. This fact caused a higher motivation for learners and led to satisfying conversations with native speakers at the location.

Dearman and Truong [5] also used a micro-learning approach for second language learning. They developed a smartphone live wallpaper displaying second language vocabulary (see Figure 2). Whenever users looked at their phone, vocabulary was presented. In their study, the researchers compared context-dependent versions with a context-independent one. The context was detected through the users' GPS location. Participants reported that connecting vocabulary with related objects in the surroundings helped their learning. Study results showed that participants had greater recall of vocabulary when learning with a context-dependent version.

Another micro-learning approach was made by Beaudin et al. [2]. They took an ubiquitous-computing approach, attached RFID tags to objects and used motion sensors for context detection. The system supports second language learners over a longer period of time. It was therefore designed to be always switched on, but keep hands free. The system senses the objects a user is interacting with and provides vocabulary or phrases related to the object or action. For their prototype, Beaudin et al. used a mobile phone app and equipped a household with sensors into which participants moved. In their study, they evaluated the acceptance and usability of such a system. Based on participants' feedback, the system was acceptable even after extended usage and Beaudin et al. could develop design attributes for future systems.

Also using ubiquitous computing, Hsieh et al. [9] presented a mobile system for English learning in a campus environment. Learning material was displayed based on the users location on the campus. To sense the location, wireless access points on campus were used.

Cui and Bull [4] developed a highly sophisticated system for context-dependent language learning. The mobile application takes the users' knowledge state and location into account. Multiple-choice questions at the beginning help the system to define the users' knowledge. By adding parameters to the location, the users are able to define the context more specifically, like concentration level or likelihood of interruption as well as the time frame. According to the given parameters, adequate course material is presented to the learners. Unfortunately, the system was never evaluated. Even though the context is well determined, there is hardly any impact on the content itself. In this case, there is no authentic activity created.

3.2 Augmented-Reality-Supported Learning

Through the massive developments in the field of Augmented Reality (AR) in the past years, AR is available on a larger number of devices, such as smartphones and tablets. Azuma [1] defines AR systems: (1) As combination of the real and the virtual world, (2) responding in real time and (3) anchored in a three-dimensional space. Per definition, AR senses many contextual components and is able to respond accordingly. Besides latest popularity, AR is well fit to support learners. In his meta-review, Radu [18] summarizes the benefits and limitations of AR for learning. Among other things, AR-based approaches can support content understanding, improve collaboration and increase learner motivation. Furthermore, long-term memory retention is supported, of which language learning may benefit especially.

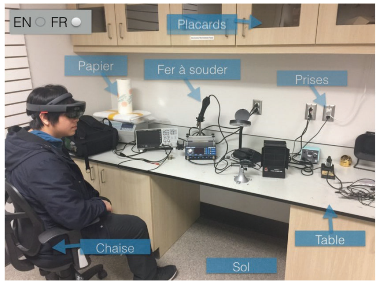
Focusing on motivational aspects, Taskiran et al. [22] developed four AR-based games for foreign language lectures. Students enjoyed enriched learning materials. Their study showed that AR games motivated participants. Furthermore, the games supported collaboration between participants, which led to an increase of retention while speaking and listening. The games took place in a classroom, but each game created a different context participants could empathize with. The vocabulary they learned depended on the game context.

In their literature review, Schmitz et al. [20] evaluated the potential of AR games for learning. They concluded that a motivational effect as well as enhanced collaboration is visible. However, most approaches, comparable to Taskiran et al. [22], do not focus on quantitative data about learning outcomes.

The work of Ibrahim et al. [10] focuses on the learners' physical surroundings and evaluates the system compared to traditional learning. They developed *ARbis Pictus*, an AR system for foreign language learning. An app for the *HoloLens* (an AR head-mounted display) labels real-world objects with virtual translations in the foreign language (see Figure 3a). Learners have the possibility to interact with labels and listen to audio recordings to find out how words are pronounced. Ibrahim et al. evaluated the system in comparison with traditional flashcards. The study showed that the context-based AR approach is more effective and preferred by participants compared to flashcards. This first approach was evaluated under test conditions, but for future systems, the team is testing machine-learning options for automatic labeling. In this case, the system would use the context of the learner in a very advanced way to support learning.

Also using the HoloLens, Strzys et al. [21] developed an app to visualize sensor data of a physics experiment. With the help of an infrared camera, temperature values of an experiment are recorded and overlaid onto the setup in real-time. Therefore, physics students can evaluate the effect of different materials through examination of temperature. In this case, the application is aware of the physical conditions of the context.

Wong et al. [26] developed AR trails to teach students the abstract concepts of integrity and ethics. The trails are created as additional part of the curriculum about academic integrity, ethics and intellectual property. Trough these trails, students should be motivated to gain a deeper understanding of the topic and be encouraged to discuss with peers. The trails consist of different scenarios triggered by a variety of technologies like QR-Code scanning, GPS, image recognition or Bluetooth. All scenarios have a direct relationship to the location or object that triggered an action. At a *library bock return box*, for example, a scenario is presented in which a book is not returned accordingly. Wong et al. researched usefulness of AR trails to educate students in academic integrity and ethics. They found out that participants were very engaged in the contextual AR learning scenarios. Also, students were motivated by the fact that scenarios took place in situations common to students everyday live. Comparing pre- and post-trail open questionnaires, participants used more specific terms about the topic in the latter. Connecting scenarios with student-relevant context seems to have had a positive effect on the learning outcome. Unfortunately, no control group was used to research differences between AR and classroom learning.



(a) Mock-up of the App *Arbis Pictus* by Ibrahim et al. [10]

(b) Study setup of *Berlin Kompass* by Kallioniemi et al. [12]

3.3 Virtual-Reality-Supported Learning

As mentioned before, besides location, time, surroundings and atmosphere are also important to context. Nonetheless, visiting relevant places may be challenging for learners. For example, it may be hard for students to visit a nuclear power plant to learn about nuclear fusion, or to listen to a speech of *Cicero* at the *Forum Romanum*. Virtual Reality (VR) can empower students to visit locations and situations in virtual worlds. Besides the physical environment, the social environment can be reconstructed as well. Kallioniemi et al. [12] and Garcia et al. [7] both use VR for foreign language learning.

In their paper, Kallioniemi et al. [12] present a virtual environment for foreign language learning. Using a collaborative multimodal system, two learners navigate through a virtual representation of Berlin (see Figure 3b). Both learners are placed in remote locations and can only reach their goal if they work together and communicate. In a study, Kallioniemi et al. investigated that the system is suited for authentic language learning scenarios because of the immersion. However, there was no focus on measuring learning or quality of conversation. They suggest to implement speech analysis in their future approach to measure quality and provide real-time feedback to learners.

Garcia et al. [7] demonstrate in their showcases how learners can study Spanish vocabulary in VR. While exploring the virtual environment, learners have to identify matching items to provided vocabulary. Garcia et al. also use gamification to additionally motivate learners. They report that their participants preferred the VR learning compared to traditional methods.

4 Context Analysis

To adapt learning to learner context, the presented examples use many different ways. The type of used context and complexity varies strongly. Determining location is quite easy today, but sensing surroundings or even users' moods is more challenging. To get a better overview what research focused on so far, the context of prior examples is analyzed. For analyzing, Schmidt et al. 's framework is used [19].

4.1 Physical Environment

A very simple indicator for users' context is the physical environment they are in. Google maps, for example, can tell a lot about users' context. Not only their location is known but there is also information about the location available, such as the type of location, frequency of use and reviews from other users, to name a few.

4.1.1 Location

Schmidt et al. [19] mention in their paper that location is very often used. Location is very easy to identify and to respond accordingly. It can be sensed with GPS or using wireless access points. Based on the information, a general guess of the user's surroundings can be made. Many of the provided examples [6, 5, 9] use location as the only context determination. Wong et al. [26] use location, but also provide the possibility to identify locations based on typical or tagged objects through Bluetooth and image detection. Kallioniemi et al. [12] use location to transfer the user to a different place to create an authentic activity in VR. The Japanese polite expression recommender system by Ogata et al. [17] also uses the location as a important context.

4.1.2 Objects

A few examples dive deeper into the location by using surroundings or objects. Beaudin et al. [2] are the only representatives of the mobile learning approaches. Their intelligent home is very sophisticated and provides a lot of information about the user's actions. Therefore, the system is able to adapt learning content in a detailed way. In contrast to mobile learning, all provided AR approaches somehow use objects. Taskiran et al. [22] use objects like photos or newspaper articles for their AR games. The approach of Strzys et al. [21] is completely orientated on single objects. Location as well as surroundings are irrelevant after the object is detected. Ibrahim et al. [10] use location to detect surrounding objects. In their future approach, they try to identify objects independent from the location by image detection and machine learning. Dealing with a similar problem of object identification, Garcia et al. [7] used a VR location where objects can be detected without machine learning.

4.2 Human Factors

Human Factors like the user's mood, social surroundings or intentions play a huge roll in identifying a user's context. However, these components are hard to find in context-based learning approaches.

4.2.1 User Tasks

Knowing users' intentions or the tasks they have to solve is relevant for the context the users are in. Based on the intelligent home, Beaudin et al. [2] have the possibility to create context, based on user movement or conclude intentions out of used objects. Ogata et al. [15] or Taskiran et al. [22] make up tasks and try to create them as authentically as possible.

4.2.2 Social Environment

Social environment has a huge impact on the users' context. Therefore, it is important to gain a better understanding of this part of the users' context. It is also an important factor for cognitive apprenticeship. A not very technical but efficient method to control the social environment is used by Ogata et al. [15]. By creating tasks with social interaction, learners have to deal with other persons and solve tasks in very authentic situations. In the Japanese polite expression recommender [17], the social context is the key to the system. However, the social environment has to be entered manually by the users. Kallioniemi et al. [12] and Taskiran et al. [22] developed their systems to encourage collaboration between users. In both approaches, users have to communicate in the created contexts to solve their tasks.

4.3 Interactivity

In authentic activities, there is the possibility to interact with the context and manipulate it. Gaining knowledge relies to a large extent on contextual feedback. Interaction can take place between users or between a user and objects. Approaches using social environment as context create interactivity between humans. However, besides human feedback, none of these systems provides automatic feedback based on the context. Being aware of that fact, Kallioniemi et al. [12] made the suggestion to implement speech analysis in their system. Strzys et al. [21] provided an interesting example how feedback based on human-object interaction can be displayed. Students have the possibility to manipulate the objects' conditions. The consequences of this manipulation are overlaid in AR on the objects. Learners get a direct feedback and can therefore observe the direct outcome of their action.

Cui and Bull [4] developed an approach combining physical environment as well as human factors for context creation. They let users choose and rate the context regarding location and social surrounding. Using this non technical way, they get very detailed information of the users' context. Their system uses this information only for adapting the quality and quantity of the learning content, but many other usages seem possible.

Analyzing how different examples use context also reveals another finding. The majority of the approaches try to detect a given context and respond accordingly. These approaches use technology to sense the situation the user is in. A few approaches try to create a new context. These approaches use technology to transfer the learner into an authentic activity.

To summarize the used contexts, most approaches use location. A few examples show the usage of objects the learner interacts with. Mainly AR-based systems use object detection. Human factors are rarely considered as contexts in the mentioned approaches. If used, human factors mainly focus on the social environment. Only one paper deals with a combination of human factors and physical environment. All implemented methods to detect the users context didn't go on a deep level. Physical factors like time of day, motion profiles or learner's habits haven't been used. Surprisingly, no example viewed the users themselves as relevant for context. The users' mental and physical state, both important factors for learning, was not even once considered as relevant to adapt learning content.

5 Discussion

All presented approaches for technology-supported context-based learning have promising findings. Modern technology, especially ubiquitous and mobile

computing as well as the Internet, support learning in authentic activities. These techniques enable learners in real time to access information whenever needed and wherever they are. By modifying the educational curricula and adapting them to students' contexts, an efficient way for learning outside the classroom seems possible. Moreover, as mentioned in Section 4, the full potential of context isn't used. This may have two reasons: (1) Results with context mainly limited to location are already satisfying or (2) the used technology was not as advanced as it is today. This fact applies especially to approaches made in the first decade of this century. Furthermore, the context's complexity increases when more details are considered.

Besides promising findings, this review also reveals limitations of prior research. Context-based technology-supported learning is mainly researched in language learning. Ten out of twelve presented examples deal with second language learning. Certainly, language learning is well suited for research. Learning can take place more or less everywhere and vocabulary is easily adaptable to the context. Furthermore, learning progress is easy to monitor. Learning physics for example, as described by Strzys et al. [21], is limited to locations or objects a student has to be close to. However, the question how well other areas of education can be covered with technology-supported context-based learning has to be answered.

An other fact is that most approaches focus on user experience in their studies. That means that most approaches study motivation, enjoyment or practicability. Only a few systems are evaluated regarding the learners success compared to traditional methods. This may be due to the fact, that measuring learning success is not trivial. Especially in these new approaches of learning, it is complicated to find traditional equivalents to compare with. In addition, the *Novelty Effect* has to be taken into account [25]. New technological approaches tend to cause excitement and enthusiasm and can affect learning positively.

Most of the studies were carried out in Asia. To overcome cultural differences, these studies should be evaluated also in other cultures.

All the presented approaches were evaluated under test conditions. In most cases, the learning content as well as the possible differences in the context were limited. To adapt the ideas of the presented approaches into school's daily routines, systems need to be adapted to a mass of users. Also, the context-related learning content has to be created in an automated way.

Technology-supported context-based learning offers a lot of benefits to the educational system. However, further research has to overcome the limitations until these systems can be adapted.

6 Future Work

As mentioned in the previous section, there are a few challenges to still focus on. By looking at limitations of the described approaches, starting points for future work can be determined.

At first, systems must be designed and tested for a bigger user group and a wider variety of educational fields. Key factor for the implementation is automatic content creation. If learning content can be adapted to the context automatically, a wider test range with more students and more context factors can be set up. Especially latest developments in machine learning and artificial intelligence provide interesting starting points for automated context-dependent content creation. Furthermore, it should be researched if technology-supported context-based learning provides benefits for other subjects as, for example, art, history or mathematics.

One strength of technology-supported learning is revealing invisible processes, as they occur, for example, in physics or chemistry. VR and AR are ideal technologies to enable learning by observation. Adding interactivity to objects enables students to get feedback based on their interactions. Therefore, *learning by doing* can be used risk free and even in abstract scenarios. Research should evaluate possible applications for higher education.

Human factors for context are mostly left out by researchers. Collaboration and social interaction are used in some research as part of authentic activities. However, there are only few approaches to technically support social interaction. Furthermore, no approach considers the learners themselves as relevant for context. Smartphones and wearables, which are always present, provide constant information about the users' conditions. This data should be used to adapt learning content for the users' current mental and physical state. For example, micro learning could take place in situations where users seems to be bored. The amount of motivation users need may depend on their mood, and their thinking capacity varies with their ability to concentrate. All these factors could be considered in a learning support system.

7 Conclusion

Learning in classrooms separates learning material from its natural context. However, researchers from educational science claim that context supports knowledge gaining enormously. This paper outlines technology-supported context-based learning approaches. It summarizes context-aware mobile learning, Augmented-Reality- and Virtual-Reality-supported learning. The analyzed context that is used in these approaches reveals research gaps and therefore starting points for future research. Especially the human component of context is hardly studied so far. Latest developments in machine learning and artificial intelligence enable a lot more possibilities to adapt learning content based on the context. Many new aspects can be studied. Therefore, it will probably last a few more years until students can learn outside classrooms.

References

1 Ronald T Azuma. A Survey of Augmented Reality. *Presence: Teleoperators & Virtual Environments*, 4(August):355–385, 1997. arXiv:arXiv:1011.1669v3, doi:10.1162.

2 Jennifer S. Beaudin, Stephen S. Intille, Emmanuel Munguia Tapia, Randy Rockinson, and Margaret E. Morris. Context-sensitive microlearning of foreign language vocabulary on a mobile device. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4794 LNCS:55–72, 2007. doi:10.1007/978-3-540-76652-0_4.

3 John Seely Brown, Allan Collins, and Paul Duguid. Situated cognition and the culture of learning. *Subject Learning in the Primary Curriculum: Issues in English, Science and Mathematics*, pages 288–305, 2005. doi:10.4324/9780203990247.

4 Yanchun Cui and Susan Bull. Context and learner modelling for the mobile foreign language learner. *System*, 33(2):353–367, 2005. doi:10.1016/j.system.2004.12.008.

5 David Dearman and Khai N. Truong. Evaluating the implicit acquisition of second language Vocabulary using a live Wallpaper. *Conference on Human Factors in Computing Systems - Proceedings*, pages 1391–1400, 2012. doi:10.1145/2207676.2208598.

6 Darren Edge, Elly Searle, Kevin Chiu, Jing Zhao, and James A. Landay. MicroMandarin: Mobile language learning in context. *Conference on Human Factors in Computing Systems - Proceedings*, pages 3169–3178, 2011. doi:10.1145/1978942.1979413.

7 Sarah Garcia, Ronald Kauer, Denis Laesker, Jason Nguyen, and Marvin Andujar. A virtual reality experience for learning languages. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, page INT039. ACM, 2019.

8 Gerhard Gassler, Theo Hug, and Christian Glahn. Integrated Micro Learning - An outline of the basic method and first results. *Interactive Computer Aided Learning*, pages 1–7, 2004.

9 Hsin-Chun Hsieh, Chih-Ming Chen, and Chin-Ming Hong. Context-aware ubiquitous english learning in a campus environment. In *Seventh IEEE international conference on advanced learning technologies (ICALT 2007)*, pages 351–353. IEEE, 2007.

10 Adam Ibrahim, Brandon Huynh, Jonathan Downey, Tobias Hollerer, Dorothy Chun, and John O'Donovan. ARbis Pictus: A study of vocab-

ulary learning with augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 24(11):2867–2874, 2018. `arXiv:1711.11243`, `doi:10.1109/TVCG.2018.2868568`.

11 Pauli Kaikkonen. Authentizität und authentische Erfahrung in einem interkulturellen Fremdsprachenunterricht. *Informationen Deutsch als Fremdsprache*, 29(1):3–12, 2017. `doi:10.1515/infodaf-2002-0102`.

12 Pekka Kallioniemi, Laura-Pihkala Posti, Jaakko Hakulinen, Markku Turunen, Tuuli Keskinen, and Roope Raisamo. Berlin Kompass: Multimodal Gameful Empowerment for Foreign Language Learning. *Journal of Educational Technology Systems*, 43(4):429–450, 2015. `doi:10.1177/0047239515588166`.

13 Kalle Lyytinen and Yoo Youngjin. Issues and Challenges in Ubiquitous Computing. *Communications of ACM*, 45(12):181–201, 2002. `doi:10.4018/978-1-5225-3878-3.ch012`.

14 George A. Miller and Patricia M. Gildea. How children learn words. *Scientific American*, 257(3):94–99, 1987. Last accessed February 22th, 2020. URL: `http://www.jstor.org/stable/24979482`.

15 Hiroaki Ogata, Gan Li Hui, Chengjiu Yin, Takahito Ueda, Yasuko Oishi, and Yoneo Yano. LOCH: supporting mobile language learning outside classrooms. *International Journal of Mobile Learning and Organisation*, 2(3):271–282, 2008. `doi:10.1504/IJMLO.2008.020319`.

16 Hiroaki Ogata and Yoneo Yano. How ubiquitous computing can support language learning. *Proc. of KEST 2003*, pages 1–6, 2003.

17 Hiroaki Ogata and Yoneo Yano. Context-aware support for computer-supported ubiquitous learning. In *The 2nd IEEE International Workshop on Wireless and Mobile Technologies in Education, 2004. Proceedings.*, pages 27–34. IEEE, 2004.

18 Iulian Radu. Augmented reality in education: A meta-review and cross-media analysis. *Personal and Ubiquitous Computing*, 18(6):1533–1543, 2014. `doi:10.1007/s00779-013-0747-y`.

19 Albrecht Schmidt, Michael Beigl, and Hans W. Gellersen. There is more to context than location. *Computers and Graphics (Pergamon)*, 23(6):893–901, 1999. `doi:10.1016/S0097-8493(99)00120-X`.

20 Birgit Schmitz, Roland Klemke, and Marcus Specht. An analysis of the educational potential of augmented reality games for learning. *CEUR Workshop Proceedings*, 955(October):140–147, 2012.

21 M P Strzys, S Kapp, M Thees, P Lukowicz, P Knierim, A Schmidt, and J Kuhn. Augmenting the thermal flux experiment: a mixed reality approach with the HoloLens. *The Physics Teacher*, 55(6):376–377, 2017. `arXiv:arXiv:1709.01342v1`.

22 Ayse Taskiran. The effect of augmented reality games on English as foreign language motivation. *E-Learning and Digital Media*, 16(2):122–135, 2019. `doi:10.1177/2042753018817541`.

23 John Traxler and Agnes Kukulaska-Hulme. *Mobile Learning the next Gerneration.* Routledge, New York, New York, USA, 2016.

24 L. Van Lier. *Interaction in the Language Curriculum: Awareness, Autonomy, and Authenticity.* Applied linguistics and language study. Longman, 1996. Last accessed February 22th, 2020. URL: `https://books.google.de/books?id=B8QbAQAAIAAJ`.

25 John D Wells, Damon E Campbell, Joseph S Valacich, and Mauricio Featherman. The Effect of Perceived Novelty on the Adoption of Information Technology Innovations : A Risk / Reward Perspective. *Decision Sciences*, 41(4):813–843, 2010.

26 Eva Y.W. Wong, Theresa Kwong, and Mark Pegrum. Learning on mobile augmented reality trails of integrity and ethics. *Research and Practice in Technology Enhanced Learning*, 13(1), 2018. `doi:10.1186/s41039-018-0088-6`.

Visualisieren von physiologischen Daten

Alexander Hiesinger

Ludwig-Maximilians-Universität München, München, Deutschland
`alexander.hiesinger@campus.lmu.de`

Zusammenfassung

Heutzutage werden immer mehr physiologische Daten über Personen gesammelt. Die Daten alleine nutzen jedoch nicht viel, sie müssen aufbereitet und visualisiert werden. Doch wie sieht eine entsprechenden Visualisierung von physiologischen Daten aus? Ein Sportler benötigte andere Informationen über seinen Körper und seinen Leistungsstand, als ein Kind, welches den Aufbau des Körpers und seiner inneren Organen lernt. Es gibt verschiedene Arten der Visualisierung von Daten, je nach Benutzer und Anwendungsgebiet muss die richtige Wahl getroffen werden. Werden Daten falsch visualisiert kommt es zu Unverständnis dem Benutzer gegenüber und er wird in die Irre geführt. Worauf bei der Visualisierung geachtet werden muss, welche Arten der Visualisierung für physiologische Daten es gibt und was überhaupt physiologische Daten sind wird in dieser Arbeit näher erläutert.

2012 ACM Computing Classification Human-centered computing → Visualisation → Visualisation systems and tools

Keywords and phrases HCI, physiological data, visualisation

1 Einleitung

In der heutigen Zeit werden immer mehr Daten über einen selbst gesammelt. Koordinaten mit Orten an denen man war, die Zahlungsart an der Kasse oder der Verlauf des Webbrowsers, um nur ein paar Beispiele zu nennen. Vermehrt werden auch Daten über den eigenen Körper gesammelt. Das Smartphone in der Hosentasche oder die Smartwatch am Handgelenk [16] sammeln fleißig Daten wie beispielsweise die zurückgelegten Schritte oder den Herzschlag. Mit den gesammelten Daten kann der Mensch so aber nichts anfangen. Es sind einfach zu viele und jede Sekunde werden es mehr und mehr. Da der Mensch nur eine beschränkte Wahrnehmung besitzt, müssen die gesammelten Daten gefiltert und visualisiert werden, dadurch erhält der Anwender schnell und einfach die wichtigsten Informationen kurz und prägnant. Was bedeutet jedoch Visualisieren von Daten? Visualisierung bedeutet, das Sichtbarmachen von Inhalten, in Form eines Bildes oder Textes beziehungsweise einer Kombination davon. Es wird verwendet, um etwas anschaulich darzustellen und zu verdeutlichen. Die

© Alexander Hiesinger;
licensed under Creative Commons License CC-BY
Cite as: Alexander Hiesinger. Visualisieren von physiologischen Daten. In *5th Seminar on Ubiquitous Interaction (UBIACTION 2020)*. Editors: Florian Lang, Pascal Knierim, Jakob Karolus, Fiona Draxler, Ville Mäkelä, Francesco Chiossi, Luke Haliburton, Matthias Hoppe, Albrecht Schmidt. January 27, 2020. Munich, Germany. pp. 4:1–4:14.

Einsatzgebiete der Visualisierung sind sehr vielfältig. Es findet beispielsweise Anwendung beim Moderieren, Präsentieren oder auch in der Wissensvermittlung (Unterricht). Die Visualisierung ist jedoch nicht immer leicht. Es wird unwissentlich die falsche Darstellungsform gewählt oder wissentlich werden Daten außer acht gelassen, wodurch sie falsch wiedergegeben werden. Das führt zu Unverständnis dem Benutzer gegenüber, der aus den visualisierten Daten seine Informationen gewinnt. Die Kunst, bei der Visualisierung von Daten, ist die Wahl der richtigen Fragestellung und Darstellungsform. Warum brauchen wir das ganze Sammeln und Visualisieren von physiologischen Daten überhaupt? Visualisierungen sind in der Personalinformatik wichtig [11], sie helfen dem Benutzer, sein Verhalten zu reflektieren und Erkenntnisse über sein Verhalten zu gewinnen, das ermutigt sie dazu, weiter ihre Ziele zu verfolgen [12].

Profisportler messen ihre Leistungen um ihren Fortschritt über einen längeren Zeitraum zu dokumentieren und ihre Schwächen zu erkennen. Durch eine kontinuierliche Überwachung und Auswertung der Daten die ihr Körper liefert, sind sie in der Lage, sich optimal auf den nächsten Wettkampf vorzubereiten und Verletzungen vorzubeugen. Für einen Marathonläufer ist zum Beispiel die Pace (Minuten pro Kilometer) interessant, da er sie mit alten Werten Vergleichen kann und sieht wie und ob er sich verbessert hat. Aber nicht nur Profisportler profitieren von einer Visualisierung der Körperdaten. Es ist auch für Leute interessant die Sport betreiben indem sie beispielsweise nach der Arbeit eine Runde laufen oder ins Fitnessstudio gehen. Das Smartphone, mit einer entsprechenden App ausgestattet, bietet eine einfache Auswertung, es wird keine Workstation benötigt und es liefert die Visualisierungen in Echtzeit [3]. Viele Apps wie zum Beispiel Fitbit[1], Dailyburn[2], Runkeeper[3] oder Apple Health App[4] die physiologische Daten ihres Benutzers Auswerten, zeigen nicht nur die Vitalität des Benutzers, sie geben auch Tipps zur Ernährung und motivieren mit einem Erfolgssystem. Die freigeschalteten Erfolge können mit Familie und Freunden geteilt werden um den eigenen Fortschritt zu präsentieren. Der Anwender kann sich selbst etwas beweisen, indem er versucht seine bisherigen Leistungen zu übertreffen oder gegen seine Freunden in einem Wettkampf antritt. Die steigende Nachfrage nach Body-Tracking-Technologien erhöht auch das Interesse an der Erforschung von "embodied interaction" [7].

In der Medizin wird dadurch die Forschung unterstützt. Die Daten dienen dem Erforschen von Krankheiten und Verhaltensmustern, da die Ärzte die Veränderungen und Reaktionen im Körper messen, darstellen und auswerten können. Ein Patient, beispielsweise ein Diabetiker zeichnet seine Insulin Zufuhr auf und kann die Ergebnisse mit seinem Arzt teilen und besprechen. Aus den so gewonnen Erkenntnissen können anschließend neue Behandlungsmethoden abgeleitet werden.

Die Visualisierung von physiologischen Daten kann auch sehr gut in der Bildung zum Einsatz kommen [2], hierauf wollen wir näher eingehen. Kinder lernen indem sie Dinge ausprobieren, fühlen und erforschen. Im Gegensatz zu Fingern, Armen, Zehen und anderen äußeren Körperteilen ist das Erlernen der Position, Struktur und Funktion von inneren Körperteilen eine Herausforderung [23, 18]. Die inneren Organe sind unter Schichten von Haut, Muskeln und Gewebe verborgen und arbeiten ohne bewusste Gedanken. Das macht es schwierig für Kinder die innere Funktionsweise ihres Körpers zu verstehen. Das Wissen über den Körper ist jedoch wichtig, je früher Kinder ihren Körper kennen lernen desto eher achten sie auf sich selbst und ihre Gesundheit [18].

Das Problem der Visualisierung von physiologischen Daten ist die richtige Wahl der Form. Wie bereitet man dem Nutzer am besten physiologische Daten auf, damit er sie schnell und einfach versteht und seinen Nutzen aus den so neu gewonnen Erkenntnissen ziehen kann? Eine häufig genutzte Form der Visualisierung ist das Darstellen von Grafiken und Text. Es gibt aber auch noch andere Möglichkeiten der Visualisierung, wie zum Beispiel die haptische Visualisierung. In den Folgenden Kapiteln werden verschiedene Arten der Visualisierung von physiologischen Daten gezeigt.

2 Physiologische Daten

Damit physiologische Daten visualisiert werden können, müssen sie erst einmal erfasst werden. Physiologische Daten werden aus Biosignalen gewonnen, das sind Signale von lebenden Organismen. Sie liefern Informationen über die biologischen und physiologischen Strukturen des lebenden Organismus und deren Dynamik. Es gibt verschiedene Art von Biosignalen die bei einem lebenden Organismus erfasst werden können, im folgenden sind einige Beispiele von Biosignale aufgelistet [17, 4].

- **Bioelektrische Signale:** Signale, die von Nerven und Muskeln stammen.
- **Elektrische Leitfähigkeit :** Die Variation der elektrischen Leitfähigkeit der Haut, ändert sich z.B. wenn die Haut schwitzt bei körperlicher Anstrengung, da sie primär durch die Aktivität der Schweißdrüsen beeinflusst wird.
- **Bioimpedanz Signale:** Der Widerstand, der beim Anlegen eines kleinen Wechselstroms an das Gewebe gemessen wird. Einsatzgebiet ist die Be-

---

[1] www.fitbit.com
[2] www.dailyburn.com
[3] www.runkeeper.com
[4] www.apple.com/lae/ios/health/

stimmung der Körperzusammensetzung von Fettmasse, Körperwasser und Muskelmasse.
- **Bioakustische Signale:** Durch Veränderungen im Körper, beispielsweise Herzfrequenz, Durchblutung oder Atemfrequenz, entstehen Geräusche die mit Mikrofonen erfasst werden.
- **Biooptische Signale:** Veränderung der optischen Eigenschaften (für das menschliche Auge teilweise nicht sichtbar) eines Körperteils, z.B. Messung des Herzschlags mit optischen Sensoren, an der Unterseite einer Smartwatch, durch zählen der Impulse in den Adern.

Ein paar dieser Signaltypen können, mit Hilfe von im Handel erhältlichen Geräten wie Smartphones und Smartwatches, erfasst werden. In modernen Smartphones und Smartwatches sind Gyrosensoren verbaut, die nicht nur der Ausrichtung des Displays in hoch und quer Format dienen, sondern sie messen auch die Beschleunigung und Richtung einer Bewegung. Im Fall der Smartwatch ist beispielsweise auch noch ein Herzschlag Sensor auf der Unterseite angebracht. In Kombination der beiden Sensoren können so aus den gewonnenen Daten Rückschlüsse auf die zurückgelegte Strecke und die Intensität eines Lauftrainings gezogen werden.

3 Visualisierung

> **vi · su · al · ize**
>
> **1.** to form a mental image of
> **2.** to make visible
>
> — The American Heritage Dictionary [6]

Visualisierungen helfen dem Menschen zu denken, sie formen ein mentales Bild, machen Zusammenhänge sichtbar, reduzieren die Belastung des Arbeitsspeichers im Gehirn und nutzen die Stärken der menschlichen Wahrnehmung. Bei der Visualisierung von Daten, dem Sichtbarmachen von Inhalten und Zusammenhängen, muss stets auf die richtige Wahl der Darstellung geachtet werden. Soll eine Grafik verwendet werden, nur Text oder eine Kombination aus beidem? Je nach Datentyp und Anwendungsgebiet ergeben sich andere Anforderungen an die Visualisierung, diese Anforderungen gelten auch für die Visualisierung von physiologischen Daten. Wird eine falsche Form der Visualisierung gewählt, wird der Benutzer in die Irre geführt, falsch informiert und es kommt zu Unverständnis. Bei der klassischen grafischen Visualisierung [22] kommen verschiedene Arten von Grafiken zum Einsatz, wie zum Beispiel Balkendiagramm, Kuchendiagramme oder die Karte eines U-Bahnnetzes. Sie informieren den Benutzer über die wesentlichen Inhalte eines Datensatzes

und erleichtern das Vergleichen mit anderen Werten. Es besteht auch die Möglichkeit Visualisierungen so zu gestalten, dass der Nutzer sie anfassen kann, um haptisches Feedback zu erhalten. Im folgenden werden verschiedene Arten der Visualisierung von physiologischen Daten gezeigt. Die ersten beiden Visualisierungen kommen im Bildungsbereich zum Einsatz, die letzte der drei ist eine abstrakte Darstellung von physiologischen Daten.

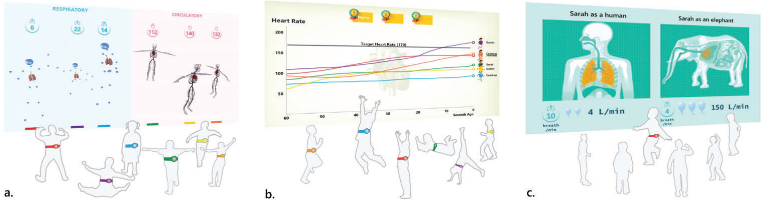


Abbildung 1 *SharedPhys* [10] kombiniert physiologische Sensorik, Ganzkörperinteraktion und Echtzeit-Großbildvisualisierungen, um eine neu Art von Körperinteraktionen und Lernerfahrungen zu schaffen. Zu sehen sind die drei interaktiven *SharedPhys*-Prototypen: **(a)** Magischer Spiegel, **(b)** Bewegte Grafiken und **(c)** Tier-Avatar.

3.1 Mixed-Reality Visualisierung

SharedPhys ist eine Mixed-Reality Visualisierung, die die virtuelle Realität (VR) mit der realen Welt verschmelzen lässt [10, 14] (siehe Abbildung 1). Sie ermöglicht physiologische Echtzeit-Sensorik eng mit Ganzkörperinteraktionen zu kombinieren, um eine neue Form der verkörperten Interaktion und des kollaborativen Lernens zu erschaffen. Mit dem System interagiert der Benutzer über Körperbewegung, Gestik und Position als auch über die sich verändernde Physiologie. Hierbei werden die physiologischen Daten des Körpers in Echtzeit erfasst, analysiert und weiterverarbeitet. Der Benutzer wird in den Prozess aktiv mit einbezogen, was neue Möglichkeiten für Feedbackschleifen und spielerische Experimente schafft. Im Folgenden werden drei Prototypen von *SharedPhys* genauer erläutert. Der Magischen Spiegel (siehe Abblindung 1**a**) ermöglicht dem Benutzer das menschlichen Atmungs- und Kreislaufsystems zu erkunden und näher kennenzulernen, sowohl die Position, Form und Größe relevanter interner Körperteile als auch deren Funktionen. Die Bewegten Grafiken (siehe Abblindung 1**b**) zeigen die Beziehung zwischen dem menschlichen Atmungs- und Kreislaufsystems und körperlicher Aktivität. Hierbei kommen Liniendiagramme für die Visualisierung zum Einsatz um die Echtzeit-Herzfrequenzen von bis zu sechs Benutzern in den letzten 60

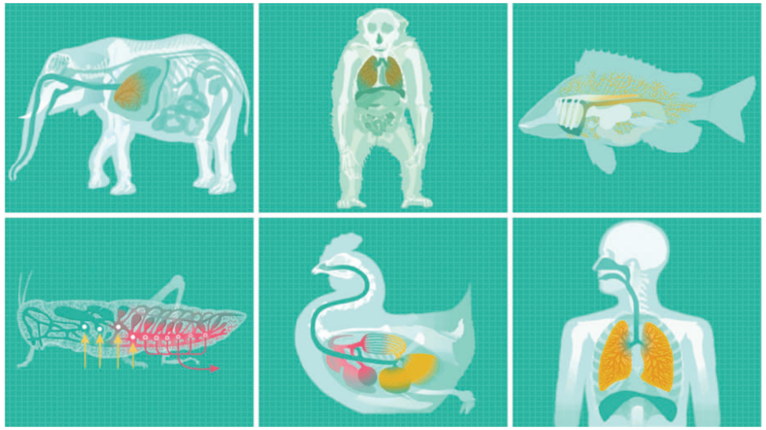


Abbildung 2 *SharedPhys* [10]: Im Tier-Avatar stehen sechs verschiedene Avatare zur Auswahl. Sie reagieren auf die empfundene Physiologie des Benutzers, diese wird in die Form des ausgewählten Avatars eingepasst und visualisiert so die anatomischen Abläufe.

Sekunden darzustellen. Die Visualsierung von *SharedPhys* ermöglicht sowohl zeitliche Vergleiche, beispielsweise “Wie verändert sich meine Herzfrequenz über die Zeit? “ als auch soziale Vergleiche mit anderen Benutzern, “Wie verhält sich meine Herzfrequenz im Vergleich zu Maya? “. Mithilfe des Tier-Avatars (siehe Abblindung 1**c**) werden dem Benutzer die biologischen Abläufe von Lebewesen genauer erklärt. Am Anfang wird zwischen einem von sechs zur Auswahl stehenden Lebewesen, Elefant, Schimpanse, Fisch, Heuschrecke, Huhn oder menschliches Kind gewählt (siehe Abbildung 2). Anschließend soll der Benutzer über seinen ausgewählten Avatar nachdenken und versuchen sich mithilfe von Geräuschen und Bewegungen in dessen Rolle zu versetzen. Die Visualisierung zeigt beispielsweise welcher Avatar Löcher in seinem Körper zum Atmen benutzt oder welcher gleichzeitig ein- und ausatmen kann. Die Mixed-Reality Visualisierung zeigt das Zusammenspiel von Aktion (z.B. laufen) und Reaktion (z.B. erhöhte Herzfrequenz). Die Informationen werden aber nur digital bereitgestellt, eine gute Ergänzung dazu stellt die haptische Visualisierung dar, die im folgenden Abschnitte genauer erläutert wird.

3.2 Haptische Visualisierung

Bei der haptischen Visualisierung agiert der Benutzer mit greifbaren Interfaces. Er kann die Visualisierung anfassen, sie erleben, sie fühlen. Die Gestaltung

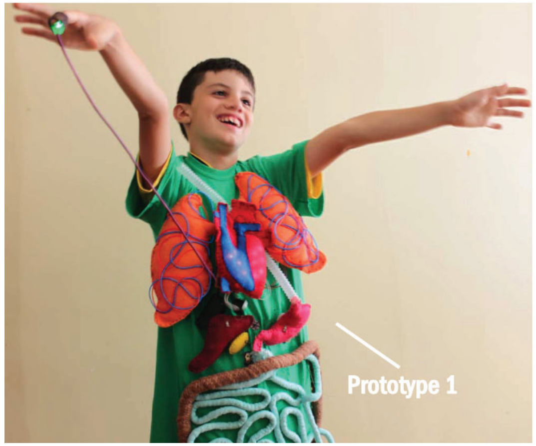


Abbildung 3 *BodyVis* [15] ist ein speziell designtes, maßgeschneidertes tragbares Textilhemd, das biometrische Sensorik und interaktive Visualisierung kombiniert. Es reagiert auf die physiologischen Phänomene seines Trägers und visualisiert die Körperdaten. Zu sehen ist der erste Prototyp.

von haptischen Visualisierungen erfordert nicht nur die Gestaltung des Digitalen, sondern auch des Physischen und deren Wechselbeziehung. Außerdem müssen neue und passende Interaktionstypen entwickelt werden, die sich als Ganzkörper-, Haptik- und Rauminteraktion charakterisieren lassen [9]. Interaktionsobjekte, die auf den Erfahrungen des Benutzers mit der realen Welt beruhen, senken die Aktivitätsschwelle. Sie eignen sich hervorragend zur Unterstützung von Face-to-Face-Sozialinteraktion, was sich auch in einer beträchtliche Anzahl von Systemen widerspiegelt [1, 20, 21]. *BodyVis* ist beispielweise eine solche haptische Visualisierung. Es ist ein speziell designtes, maßgeschneidertes, tragbares Textilhemd, das biometrische Sensorik und interaktive Visualisierung kombiniert [15]. Mit *BodyVis* werden innere Organe und Funktionen des menschlichen Körpers sichtbar gemacht, die sonst unsichtbar sind (siehe Abbildung 3). Die physiologischen Phänomene des Trägers werden auf der externen Anatomie des Textilhemdes visualisiert. Dadurch erlangen der Träger und die Betrachter einen einzigartigen Blick auf den inneren Körper. *BodyVis* ist mit einem herausnehmbaren Organsystem ausgestattet. Einzelnen Organe wie das Herz, die Leber, die Lunge und Teile des Magens können mit Hilfe von Magneten entfernt und wieder angebracht werden. Der Träger oder anderen Personen können die Organe dynamisch entfernt und wieder befestigten, um die verschiedenen Schichten des Körpers zu untersuchen (siehe Abbildung 4). Wird ein Organ entfernt, hört es auf zu arbeiten und die Animation stoppt, startet aber automatisch, wenn es wieder befestigt wird.

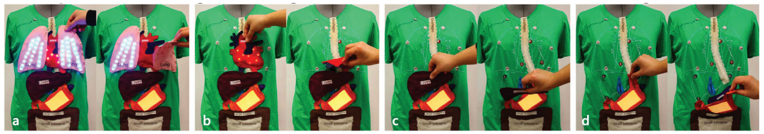


Abbildung 4 *BoyVis* [15]: Mit Hilfe von Magneten können die einzelnen Organe, **(a)** die Lunge, **(b)** das Herz, **(c)** die Leber und **(d)** ein Teil des Magens entfernt und wieder angebracht werden. Dadurch werden die Lernenden in die Lage versetzt, die verschiedenen Organschichten des menschlichen Körpers zu erkunden.

3.3 Abstrakte Visualisierung

Im freien Handel gibt es viele verschiedene Fitnesstracker und Apps, wie zum Beispiel Fitbit[5], Dailyburn[6] oder Runkeeper[7] zu erwerben. Die Meisten von ihnen verwenden Diagramme (Balken-, Linien-, Kuchendiagramme) um die körperlichen Aktivitäten ihres Trägers darzustellen. Es gibt auch Lösungsansätze die keine Diagramme nutzen, welche erfolgreich dazu eingesetzt werden können Informationen über physiologische Daten zu vermitteln. Im Gegensatz zu herkömmlichen Visualisierungen [22] sind sie stimmiger, ansprechender vom Design und motivierender (z.B. Blumen [5], Fische [13]). Der Vorteil von lebenden Metaphern ist, dass sie den Nutzer belohnen wenn er sich körperlich aktiv betätigt hat, indem die Fische fröhlich sind und die Blumen blühen. Allerdings können solche Visualisierungen auch demotivierend sein, da die Fische traurig sein können und die Blumen verwelken, wenn der Nutzer sich nicht körperlich betätigt hat.

Einen neuen Ansatz zur Visualisierung von physiologischen Daten bietet *Spark* (siehe Abbildung 5). Es ist ein informatives Kunstdisplay, das zum Nachdenken über körperliche Aktivität anregt [8]. Die körperliche Aktivität des Benutzers werden auf eine abstrakte Art visualisiert, dadurch versucht es neutraler zu erscheinen und ist potenziell weniger entmutigend als lebende Metaphern. *Spark* stellt dem Benutzer fünf verschiedene Visualisierungen zur Auswahl, sie sollen ihm die Wahl lassen welche Art der Darstellung am ansprechendesten ist und ihn am meisten motiviert. In der gezeigten Abbildung 5 werden die Schritte eines Nutzers über einen ganzen Tag hinweg visualisiert.

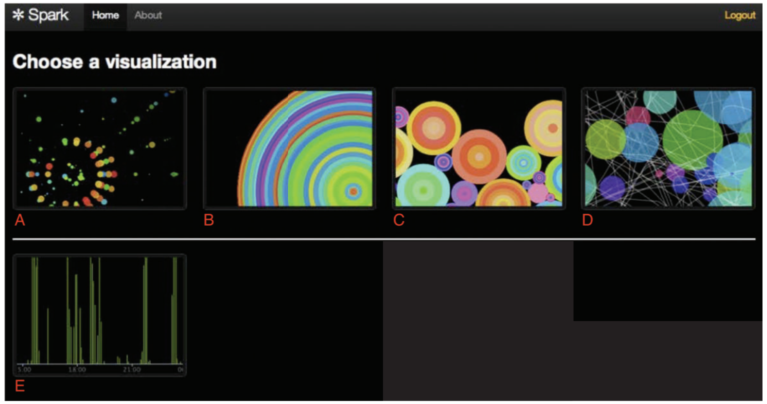


Abbildung 5 Auf dem Hauptbildschirm von *Spark* [8] trifft der Benutzer eine Auswahl zwischen den fünf verschiedenen Visualisierungen. **(a)** Spirale, **(b)** Ringe, **(c)** Eimer, **(d)** Pollock und **(e)** Säule.

- **Spirale:** In der Mitte des Bildschirms wird alle fünf Minuten ein Kreis erzeugt, der die Schrittanzahl darstellt. Dieser neue Kreis schiebt die älteren Kreise spiralförmig nach außen. Die jüngsten Schritte sind bei dieser Visualisierung in der Mitte, die ältesten am weitesten außen.
- **Ringe:** Ringe ist inspiriert durch Baumringe. Alle fünf Minuten wird im Zentrum ein neuer Ring erzeugt, so wächst die Visualisierung im Laufe des Tages an. Im Gegensatz zur Spiralenvisualisierung zeigt Ringe nur Momente der Aktivität an.
- **Eimer:** Beim Eimer fallen alle fünf Minuten Kreise von oben herab, die die entsprechende Anzahl an zurückgelegten Schritten widerspiegeln. Die Idee dabei ist möglichst große Kreise zu erzeugen, damit sich der Eimer füllt und das Tagesziel erreicht wird, sobald die Kreise die obere Kante berühren.
- **Pollock:** Inspiriert vom Künstler Jackson Pollock, bewegt sich eine durchgehenden weißen Linie zufällig über den Bildschirm. Alle fünf Minuten erscheint dann ein Kreis in Relation zur Schrittanzahl an der Stelle, an der die Linie in diesem Moment gezeichnet wird.
- **Säulen:** Die Säulen stellen ein Balkendiagramm dar. Hier kann der Nutzer genaue Werter und den zeitlichen Verlauf seiner Aktivität ablesen.

---

[5] www.fitbit.com
[6] www.dailyburn.com
[7] www.runkeeper.com

Spark verwendet verschiedene Farben, sie spiegeln die Intensität der Aktivität wieder. Blau und Grün stehen für lässiges Gehen, Gelb und Rot für zügiges Gehen und Violett und Pink für laufen. Je größer der Kreis ist, desto mehr Schritte wurden in diesen fünf Minuten zurückgelegt. Durch Anklicken eines Kreises auf dem Display, ist es dem Benutzer möglich genauere Informationen zu erhalten, es werden der Zeitstempel und die Anzahl der Schritte eingeblendet. Eine Abstrakte Visualisierung von physiologischen Daten bietet die Möglichkeit dem Nutzer Informationen, einfach, stimmig und ansprechend vom Design zu vermitteln und dabei motivierend zu sein.

4 Anwendungszwecke

Im Folgenden werden verschiedene Anwendungsgebiete aufgezeigt in denen die Visualisierung von physiologischen Daten zum Einsatz kommt. Dabei dienen die zuvor vorgestellten Visualisierungen als Beispiele.

4.1 Lehre

Die inneren Organe des Menschen sind nicht sichtbar, was es Kindern erschwert die Größe, Position und Funktion der Organe zu erlernen. In der Schule wird die Anatomie im Zusammenspiel mit der physiologischen Aktivität mit Hilfe von Büchern oder 3D-Modellen gelehrt. Die Visualisierung der physiologischen Daten bietet den Vorteil, dass sie sich in Echtzeit anpassen und verändern (z.B. *SharedPhys*), wohingegen Bücher nur Bilder und Text zeigen. Eine enge Verbindung zwischen physischer Aktivität, Wahrnehmung und Visualisierung in einer Umgebung, ermöglicht es Kindern leichter die Abläufe im Körperinneren zu verstehen und zu lernen. Mit dem Visualisierungsansatz des *BodyVis* Anzugs wird es Kindern ermöglicht die inneren Organe des menschlichen Körpers zu verstehen, Abläufe und Zusammenhänge zu erkennen und zu lernen. Die Ursachen-Wirkungs-Beziehung zwischen den einzelnen Organen und physischen Aktivitäten kann mit *SharedPhys* spielerisch erforscht werden, zudem wird die soziale Interaktion mit den Mitschülern gefördert [19]. Diese beiden Arten der Visualisierung von physiologischen Daten eignen sich hervorragend für den Einsatz im Unterricht.

4.2 Überwachung des Körpers

Das Überwachen des eigenen Körpers und seiner Aktivität wird immer beliebter in der heutigen Zeit. Profisportler überwachen ihren Herzschlag, die Atemfrequenz und weitere physiologische Werte ihres Körpers ständig, um ihre Leistung im Auge zu behalten und frühzeitig auf Veränderungen reagieren zu können. Ein Arzt überwacht den gesundheitlichen Zustand seiner Patienten und kann aus den Visualisierungen und Werten ableiten wie die körperliche Verfassung ist. Immer mehr Leute nehmen die Möglichkeit der Körperüberwachung wahr, jedoch sind die Visualisierungen der physiologischen Daten die bei Profisportlern und Ärzten die zum Einsatz kommen zu komplex. Für einen normalen Benutzer sind sie also nicht zu gebrauchen, neue Visualisierungen müssen geschaffen werden, die eine breite Masse an Menschen versteht. Hier kommt *Spark* ins Spiel, es visualisiert die Anzahl der Schritte im Verlauf des Tages auf eine einfache Art und Weise und gibt dem Nutzer Feedback zu seiner physischen Aktivität. Diese Art von Visualisierung ist einfach zu verstehen, soll den Nutzer dazu animieren seine Leistungen zu verbessern und dabei helfen Tagesziele, wie zum Beispiel 10.000 Schritte zu erreichen.

5 Schlussfolgerung

Die Visualisierung von physiologischen Daten über den menschlichen Körper ist nicht immer einfach und kann auf unterschiedlichste Arten erfolgen. Bei der Wahl der richtigen Visualisierung, müssen sowohl Anwendungsgebiet als auch Einsatzzweck mit berücksichtigt werden. Eine entscheidende Rolle bei der Wahl der richtigen Visualisierung spielt dabei der Kosten und Nutzen Faktor. Aufwendig zu gestaltende Formen der Visualisierung von physiologischen Daten wie der *BodyVis* Anzug sind in der Herstellung aufwendig und teuer. Jedoch besteht sein Mehrwert in der spielerischen und einfachen Art und Weise, Kindern ihren Körper näher zu bringen. Je früher Kinder ihren Körper kennen lernen desto eher achten sie auf sich selbst und ihre Gesundheit [18], ein gesunder Ernährungsstil kommt auch der Umwelt zu gute. Fitness Tracker gibt es mit unterschiedlichem Funktionsumfang was die Visualisierung und die Aufnahme der Daten betrifft. Für kleines Geld gibt es bereits Fitness Tracker, die die Schritte des Benutzers zählen und basierend darauf verschiedene Grafiken ausgeben über die Schrittanzahl und die verbrannten Kalorien. Investiert der Nutzer mehr Geld in einen Fitness Tracker, werden im auch beispielsweise Herzschlagmessungen und Schlafüberwachung geboten. Die Visualisierungen zeigen dabei anschaulich körperliche Aktivitäten und motivieren den Benutzer auf unterschiedlichste Art und Weise.

6 Zusammenfassung und Ausblick

Die Visualisierung von physiologischen Daten stellt eine Herausforderung für die Human Computer Interaction (HCI) dar. Wird die Visualisierung falsch gewählt, führt sie in die Irre und es kommt zu Unverständnis dem Benutzer gegenüber. Eine gute Visualisierung, die für ihren Einsatzzweck passend gestaltet ist, bietet einen erheblichen Mehrwert bei der Analyse von physiologischen Daten und trägt zu dem verstehen des menschlichen Körpers bei. In der Lehre wird die digitale Visualisierung mit neuen Technologien, wie Virtual Reality (VR) und Augmented Reality (AR) aber auch die haptische Visualsierung das Lernen erleichtern und Kindern neue Möglichkeiten bieten ihren Körper kennenzulernen. Im alltäglichen Leben unterstützen und motiviert sie die Menschen bei ihren sportlichen Zielen und in der Forschung trägt sie zu neuen Erkenntnissen bei. Mit einer Kombination aus verschiedenen Visualsierungsformen können interessante Einblicke in den physiologischen Ablauf des menschlichen Körpers gewonnen werden.

Literatur

1 Ernesto Arias, Hal Eden, and Gerhard Fischer. Enhancing communication, facilitating shared understanding, and creating better artifacts by integrating. In *Physical and Computational Media for Design, Proc. of DIS '1997, ACM*, pages 1–12. Press, 1997.

2 Matt Bower and Daniel Sturman. What are the educational affordances of wearable technologies? *Computers & Education*, 88:343–353, 2015.

3 Christian Breitwieser, Oliver Terbu, Andreas Holzinger, Clemens Brunner, Stefanie Lindstaedt, and Gernot R Müller-Putz. iscope–viewing biosignals on mobile devices. In *Joint International Conference on Pervasive Computing and the Networked World*, pages 50–56. Springer, 2012.

4 Arnon Cohen. Biomedical signals: Origin and dynamic characteristics; frequency-domain analysis. In *The biomedical engineering handbook*, volume 2, pages 951–74. CRC Press, 2000.

5 Sunny Consolvo, David W McDonald, and James A Landay. Theory-driven design strategies for technologies that support behavior change in everyday life. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 405–414. ACM, 2009.

6 The American Heritage Dictionary. Visualize. visited (06.01.2020). URL: `https://ahdictionary.com/word/search.html?q=visualize&submit.x=60&submit.y=12`.

7 Paul Dourish. *Where the action is: the foundations of embodied interaction*. MIT press, 2004.

8 Chloe Fan, Jodi Forlizzi, and Anind K Dey. A spark of activity: exploring informative art as visualization for physical activity. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 81–84. ACM, 2012.

9 Eva Hornecker and Jacob Buur. Getting a grip on tangible interaction: A framework on physical space and social interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 437–446, New York, NY, USA, 2006. ACM. `doi:10.1145/1124772.1124838`.

10 Seokbin Kang, Leyla Norooz, Vanessa Oguamanam, Angelisa C. Plane, Tamara L. Clegg, and Jon E. Froehlich. Sharedphys: Live physiological sensing, whole-body interaction, and large-screen visualizations to support shared inquiry experiences. In *Proceedings of the The 15th International Conference on Interaction Design and Children*, IDC '16, pages 275–287, New York, NY, USA, 2016. ACM. `doi:10.1145/2930674.2930710`.

11 Ian Li, Anind Dey, and Jodi Forlizzi. A stage-based model of personal informatics systems. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 557–566. ACM, 2010.

12 Ian Li, Anind K. Dey, and Jodi Forlizzi. Understanding my data, myself: Supporting self-reflection with ubicomp technologies. In *Proceedings of the 13th International Conference on Ubiquitous Computing*, UbiComp '11, pages 405–414, New York, NY, USA, 2011. ACM. `doi:10.1145/2030112.2030166`.

13 James J Lin, Lena Mamykina, Silvia Lindtner, Gregory Delajoux, and Henry B Strub. Fish'n'steps: Encouraging physical activity with an interactive computer game. In *International conference on ubiquitous computing*, pages 261–278. Springer, 2006.

14 Paul Milgram and Fumio Kishino. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12):1321–1329, 1994.

15 Leyla Norooz, Matthew Louis Mauriello, Anita Jorgensen, Brenna McNally, and Jon E Froehlich. Bodyvis: a new approach to body learning through wearable sensing and visualization. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1025–1034. ACM, 2015.

16 Blaine Reeder and Alexandria David. Health at hand: A systematic review of smart watch uses for health and wellness. *Journal of Biomedical Informatics*, 63:269 – 276, 2016. `doi:10.1016/j.jbi.2016.09.001`.

17 Albrecht Schmidt. Biosignals in human-computer interaction. *interactions*, 23(1):76–79, 2016.

18 Cheryl K Schmidt. Development of children's body knowledge, using knowledge of the lungs as an exemplar. *Issues in Comprehensive Pediatric Nursing*, 24(3):177–191, 2001.

19 Scott S Snibbe and Hayes S Raffle. Social immersive media: pursuing best practices for multi-user interactive camera/projector exhibits. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1447–1456. ACM, 2009.

20 Danae Stanton, Victor Bayon, Helen Neale, Ahmed Ghali, Steve Benford, Sue Cobb, Rob Ingram, Claire O'Malley, John Wilson, and Tony Pridmore. Classroom collaboration in the design of tangible interfaces for storytelling. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 482–489. ACM, 2001.

21 Hideyuki Suzuki and Hiroshi Kato. Interaction-level support for collaborative learning: Algoblock—an open programming language. In *The first international conference on Computer support for collaborative learning*, pages 349–355. L. Erlbaum Associates Inc., 1995.

22 Edward R Tufte. *The visual display of quantitative information*, volume 2. Graphics press Cheshire, CT, 2001.

23 JA Vessey, Karin Bannerot Braithwaite, and Marie Wiedmann. Teaching children about their internal bodies. *Pediatric nursing*, 16(1):29–33, 1990.

Mensch - Roboter - Interaktion

Lisa Gärttner

Ludwig-Maximilians-Universität, München, Deutschland
`Lisa.Gaerttner@campus.lmu.de`

Zusammenfassung

In dieser Arbeit wird die Interaktion zwischen Mensch und Roboter, im speziellen Assistenzroboter, ausgearbeitet. Roboter werden immer intelligenter, menschenähnlicher und sind so auch als mitarbeitende Unterstützer hilfreich. Deshalb sollen sogenannte Assistenzroboter beispielsweise in körperlich anstrengenden oder unterbesetzten Arbeitsbereichen eingesetzt werden. Eine dieser Zuständigkeiten kann die Pflege von älteren Menschen betreffen, da, bedingt durch den demographischen Wandel, oft zu wenig Personal vorhanden ist und die notwendige Hilfe somit nicht gewährleistet werden kann[7]. Dabei ist es wichtig, die Zusammenarbeit oder sogar das Zusammenleben für den Menschen so angenehm wie möglich und die Kommunikation natürlich und intuitiv zu gestalten. So können auch Personen, die nicht speziell ausgebildet sind, dieses im Grunde technische System steuern. Dazu wird im folgenden auf die Grundbestandteile der Kommunikation, von Sprache, Körpersprache, Mimik und Gestik und der Einfluss von Empathie und Emotionen eingegangen und entsprechende Beispiele zur Veranschaulichung dargelegt. Ein Teilergebnis wird die aus den oberen Punkten folgende gesellschaftliche Akzeptanz humanoider Roboter zeigen.

2012 ACM Computing Classification Human-centered computing→ Human computer interaction(HCI)→Interaction techniques

Keywords and phrases Mensch - Roboter - Interaktion; Mensch - Roboter - Kommunikation; Mensch-Roboter-Kollaboration; Emotion; Emphatie; Gestik; Mimik; Soziale Akzeptanz

1 Einleitung

In Literatur und Science- Fiction Filmen wird seit Jahrzehnten eine Koexistenz von Mensch und Roboter prophezeit. Unabhängig von Kritik, Furcht und schierer Begeisterung der Bevölkerung, arbeitet die Forschung daran diesem Traum so nahe wie möglich zu kommen. In vielen Bereichen des Lebens werden wir bereits erfolgreich von intelligenten Systemen, künstlicher Intelligenz und auch Robotern unterstützt. Dennoch ist hier dem schöpferischen Drang des Menschen keine Ende gesetzt. Durch immer weiter entwickelter und verfeinerter

© Lisa Gärttner;
licensed under Creative Commons License CC-BY

Cite as: Lisa Gärttner. Mensch - Roboter - Interaktion. In *5th Seminar on Ubiquitous Interaction (UBIACTION 2020)*. Editors: Florian Lang, Pascal Knierim, Jakob Karolus, Fiona Draxler, Ville Mäkelä, Francesco Chiossi, Luke Haliburton, Matthias Hoppe, Albrecht Schmidt. January 27, 2020. Munich, Germany. pp. 5:1–5:18.

Technik im Bereich der Sensorik und der steigenden Komplexität von künstlicher Intelligenz, scheint das technische Pendant zum Menschen in Reichweite zu sein. Die Vergößerung der technologischen Bandbreite verspricht nebenbei auch eine Erweiterung in den Einsatzgebieten eines Roboters. So können diese in körperlichen Kontakt mit uns Menschen treten ohne ein Verletzungsristko darzustellen. Aber im Hinblick auf ein mögliches Zusammenleben, ist nicht nur die körperliche Unversehrtheit ausschlaggebend, sondern auch ein intuitiver sowie vertrauensvoller Umgang der beiden Parteien. Denn nehmen beispielsweise sogenannte Assistenzroboter einen festen Platz in unserer Gesellschaft ein, werden auch technisch unerfahrene, nicht speziell ausgebildete Menschen mit diesen Maschinen in Austausch treten müssen. Die folgende theoretische Arbeit fokussiert sich im Besonderen auf die Kommunikation zwischen Mensch und Roboter, sowie Roboter und Mensch. Durch gezielten Einsatz von verbaler Sprache und Körpersprache, wie Mimik und Gestik, soll eine emotionale oder sogar empathische Bindung zwischen Mensch und Roboter enstehen. Dadurch und durch ein humanoides Erscheinungsbild soll so nebenbei die gesellschaftliche Akzeptanz erhöht werden.

2 Mensch - Roboter - Interaktion

Der Begriff "Mensch-Roboter-Interaktion" ist vielfach verwendet und kann unterschiedliche Kernbereiche mit einschließen. Die rudimentärste Art der Mensch-Roboter-Interaktion ist das simple Ein- und Ausschalten der Maschine. Weiter geht hier die Mensch-Roboter-Kollaboration, in der beide gleichzeitig in einem Kollaborationsraum zusammenarbeiten und nicht mehr räumlich getrennt sind [7]. Inzwischen gibt es jedoch ein wesentlich breiteres Feld von unterschiedlichen Roboterarten. Darunter finden sich Serviceroboter, die teil- oder vollautomatisch Aufgaben für den Menschen verrichten. Assitenzroboter, die ein autonomes, sicheres und interaktionsfähiges System darstellen, das zudem lernen und Verhaltensweisen steuern kann. Eine Spezialisierung stellen hierbei Personal-Service- Roboter dar. Diese verfügen zusätzlich über eine menschliche Interaktionsfähigkeit, wodurch sie noch intuitiver bedient werden und den Arbeitsplatz oder Lebensraum mit dem Menschen teilen können. Ein eigenes Kapitel in dieser Arbeit nehmen die humanoiden Roboter ein, die mit ihrem Aussehen und ihrem Verhalten besonders dem Menschen ähneln sollen [7]. Durch die Entwicklung dieser verschiedenen angepassten Roboterarten steigen auch die Einsatzgebiete, in denen eine filigranere Art und Weise der Mensch-Roboter-Interaktion stattfinden muss [17, S.12].

2.1 Ein Rückblick

Die Mensch-Roboter-Interaktion wurde im Zuge der Industrie 4.0[1], in der der Einsatz von Produktionsrobotern erhöht wurde und somit auch die Zusammenarbeit derer mit dem Menschen, zur Handhabungs- und Sicherheitsfrage. Die Roboter können basierend auf einem vorimplementiertes Programm teilautonom Arbeitschritte schneller und effizienter ausführen als ihre menschlichen Kollegen. Inzwischen sind Industrieroboter in automatisierten Fertigungsanlagen zum gängigen Verfahren geworden und nehmen dem Menschen körperlich schwere oder gefährliche Arbeitsschritte ab. Industrieroboter unterscheiden sich in der Definition von anderen Assistenzrobotern, denn sie interagieren nicht direkt mit dem Menschen und können für diesen innerhalb eines gewissen Radius sogar lebensbedrohlich sein. Um den Mensch körperlich zu schützen, wird der Tätigkeitsradius unteranderem mit einem Sicherheitsbereich gekennzeichnet oder sogar mittels eines Käfigs begrenzt.

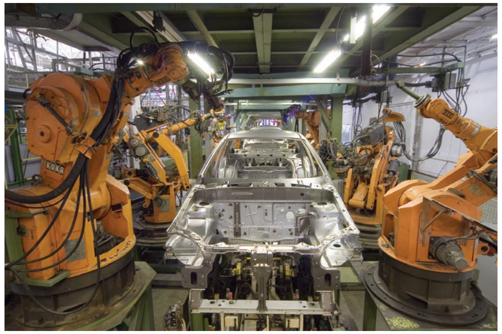


Abbildung 1 Industrieroboter im Sicherheitsbereich, Mixabest[2], 2007

---

[1] Laut Arbeitskreis Industrie 4.0 versteht man darunter »eine Vernetzung von autonomen, sich situativ selbst steuernden, sich selbst konfigurierenden, wissensbasierten, sensorgestützten und räumlich verteilten Produktionsressourcen (Produktionsmaschinen, Roboter, Förder- und Lagersysteme, Betriebsmittel) inklusive deren Planungs- und Steuerungssysteme« (Umsetzungsempfehlungen für das Zukunftsprojekt Industrie 4.0) [8].

[2] Industrieroboter IR 160/60, IR601/60, Mixabest / CC BY-SA (https://creativecommons.org/licenses/by-sa/3.0), Veröffentlicht: (https://commons.wikimedia.org/wiki/File:KUKA_Industrial_Robots_IR.jpg)

2.2 Neue Einsatzgebiete

Aufgrund verbesserter Sicherheitsmaßnahmen ist die engere Zusammenarbeit zwischen Mensch und Roboter möglich und findet so in weiteren Bereichen Einsatz. Assistenzroboter oder auch Service-Roboter, wobei hierzu überlappende Definitionen in der Literatur zu finden sind, können in Zukunft unter anderem in humanitären Einrichtungen Verwendung finden. In unserer immer älter werdenden Gesellschaft zeigt die Zahl der Plegekräfte nicht in dem Maße an, dass damit die Betreuung von Menschen mit eingeschränkter Mobilität und Selbstständigkeit leistbar wäre [7, 3, S.219]. Begründet durch den starken demographischen Wandel, konzentriert sich ganz besonders Japan auf die Entwicklung von Robotern und entsprechender künstlichen Intelligenz, um die Altenpflege aufrecht erhalten zu können. Geschätzt wird, dass dort 2025 rund 370.000 Pflegekräfte fehlen werden [13]. In diesem Punkt können sogenannte Personal-Service- Roboter oder Assistenzroboter Abhilfe schaffen. Aber auch als Enduitainment-Produkte[3] [17], die Unterhaltung und Lernen spaßhaft vereinen, finden sie zunehmend Anklang. Für diese Bereiche oder in der Zusammenarbeit mit Kindern ist die intuitive Interaktion mit den Assistenzrobotern besonders wichtig. Ein menschliches Aussehen kann zudem hilfreich sein, um die Hemmschwelle besonders niedrig zu halten. Diese Art von Robotern werden als humanoide Roboter bezeichnet, welche in einem späteren Kapitel behandelt werden.

3 Mensch - Roboter - Kommunikation

In Arbeitsbereichen, in denen viel Kontakt zwischen Mensch und Roboter besteht, kann nicht davon ausgegangen werden, dass die Personen entsprechnde technische Vorkenntnisse haben. Also ist es erstebenswert die Kommunikation zwischen den Parteien so intuitiv wie möglich zu gestalten. Kommunikation ist zwar ein vielfältig verwendeter Begriff, der aber nur schwer zu definieren ist [15, S.20]. Kommunikation ist unteranderem durch einen Inhalts- und einen Beziehungsaspekt kategorisierbar. Der fundamentale Hauptgrund einer Kommunikation ist Informationen auszutauschen, dieser repräsentiert der Inhaltsaspekt. Allerdings wird über andere Faktoren, wie Gestik und Mimik oder auch Tonfall, die reine Information unterschiedlich interpretiert und der Beziehungsaspekt auch anders verstanden. So wird eine eigentlich positive Information plötzlich negativ aufgenommen, wodurch die gesamte Kommunikation in eine andere Richtung geführt wird. Somit beinhaltet ein kommunikativer Austausch auch immer das Konzept der Ursache und Wirkung.

---

[3] Enduitainment: Kombination Education und Entertainment

Watzlawick [16] definiert diese Reaktionskette mit drei Regeln:
Zum einen gibt jeder Partner einer Interaktion eine Struktur in die Beziehung, zum anderen entsteht eine Reiz-Reaktionskette und als dritter Punkt wird der Reiz an sich als Kommunikation anerkannt. Diese Abfolge verläuft kreisförmig und es gibt somit keinen Anfangspunkt, wodurch die absolute Ursache unklar ist. Das System kann beispielsweise zerstört werden, wenn die Kommunikationspartner "aneinander vorbei reden", da sie unterschiedliche Basisinformationen über den Inhalt haben. Die zwischen"menschliche" Beziehung kann die Kommunikation der Interagierenden zudem beeinflussen. So ist es möglich, dass sich das Verhältnis komplementär oder symmetrisch ausprägt. In der komplementären Beziehung ergänzen sich die unterschiedlichen Verhaltensweisen, wobei es dadurch zur Unterordnung eines Partners kommt. In der symetrischen Beziehung versuchen die Partner auf der gleichen Augenhöhe zu bleiben.
In der Übertragung auf die Mensch-Roboter-Kommunikation ist also die komplementäre Kommunikation nur dann sinnvoll, kann der Roboter das Gespräch so interpretieren und dementsprechend reagieren, sodass er sich als untergeordneter Kommunikationspartner versteht und das menschliche Gegenüber somit ein Gefühl der Sicherheit im Umgang mit dem Roboter bekommt. Im folgenden wird die verbale Komponente, die Sprache, genauer betrachtet.[16]

3.1 Sprache

Die Mensch-Roboter-Kommunikation setzt sich aus natürlicher Sprache und künstlicher Sprache zusammen [17, S.61]. Kommunikation ist, wie im vorherigen Kapitel bereits gezeigt wurde, kein einseitiges Phänomen. Es wird auch von einem Kommunikationsraum gesprochen. Dieser wird zum einen vom Menschen und zum anderen vom Roboter aufgespannt.
Hierzu gehören, die für die Mensch-Roboter-Kommunikation ausschlaggebenden Punkte:

- 1. Die Tonausgabe: Tonsequenzen können unterschiedliche Kommunikationsparameter beinhalten, welche unterbewusst von den Nutzern wahrgenommen und verarbeitet werden. Dies führt wieder auf den vorhin beschriebenen Beziehungsaspekt zurück. Die Tonlage oder auch die Betonung in den Wörtern und Sätzen bestimmen, ob eine Aussage als positiv freundlich, negativ unfreundlich oder beispielsweise auch ironisch oder sarkastisch wahrgenommen wird. Dadurch wird wiederum der gesamte Kommunikationszyklus in die entsprechende Richtung geleitet und beeinflusst, ob wir unser Gegenüber als sympathisch oder unsympathisch empfinden.

- 2. Die Sprachsynthese: Für eine intuive und angenehme Kommunikation mit einem Roboter ist es von Vorteil, wenn dessen Stimme und Sprache so natürlich wie möglich designed ist. Da die Gestaltung von technischen Sprachen an sich schon sehr komplex ist und die dementsprechende Spracherkennung umso mehr, existieren unterschiedliche Forschungsansätze. Kismet (2002) beispielsweise simuliert das Verhalten eines Kindes, weshalb keine fließende Sprache erwartet wird und die Lautgestaltung einen trivialeren Ansatz erlaubt. Zudem steigert das kindliche Verhalten die soziale Akzeptanz des Roboters und somit auch die Empathie ihm gegenüber.

- 3. Die Stimmqualität: Für die akustische Verständlichkeit ist die Qualität der Roboterstimme wichtig. Allerdings ist fraglich, ob sich an der reinen Verständlichkeit etwas ändert, wenn es sich um eine roboterhafte Stimme oder eine menschliche Stimme handelt. Bei einer Roboterstimme ist der technische Hintergrund offensichtlicher, wobei bei einer natürlichen Stimme die Technik erst beim Auftreten von Fehlern während der Kommunikation erkennbar werden dürfte und diese betont. [17]

Wird hier auf die kulturellen Gepflogenheiten im Sprachgebrauch auch Rücksicht genommen, spielen zudem die Volkalisierung und der unterschiedliche Gebrauch von Tonfall und Satzmelodie eine große Rolle in der beasichtigten Kommunikation. Da die technische Sprachsynthese diese Feinheiten aber noch nicht zulässt wird in dieser Arbeit darauf nicht weiter eingeangen. [1]

3.2 Körpersprache

Zur Kommunikation gehören auch non-verbale Ausdrücke, wie Körpersprache, Mimik und Gestik. Hierbei sollten ensprechende Roboter nicht nur ihre eigene Kommunikation durch Gesten unterstützen können, sondern diese von ihrem Gegenüber auch erkennen und deuten lernen. [7, S.218]

Die Gestik

Die Verwendung von Arm-, Hand-, Finger- oder Kopfgesten kann die sprachliche Aussage verstärken oder ersetzen. [7, S.216]

Dabei wird zusätzlich zwischen Embleme, mit Sprache verknüpften Gesten und emotionalen Gesten unterschieden. Embleme sind eine direkte verbale Übersetzung des Gesprochenen. Beispielsweise das Heranwinken einer Person, wenn diese gerade dazu aufgefordert wird zu einem zu kommen. Aber auch Achselzucken, Kopfnicken oder das abwehrende Haltzeichen mit der flachen Hand fallen in diesen Bereich. Weitere sind illustrierte Untersteichungen, wie das Kreisen einer Hand oder eines Fingers, um eine Drehbewegung nachzuempfinden. Emotionale Gesten sind nur schwer zuzuordnen und haben oft einen persönlichen individuellen Hintergund. Im Allgemeinen gibt es bei der Gestenverwendung und Interpretation, im Vergleich zur Mimik, wesentlich größere kulturelle Unterschiede, die eine Verallgermeinerung von deren Einsatz in der Mensch-Roboter-Kommunikation eingrenzen. 1986 wurden von Creider [5] vier ostafrikanische Kulturen und 1972 von Saitz und Cervenka [14] die USA und Kolumbien auf die Unterschiede in ihren Gestiken untersucht. Interessanterweise stimmte die Verwendung der vier afrikanischen Ländern zu 65% mit den nordamerikanischen und zu 73% mit den südamerikanischen Gesten überein. Darunter sind:

- Zeigen
- Achselzucken
- Kopfnicken
- Klatschen
- Heranwinken
- Winken
- Halt-Zeichen
- Schulterklopfen
- Daumen nach unten
- Umreißen der weiblichen Figur
- Kopf mit aufgelegter Handfläche zur Seite neigen(schlafen)
- mit flacher Hand die Größe eines Kindes zeigen [1]

Um die Interaktion mit einem Roboter dennoch durch Arm-, Hand- und Körperbewegungen aufzulockern und so weniger roboterhaft zu gestalten, bieten sich folglich vor allem Gesten, an die direkt mit dem Gesagtem verküpft sind. Da die oben genannten Embleme den asiatisch Raum außen vor lassen, ist in dem Zusammenhang zu prüfen, ob die Gesten auch dort verständlich sind und gleiche Bedeutungen haben.

Die Mimik

Die Mimik ist im Vergleich zur Gestik global eindeutiger verwendet, was ein breiteres Spektrum der Kommunikationuntermalung erlaubt.

Die Mimik oder das Mienenspiel beruht auf Änderungen des Gesichtsausdrucks, wodurch die sprachliche Aussage betont oder verstärkt werden kann. Auch der Gemütszustand kann durch die Mimik verdeutlicht werden. Das einfache symetrische Anheben unserer Mundwinkel symbolisiert in unserer Kultur zum Beispiel ein Lächeln und somit ein positives Gefühl. Wobei das einseitige verziehen eines Mundwinkels eher einen kritischen Ausdruck vermittelt. [7, S.219]

Befinden wir uns in einem Gespräch und der Gesprächspartner schaut uns währenddessen nicht an, lässt es vermuten, dass uns dieser nicht die entsprechende Aufmerksamkeit zukommen lässt. Über den Blick kann aber auch eine Gesprächsperson chrakterisiert werden. Richtet jemand seinen Blick auf den Boden, wird er als schüchterner oder unsicherer Mensch wahrgenommen. Hält er regen Blickkontakt selbstbewusst. Diese Eindrücke können auch auf die Mensch- Roboter- Interaktion übertragen werden.

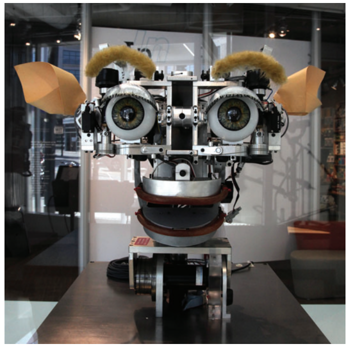


Abbildung 2 Mimikdarstellung, Rama[4], 1996

Versuchsbeispiele in diesem Bereich stellen Kismet, der Roboterkopf von iCub und Armar 3 dar. Bei Kismet wurde das Hauptaugenmerk auf die Emotionsdarstellung gelegt. In dem Vorführvideo der unterschiedlichen Mimiken, werden die Hauptbestandteile deutlich, die für die Visualisierung der Gemütszustände verwendet wurden. Kismet besitzt neben Augäpfeln und Augenlieder, die ganz oder nur halb geöffnet werden sowie Augenbrauen, die sich entsprechend des Gesichtsausdruckes nach oben bewegen oder ihre Drehung ändern. Der Mund und die Ohren unterstreichen die sieben Gefühlskategorien ruhig, verärgert, traurig, interessiert, angewidert, glücklich und überrascht. [2] Kismet wendet auch mimische Kommunikationsparameter über seine Augen, Augenbrauen, Ohren und Mund an.[17] Dennoch ist bei Kismet klar zu erkennen, dass es sich um einen Roboter handelt, da die technischen Bestandteile offen liegen.

---

[4] Kismet, Rama / CC BY-SA 3.0 FR (https://creativecommons.org/licenses/by-sa/3.0/fr/deed.en), Veröffentlicht: https://commons.wikimedia.org/wiki/File:Kismet-IMG_6008.jpg

Sophia der humanoide Roboter

Sophia ist ein humanoider Roboter von Hanson Robotics, dessen Gesicht dem eines Menschen stark nachempfunden ist. Die Latexstruktur, die die Technik überdeckt bilden beim Lachen kleine Fältchen und in in den Augen sind Kameras integriert, die zum Erkennen von Mimik und Emotionen des Gegenüber zu erkennen. Sophia kann durch diese Zusatzfunktionen im Vergelich zu Kismet 55 Gesichtausdrücke mehr, also insgeamt 62 Emotionen ausdrücken. Zudem hält sie mit ihrem Gesprächspartner Bilckkontakt und bewegt ihre Gesichtsmuskulatur, genauso wie es ein Mensch tun würde. An ihrem Torso sind Arme angebracht mit denen sie gestikulieren kann, um so ihre Worte zu unterstreichen. [11]

Abbildung 3 Beispiel für ein Gesichtsausdruck bei Sophia, International Telecommunication Union[5], 2017

---

[5] Sophia, Hanson Robotics Ltd., speaking at the AI for GOOD Global Summit, ITU, Geneva, Switzerland; International Telecommunication Union CC BY (https://creativecommons.org/licenses/by/2.0), Veröffentlicht: (https://www.flickr.com/photos/itupictures/35008372172)

4 Empathie und Emotionen

“Die Bezeichnung emotionale Intelligenz prägten die Psychologen Salovey und Mayer im Jahre 1990. Sie beschreiben emotionale Intelligenz als Form der sozialen Intelligenz, die die Fähigkeit mit einbezieht, eigene Gefühle und Gefühle anderer zu überwachen, diese zu unterscheiden und zu verwenden, um das eigene Denken und Handeln zu leiten” [6]. Diese komplexe Art von Selbstreflexion, die sogar so manch einem Mitmenschen schwerfallen dürfte und in der Robotik eine ausgereifte künstliche Intelligenz vorraussetzt, liegt in dieser Arbeit nicht im Fokus. Fraglich ist dennoch, ob sich durch den passenden optischen Ausdruck von Emotionen, eine Art Empathie zu einem Roboter hergestellt oder Empathie von einem Roboter ausgehend empfunden werden kann [17]. Was würde also einen Roboter sozial oder empathisch wirken lassen und sogar Empathie gegenüber des Roboters hervorrufen? Werden die bereits ausgearbeitet Faktoren zusammengetragen, lässt sich bereits sagen, dass dazu eine Kombination aus Sprache, Gestik und Mimik erforderlich ist. Wird also angenommen es handelt sich um einen Roboter mit menschenähnlichem Gesicht, um Mimik auszudrücken, und Armen, um gestikulieren zu können, sowie einer Sprachfunktion, sollte er in der Lage sein mit dem gleichen Verhalten zu reagieren, wie es ein Mensch tun würde.

4.1 Encodieren von Emotionen

Die Emotionalität eines Roboters besteht im Unterschied zur menschlichen Emotionalität nicht an sich[6]. Nur durch die menschliche Interpretation des Verhaltens während einer Interaktion, kann emotionale Kompetenz begriffen werden und somit eine Roboteremotionalität authentisch nachempfunden werden [6]. Im Umkehrschluss bedeutet das, dass der Roboter gezielt Ausdrucks- und Verhaltensweisen des Menschen kopieren muss, um eine Illusion des Emotionsausdruckes zu kreieren. Wie bereits im vorherigen Kapitel anhand des Roboterkopfes Kismet erarbeitet wurde, ist die Mimik und Gestik ein erheblicher Bestandteil in der Kommunikation von Emotionen. Lächelt ein Mensch wird er als freundlich empfunden und es entsteht ein positives Gefühl während der Kommunikation. Genau dasselbe System lässt sich auch auf Roboter übertragen und auch in vereinfachter technisch anmutenden Form, wie bei Kismet, verständlich darstellen. Durch Tonlage und Satzmelodie kann dies zusätzlich unterstützt werden.

---

[6] ein Roboter hat kein subjektives Empfinden und zeigt keine physiologischen Reaktionen

4.2 Decodieren von Emotionen

Roboter müssen aber nicht nur eigene Emotionen simulieren können, um mit ihrer Umwelt auf dieser Paraebene zu kommunizieren. Sie müssen auch fähig sein die Absichten und Gefühlsausdrücke von Interaktionspartnern richtig zu interpretieren und dementsprechend zu reagieren [4]. Die Wahrnehmung kann wie bei dem Roboterkopf von iCube passieren. Dieser verwendet in den Augen verbaute Kameras, die die Umgebung aufzeichnen. Ein ähnliches auf Kameras basiertes System wird bei Armar und Sophia eingesetzt. Die weniger triviale Frage ist, wie diese aufgezeichneten Bilder sinnvoll verarbeitet werden können. Sophia ist hierzu mit einer Datenbank für Mimikerkennung verknüpft, die ihr hilft ein Gesicht und den menschlichen Gesichtausdruck zu erkennen. Durch Sophias zusätzliche Fähigkeit Sprache zu verstehen, lernt sie immer weiter dazu je mehr sie sich mit Menschen unterhält [11]. Komplexe Interpretationen von menschlichen Formulierungen stellen gegenwärtig noch ein Problem dar. So sind für Sophia Lügen nur schwer zu erkennen. Wird allerdings eine Unstimmigkeit wahrgenommen, löst sie dieses Problem durch das Hinterfragen der entsprechenden Antworten. Wird sie dennoch weiterhin angelogen, können die Unwahrheiten nicht identifiziert werden. [9]

4.3 Humanoide Assistenzroboter

Humanoide Roboter sind Robotersysteme, die in ihrem Aussehen und in ihrem Verhalten Menschen nachempfunden sind. Im Gegensatz zur Industrieroboter sind die Anforderungen an diese Roboter darauf ausgerichtet, dass Mensch und Maschine im gleichen Arbeitsraum sicher miteinander agieren können. [7, S.213]

Abbildung 4 Humanoider Roboter Pepper, Nesnad[7], 2016 (links) und Softbank Robotics[8], 2016 (rechts))

Wie bereits im vorherigen Teil erarbeitet, ist die Mimik ein essentieller Bestandteil der Kommunikation. Nach Zeller [17] gibt es drei hervorzuhebende Vorteile in der interaktion mit einer menschlichen Schnittstelle: zum einen die "Persistente Nutzerinvolvierung", dies bedeutet, dass die Nutzer die Interaktion länger und aufmerksamer nutzen als Textinteraktion. Zum anderen ist die Konversation in der Qualität höher, die hier "Qualitative Nutzerinvolvierung" benannt wird. Im vorherigen Kapitel zu dem Thema "neue Einsatzgebiete" wurde bereits auf die intutitivere Interaktion mit Assistenzrobotern in Altersheimen oder zusammen mit Kindern angesprochen. Die "Optimierte Reaktionssteuerung" beschreibt genau dieses Phänomen der sozialen und somit verbesserten Interaktion. [17]

Humanoide Roboter haben einen Körperbau ähnlich den eines Menschen, mit Armen, Beinen, Kopf ect.. Dadurch, dass so auch Körpersprache, Mimik und Gestik menschlicher dargestellt werden können, kann der Mensch diese besser interpretieren und verstehen, da er die Verhaltensweisen übertragen kann.

---

[8] Pepper, Nesnad / CC BY (https://creativecommons.org/licenses/by/4.0), Veröffentlicht: https://commons.wikimedia.org/wiki/File:Pepper-inamall-aug27-2016.jpg

[8] Pepper the Robot, Softbank Robotics Europe / CC BY-SA (https://creativecommons.org/licenses/by-sa/4.0), Veröffentlicht: https://commons.wikimedia.org/wiki/File:Pepper_the_Robot.jpg

Gerke hat die Anforderungen an einen humnoiden Roboter in ihrem Buch [7] genauer zusammengefasst. Dies ist in Abbildung 5 zitiert.

Anforderung	Beschreibung
autonome Navigation	Roboter sollen in der Lage sein, eigenständig und unabhängig zu agieren. Sie sollen Hindernissen ausweichen und Ziele eigenständig anlaufen oder anfahren können.
Menschenähnlichkeit	Absenkung der Hemmschwelle beim Umgang mit dem Roboter, humanoide Roboter sollten Gesten und die Mimik des Menschen nachahmen können. Ihr Aufbau sollte dem des Menschen entsprechen.
Lernfähigkeit	Humanoide Roboter sollen neue Fähigkeiten erwerben können. Die taktile und visuelle Exploration erfordert kognitive Fähigkeiten.
Kooperationsfähigkeit, d. h. der Roboter soll zusammen mit dem Menschen Aufgaben erledigen	
Mobilität	Humanoide Roboter sollten möglichst auch auf unebenem Gelände operieren können.
Interaktionsfähigkeit	Interaktion mit dem Roboter über Sprache, Gestik und Haptik.
Sicherheit	Zum Schutz des Menschen muss das System in einem sicheren Zustand sein. Es sollte bei Auftreten äußerer Kräfte nachgiebig sein und die Oberfläche sollte möglichst aus einem nachgebenden und dämpfenden Material bestehen.
Personenerkennung	Roboter sollen sich in Richtung von erkannten Personen ausrichten können bzw. den Kopf schwenken können Dabei wird die Erkennung der Beine, des ganzen Körpers und des Gesichtes erforderlich. Die Entfernung zur Person muss berechnet werden. Bekannte Personen sollten wiedererkannt werden.
Spracherkennung und Sprachsynthese	Roboter sollten in der Lage sein, Sätze aus einer für den Menschen normalen Entfernung zu verstehen. Außerdem erwarten Bediener, dass sie Sätze bilden können.
Personenverfolgung	Humanoide Roboter sollten in der Lage sein, Personen zu folgen. Die Entfernung zu erkannten Personen muss berechnet werden und bei Bewegung der Personen online ermittelt werden.
Erkennen und gefühlvolles Greifen von Objekten	Objekte müssen dreidimensional vermessen und erkannt werden. Der Kopf muss zu den Händen ausgerichtet werden. Die Hände und Finger müssen visuell verfolgt werden. Die Bewegungsgrenzen der Gelenke müssen berücksichtigt werden. Im Sinne einer taktilen Exploration sollten unbekannte Objekte abgefühlt werden können.

Abbildung 5 Anforderungen an einen humanoiden Roboter von Gerke, entnommen aus [7].

5 Gesellschaftliche Akzeptanz von Assistenzrobotern

Bei dem Gedanken an Robotern, die uns Arbeit abnehmen oder unsere Arbeit erleichtern, seien es Industrie-, Assistenz-, oder Serviceroboter, kommt bald die Frage auf, wie unser Leben dadurch beinflusst und verändert wird. Ab wann ersetzen sie uns zum Beispiel am Arbeitsplatz und ist es sozial und moralisch fragwürdig, wenn sie in Pflege-und Betreueungstellen eingesetzt werden? Nach einer Umfrage zur Akzeptanz am Arbeitsplatz ist zu beobachten, dass junge Menschen positiver gegenüber robotiver Arbeitskollegen eingestellt sind als Ältere. Dies ist elementär durch die Angst begründet, nicht die ausreichenden Vorkenntnisse zur Kooperation mit Robotern zu haben oder erlangen zu können [10]. Die unten dargestellte Grafik visualisiert eine Umfrage zur zukünftigen Arbeitsteilung von Mensch und Roboter. 70% der Befragten gehen demnach davon aus, dass der Roboter in Eigenrechie körperlich anstrengende Tätigkeiten übernehmen wird. Kommunikation und Teamarbeit sehen 73% ausschließlich dem Menschen vorbehalten. In der Mensch-Roboter-Kollaboration wird der größte Prozentsatz der Befragten im Bereich der Kreativität und Problemlösung verortet. Wobei diese Tätigkeiten auch ein höhes Maß an Transferdenken und Kommunikation der Mitarbeitenden erfordert.

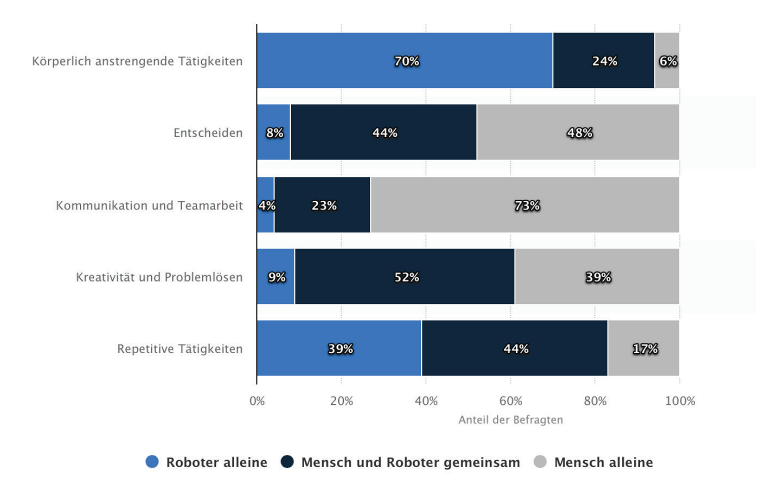


Abbildung 6 Umfrage zur zukünftigen Arbeitsteilung von Mensch und Roboter aus dem Jahr 2018 [12].

Eine weitere Studie zur Akzeptanz im direkten Umgang, beschäftigte sich mehr mit dem Inhalt eines Gespräches und den physichen und psychischen Nebenwirkungen. In dieser wurde eine Gruppe von Studenten angehalten sich mit dem humanoiden Roboter Sophia zu unterhalten. Während des Gesprächs wurde der Puls der Studienteilnehmer gemessen und sie sollten im Nachhinein ihre Gefühlslage beschreiben. Im Laufe der Interaktion mit Sophia senkte sich der Puls der Studenten und, wie es Dr. Julia Mossbridge beschrieb,"redeten die meisten sehr offen über ihre Gefühle". Die Tatsache, dass ein Mensch möglicherweise sehr aufgeschlossen intime Dinge mit einem Roboter teilt, mag ihr nach daran liegen, dass ein Roboter nicht urteilt und somit die Hemmschwelle sinkt [10]. Dieses Phänomen könnte pflegebedürftigen Menschen im Krankenhaus oder im Altersheim zugutekommen. Hier müssen Pfleger den Patienten oft körperlich sehr nahe kommen, was für den zu Pflegenden sehr beschämend sein kann [4].

6 Fazit und Ausblick

Die Forschung in der Mensch-Roboter-Kommunikation steckt bislang in den Kinderschuhen, sodass dort in den kommenden Jahren erhebliche Entwicklungen, auch hinsichtlich der stark expandierenden Einsatzgebieten, zu erwarten sind. Für reibunglose Interaktion und Kommunikation ist eine komplexere künstliche Intelligenz notwendig, die ihre Antworten nicht aus einer faktenbasierten Datenbank bezieht. So wie es mit den meisten technischen Fortschritten ist, wird sich aber auch der Mensch über die Zeit immer mehr an technisch bedingte Verhaltensweisen anpassen und diese verstehen lernen, sodass die Kommunikation und die Interaktion zunehmend natürlicher und selbstverständlicher erscheinen wird. Genauso wie das Smartphone zu einem ständigen Begleiter wurde, ist es durchaus wahrscheinlich, dass Assistenzroboter eine ähnliche Rolle einnehmen können und wir täglich mit diesen kollaborieren, interagieren und kommunizieren. Dennoch werden wie bereits im vorherigen Kapitel angesprochene rechtliche Fragen auftauchen. Was passiert zum Beispiel, wenn ein Roboter einen Menschen verletzt oder auch anders herum? Und haben Roboter tatsächlich Anspruch auf eine Staatsbürgerschaft? Diese Fragen müssen analog zum technischen Fortschritt im Bereich Robotik geklärt werden und im Gegenzug Faktoren, wie die gesellschaftliche Akzeptanz zusätzlich beeinflussen.

Literatur

1 Michael Argyle. *Körpersprache & Kommunikation*. Jufermann Verlag Paderborn, 2013.

2 Jürgen Blume. Methoden und anwendungen zur intuitiven mensch- roboter - interaktion. 2014.

3 Hans Buxbaum. *Kollaborierende Roboter in der Pflege -Sicherheit in der Mensch-Menschine-Schnittstelle*. O.Bendel(Hrsg.), Pflegeroboter; Springer Gabler, 2018.

4 Martina Čaić, Gaby Odekerken-Schröder, and Dominik Mahr. Service robots: value co-creation and co-destruction in elderly care networks. *Journal of Service Management*, 29(2):178–205, 2018.

5 Chet A. Creider. Towards a description of east african gestures. *Sign Language Studies*, (14):1–20, 1977.

6 Natalia Esau. *Emotionale Aspekte der Mensch-Roboter-Interaktion und ihre Realisierung in verhaltensbasierten Systemen*. Shaker, 2009.

7 Wolfgang Gerke. *Technische Assitenzsysteme: vom Roboter zum Roboterassistenten*. Walter de Gruyter GmbH, Berlin/München/Boston, 2015.

8 Fraunhofer Gesellschaft. Fraunhofer produktion und dienstleistung - industrie 4.0. Eingesehen am 22.12.2019, URL: `https://www.fraunhofer.de/de/forschung/forschungsfelder/produktion-dienstleistung/industrie-4-0.html`.

9 The Red Bulletin Innovator. Liebevolle künstliche intelligenz: Dieser roboter wurde programmiert um liebe zu lehren, 16.08.2018. Eingesehen am 20.11.2019, URL: `https://www.business-punk.com/2018/08/sophia-roboter/2/`.

10 Julian Stubbe, Johannes Mock, Steffen Wischmann. Akzeptanz von service robotern, 2019. Eingesehen am 15.12.2019, URL: `https://www.digitale-technologien.de/DT/Redaktion/DE/Downloads/Publikation/PAiCE_Servicerobotik_Studie.pdf`.

11 Edwin Kee. Sophia the robot was an attention grabber at SXSW interactive, 20.03.2016. Eingesehen am 28.02.2020, URL: `https://www.ubergizmo.com/2016/03/sophia-the-robot-was-an-attention-grabber-at-sxsw-interactive/`.

12 T. Käufer. Umfrage zur zukünftigen arbeitsteilung von mensch und roboter 2018, 20.06.2018, 10.06.2018. Eingesehen am 27.12.2019, URL: `https://de.statista.com/statistik/daten/studie/869849/umfrage/umfrage-zur-zukuenftigen-arbeitsteilung-von-mensch-und-roboter/`.

13 Deutscher Ärzteverlag GmbH. Japan setzt verstärkt auf pflegeroboter und künstliche intelligenz, 18.04.2019.

14 Robert L. Saitz and Edward J. Cervenka. *Handbook of gestures: Colombia and the United States*. Walter de Gruyter GmbH & Co KG, 2019.

15 Doris Ternes. Kommunikation-eine schlüsselqualifikation. *Junfermann, Paderborn*, 2008.

16 Paul Watzlawick. Die 5 axiome von paul watzlawick. Eingesehen am 28.02.2020, URL: `https://www.paulwatzlawick.de/axiome.html`.

17 Frauke Zeller. *Mensch-Roboter-Interaktion: eine sprachwissenschaftliche Perspektive*. Kassel University Press, 2005.

Physical Human-Robot Interaction

Julian Bernhard Lorenz

Ludwig-Maximilians-Universität München, München, Deutschland
`Julian.Lorenz@campus.lmu.de`

Zusammenfassung

As human-robot interaction (HRI) and especially physical human-robot interaction (pHRI) are fast developing fields, an overview of important aspects of the topics can be helpful in understanding the rising challenges. In order to gain structured knowledge, we firstly introduce the reader to the science of HRI simply by presenting a taxnomoy for different use cases and offering a classification of robot user interfaces. We then place the special case of pHRI in our taxnonomy and further identify and explain special requirements and challenges in pHRI.

2012 ACM Computing Classification Human-centered computing → Human computer interaction (HCI)

Keywords and phrases Robots, HCI, Human-Robot Interaction, physical Human-Robot Interaction

1 Introduction

The field of robotics continues to evolve faster and faster. A critical factor is the colossal decrease in expenses of hardware that computes, senses and actuates [27]. For instance, autonomous driving is a current topic of interest and has made huge steps in the last decade. Not less important among other influences are late advances in robotic autonomy, which have improved the ability of automated systems to act in unstructured and uncertain environments [24]. Robots will and are already providing services at our workplaces and in our homes to directly assist humans. Especially robots that offer physical interaction can increase human capabilities in terms of force, speed and precision. Moreover, they can improve the general quality of live and reduce stress [15].To accomplish the task, a robot has to interact face to face with a human individual while performing his duty [27].

According to the United Nations, robotics can be grouped in three major categories: industrial robotics, professional service robotics, and personal service robotics. The different groups are mostly defined by their application domains and each of them represents different technologies. Industrial robots,

© Julian Bernhard Lorenz;
licensed under Creative Commons License CC-BY

Cite as: Julian Bernhard Lorenz. Physical Human-Robot Interaction. In *5th Seminar on Ubiquitous Interaction (UBIACTION 2020)*. Editors: Florian Lang, Pascal Knierim, Jakob Karolus, Fiona Draxler, Ville Mäkelä, Francesco Chiossi, Luke Haliburton, Matthias Hoppe, Albrecht Schmidt. January 27, 2020. Munich, Germany. pp. 6:1–6:14.

which were the first actually being engineered and therefore exist for the longest time, have three main elements: they manipulate their physical environment; they are computer controlled and they operate in industrial settings. Further, industrial robots do not interact directly with people and operate at the lowest level of autonomy. Industrial settings are engineered in a way that enable a robot to execute his task in an almost mechanical way. Professional service robots on the other hand assist people in favor of their goals and do mostly not work in industrial environments. Other then industrial robots, they manipulate, navigate their physical surrounding and often share the same physical space with humans [27]. An example would be the Helpmate robot, which transports different things in a hospital [18]. Personal service robots help or engage individuals in domestic settings or in recreational activities. The field of service robotics both professional and personal is rising fast and it is important to mention that many of these robots interact with people who mostly have no special skills in using or working with a robot. Therewith, enabling autonomy is a core focus of robotics research [27].

The relatively new human-robot interaction results in a number of rising research and design challenges especially when it comes to physical contact. In this paper, the most important aspects of physical human-robot interaction (pHRI) are analyzed and concluded.

2 The Science of Human-Robot Interaction

Human-robot interaction is getting more intense and new challenges arise.For example, how to ensure sustainable and productive interaction? How to satisfy the need of a deeper understanding of human relationships with robots? How to fulfill privacy and security concerns and how to ensure safety? In case of direct interaction between humans and robots, we still need to learn about how robots with different capabilities, functions, and representations can work with each other as well as with humans [17]. In this chapter, in order to give an introduction into the field of research and to determine the part of human-robot interaction that this paper handles, a taxonomy for robotic use cases is presented. Further, robotic user interfaces and user experience in the context of robotics are being explained and reviewed.

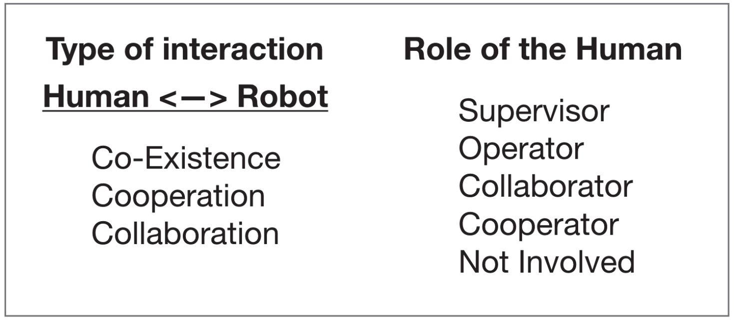
2.1 Taxonomy for use cases

Especially when pHRI is demanded, smart robotic solutions are needed to offer optimal support for the user. In order to accomplish this challenge, behavioural adaptivity of robots in difficult or unusual situations is important.

In today's life, an enormous variety of robots in relation to their shape, usage and form of interaction can be found. To compare them in a structured and logic way, a taxonomy that is based on various human-robot-interaction scenarios is required. Further, the taxonomy should be specific enough to offer an analysis of outstanding features. The needed taxonomy can be found in Onnasch's et al. "Mensch-Roboter-Interaktion - Eine Taxonomie für alle Anwendungsfälle", which will be summarized in the following [22]. Further, Onnasch et al. figured out three classification clusters, which are the basic elements in the taxonomy.

2.1.1 Interaction Classification

Interaction Classification means the hierarchic structure within the HRI. It is dominated by two variables: The kind of interaction and the role of the human during this interaction. The goal task consists figuring out the position or role of a human within the interaction. As shown in Figure 1, the type of interaction can be described as collaborational, cooperational or co-existent. [26].



Type of interaction
Human <—> Robot

Co-Existence
Cooperation
Collaboration

Role of the Human

Supervisor
Operator
Collaborator
Cooperator
Not Involved

Abbildung 1 Interaction Classification [22]

Co-existent interaction describes an episodic meeting of human and robot, in which the targets don't follow any joint goal. Much more, the involved entities try to get out of each others way. An example would be visitors in a hospital who walk past a transport robot: A direct interaction is being avoided, both of the parties just try to get behind each other without a collision.

In contrast, cooperation and collaboration stand for close interaction between human and robot in order to reach a common goal. To have a closer look, cooperation means working together on a project but with different and disjointed single tasks. Collaboration on the other hand means working directly together. The achievement literally depends on success of the interaction, which can also be physically. To reach a specific goal, cordinational requirements must be fulfilled [22].

Further, there are five distinct roles in interaction for a human. The supervisor is monitoring the robot and gives instructions. The operator controls and manages the robot. Both collaborator and cooperator are co-working with a robot, whereas collaboration means shared responsibilities and cooperation means that a human is still in charge. Finally, a not involved does not offer any HRI and is therewith not in the field of interest [22].

2.1.2 Robotic Classification

In HRI, the shape and mode of operation of a robot have a big impact. The taxonomy of robotic classification deals with questions and attributes related to it's work context.

Robotic Task	Operation Area	Morphology	Level of Autonomy: Attributes
Information Exchange	Industry	Humanoid	Information Recording
Discharge	Service	Zoomorphic	Information Processing
Precision	Personal Service	Functional	Decision Making
Transport			Action Performance
Manipulation			

Abbildung 2 Robot Classification [22]

Figure 2 shows an overview of important aspects in robot classification. Further, a robotic task can be split in five different types:

- Information Exchange: The robot is collecting and reproducing information for a human. This is used when the user is not able to explore a specific environment (e.g. the deep ocean).
- Discharge: The robot is used for physical tasks in order to support humans (e.g. an exoskeleton is helping someone to lift heavy objects).
- Precision: The robot fulfils particularly delicate tasks, in which he is more precise than a human (e.g. in surgery).
- Transport: The robot transports objects from one to another place (e.g. in logistics)
- Manipulation: The robot is physically changing its surrounding (e.g. a welding robot)

Another important attribute is the morphology of a robot. It is essential for the users expectation about the skills of a robot and also shows him to to communicate with the machine. Thus, the morphology of a robot should be used to set up specific associations for a human. This can result in an intuitive interaction. As an example, humanoid robotics call human associations, which will enhance a person to interact in an interpersonally way with the robot. Further, the user can expect natural communication and will probably just talk to the robot in the assumption to get an answer [22].

Also of interest is a robots level of autonomy. The four key attributes are: Information recording, information processing, decision making and action performance [28]. The level of autonomy has four steps rising from low to high and represents the human intervention in an abstract way. For instance, a robot with low autonomy level needs maximum help and guiding by a a human master. Autonomy also alludes to a robot's skill to accommodate variations in its environment [27]. In summary, autonomy is a determining factor for HRI.

2.1.3 Team Classification

The team classification classifies the structure and type of a human-robot collaboration. It can be divided in four parts (Figure 3).

Spatial Proximity	Temporal Proximity	Team Compositon Human Robot Ratio	Communication Channel
Touching	Synchronous		Human -> Robot:
Nearly	Asynchronous	NH = NR	Electronic
Guiding		NH > NR	Mechanically
Pass By		NH < NR	Acoustically
Avoiding		(Number Human = NH Number Robots = NR)	Optical
Remote-Controlled			
			Robot -> Human:
			Mechanically
			Acoustically
			Visually

Abbildung 3 Team Classification [22]

The spatial proximity indicates the distance between the working space of a robot and a human. It is described by six attributes [25]:

- Touching: Human and robot share the same working space. Physical human-robot interaction is possible.
- Nearly: Human and robot share the same working space. Physical human-robot interaction is impossible.
- Guiding: Human and robot are in stable physical contact (over a longer period of time).
- Pass by: The working space of human and robot can partially or completely overlap. Physical contact is avoided.

- Avoiding: Human and robot do not share the same working space and avoid physical contact.
- Remote-controlled: Human and robot are physically in the same working space. The robot gets remotely controlled by the human.

Beneath the spatial proximity, the temporal proximity provides information on synchronous or asynchronous working type. Synchronous means that human and robot work at the same time. Asynchronous means that human and robot work at different times (e.g.: robot at night, human during the day) [22].

Another variable for the team classification is the team composition. It has three manifestations: The number of humans and robots is equal (NH = NR), the number of humans is greater than the number of robots (NH > NR) and the number of humans is smaller than the number of robots (NH < NR) [25].

The fourth attribute is describing the communication channel in which the interaction takes place. Decisive is on which channel the robot gets information by the human and on the other hand on how the human receives information from the robot [22].

2.2 Robotic UI and UX

Many possible clients of robots have never had any past involvement with robots, nor have they at any point read a manual about any robotics. All their insight into robots originates from interaction in the field. A decent user experience (UX) is critical for the user acceptance of the robot. Summarizing, the user interface (UI) and user experience are key factors for product management, requirements specifications, customer engagement, design of the robot and interface, and customer testing [5]. Further, robotic user interface does not mean the interface to a robot, but the robot being the interface. In the following, a classification of robotic user interfaces based on their usage and skills is given [3].

2.2.1 Toy Scale

The robot does not aim to help a human solve a problem, nor does he work effective or efficient in finding a task solution. The robot's goal is to entertain the client and therefore, a special user interface must be built. As an example, Sony's "Aibo" can be named. It's only purpose is to entertain the owner and "bring warmth and delight to the everyday lives of its owners" [1]. Further, "Aibo" is shaped like a puppy dog and therefore has no practical value at all.

2.2.2 Remote Control - Autonomous Scale

The autonomous scale can mostly be copied from our robot classification in 2.1.2. Some robots are remotely controlled and not able to act independent, while others control their functions on their own and do not require human instructions. Robotic autonomy therefore often comes with the development of artificial intelligence [3]. Especially for service robots autonomy is a very important section as a robot must perform a specific task he is asked for [27]. For example, a robotic vacuum cleaner must not disturb its owner while cleaning the house. Therefore, he acts autonomous and even recharges itself on its own.

2.2.3 Reactive - Dialogue Scale

Reactive acting robotics wait for a user event, like touching the robot, which will create an event or a certain reaction. Logically, the following indirect interaction often follows "strict turn taking patterns, in which either the robot or the human fills in the role of the dialogue manager" [3]. Indirect interaction means a single information flow and is therefore used for professional service robots. The most common example would be the master-slave interface, in which the robot exactly duplicates physical movement of its human master [27]. On the other hand, direct dialogue systems are based on a bi-directional communication flow and allow mixed initiative. Subsequent, they must not follow strict turn taking patterns [3]. Direct interaction is a common design pattern in personal service robotics and gains more and more popularity, as rising technical possibilities offer the chance to make human-robot interaction equal footing in both directions.

2.2.4 Anthropomorphism Scale

To develop socially interactive robots, which will be used in everyday environments as personal or professional service robots, a humanlike design is a crucial part in acceptance [9]. Further, it is assumed that people prefer a humanlike interaction with robots [10]. This suggests anthropomorphic parts of a robot's physical shape, using human facial expressions and other social indications as well as natural humanlike communication (e.g. speech, gestures). The job of humanoid attribution in robotics is not to fabricate a fake human but to instead take advantage of it as a component through which social communication can be encouraged [8]. Hanson Robotics human-like robot "Sophia" is a good example for recent progress in this field of robotic user interfaces and gives a hint on what the industry may provide in the future.

3 Physical Human-Robot Interaction

Physical human-robot interaction (pHRI) does not automatically occur in every kind of HRI, but under certain circumstance, it's an important part of it. The taxonomy from 2.1 is used to specify the interaction type in which pHRI takes place. The type of interaction can be named as collaborational, with a human in the role of the collaborateur. Aditionally, operation areas are service and industry robotics. As robotic tasks, discharge, precision, transport and manipulation can be named. The spatial proximity is touching and guiding, temporal proximity is synchronous.

It further is important so say that pHRI should follow the same rules in communication as human-human interaction. Physical interaction between humans relies upon the cooperative force and past experience of individuals [23]. Humans constantly build interaction patterns upon a base of experience in the world through repeated interactions and are therewith able to build internal models that can for example anticipate the weight of an object based upon its size and predict the force needed to lift that object [19, 11]. In summary, it is important to mention that the performance of a cooperation is subsequently rising with a humanlike acting robot [23]. Beneath safety and dependability concerns, which are crucial in pHRI, a key factor is to implement the technical aspects of robotics in a way that imitates human behavior.

3.1 Special Requirements in pHRI

In the following, different approaches, important for physical human-robot interaction, will be introduced. Compliant behavior is an important need for robots to interact with humans [4]. Furthermore, active compliance, meaning compliance in relation to control design, is necessary for reaching the goal of safe and intuitive physical interaction [7]. Figure 4 shows a hierarchical feedback loop by Khoramshahi et al., which maps most control approaches for pHRI [16]. The final robotic behavior must then deliver compliance at different levels:

- Compliance at the task-level: Adaptation in execution that comply with the intention of a robot's human partner are possible [6].
- Compliance at the motion-level: Variations in motion while executing a particular task are given [4].
- Compliance at the force-level: While executing a specific motion, the robot remains compliant toward small perturbations due to the external forces [12].

With the goal of ensuring smooth transition and to avoid delay during an interaction, intent detection is important. A robot should know aspects of a planned action from a user in order to assist appropriately. Losey et al. [20] split the intent detection in three parts as showed in figure 5, which will be explained in the following: Defining intent, measuring intent and interpreting intent.

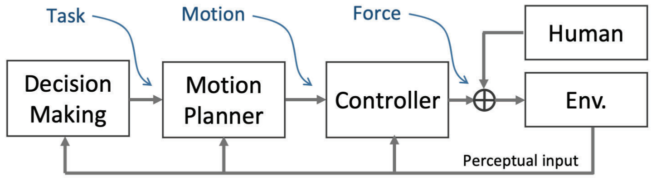


Abbildung 4 Design approach to pHRI. After deciding on a task, motions are planned and forces applied [16].

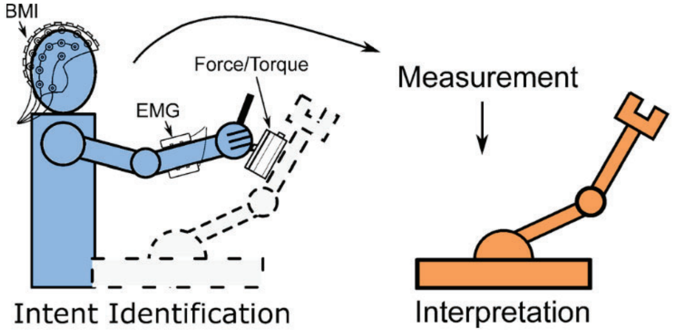


Abbildung 5 The different steps for to transfer user intent to a robot [20].

For defining a user's intent, the human activities have to be found that can be seen as data input. It is a quite complex task, involving brain activity, the musculoskeletal system and possible environmental interactions that give a hint on the intent. Therefore, the most common states of intent are represented through different characteristics, that reflect on the human system [20]. For a certain robotic task though, not all of them are important. The most simple approach to define intent is to represent it in a binary way, which allows the robotic system to just select from a set of predefined motions. The human intent is therewith defined as a trigger to initiate a certain movement, which is often interpreted by a brain-machine interface [21]. Other forms interpret user intent through muscle activation signals or predicted forward paths over a short time period [13]. After defining the intent, it hast to be measured. For measuring intent, different approaches for different definitions come to action. A brain-machine interface will for instance be mostly using the technique of electroencephalography (EEG) as an instance. For measuring the activation of muscles on the other hand, myography will take place. The last challenge of how a certain measurement has to be measured by a robot and further how it is included in a robots control structure is then called interpreting intent [20].

Another meaningful aspect in pHRI is machine learning. Machine learning helps a robot to adjust its behavior over a certain time in order to improve the cooperation and its features. Ikemoto et al. therefore conducted an experiment in which they tested a certain learning scheme by repeating a physical interaction between a human and a robot for several times [14]. After collecting training data, Ikemoto et al. then applied a Gaussian mixture model. Furhter, they adjusted the model in every loop of the interaction. This results in a new acting model for the robot during every iteration of the experiment. The measurements and results after the completion of the study are: Smoother transitions between robot and human can be found, as well as more efficient motion and a generally decreased interaction burden for the human [14].

3.1.1 Safety and Dependability

When robots are in the presence of our daily life, dependability must be ensured. Considering unexpected behavior of people, that can possibly always happen, robotics have to react in a certain way in order to prevent severe injuries [7]. Therefore, offering a safe end dependable pHRI is an important aspect of robotics research. Nevertheless, dependability is assailable through various potential failures or untreated aspects in sensors, control systems and software architecture. Furthermore, it can be split the following parts [2]:

- Safety: Means to ensure the non-appearance of severe consequences on the people in the HRI environment and the environment itself. Guaranteed safety while offering today's robot performance standard in speed and accuracy is still an open challenge to solve for science [29]. Nevertheless, there are existing ways to reduce the instantaneous severity of impacts. The easiest way to fulfill the task comes with "passive " safety. It means adaption of a mechanical design for the special purpose. Possible actions are: using lightweight or viso elastic materials, springs, rubber coverings and artificial skin [7]. Beneath the mechanical design, electronics and software design should also be taken to action. For instance, safety plans for anticipation and reaction to an occurring collision can be implemented. For instance, through the combined utilization of external/internal robot sensors and hardware/software, the manipulator operation can be smartly monitored, supervised and controlled [7].
- Availability and Reliability: Means to ensure preparedness for a demanded service. The achievement of the goal in this aspects relies in the execution and completion of its intended task. Availability/reliability and safety contradict each other, as the safest robot possible would probably do nothing then exist [7].
- Integrity: Avoidance of unintentional system alterations. Without integrity in robotics, there can be no availability nor reliability nor safety. Integrity is given when all physical and logical resources work reliable and if needed, protection mechanisms for the robot against system error's are implemented [7].
- Maintainability: Means the possibility to execute maintenance, modifications and repairs [7].

4 Conclusion and Outlook

An overview over the field of human-robot interaction plus recent research findings when it comes to physical contact in cooperation have been given. The large topic was split into different parts and a taxnomoy classifies different robotics for different use cases. As science in pHRI is still relatively new and continously developing, the goal was to determine key factors that differ pHRI from HRI and explain them. Further, the deployment of robotics in our daily life only seems to be a question of time, since the desire for automization of everyday life and support for difficult tasks in industrialized countries plus the reducing of costs and progress in science and technology are taking action. HRI, especially pHRI is a field of constant alteration and limits are not nearly reached yet. Some concepts are not fully sophisticated and therefore need to be fully developed. Moreover, human acceptance and trust for robotics has to be increased. In the future, there will be huge developments in robotic, but it is still a long way ahead before robots will be fully integrated in our daily lives.

Literatur

1 Sony's aibo. `https://us.aibo.com`. Accessed: 2020-01-01.

2 Algirdas Avizienis, J-C Laprie, Brian Randell, and Carl Landwehr. Basic concepts and taxonomy of dependable and secure computing. *IEEE transactions on dependable and secure computing*, 1(1):11–33, 2004.

3 Christoph Bartneck and Michio Okada. Robotic user interfaces. In *Proceedings of the Human and Computer Conference*, pages 130–140. Citeseer, 2001.

4 Aude Billard. On the mechanical, cognitive and sociable facets of human compliance and their robotic counterparts. *Robotics and Autonomous Systems*, 88:157–164, 2017.

5 Rodney Brooks. A brave, creative, and happy hri. *ACM Transactions on Human-Robot Interaction (THRI)*, 7(1):1, 2018.

6 Antoine Bussy, Pierre Gergondet, Abderrahmane Kheddar, François Keith, and André Crosnier. Proactive behavior of a humanoid robot in a haptic transportation task with a human partner. In *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*, pages 962–967. IEEE, 2012.

7 Agostino De Santis, Bruno Siciliano, Alessandro De Luca, and Antonio Bicchi. An atlas of physical human–robot interaction. *Mechanism and Machine Theory*, 43(3):253–270, 2008.

8 Brian R Duffy. Anthropomorphism and robotics. *The Society for the Study of Artificial Intelligence and the Simulation of Behaviour*, 20, 2002.

9 Julia Fink. Anthropomorphism and human likeness in the design of robots and human-robot interaction. In *International Conference on Social Robotics*, pages 199–208. Springer, 2012.

10 Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. A survey of socially interactive robots. *Robotics and autonomous systems*, 42(3-4):143–166, 2003.

11 Andrew M Gordon, Goran Westling, Kelly J Cole, and Roland S Johansson. Memory representations underlying motor commands used during manipulation of common and novel objects. *Journal of neurophysiology*, 69(6):1789–1796, 1993.

12 Neville Hogan. On the stability of manipulators performing contact tasks. *IEEE Journal on Robotics and Automation*, 4(6):677–686, 1988.

13 Cunjun Huang, Glenn S Wasson, Majd Alwan, Pradip Sheth, and Alexandre Ledoux. Shared navigational control and user intent detection in an intelligent walker. In *AAAI Fall Symposium: Caring Machines*, pages 59–66, 2005.

14 Amor Minato Jung Ishiguro Ikemoto, Shuhei. Physical human-robot interaction: Mutal learning and adaptation. *IEEE robotics & automation magazine*, 11(4):24–35, 2012.

15 Oussama Khatib, Kazu Yokoi, Oliver Brock, K Chang, and Arancha Casal. Robots in human environments: Basic autonomous capabilities. *The International Journal of Robotics Research*, 18(7):684–696, 1999.

16 Mahdi Khoramshahi and Aude Billard. A dynamical system approach to task-adaptation in physical human–robot interaction. *Autonomous Robots*, 43(4):927–946, 2019.

17 Sara Kiesler and Michael A Goodrich. The science of human-robot interaction. *ACM Transactions on Human-Robot Interaction (THRI)*, 7(1):9, 2018.

18 Steven J King and Carl FR Weiman. Helpmate autonomous mobile robot navigation system. In *Mobile Robots V*, volume 1388, pages 190–198. International Society for Optics and Photonics, 1991.

19 John W Krakauer, Maria-Felice Ghilardi, and Claude Ghez. Independent learning of internal models for kinematic and dynamic control of reaching. *Nature neuroscience*, 2(11):1026, 1999.

20 Dylan P Losey, Craig G McDonald, Edoardo Battaglia, and Marcia K O'Malley. A review of intent detection, arbitration, and communication aspects of shared control for physical human–robot interaction. *Applied Mechanics Reviews*, 70(1):010804, 2018.

21 David P McMullen, Guy Hotson, Kapil D Katyal, Brock A Wester, Matthew S Fifer, Timothy G McGee, Andrew Harris, Matthew S Johannes, R Jacob Vogelstein, Alan D Ravitz, et al. Demonstration of a semi-autonomous hybrid brain–machine interface using human intracranial eeg, eye tracking, and computer vision to control a robotic upper limb prosthetic. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 22(4):784–796, 2013.

22 Linda Onnasch, Xenia Maier, and Thomas Jürgensohn. *Mensch-Roboter-Interaktion-Eine Taxonomie für alle Anwendungsfälle*. Bundesanstalt für Arbeitsschutz und Arbeitsmedizin Dortmund, 2016.

23 Kyle B Reed and Michael A Peshkin. Physical collaboration of human-human and human-robot teams. *IEEE Transactions on Haptics*, 1(2):108–120, 2008.

24 Stuart J Russell and Peter Norvig. *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited,, 2016.

25 Jean C Scholtz. Human-robot interactions: Creating synergistic cyber forces. In *Multi-Robot Systems: From Swarms to Intelligent Automata*, pages 177–184. Springer, 2002.

26 D Surdilovic, A Bastidas-Cruz, J Radojicic, and P Heyne. Interaktionsfähige intrinsisch sichere roboter für vielseitige zusammenarbeit mit dem menschen, 2018.

27 Sebastian Thrun. Toward a framework for human-robot interaction. *Human-Computer Interaction*, 19(1):9–24, 2004.

28 Christopher D Wickens, Justin G Hollands, Simon Banbury, and Raja Parasuraman. *Engineering psychology and human performance*. Psychology Press, 2015.

29 Michael Zinn, Bernard Roth, Oussama Khatib, and J Kenneth Salisbury. A new actuation approach for human friendly robot design. *The international journal of robotics research*, 23(4-5):379–398, 2004.

Human Drone Interaction

Jan Paolo Holinski

Ludwig-Maximilians-Universität München, München, Deutschland
`P.Holinski@campus.lmu.de`

Abstract

Unmanned quadrocopters, colloquially often called "drones", have become more prominent in recent years. These small, flying robots enable many new and exciting applications, which were previously not possible. Although the field of human drone interaction is relatively young, there already exists a wide variety of research, which is hard to get a hold of. In order to gain a succinct overview, we survey and summarize recent publications in the field and develop a taxonomy, which we then use to classify the presented works. We also identify common problems and issues when working with drones. Based on our analysis of issues and the classification of existing efforts, we then determine areas for future research, hinting at developments to come.

2012 ACM Computing Classification Human-centered computing → Human computer interaction (HCI)

Keywords and phrases Drones, HCI, Human-Robot Interaction, Human-Drone Interaction.

1 Introduction

Drones have become more reliable, smaller and cheaper over the last couple years, making them a readily available consumer grade technology. Together with advancements in machine learning and image recognition such as conversational user interfaces and face and object recognition, this opens up new ways of interaction.

In the following work we will first give an overview of the broad spectrum of applications that drones have been used or proposed to be used for and survey recent research in section 2. We then give a taxonomy to classify existing research and projects involving drones in section 3. Afterwards, in section 4, we identify common issues and problems that show up again and again in the different works. Based on the taxonomy given in section 3 we identify areas for future research in section 5 and conclude our work in section 6.

While there have been prior attempts at giving an overview of the current state of research in Human-Drone Interaction[27, 13], none of them are as

© Jan Paolo Holinski;
licensed under Creative Commons License CC-BY

Cite as: Jan Paolo Holinski. Human Drone Interaction. In *5th Seminar on Ubiquitous Interaction (UBIACTION 2020)*. Editors: Florian Lang, Pascal Knierim, Jakob Karolus, Fiona Draxler, Ville Mäkelä, Francesco Chiossi, Luke Haliburton, Matthias Hoppe, Albrecht Schmidt. January 27, 2020. Munich, Germany. pp. 7:1–7:16.

exhaustive as this contribution.

2 A Survey of Recent Research in Human Drone Interaction

Drones are very versatile and have been used in a wide variety of contexts such as architecture and construction [33], search and rescue (SAR) [32, 24, 6], sports and recreation [15, 26, 28], navigation and guidance [2, 3, 22, 10, 5], visualization [19, 14, 4], and virtual reality (VR) [17, 20].

With all these new applications there is the question of how to best interact and communicate with the drone.

In some areas drones are mostly just a cheaper replacement for traditional helicopters and the concrete applications as well as the human involvement in the task did not change much with the introduction of drones. In other areas however, drones enable truly new and different applications, which were previously not possible, for example because traditional helicopters were too big, or the risk involved for the human pilot was not acceptable. These new applications require new interactions and are therefore the main focus of this work.

Cacace et al. propose to use multiple drones in SAR missions to locate missing persons[6]. Because the rescuer is heavily involved in the rescue themselves, they cannot allocate their full attention and cognitive resources to operating the drone. The authors therefore suggest fairly autonomous drones together with a combination of hand gestures, voice control, and a touchscreen, to operate the drone. The operator could for example say “go there”, together with a pointing gesture, to move a drone, or mark an area on the touchscreen to have the drone autonomously search it using a sweep pattern.

The use of a drone as a jogging companion is first introduced in [15] and then further evaluated in [28]. The authors conducted a study where they had their participants jog along a quadrocopter, that was following a path of preprogrammed waypoints.

Many papers experiment with using drones as navigational aids. Avila et al. [2, 3] try to help visually impaired people navigate a space by using drones. In both cases a small drone leads the way for a visually impaired or blind user. The user can locate the drone through the distinct sound of its rotors and in the case of [3] through tactile feedback, provided by a cable, that connects the quadrocopter to a battery pack carried by the user. In a user study the authors found that the quadrocopter based approach significantly reduced navigation errors when compared to audio instructions such as “turn left” and “go straight” provided through headphones.

The work by Colley et al. [10] is aiming at sighted users and explores how to use the movement of the drone itself to communicate navigation instructions, such as making a turn or crossing a road. The authors develop several “drone gestures” (e.g. rotating in place about the vertical axis, tilting and moving sideways, hovering in place), which they then evaluate in an online study and a field study. While some gestures were universally understood, others lead to different results, both between the online and the field study, as well as between subjects in the field study.

The applications in [22] and [5] use a mobile projector attached to the drone to augment the physical environment of the user with extra information for navigation.

The quadrocopter in [5] can project maps on the ground to help users orient themselves. In a first indoor study using a stationary projector without a drone the authors evaluated two approaches using a smartphone to interact with the map and select points of interest (POIs). In one approach all navigation and interaction was done on the phone’s touchscreen, whereas in the other approach, panning was achieved by physically moving the phone. In a follow-up field study, involving a real drone, the authors further developed the second approach using a depth sensor for hand tracking, obviating the need for a smartphone. Users could instead interact with the map using hand gestures. Interaction proved difficult at times, because the drone carrying both the projector and the hand tracker was not always perfectly stable.

Knierim et al. use the projector not to project maps, but directions in the form of arrows directly in the natural environment (see Figure 1). In a user study where the subjects had to navigate a given path using the directions as projected by the drone or displayed on a smartphone, the authors could show, that in the drone scenario, participants could better recall POIs along the route, suggesting that the approach allows to pay better attention to one’s surroundings[22].

Drones have been used in more ways to enhance or augment the user’s surroundings, both in VR and the physical world. In VR, Knierim et al. and Hoppe et al. use drones to provide haptics for entities, such as insects or fish, in the virtual scene[20, 17]. The drone is equipped with an accurate motion tracking system to precisely position it in the room, based on the location of the virtual entity in the virtual environment. In addition, different haptic extensions are attached to the drone, which can then be safely touched by the user. This allows the user to physically interact with the drone, as a proxy for the virtual entities, as can be seen in Figure 2. As a result the perceived level of realism increases, when compared to no haptic feedback, or vibrotactile feedback delivered through a handheld controller.

Outside of VR drones are often used as a way to position user interface elements in 3D space. In many applications the drone carries a display, which

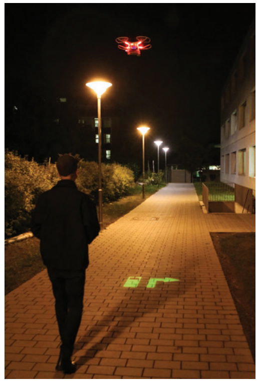


Figure 1 A drone projecting navigation instructions on the floor. Image from [22]

can then be used to show high resolution pictures and video [30, 14, 34]. The display does not necessarily have to be flat and square, but can also be cylindrically bent around the drone [14] or even spherical [34] to provide a greater viewing angle. As already mentioned before, the drone can also project images onto surfaces in its surroundings [22, 5, 25]. Compared to handheld displays, displays carried or projected by drones offer a hands-free experience and can enable interaction with a group of people. They can also dynamically adapt the position of the displayed content in the real world, which is impossible with traditional digital signage.

The drone does not necessarily have to carry a high resolution display however. Sometimes a single light-emitting diode (LED) is enough. Combining hundreds of drones with a single RGB-LED each, yields a three-dimensional voxel display [19]. This also works on a smaller scale [14, 4]. To control the position and color of the drones in [19], the authors developed a special purpose control software, which allows preprogramming different flightpaths and animations.

In [4] the interaction is more immediate, using a 3D wand input device, which allows selection, rotation and translation of voxels, through midair gestures. Kosch et al. also use a handheld controller, to position a drone in 3D space with a pointing motion [23].

In [14] and [21] the user can directly interact with the system by touching or grabbing a drone, which is encased in a protective cage, to prevent injury from getting into contact with the spinning rotor blades. The drones thereby practically become the user interface – a so called tangible user interface (TUI) – and multiple drones can react to a single drone getting manipulated by the user, as demonstrated in Figure 3.

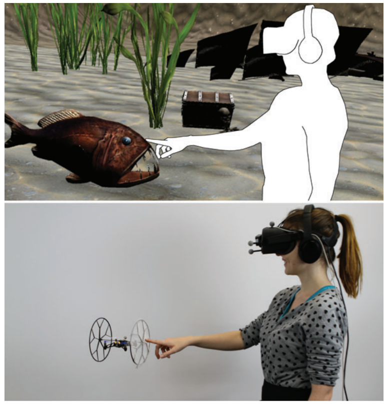


Figure 2 A drone providing haptic feedback as a proxy for the anglerfish in VR. Image from [17]

Multiple papers explore the use of gestures to control quadrocopters. Ng and Sharlin propose using gestures inspired by falconry, which allow intuitive interaction with the drone and make the user perceive the drone as a companion or pet [29]. In order to find universal gestures for common actions, Cauchard et al. asked the participants of their research study to execute a series of tasks involving a drone. Participants had full freedom to invent new gestures and ways of interaction to accomplish the tasks. An example of user-defined gestures as determined through the study can be seen in Figure 4.

In addition to hand and body gestures, Fernandez et al. also examined the use of visual tracking markers, and voice commands to communicate with the drone. Different tracking markers could communicate different commands, such as “hover”, “take off”, or “follow me”. Similarly, different voice commands were provided.

Hansen et al. study the use of gaze to control the drone [16]. In an experimental setup, four different mappings from gaze direction to drone motion were evaluated. Specifically, all combinations of mapping the vertical axis of the user’s gaze to altitude control or forward velocity, and the horizontal axis of the user’s gaze to rotation or lateral movement, were examined.

Erat et al. also use gaze to direct a drone in their sample application, although in a less explicit way [11]. The drone is flying in an indoor environment, separated from the operator through a wall. The drone is scanning its environment with an attached camera. A virtually reconstructed environment from the drone’s sensor feed is now overlaid on top of the real environment on

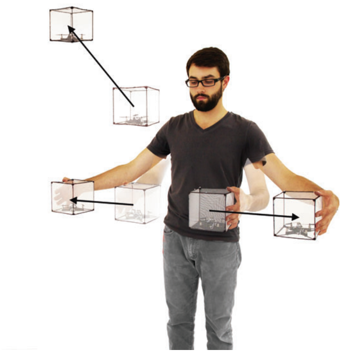


Figure 3 A user resizing a compound object using a bimanual pinch gesture, by moving 2 of the total 3 drones. Image from [14]

a head-mounted display (HMD) worn by the operator, providing a “virtual hole” in the wall. This form of augmented reality (AR), where the user can “see through walls”, allows control of the drone from an outside (exocentric) viewpoint, even when there is no direct line-of-sight between the operator and the drone.

Given this “X-ray vision”, the user can now steer the drone by focusing on a point in the scene, which causes the drone to fly closer to that point, to get a better view. In an alternative interaction technique, the user can “grasp” a virtual representation of the drone, as if it were within their arm’s reach, and place it somewhere else in the scene. This then causes the physical drone to assume the virtual drone’s position.

3 A Taxonomy for Classifying Applications Involving Drones

In the following section we give a taxonomy for classifying the aforementioned applications and for identifying potential future research.

We establish multiple dimensions, in part inspired by [27], along which the applications can be distinguished. Our classification can be seen in Tables 1-6. Not all dimensions are applicable to all applications and some applications do not fall on a single point along a dimension. We therefore omit applications or list them multiple times as necessary. When there are multiple papers, which classify the same along a dimension, we highlight one, of which we think that it represents the corresponding characteristic very well, in **bold**. We recommend looking at the highlighted papers first, to get a good overview of the different dimensions and their values.

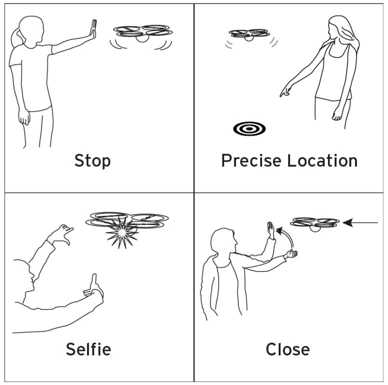


Figure 4 User-defined gestures for accomplishing common tasks. Image from [7]

3.1 Human-to-Drone Interaction

One dimension is the form of interaction with the drone, i.e. the way of controlling the drone.

The traditional way of controlling a drone is through a remote control. The controller does not necessarily have to be a classical one with two joysticks, but can for example also take the form of the tangible controller presented in [23]. The drone can also be controlled directly by touching and pushing or grasping and moving it, as is the case in the various TUI applications [14, 21]. Another way of control is through the use of gestures. These can be finger gestures [6, 12], arm gestures [29], foot gestures [25], or body gestures [7]. Voice control presents another possibility [7, 12]. In some applications the user has no explicit control over the drone at all, and merely observes or participates.

remote control	touch	gestures	gaze	voice	none
[4], **[23]**	**[14]**, [21]	[6], [12], [29], [25], **[7]**	**[16]**, [11]	**[6]**, [12], [7]	**[17]**, [20]

Table 1 Classification with regard to human-to-drone interaction

3.2 Drone-to-Human Interaction

Complementary to the question how the operator controls the drone, is the problem of how the drone conveys information to the operator and bystanders. In general, the drone can use visual, auditory, and tactile communication channels, the most information dense of which is the visual channel.

Attaching a display or portable projector to the drone allows displaying high resolution graphics on the drone or in the surroundings [30]. The drone can also convey information through its position alone, as is the case in the TUI applications [14, 21, 4]. Light-emitting diodes (LEDs) can be used to make spotting the drone's location easier. In addition to the position, the drone can also use movement, to convey information. A drone moving fast along a straight trajectory conveys a different mood from a drone that zigzags through the air like a bumblebee [8]. In terms of non-visual communication channels, the drone naturally conveys information through the sound of its spinning rotors [2, 3]. It can additionally play sound through portable speakers. Tactile feedback is provided in the presented VR and TUI applications. When coupled through a leash, the drone can also exert a pulling force [3]. Moreover, the drone can of course send information, such as for example telemetry data, through a wireless downlink to a separate device, to be displayed remotely [6, 11]. Since this case only marginally involves the drone, we do not examine it further.

display	projector	LEDs	position	movement	audio	touch
[30], **[34]**	[25], **[22]**, [5]	[14], [4], **[19]**	[21], [14], **[4]**, [19]	[10], **[8]**	[2], **[3]**, [14]	**[17]**, [20], [21], [14], [3]

Table 2 Classification with regard to drone-to-human-interaction

3.3 Positional Control

Another way to categorize applications is by looking at how the drone navigates in space. We differentiate five different categories, adapted from [13].

The simplest form of positional control is the one implemented with traditional remote controls, namely direct velocity control. The operator directly controls the velocity of the drone along the 3 spatial and up to 3 rotational axes. The control does not necessarily have to be through a handheld remote control though. It can also happen for example through voice commands, such as "move forward slowly".

A slightly more advanced form is position relative control, where the operator gives a translation or rotation vector relative to the drones current position. Without a global frame of reference, the magnitude of movement has to be reconstructed by integrating over the velocity, which is prone to accumulate error. With the availability of some global frame of reference, it becomes possible to give the drone absolute coordinates or waypoints, which it should fly to. If the drone is able to track moving objects in its surroundings, it can position itself relative to these objects, for example following a tracking marker at a distance of 10m. The most advanced form of positional control is task based positional control. With this form of control the drone decides its trajectory based on the requirements of an abstract task. This usually involves path planning, object recognition, and collision avoidance.

direct velocity control	position relative control	absolute position control	object relative control	task based control
traditional remote controls, [12], **[16]**, [6]	**[12]**, [29], [7]	[6], [15], [28], **[17]**, [20], [19], [14], [21], [4], [23], [11], [7]	**[12]**, [29], [7]	[3], [2], **[6]**, [10] [29], [11], [7]

Table 3 Classification with regard to positional control

3.4 Cardinality

While some applications only require a single drone, others only work with multiple drones, or even an entire swarm of drones. In a swarm the drones are generally not distinguishable. The boundary between multiple drones and a drone swarm is not always clear, but as a rule of thumb, any number of drones above low double digits can be considered a swarm.

single	multiple	swarm
[15], [28], [17], [20], [2], **[3]**, [22], [10], [5], [25], [29], [7], [11]	[6], **[14]**, [4], [21], [17], [20]	[14], [4], [21], **[19]**

Table 4 Classification with regard to cardinality

3.5 Function

Drones can be distinguished by their function. A drone can function as a sensor, for example when used as a TUI element. Or, a drone can function as an actor, for example when providing haptic feedback in VR. A drone can also function as an anchor, allowing attached objects to be positioned arbitrarily in 3D space.

sensor	actor	anchor
[14], [21]	[3], **[17]**, [20], [14], [21]	**[22]**, [5], [14], [19], [30], [34], [25]

Table 5 Classification with regard to drone function

3.6 Environment

Many applications are specific to an environment. On a high level, we can distinguish indoor from outdoor use, while on a more detailed level we can for example further categorize indoor spaces into large open spaces, such as halls, medium sized spaces, such as apartments, and small spaces, such as corridors or ventilation shafts. We can also further subdivide outdoor sites into urban, rural, and natural environments.

indoor	outdoor
[17], [20], [2], [3], [5], [30], [14], [21], [4], **[11]**	**[6]**, [15], [28], [22], [10], [5], [19], [30]

Table 6 Classification with regard to the drone environment

4 Common Issues and Challenges

Throughout all the presented applications there are a few common issues which occur again and again.

One prevalent problem is the amount of noise emitted from a flying drone. While the guidance applications for visually impaired users actually exploit this fact to provide audible feedback of the drone's location [2, 3], in most cases the harsh sound of the propellers poses a limitation [10, 21, 30, 9, 5]. The problem can be mitigated by "reducing mechanical imbalances associated with off-the-shelf hardware" [14], or, in a controlled environment, through the use of noise-canceling headphones [17].

Multiple papers list insufficient accuracy as a source of problems [5, 30, 14]. Sometimes small movements caused by turbulence and inaccurate sensor measurements can have positive effects however, giving the impression of the quadrocopter as a lively companion[28].

Another common limitation is the poor battery life of drones [10, 14, 30]. An already short flight time drops even further, if the drone has to carry additional weight or supply power to attached devices, such as displays or projectors. Proposed solutions are the use of inductive recharge stations [14], on-the-fly battery pack replacement or solar charging [30]. Avila Soto et al. solve the issue in [3] by powering the drone through an external battery pack, connected to the drone through a leash. This increases the flight time to approximately 5 hours.

A major concern when interacting with drones is safety [28, 1, 10]. The fast spinning rotor blades of sufficiently large drones have a dangerous potential to injure or even kill a human. It is therefore paramount, that the system is designed in a safe manner and this is appropriately communicated to the user. Encasing the drone in a protective cage is very helpful for this [1].

Since drones often carry a camera they can pose a serious threat to people's privacy. Even if the camera is not necessarily used to take pictures or record video, but only to help the drone orient itself, it is not immediately apparent to bystanders, that they are not being filmed [9]. When flying close to private property the loud noise of the drone can also present an intrusion into private space.

5 Future Research Trends

Given our classification in section 3, we can identify areas that have not yet been explored and are well suited for future research.

Most projects focus on the control of a single drone. The projects involving a swarm of drones use fairly traditional ways of control, such as a handheld controller [4] or special purpose control software to be operated by experts [19]. More research will be necessary to develop intuitive and approachable ways for controlling large swarms of drones, for example through the use of gestures.

Almost all presented applications assume a single operator. It will be interesting to evaluate how multiple operators can work together, controlling multiple drones, or even sharing control over a single drone.

All presented projects, even when incorporating multiple drones, focus on interacting with their system in isolation. Should applications involving drones get more widely adopted, it will be inevitable that drones from different systems will encounter each other and have to interact. In that scenario it will not only be relevant for the drones to decide how to react to other drones, but also how to communicate that decision to the operator.

Many papers use fairly low-fi prototypes to evaluate their theories, often using a human drone operator to mock some functionality of the proposed system. This works well in an early exploratory phase and for prototypes of limited complexity. When testing systems involving multiple drones, or systems, that require high precision however, this approach in insufficient. To overcome these limitations, we see a professionalization of prototyping methods and the advent of frameworks for developing drone applications [18].

Although there already has been some research on the perception of privacy around drones [9] and the acceptance of drones in public spaces[31], more work is necessary to better understand the design requirements and help inform decision making when it comes to developing regulations.

6 Conclusion and Summary

We have given an overview of recent research in the field of human drone interaction and provided a taxonomy to classify different applications. The large variety of applications shows the general utility of drones and encourages additional exploration. We classified the presented works based on our framework and identified common issues and problems. Using this information, we determined future research trends. We see major opportunities in exploring the control of swarms of drones, as well as the interaction between multiple drones. We identified developments to professionalize prototyping for research purposes using standard frameworks. While the current state of the art shows promising progress, there is still further research required into the ethics of drone usage and general acceptance of drones in the population to inform the necessary regulation and make drones commonplace.

References

1 Parastoo Abtahi, David Y. Zhao, Jane L. E., and James A. Landay. Drone near me: Exploring touch-based human-drone interaction. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 1(3), September 2017. 10.1145/3130899. URL `https://doi.org/10.1145/3130899`.

2 Mauro Avila, Markus Funk, and Niels Henze. Dronenavigator: Using drones for navigating visually impaired persons. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*, pages 327–328. ACM, 2015.

3 Mauro Avila Soto, Markus Funk, Matthias Hoppe, Robin Boldt, Katrin Wolf, and Niels Henze. Dronenavigator: Using leashed and free-floating quadcopters to navigate visually impaired travelers. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 300–304. ACM, 2017.

4 Sean Braley, Calvin Rubens, Timothy R Merritt, and Roel Vertegaal. Griddrones: A self-levitating physical voxel lattice for 3d surface deformations. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, page D200. ACM, 2018.

5 Anke M Brock, Julia Chatain, Michelle Park, Tommy Fang, Martin Hachet, James A Landay, and Jessica R Cauchard. Flymap: Interacting with maps projected from a drone. In *Proceedings of the 7th ACM International Symposium on Pervasive Displays*, page 13. ACM, 2018.

6 Jonathan Cacace, Alberto Finzi, Vincenzo Lippiello, Michele Furci, Nicola Mimmo, and Lorenzo Marconi. A control architecture for multiple drones operated via multimodal interaction in search & rescue mission. In *2016 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, pages 233–239. IEEE, 2016.

7 Jessica R Cauchard, Kevin Y Zhai, James A Landay, et al. Drone & me: an exploration into natural human-drone interaction. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, pages 361–365. ACM, 2015.

8 Jessica R Cauchard, Kevin Y Zhai, Marco Spadafora, and James A Landay. Emotion encoding in human-drone interaction. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 263–270. IEEE, 2016.

9 Victoria Chang, Pramod Chundury, and Marshini Chetty. Spiders in the sky: User perceptions of drones, privacy, and security. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 6765–6776, New York, NY, USA, 2017. ACM. ISBN 978-1-4503-4655-9. 10.1145/3025453.3025632. URL `http://doi.acm.org/10.1145/3025453.3025632`.

10 Ashley Colley, Lasse Virtanen, Pascal Knierim, and Jonna Häkkilä. Investigating drone motion as pedestrian guidance. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia*, pages 143–150. ACM, 2017.

11 O. Erat, W. A. Isop, D. Kalkofen, and D. Schmalstieg. Drone-augmented human vision: Exocentric control for drones exploring hidden areas. *IEEE Transactions on Visualization and Computer Graphics*, 24(4):1437–1446, April 2018. ISSN 2160-9306. 10.1109/TVCG.2018.2794058.

12 Ramon A Suarez Fernandez, Jose Luis Sanchez-Lopez, Carlos Sampedro, Hriday Bavle, Martin Molina, and Pascual Campoy. Natural user interfaces for human-drone multi-modal interaction. In *2016 International Conference on Unmanned Aircraft Systems (ICUAS)*, pages 1013–1022. IEEE, 2016.

13 Markus Funk. Human-drone interaction: Let's get ready for flying user interfaces! *Interactions*, 25(3):78–81, April 2018. ISSN 1072-5520. 10.1145/3194317. URL `http://doi.acm.org/10.1145/3194317`.

14 Antonio Gomes, Calvin Rubens, Sean Braley, and Roel Vertegaal.

Bitdrones: Towards using 3d nanocopter displays as interactive self-levitating programmable matter. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 770–780. ACM, 2016.

15 Eberhard Graether and Florian Mueller. Joggobot: A flying robot as jogging companion. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '12, pages 1063–1066, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1016-1. 10.1145/2212776.2212386. URL `http://doi.acm.org/10.1145/2212776.2212386`.

16 John Paulin Hansen, Alexandre Alapetite, I. Scott MacKenzie, and Emilie Møllenbach. The use of gaze to control drones. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ETRA '14, pages 27–34, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2751-0. 10.1145/2578153.2578156. URL `http://doi.acm.org/10.1145/2578153.2578156`.

17 Matthias Hoppe, Pascal Knierim, Thomas Kosch, Markus Funk, Lauren Futami, Stefan Schneegass, Niels Henze, Albrecht Schmidt, and Tonja Machulla. Vrhapticdrones: Providing haptics in virtual reality through quadcopters. In *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*, pages 7–18. ACM, 2018.

18 Matthias Hoppe, Marinus Burger, Albrecht Schmidt, and Thomas Kosch. Dronos: A flexible open-source prototyping framework for interactive drone routines. In *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia*, MUM '19, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450376242. 10.1145/3365610.3365642. URL `https://doi.org/10.1145/3365610.3365642`.

19 Horst Hörtner, Matthew Gardiner, Roland Haring, Christopher Lindinger, and Florian Berger. Spaxels, pixels in space. In *Proceedings of the International Conference on Signal Processing and Multimedia Applications and Wireless Information Networks and Systems*, pages 19–24, 2012.

20 Pascal Knierim, Thomas Kosch, Valentin Schwind, Markus Funk, Francisco Kiss, Stefan Schneegass, and Niels Henze. Tactile drones-providing immersive tactile feedback in virtual reality through quadcopters. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 433–436. ACM, 2017.

21 Pascal Knierim, Thomas Kosch, Alexander Achberger, and Markus Funk. Flyables: Exploring 3d interaction spaces for levitating tangibles. In *Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction*, pages 329–336. ACM, 2018.

22 Pascal Knierim, Steffen Maurer, Katrin Wolf, and Markus Funk. Quadcopter-projected in-situ navigation cues for improved location awareness. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 433. ACM, 2018.

23 Thomas Kosch, Markus Funk, Daniel Vietz, Marc Weise, Tamara Müller, and Albrecht Schmidt. Dronectrl: A tangible remote input control for quadcopters. In *The 31st Annual ACM Symposium on User Interface Software and Technology Adjunct Proceedings*, pages 120–122. ACM, 2018.

24 Lorenzo Marconi, Claudio Melchiorri, Michael Beetz, Dejan Pangercic, Roland Siegwart, Stefan Leutenegger, Raffaella Carloni, Stefano Stramigioli, Herman Bruyninckx, Patrick Doherty, et al. The sherpa project: Smart collaboration between humans and ground-aerial robots for improving rescuing activities in alpine environments. In *2012 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, pages 1–4. IEEE, 2012.

25 Mikhail Matrosov, Olga Volkova, and Dzmitry Tsetserukou. Lightair: A novel system for tangible communication with quadcopters using foot gestures and projected image. In *ACM SIGGRAPH 2016 Emerging Technologies*, SIGGRAPH '16, pages 16:1–16:2, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-4372-5. 10.1145/2929464.2932429. URL `http://doi.acm.org/10.1145/2929464.2932429`.

26 Sven Mayer, Pascal Knierim, PW Wozniak, and Markus Funk. How drones can support backcountry activities. In *Proceedings of the 2017 natureCHI workshop, in conjunction with ACM mobileHCI*, volume 17, page 6, 2017.

27 Silvia Mirri, Catia Prandi, and Paola Salomoni. Human-drone interaction: state of the art, open issues and challenges. In *Proceedings of the ACM SIGCOMM 2019 Workshop on Mobile AirGround Edge Computing, Systems, Networks, and Applications*, pages 43–48. ACM, 2019.

28 Florian 'Floyd' Mueller and Matthew Muirhead. Jogging with a quadcopter. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 2023–2032, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3145-6. 10.1145/2702123.2702472. URL `http://doi.acm.org/10.1145/2702123.2702472`.

29 Wai Shan Ng and Ehud Sharlin. Collocated interaction with flying robots. In *2011 Ro-Man*, pages 143–149. IEEE, 2011.

30 Stefan Schneegass, Florian Alt, Jürgen Scheible, Albrecht Schmidt, and Haifeng Su. Midair displays: Exploring the concept of free-floating public displays. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '14, page 2035–2040, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450324748. 10.1145/2559206.2581190. URL `https://doi.org/10.1145/2559206.2581190`.

31 Mauro Avila Soto and Markus Funk. Look, a guidance drone! assessing the social acceptability of companion drones for blind travelers in public spaces. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility. ACM, New York, NY, USA*, pages 417–419, 2018.

32 Sonia Waharte and Niki Trigoni. Supporting search and rescue operations with uavs. In *2010 International Conference on Emerging Security Technologies*, pages 142–147. IEEE, 2010.

33 Jan Willmann, Federico Augugliaro, Thomas Cadalbert, Raffaello D'Andrea, Fabio Gramazio, and Matthias Kohler. Aerial robotic construction towards a new field of architectural research. *International journal of architectural computing*, 10(3):439–459, 2012.

34 Wataru Yamada, Kazuhiro Yamada, Hiroyuki Manabe, and Daizo Ikeda. Isphere: Self-luminous spherical drone display. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, UIST '17, page 635–643, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450349819. 10.1145/3126594.3126631. URL `https://doi.org/10.1145/3126594.3126631`.

Interaction with Real-Life Objects in Augmented Reality in an Educational or Gaming Context

Julian Preissing

Ludwig-Maximilians-Universität München, München, Deutschland
`julian.preissing@campus.lmu.de`

Abstract

Augmented Reality (AR) offers a very intuitive way to add information to your surroundings by enhancing the vision. To make good use of the extra information, it is mandatory to have an intuitive way to interact with the AR application. The most intuitive interaction that can be used is the interaction with real-life objects. As AR is able to track the interaction with real-life objects, it can also be used to interact with AR applications. In this paper, the state of the art and future possible use cases regarding the interaction with real-life objects in AR were evaluated with a special focus on gaming and educational applications. After carefully reviewing the state of the art, it became clear that there are definitely use cases that can benefit from the interaction with real-life objects in AR, especially in the educational context.

2012 ACM Computing Classification Human-centered computing → Human computer interaction (HCI)

Keywords and phrases Augmented Reality; Real-Life Object Interaction.

1 Introduction

Augmented Reality (AR) offers an intuitive way to add information to your surroundings by enhancing the vision. Sophisticated AR applications can understand the surroundings and provide information that is tailored to what you see. To make good use of the extra information, it is important that the handling of AR devices does not consume too much attention or else the benefits are outweighed. A good way would be to use interaction tracking. Devices like Microsoft's Kinect make it easy to implement interaction tracking. Also, most Virtual Reality (VR) applications make use of this method, but it is still rare in AR applications. When enhancing one's vision in AR applications, the focus should be on the extra information provided to make good use of it

© Julian Preissing;
licensed under Creative Commons License CC-BY

Cite as: Julian Preissing. Interaction with Real-Life Objects in Augmented Reality in an Educational or Gaming Context. In *5th Seminar on Ubiquitous Interaction (UBIACTION 2020)*. Editors: Florian Lang, Pascal Knierim, Jakob Karolus, Fiona Draxler, Ville Mäkelä, Francesco Chiossi, Luke Haliburton, Matthias Hoppe, Albrecht Schmidt. January 27, 2020. Munich, Germany. pp. 8:1–8:14.

and the handling should be intuitive. A self-explanatory interaction is given with real-life objects that are used every day. To interact with these objects, it is not necessary to focus. If AR applications can track the interaction of real-life objects, a user can focus completely on the content that is provided via AR.

2 Augmented Reality Applications

2.1 Use Cases

The development of AR entertainment and educational applications has highly increased in the in the last five years. With more usage of AR in these sectors, there are more use cases which require different interaction methods [18]. The biggest difference in an application is whether it is designed for one user or for multiple users.

2.1.1 Single User

Most computer applications are designed to be used by one person at a time [9]. When working with a single-user AR application, the interaction is only between the user and the device. The focus is on the information provided by the device. The goal of single-user applications is to present information to the user or to entertain the user.

2.1.2 Collaborative AR

Multi-user applications, AR or not, require special programming dimensions. Patterson defined the following three dimensions in 1991 [26]:

- Concurrency: coping with parallel activities.
- Abstraction: separate the user interface from an underlying application abstraction.
- Roles: characterize users and customize the user-interface appropriately depending on roles.

While single-user applications may also handle some of these dimensions, they are not as essential as they are in multi-user applications according to Patterson [26]. There is an undeniable extra effort in designing multi-user applications but the benefits could outweigh the extra costs. Designing AR applications for more than one user creates a virtual space that can be accessed by various users simultaneously in order to collaborate. Conventional collaborative technology faces many problems, especially for interacting with spatial content. Normal face-to-face collaboration makes use of speech, gesture, gaze and non-verbal cues in order to communicate in the clearest way possible. Real objects also play an important role, particularly in design and spatial collaboration tasks. Augmented Reality techniques offer a new approach that can make use of the advantages of conventional face-to-face collaboration in the shared virtual space [2]. By augmenting the reality remote people can work together in the same virtual augmented place. A feeling of co-location can be created [22]. According to Mark Billinghurst, the greatest potential use for Augmented Reality is the development of new types of collaborative interfaces [2].

2.2 Display techniques

There are various ways to display Augmented Reality. The two main methods are either a handheld device or a projection-based solution. They both have advantages and disadvantages. Which solution is better depends on the use case.

2.2.1 Handheld AR

Handheld devices are very common. Most people, especially people that would want to use AR, have a smartphone. It is very user-friendly as no special setup is required. They are most fitting for untrained users in unconstrained and nonsupervised environments [32]. The only restriction is that one hand has to be used to hold the device.

2.2.2 Projection Based AR

On the other hand, projection-based AR offers a wider range of possibilities. Without a handheld device, your hands are free and can be used; either to interact with your AR device or to use them to interact with the world. For example, in cars, it is mandatory that you have both of your hands available, making a projection-based Head-Up Display (HUD) the better choice compared to an application on your phone. In an HUD, information gets projected onto the windshield into the field of vision. This can reduce stress and tension of the driver [14]. One of the other most known projection-based AR Displays are Head-Mounted Displays like the Microsoft Holo Lens. Sci-Fi Movies have often predicted a future with glasses that enhance your vision and are very subtle and easy to use. This is not quite the reality because right now these glasses are mostly used by professionals and are not really suitable for everyday private usage due to their price and design.

2.3 Augmented Reality used in Learning Applications

An interesting use case for AR applications are learning applications. Two types of AR can be used in the learning context. There is the location-aware and vision-based approach. Location-aware AR means that the information is presented to the learner while moving around in the real world. With GPS-enabled smartphones, the information to augment the reality can then differ depending on the current location of the user. Vision-based AR presents information to the user after pointing the camera of their smartphone at a defined target object [5]. The advantage of both AR methods is that the AR context (the real world) is an environment all students are already connected to. By enhancing it via AR, students are engaged to further explore it while being presented useful educational information [16]. The three most effective ways to enhance the reality for a learning purpose are real-world annotation, contextual visualization, and vision-haptic visualization [31]. An important skill that can be learned is a language and many AR applications offer their approach to learning a language. Further, they seemed to provide a better understanding for the students. One example is the AR game MOW, which was evaluated in a case study. MOW is a word-matching game that provides visual and auditory cues via AR to help children learn how to pronounce and write animal names. The results of Barreira's case study indicated that the children that used the AR game had a better learning progress than those who used conventional methods [1].

2.4 Augmented Reality used in the Gaming Context

Game developers always strive to adapt to new technologies early on. The first commercially successful usage of AR in a game was the EyeToy® (www.eyetoy.com) for Sony Playstation 2. It was released in 2003. The EyeToy® webcam allowed the user to interact with various games by using motion or sound while augmenting their reality [4]. One of the main aspects that provides pleasure and excitement in video games and the biggest difference to movies is the ability to actively interact with the content [10]. Therefore, it is necessary to put a lot of effort into the design of the control. The control should be easy to understand and should not consume to much attention that would be taken away from the actual game content. This is where AR brings a big advantage. Controlling a game with gestures might be the most intuitive way if designed properly, making any controller obsolete. Right now there is a shift in the gaming industry, away from expensive consoles onto devices that most people already have. Games on mobile phones have surpassed all other consoles in terms of revenue [33]. The smartphone, equipped with two cameras and a microphone, already has all the requirements an AR game needs. The expansion from 2D displays to 3D spaces that is possible with AR is an an ideal opportunity for game developers to design more playful and interesting mobile AR games that embed various interaction forms [17]. One commercially successful game was Pokemon GO, released in 2016. While the interaction in AR is limited, AR is used to place the Pokemon by looking through the camera of a smartphone as seen in Figure 1. Partially thanks to this feature, the game saw great success upon release that still lasts to this day. In 2019 Pokemon Go made a revenue of 774M US$.

Figure 1 Pokemon GO: Pokemon is placed in the real world with AR [27].

3 The Interaction with Real-Life Objects

Human-computer interaction (HCI) focuses on making interaction with computers intuitive. Often, inspiration is drawn from the nature. Making something "natural" makes it easy to understand and use [21]. The most natural interaction a human can have is the interaction with real-life objects. Long before the age of computers, humans were interacting with objects. Humans interact with objects of the real world the moment they are born. 3D objects in the real world have rich interaction behaviours that we learn to understand from day one. Visual, haptic and auditory aspects of objects provide important interaction cues. The "rich interaction behaviors [. . .] include how an object deforms on contact, how its surface feels when touched, and what kinds of sounds it makes when one interacts with it" [24].

Furthermore, Hornecker divided tangible interaction into four parts [13]:

- Tangible Manipulation: The feel of a material when touching it.
- Spatial Interaction: The placement of an object in a certain space.

- Embodied Facilitation: The configurations of objects and space and how these are able to slightly direct behavior.
- Expressive Representation: The legibility and significance of the presented material.

Due to these characteristics and the possibility to integrate real-life object interaction into AR applications makes this a topic of high interest for HCI.

4 Using the Interaction of Real-Life Objects in Augmented Reality: Tangible AR

A new approach that makes use of the interaction with real-life objects is tangible AR. By using real-life objects and their intuitive interaction, AR interfaces can be developed that support seamless interaction and are very intuitive to use. In Tangible AR, the enhanced display possibilities are combined with the intuitive interaction of physical objects [3]. Compared to a computer mouse that acts as a generic and transient intermediary, "these objects are simultaneously interface, interaction object and interaction device" [13].

4.1 Possibilities

One of the main advantages that Tangible AR has to offer is the intuitiveness of interaction that enhances the experience of using an application. Another application is the combination of 2D and 3D space. By creating a virtual space that includes real-life objects but can also include virtual 2D programs, there is no need to switch between those spaces. Instead, they can be used seamlessly in the same space, offering a higher productivity [12].

4.2 Problems

While AR already enables users to manipulate digital objects with tangible physical tools, there are still some limitations and difficulties which will be discussed in the following chapter. The difficulties can be divided into technological and human factors [36].

4.2.1 Technological Challenges

The physical and the digital world are nowhere near the same. It is clear that any attempt to close a gap between two spaces that are so different will face many technological challenges. In order for computers to understand gestures of humans, they have to learn them. Giving computers a complete understanding of the human language and the real world is not yet feasible, therefore most approaches create a "new language" that consists of only parts of the human language. This restricts the intuitiveness of applications. Users have to know which gestures and commands can be interpreted by the computer in order to interact with it [8].

Another technological challenge is the ability of a system to recognize a physical object in the real world. While successful efforts have been made, they face challenges depending on the texture. An example is the Universal Media Book [11], which used a physical book as an tangible interface to present projected images. The researchers used feature points to calculate the geometry of the pages but reported to have problems with the estimation of the geometry if the content to be rendered did not have enough signal change [11]. Another tracking problem independent of texture is the hand occlusion. The system has to recognize a hand in front of an object and has to deal with it in a certain way, either detecting an interaction or ignoring the hand covering up a physical object that is part of the augmented reality.

4.2.2 Human Factors

From a human factors point of view, there are also issues that have to be considered. The design of an AR application can be complicated, resulting in a complicated user experience [36]. In order to realize a tangible AR interface, it is important to keep the usability in mind. As the technological challenges might require programming tweaks to realize the functionality, it is important to keep the interaction simple and intuitive.

5 Use Cases for AR Real-Life Object Interaction

The interaction with real-life objects offers an interesting new interaction method that can be used in AR. Both educational and gaming applications rely on good interaction methods. Applications in both sectors can make good use of tangible AR. In the following chapter, use cases for tangible AR in current applications will be explored.

5.1 Tangible AR in Educational Applications

Most educational AR applications (tangible or not) try to present the learning material in a playful manner. These applications are called serious games or edutainment [15, 20]. Serious games using AR have already been shown to provide a better learning experience compared to traditional learning techniques [1]. By adding real-life models to the experience, the learning progress can be further optimized. Getting the well-known haptic feedback of known objects combined with the additional information provided via AR seems to further strengthen the learning process. An example application that made use of tangible cubes as a user interface was evaluated in a study [15]. The goal of the game was to learn about endangered animals in a fun way. The results of the AR game were compared to a conventional learning game providing the same information. The results showed that the children enjoyed the AR game more but thought it was more complex to interact with. This further shows that there are technical challenges when implementing AR games in an easy-to-use manner as described in Chapter 4, while also proving that a tangible interface in the AR context is a viable interaction method that provides more fun [15]. Another approach to tangible AR in learning applications is the enhancement of a well-known learning device. The book can be enhanced via AR. Conventional books can create a new experience while providing the well-known feeling a book has. Additional information, animations and graphics that can be added in AR can bring the book to life and create a new exciting experience for the reader while providing more useful information and explanations [6]. An example application that enhances the book is an AR alphabet book designed for preschool children. The application enhanced the book by virtually adding animations of objects that start with the letter that is to be learned. In an evaluation of the application, the results were that the usage of AR to enhance the book "generated excitement, engagement and enjoyment during the learning process" [29].

In the future, tangible AR applications could be combined with machine learning. AR tools that are able to recognize real-life objects that are presented and provide a translation to a desired language already exist [7] and could be further optimized with advances in machine learning. Machine learning tools are already able to recognize objects by their looks [25] but efforts have to be made to optimize such an application. When working with actual objects of the surroundings, a better learning effect could be reached but to get a certain result, an application implementing a feature like this would ultimately have to be evaluated.

5.2 Tangible AR Games

As shown in the previous section, AR learning applications are able to create an exciting experience for the user. As the main goal of conventional entertainment games is to entertain the user, it is only fitting to make use of this technology in the gaming context. As shown in Chapter 2.3, AR games are broadly available nowadays and have already seen great commercial success. But not all the games make use of tangible AR. One factor might be that the necessity of a special real-life object, that eventually would have to be bought, could scare off customers. Nonetheless, efforts have been made to create and evaluate games that make use of tangible AR. One area where tangible AR makes good sense are board games. They provide physical objects that can be enhanced via AR. Molla and Lepetit already made first efforts to implement an AR board game. In their paper, they augmented the well-known board game Monopoly only by using a webcam. Their implementation created a "better immersion compared to the original board game alone and provides a different experience than a video game" [23]. With their approach, any board game could be enhanced via AR to create a new, more immersive experience for the players. Another game that was developed especially to combine computer and board games is called TableTop [35]. In an evaluation, it was concluded that the game combines the advantages of computer games and board games. Computer games provide the player with the exciting animations, while board games provide the easy interaction method [35]. Other than board games, there were also efforts made to evaluate tangible AR in computer games. Tedjokusumo, Zhou, and Winkler created and evaluated a new immersive multiplayer game system developed for virtual and augmented reality [34]. In order to evaluate the game system, they developed three game applications. Users could interact with the game world by physically moving around objects of the real world. In their evaluation, they compared the VR to the AR version. The VR version was the preferred game mode as it created an even more immersive environment. While Tangible Augmented Reality is a very effective environment for certain types of multiplayer computer games, it is not fitting for all games [35]. A good approach is to enrich an already existing game (like board games). When it comes to action games like first-person shooters, VR might be a better approach as it creates a more immersive feeling while still allowing for the same tangible interaction as AR.

5.3 Other possible use cases

Besides educational or gaming applications, there are other use cases that can make use of tangible AR. One example is the tangible AR desktop environment in which the standard 2D computer desktop is integrated into an augmented 3D space [30]. They used the physical space given by the standard office desk and enhanced the virtual space by making the user wear a head-mounted display. They could then interact with the environment in both tangible and virtual ways. In this augmented environment, the combination of 2D and 3D space created an improvement on the users' productivity according to their evaluation [30].

Another possible application is in the world of product design. Tangible Foam, a blue foam material, was used to create tangible feedback for high-fidelity prototypes. The foam is easy to form and can then be overlaid with a 3D virtual object in AR, providing the virtual product design. Augmented foam is a viable method to rapidly make high-fidelity prototypes that provide tangible feedback [19].

A possible research field for future work could be the evaluation of the

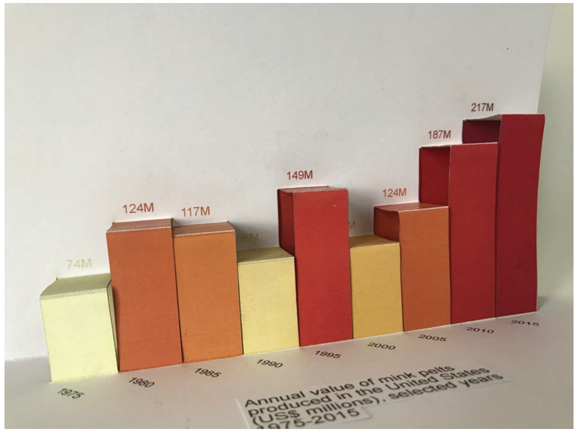


Figure 2 Kirigami Bar Chart that could be enhanced with AR [27].

combination of tangible AR with physical information visualizations. One example are Kirigami models that visualize information (as seen in Figure 2). Kirigami, the Japanese art of cutting and folding paper provides a simple method to generate physical models in a short amount of time [28]. By cutting and folding a piece of paper in certain ways, Kirigami allows for the creation of almost any three-dimensional form. When looking at the models through augmented reality, color, labeling and more information could be provided by AR while providing a tangible interface in form of the paper models.

6 Conclusion

In conclusion, it seems that tangible AR can enrich learning and gaming applications. To implement an application that makes use of tangible AR is still a technical difficulty. It is especially difficult to create an application that implements the tangibility in an intuitive way for the user to use. Applications that use tangible AR have shown to improve the learning and entertainment effect, making tangible AR a viable user interface for certain applications. The most promising results were in learning applications for children. The combination of virtual and real world seems to provide a good learning atmosphere. In regards to entertainment games, tangible AR seems like a good way to enrich games that already use physical objects, like board games. In terms of action-rich games, VR might be the better choice compared to AR, as the full immersion in VR seems to create a better user experience.

References

1 João Barreira, Maximino Bessa, Luciana C Pereira, Telmo Adão, Emanuel Peres, and Luís Magalhães. Mow: Augmented reality game to learn words in different languages: Case study: Learning english names of animals in elementary school. In *7th Iberian Conference on Information Systems and Technologies (CISTI 2012)*, pages 1–6. IEEE, 2012.

2 Mark Billinghurst and Hirokazu Kato. Collaborative augmented reality. *Communications of the ACM*, 45(7):64–70, 2002.

3 Mark Billinghurst, Hirokazu Kato, and Ivan Poupyrev. Tangible augmented reality. *ACM SIGGRAPH ASIA*, 7, 2008.

4 Wolfgang Broll, Jan Ohlenburg, Irma Lindt, Iris Herbst, and Anne-Kathrin Braun. Meeting technology challenges of pervasive augmented reality games. In *NetGames*, page 28. Citeseer, 2006.

5 Matt Dunleavy and Chris Dede. Augmented reality teaching and learning. In *Handbook of research on educational communications and technology*, pages 735–745. Springer, 2014.

6 Andreas Dünser and Eva Hornecker. Lessons from an ar book study. In *Proceedings of the 1st international conference on Tangible and embedded interaction*, pages 179–182. ACM, 2007.

7 Robert Godwin-Jones. Augmented reality and language learning: From annotated vocabulary to place-based mobile games. *Language Learning & Technology*, 20(3):9–19, 2016.

8 Matthew Gorbet, Maggie Orth, and Hiroshi Ishii. Triangles: Tangible interface for manipulation and exploration of digital information topography. pages 49–56, 01 1998.

9 Saul Greenberg. Sharing views and interactions with single-user applications. In *ACM SIGOIS Bulletin*, volume 11, pages 227–237. ACM, 1990.

10 Torben Grodal et al. Video games and the pleasures of control. *Media entertainment: The psychology of its appeal*, pages 197–213, 2000.

11 S. Gupta and C. Jaynes. The universal media book: tracking and augmenting moving surfaces with projected information. In *2006 IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 177–180, Oct 2006. `doi:10.1109/ISMAR.2006.297811`.

12 Valentin Heun, Shunichi Kasahara, and Pattie Maes. Smarter objects: using ar technology to program physical objects and their interactions. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pages 961–966. ACM, 2013.

13 Eva Hornecker and Jacob Buur. Getting a grip on tangible interaction: a framework on physical space and social interaction. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 437–446. ACM, 2006.

14 Yoonsook Hwang, Byoung-Jun Park, and Kyong-Ho Kim. Effects of augmented-reality head-up display system use on risk perception and psychological changes of drivers. *ETRI Journal*, 38(4):757–766, 2016.

15 Carmen M Juan, Giacomo Toffetti, Francisco Abad, and Juan Cano. Tangible cubes used as the user interface in an augmented reality game for edutainment. In *2010 10th IEEE international conference on advanced learning technologies*, pages 599–603. IEEE, 2010.

16 Eric Klopfer and Josh Sheldon. Augmenting your own reality: Student authoring of science-based augmented reality games. *New directions for youth development*, 2010(128):85–94, 2010.

17 Raymond Koon Chuan Koh, Henry Been-Lirn Duh, and Jian Gu. An integrated design flow in user interface and interaction for enhancing mobile ar gaming experiences. In *2010 IEEE International Symposium on Mixed and Augmented Reality-Arts, Media, and Humanities*, pages 47–52. IEEE, 2010.

18 Gun A Lee, Claudia Nelles, Mark Billinghurst, Mark Billinghurst, and Gerard Jounghyun Kim. Immersive authoring of tangible augmented reality applications. In *Proceedings of the 3rd IEEE/ACM international Symposium on Mixed and Augmented Reality*, pages 172–181. IEEE Computer Society, 2004.

19 Woohun Lee and Jun Park. Augmented foam: A tangible augmented reality for product design. In *Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'05)*, pages 106–109. IEEE, 2005.

20 Fotis Liarokapis and Sara De Freitas. A case study of augmented reality serious games. In *Looking Toward the Future of Technology-Enhanced Education: Ubiquitous Learning and the Digital Native*, pages 178–191. IGI global, 2010.

21 Fotis Liarokapis, Ian Greatbatch, David Mountain, Anil Gunesh, Vesna Brujic-Okretic, and Jonathan Raper. Mobile augmented reality techniques for geovisualisation. In *Ninth International Conference on Information Visualisation (IV'05)*, pages 745–751. IEEE, 2005.

22 Stephan Lukosch, Mark Billinghurst, Leila Alem, and Kiyoshi Kiyokawa. Collaboration in augmented reality. *Computer Supported Cooperative Work (CSCW)*, 24(6):515–525, Dec 2015. `doi:10.1007/s10606-015-9239-0`.

23 Eray Molla and Vincent Lepetit. Augmented reality for board games. In *2010 IEEE International Symposium on Mixed and Augmented Reality*, pages 253–254. IEEE, 2010.

24 Dinesh K Pai, Kees van den Doel, Doug L James, Jochen Lang, John E Lloyd, Joshua L Richmond, and Som H Yau. Scanning physical interaction behavior of 3d objects. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 87–96. ACM, 2001.

25 Giulia Pasquale, Carlo Ciliberto, Francesca Odone, Lorenzo Rosasco, and Lorenzo Natale. Teaching icub to recognize objects using deep convolutional neural networks. In *Machine Learning for Interactive Systems*, pages 21–25, 2015.

26 John F Patterson. Comparing the programming demands of single-user and multi-user applications. In *Proceedings of the 4th annual ACM symposium on User interface software and technology*, pages 87–94. Citeseer, 1991.

27 Julian Preissing. Picture taken by myself.

28 Julian Preissing. The application of kirigami for physical information visualization. 2018.

29 Dayang Rohaya Awang Rambli, Wannisa Matcha, and Suziah Sulaiman. Fun learning with ar alphabet book for preschool children. *Procedia computer science*, 25:211–219, 2013.

30 Holger Regenbrecht, Gregory Baratoff, and Michael Wagner. A tangible ar desktop environment. *Computers & graphics*, 25(5):755–763, 2001.

31 Marc Ericson C Santos, Angie Chen, Takafumi Taketomi, Goshiro Yamamoto, Jun Miyazaki, and Hirokazu Kato. Augmented reality learning experiences: Survey of prototype design and evaluation. *IEEE Transactions on learning technologies*, 7(1):38–56, 2013.

32 Dieter Schmalstieg and Daniel Wagner. Experiences with handheld augmented reality. In *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 3–18. IEEE, 2007.

33 Dean Takahashi. Superdata: Games hit 120.1 billion in 2019, with fortnite topping 1.8 billion. https://venturebeat.com/2020/01/02/superdata-games-hit-120-1-billion-in-2019-with-fortnite-topping-1-8-billion/, Accessed: 2020-02-26.

34 Jefry Tedjokusumo, Steven ZhiYing Zhou, and Stefan Winkler. Immersive multiplayer games with tangible and physical interaction. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 40(1):147–157, 2009.

35 Christiane Ulbricht and Dieter Schmalstieg. Tangible augmented reality for computer games. 04 2003.

36 Feng Zhou, Henry Been-Lirn Duh, and Mark Billinghurst. Trends in augmented reality tracking, interaction and display: A review of ten years of ismar. In *Proceedings of the 7th IEEE/ACM international symposium on mixed and augmented reality*, pages 193–202. IEEE Computer Society, 2008.

Human-Centered AI: Challenges and Opportunities

Vanessa Sarakiotis

Ludwig-Maximilians-Universität München, München, Deutschland
V.Sarakiotis@campus.lmu.de

Abstract

Intelligent systems that have Artificial Intelligence (AI) or Machine Learning (ML) algorithms integrated should support their users by making predictions and provide help for decision making tasks. This is only possible if these intelligent systems are applicable and comprehensible for the user. The information that is collected of the user's behavior influences the decision making process of the system which is likely to cause some challenges for users as well as for UX designers. Examples are the "black box" problem, imbalanced predictions, the user's fear of losing control or misaligned expectations. For this reason and because AI is affecting people's daily lives while becoming ubiquitous, there is a need for novel approaches. This paper offers an overview of the current status of human-centered AI challenges and suggestions to solve them. Useful opportunities were considered to be e.g. concrete design guidelines, collaborative decision making, explanation directives, multi- and interdisciplinary design approaches. It is obvious that not all of the proposals are adaptable or simple to combine but can offer a basic strategy to approximate the development of intelligent systems. This shared information should serve as basic knowledge for researchers from different disciplines, support the evaluation of intelligent systems and simplify future human-centered AI designs.

2012 ACM Computing Classification Human-centered computing →Ubiquitous and mobile computing

Keywords and phrases Human-centered AI, Machine learning, Adaptive user interfaces, design guidelines, explainability, transparency

1 Introduction

ML is an active research topic for about fifty years now but recently gained a lot of attention through the rising importance of big data for companies, developments in deep learning, the increasing use of analytic tools, the research on autonomous driving and especially the popularity of apps that utilize ML

© Vanessa Sarakiotis;
licensed under Creative Commons License CC-BY

Cite as: Vanessa Sarakiotis. Human-Centered AI: Challenges and Opportunities. In *5th Seminar on Ubiquitous Interaction (UBIACTION 2020)*. Editors: Florian Lang, Pascal Knierim, Jakob Karolus, Fiona Draxler, Ville Mäkelä, Francesco Chiossi, Luke Haliburton, Matthias Hoppe, Albrecht Schmidt. January 27, 2020. Munich, Germany. pp. 9:1–9:20.

for more personalized and reactive interactions. Features that are based on ML algorithms that apps have commonly integrated are for example the detection and filter of spam, the arrangement of feeds, translations, speech, gesture and face recognition, autocorrect or predictions for daily activities e.g. driving time. Additionally, due to the increasing number of companies that integrate AI and ML systems into their work and therefore decision making processes, people are nowadays directly affected as law companies can, for example, utilize predictions to pass a sentence based on actual evidence. Furthermore, CHI workshops and the establishment of human-centered AI (HAI) research institutes at Stanford University, UC Berkeley and MIT contributed to the current interest. [7, 14, 29]

This is why it is legitimate to state that *"systems that produce predictions [...] are becoming ubiquitous"* [23]. Considerations that have to be made when designing AI systems for non-experts are ensuring that these systems offer useful and usable AI [29]. It is necessary to integrate the *"individuals' differences, demands, values, expectations, and preferences"* [14] from the beginning of the design process. This is known as "Human-centered Machine Learning" or more generally "Human-centered Artificial Intelligence" *"and highlights the need to take a human-centered look at how ML solutions are impacting people"* [23].

The extended HAI framework proposes the requirement of an ethically aligned design which represents the people's mental model and offers familiarity for the user with the technology. Finally, making the developed AI systems intelligible and useful refers to human factors design. [29]

This statement is supported by Ramos et al. who arguments that the goal of this research area is to make ML accessible, understandable, intelligible and trustworthy [23]. Hence, it is crucial to provide the user with the information why the result was the output of the AI system and how exactly these results were achieved ("black box" problem) by designing AI systems with the user and human goals in mind [29]. To provide the approach of a human-centered AI system, multidisciplinary research has to be conducted including research areas such as law, social science, philosophy and psychology [29].

1.1 Definition of AI and Machine Learning

In general, AI comprises the system implementation of human-like intelligence in order to enhance the user's capabilities and support beyond human limitations with e.g. decision-making and predictions. AI has already been applied to areas such as natural language processing, game theory, and machine learning. These systems involve statistical methods for modeling and learning over time. [20]

As most of the previous literature focuses on Machine Learning algorithms, this paper will as well. A system that uses ML algorithms for predictions or decision making collects data on the user's behavior and stores it in a large database. It is then split into training and test data. Additionally, models are determined that represent how the system functions. How well the model learns and therefore performs is defined by the resulting value which should be as high as possible to allow assuming that the algorithm performs in the desired way. [28]

1.2 From human-computer interaction to human-computer integration

In recent years, the development from human-computer interaction to human-computer integration was identifiable. Instead of only focusing on the command-response interaction, AI systems can utilize the user's personal data or general contextual information and deploy it to the user's advantage. This does not mean that soon there will no longer exist human-computer interaction. On the contrary, interaction and integration should be combined *"in the broad sense of a partnership of symbiotic relationship"* [9] to offer the most valuable solutions to support and to enhance the user's capabilities. [9, 29]

The further development of human-computer integration leads us to new design and evaluation options where basic design guidelines are not sufficient anymore resulting in more complex procedures [7]. Deuschel and Scully define that AUIs automatically change their appearance based on the system's collected data on the user's behavior to adapt the interface to the assessed desires [6, 9]. Farooq and Grudin state that this also includes rethinking the minimalistic concept of making user interfaces intuitive without offering any additional explanation [9].

So far, previous literature did not provide an overview of the whole research area of human-centered AI. A few papers only focused on specific problems, psychological approaches or offered implementation examples without human factors in mind. Additionally, a lot of authors complained about insufficient research in that field. This is why the contributions of this paper are to provide an update and an overview of challenges associated with human-centered AI. This is done by presenting design guidelines for dealing with these challenges and offering implementation examples.

To achieve this goal, a literature review was conducted. For this purpose, three reviewed papers were determined as the main sources [1, 14, 29] because they included crucial principles and consequently served as a guide for where this subject should lead. The next step was to examine these paper's literature reviews and select relevant papers. Additionally, the "ACM Digital Library" was searched considering filters such as the publication year (2016 to 2019) and alternately various conference types e.g. CHI, IUI, KDD, and UBICOMP which were rated as valuable for this paper.

The paper is organized as follows: first, the challenges that user and UX designers face with AI/ML systems are focused on. To better comprehend how the then mentioned issues should be dealt with, some opportunities are presented as well as implementation examples for explanation user interfaces. Eventually, the conclusion summarizes the most crucial findings and offers an outlook for future research.

2 Challenges of AI systems

As already described above, human-computer interaction developed towards human-computer integration as AI systems learn independently from the user's behavior which rearranges the relationship between the two of them to a collaborative relationship. This can be very useful in many ways but at the same time can increase complexity for the HCI design. [29]

Therefore, human challenges, concerns, and desires should be taken into account [14].

2.1 Utilization obstacles of users

As mentioned in the introduction, many recent AI systems were designed without the human in mind. Nevertheless, as those systems should serve as an enhancement of the user's capabilities, it is necessary to integrate the user's demands into the design and development process to ensure that the system is practical, accepted, and trusted by the customers. [14, 29]

To serve this purpose, it is crucial to *"incorporate AI systems into interactive, usable, and actionable technologies that function in the natural contexts of all human stakeholders in a bias-free manner"* [14].

An obstacle that should be considered is the "black box“ problem. This problem describes that it is unclear to the user why the result of an AI system was the output and how exactly this result was achieved [29]. It is challenging making AI systems comprehensible, explainable (XAI) and intelligible for the respective user because of the statistical methods that are utilized [7, 29]. Nevertheless, these features are crucial as intuitive use is not yet provided [23]. Unfortunately, until now it is unclear what the best concept, representation or explanation type for making AI systems comprehensible is [27]. A feature that makes it even harder to understand, explore and debug the processes and the results is the AI system's ability to learn which leads to changes in its behavior over time [23]. This could result in insecurities, trust and adoption issues by the user who is not sure when to trust the system and hesitates accepting it as a support. Furthermore, the acceptance and adoption may be risked by inappropriate integration of or interaction with the user. An example is biased "thinking" caused by insufficient or distorted data. A potential consequence is inequality as the ML model is trained in a biased way. [29] However, the lack of knowledge of how to balance those biases to get a „fair“ result is not easy to achieve for real-life examples as some information is always missing [26].

Recently, there were a lot of discussions about the user being afraid of losing control to intelligent systems which continued to oppose the automated behavior of the system with the direct manipulation of the user [14]. However, the goal should be to combine both to serve the user in the most valuable way which can be described as "Collaborative Decision Making" [13, 29].

Additionally, unrealistic or misaligned expectations of the user of what the AI system is able to do can result in disappointment or abandonment, especially for sensitive contexts. This could lead to the problem that intelligent features are perceived as useless. [7]

2.2 Challenges for UX designers

At the moment there are no specific design standards to develop AI systems compared to basic design guidelines for non-intelligent systems as the "learning" of the system does not have to be considered [29]. Besides, UX designers are confronted with higher complexity as statistical methods are applied which are difficult to comprehend [7]. This is why it is even more crucial to include the design procedure early in the developing process to make the system actually usable for the ones that are supposed to use it and to avoid misalignments [15].

The „black box“ problem is not only an issue for the user but also for UX designers who do not fully understand or misconceive how ML algorithms work and what they are able to do. Traditional literature states that creative parties such as UX designers do not need to know the technical details of a system. How should they design interfaces that are intelligible and explainable for the user then? This misunderstanding probably leads to UX designers not embracing new opportunities that this research field is offering. [30]

Therefore, UX designer's desire to collaborate with AI/ML experts to understand the background processes of the systems better [7].

Another difficulty is the lack of prototyping tools for dynamic systems. In most cases, prototypes for intelligent systems would already require a lot of data in order to prevent the system from containing too many uncertainties. This is why a lot of effort is already included to only provide very basic prototypes. [7]

Speaking more generally, there is a lack of multidisciplinary research and cooperation that could improve and resolve insights for HCI design associated with AI systems [7, 29].

2.3 Expectation and trust

If users accept a system or not, strongly depends on what expectations they have and determines whether the user is satisfied [16, 19]. The accuracy of AI systems is usually vague which is why they very rarely perform without any errors. As the users are not prepared for that or just expect a perfectly working appliance, using them can result in disappointment with a negative impact on the user's trust. Giving the user more information before the first use and therefore drawing the user's attention to challenges he or she could face, can influence trust negatively too and maybe even prevent the customer from buying the product. At the same time, the trust of the user can be strengthened by communicating potential weaknesses of the system as he or she knows what to expect. When this first obstacle has been overcome, the experience can uplift the user and exceed expectations. [16]

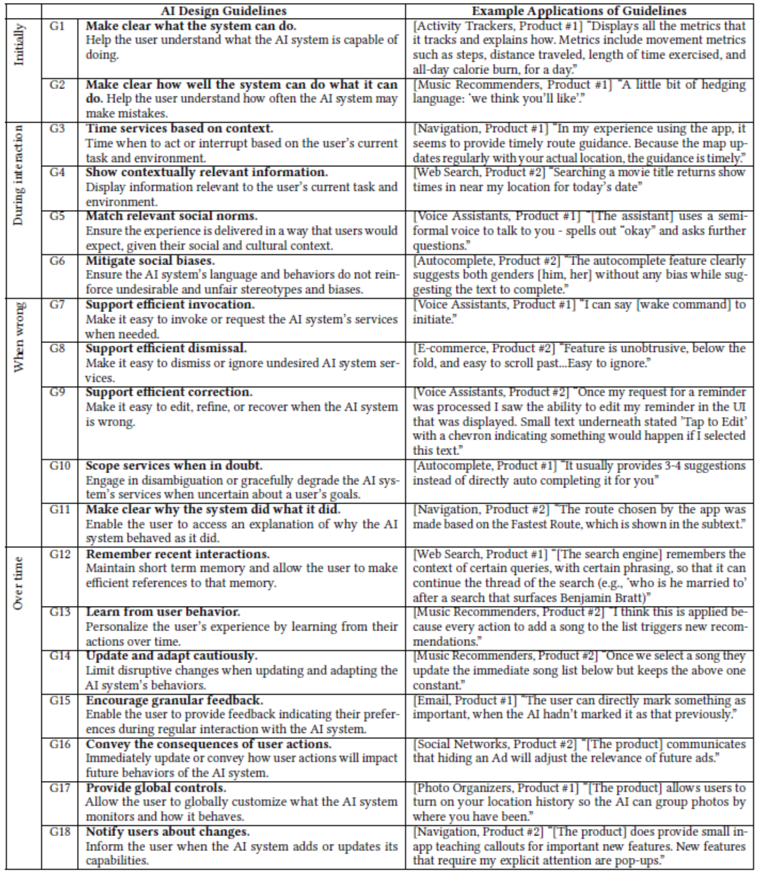
In general, previous work has shown that expectations of AI systems are influenced by *"a variety of factors including external information, knowledge and understanding, and first-hand experience"* [16]. This includes information about the system's properties by third parties and the own interpretation of its capabilities [16].

3 Design guidelines for human-centered AI systems

As the application area of AI systems and the integration of AI algorithms are likely to increase looking at the current progress, it could be very useful to introduce guidelines that are available to HCI researchers, UX designers and AI/ML developers. The guidelines should serve them as assistance regarding the design and evaluation of AI systems. The overall goal is to make those systems intelligible, understandable and trustworthy as explained in the previous chapters. [1]

3.1 Concrete guideline formulations

Intelligent systems should be useful and usable which includes explaining the user why the result was the output and how exactly these results were achieved by designing AI systems with the user and human factors in mind [14, 29]. Amershi et al. present a set of design guidelines (G) for AI systems with the focus on human-AI interaction and integration which were achieved in multiple evaluation iterations (see Table 1) [1]. In addition to the accompanying explanations below the directive headlines, example applications were listed to simplify the abstract expressions and to clarify current fields of application. To get an understanding of when to focus on which guideline in the process, the authors additionally added various workflow stages to the first left column of the table. The proposed guidelines rather apply to low-risk areas as the priorities and consequences for high-risk systems can shift. Obviously, there are appliances such as voice assistance devices that can not consider all of these guiding principles as some are just not applicable without a graphical user interface. Considerations that have to be made are: not all of these guidelines are already implemented in current devices or applications (e.g. G2, G11, G17) which refers to the challenges described previously such as the why



	AI Design Guidelines		Example Applications of Guidelines
Initially	G1	**Make clear what the system can do.** Help the user understand what the AI system is capable of doing.	[Activity Trackers, Product #1] "Displays all the metrics that it tracks and explains how. Metrics include movement metrics such as steps, distance traveled, length of time exercised, and all-day calorie burn, for a day."
	G2	**Make clear how well the system can do what it can do.** Help the user understand how often the AI system may make mistakes.	[Music Recommenders, Product #1] "A little bit of hedging language: 'we think you'll like'."
During interaction	G3	**Time services based on context.** Time when to act or interrupt based on the user's current task and environment.	[Navigation, Product #1] "In my experience using the app, it seems to provide timely route guidance. Because the map updates regularly with your actual location, the guidance is timely."
	G4	**Show contextually relevant information.** Display information relevant to the user's current task and environment.	[Web Search, Product #2] "Searching a movie title returns show times in near my location for today's date"
	G5	**Match relevant social norms.** Ensure the experience is delivered in a way that users would expect, given their social and cultural context.	[Voice Assistants, Product #1] "[The assistant] uses a semi-formal voice to talk to you - spells out "okay" and asks further questions."
	G6	**Mitigate social biases.** Ensure the AI system's language and behaviors do not reinforce undesirable and unfair stereotypes and biases.	[Autocomplete, Product #2] "The autocomplete feature clearly suggests both genders [him, her] without any bias while suggesting the text to complete."
When wrong	G7	**Support efficient invocation.** Make it easy to invoke or request the AI system's services when needed.	[Voice Assistants, Product #1] "I can say [wake command] to initiate."
	G8	**Support efficient dismissal.** Make it easy to dismiss or ignore undesired AI system services.	[E-commerce, Product #2] "Feature is unobtrusive, below the fold, and easy to scroll past...Easy to ignore."
	G9	**Support efficient correction.** Make it easy to edit, refine, or recover when the AI system is wrong.	[Voice Assistants, Product #2] "Once my request for a reminder was processed I saw the ability to edit my reminder in the UI that was displayed. Small text underneath stated 'Tap to Edit' with a chevron indicating something would happen if I selected this text."
	G10	**Scope services when in doubt.** Engage in disambiguation or gracefully degrade the AI system's services when uncertain about a user's goals.	[Autocomplete, Product #1] "It usually provides 3-4 suggestions instead of directly auto completing it for you"
	G11	**Make clear why the system did what it did.** Enable the user to access an explanation of why the AI system behaved as it did.	[Navigation, Product #2] "The route chosen by the app was made based on the Fastest Route, which is shown in the subtext."
Over time	G12	**Remember recent interactions.** Maintain short term memory and allow the user to make efficient references to that memory.	[Web Search, Product #1] "[The search engine] remembers the context of certain queries, with certain phrasing, so that it can continue the thread of the search (e.g., 'who is he married to' after a search that surfaces Benjamin Bratt)"
	G13	**Learn from user behavior.** Personalize the user's experience by learning from their actions over time.	[Music Recommenders, Product #2] "I think this is applied because every action to add a song to the list triggers new recommendations."
	G14	**Update and adapt cautiously.** Limit disruptive changes when updating and adapting the AI system's behaviors.	[Music Recommenders, Product #2] "Once we select a song they update the immediate song list below but keeps the above one constant."
	G15	**Encourage granular feedback.** Enable the user to provide feedback indicating their preferences during regular interaction with the AI system.	[Email, Product #1] "The user can directly mark something as important, when the AI hadn't marked it as that previously."
	G16	**Convey the consequences of user actions.** Immediately update or convey how user actions will impact future behaviors of the AI system.	[Social Networks, Product #2] "[The product] communicates that hiding an Ad will adjust the relevance of future ads."
	G17	**Provide global controls.** Allow the user to globally customize what the AI system monitors and how it behaves.	[Photo Organizers, Product #1] "[The product] allows users to turn on your location history so the AI can group photos by where you have been."
	G18	**Notify users about changes.** Inform the user when the AI system adds or updates its capabilities.	[Navigation, Product #2] "[The product] does provide small in-app teaching callouts for important new features. New features that require my explicit attention are pop-ups."

Table 1 Human-centered AI guidelines introduced by Amershi et al. including exemplary applications [1].

and how the system acted as it did (lack of explainability and intelligibility). This backlog amplifies the requirement of more research in that area. [1]

If the system did something wrong and therefore the output is an incorrect prediction, the system should explain why it behaved like that (G11). Sometimes this is desired even though the system behaved as expected. An example where this feature is often missing are recommender systems in the e-commerce area where customers do not understand why products they are absolutely not interested in are suggested to them. Another example is navigation devices that tell the user that the current route is the best one but do not offer any information about the reason. [1]

3.2 Collaborative decision making

Taking a closer look at G17, it proposes to give the user the opportunity to customize the behavior of the AI system. A possibility is to build verification steps into the system to provide direct manipulation by the user. Consequently, the user can intervene if the automated system shows any signs of errors or behavior that was not intended by the user. [1]

As a result, combining automated behavior with the user's direct interaction seems to offer useful opportunities and is known as "Collaborative Decision Making" [29]. Not giving the people the impression of being able to directly control the system is likely to lead to abandonment, mistrust or misuse of the AI product [5, 19].

Implementation directives are for instance that the automated services should provide true value to the system that can not be achieved by the user's interaction only and setting the timing of automated actions right to avoid too much distraction of the user. If the system is unsure about the user's intentions, it should be able to start a dialog. Furthermore, the user should be able to degrade the precision of the system and therefore enlarge the diversity of the results. The behavior of the system should additionally apply to social expectations to make sure the interaction is as intuitive as possible for the user. [13]

3.3 Explanation and transparency guidelines

Now that we know about the challenges that the user's expectations and trust involve, it is necessary to take this variability into account and provide more detailed directives that deal with explanations and transparency. This refers to G1 and G11 in particular [1]. Explanations serve the goal of enhancing the transparency of decision making systems and to *„improve system usability and overall performance“* [22].

Generally, explanations that are shown to the users should be as simple as possible but at the same time be accurate about what the system does which is called "local fidelity" [24] and is often hard to reconcile [28]. A required question to ask here is: in which situations are explanations actually necessary for the user?

Bunt et al. found out that not all users are always interested in gaining more information about the processes. While some do desire more insights, the cost of the time needed to deal with them seems too great and therefore outweighs the benefits. However, most users seem to be only interested in viewing explanations when there is a trigger or a need to read them in order to clarify misunderstandings or questions. Such triggers are for instance that the user thinks he or she can improve the interaction or enhance the understanding of odd behavior. [3]

The education and learning area, for example, provides some implementation examples of machine learning algorithms where the explanation and therefore the transparency is perceived as useful. He et al. noticed that for Massive Open Online Courses (MOOCs) the completion rate of students who enrolled for a lecture was very low (two percent). This is why an intervention was integrated into Coursera MOOCs but only for the students who were close to the pass/file threshold. This intervention contained, amongst others, the presentation of the predictions, the reasons why he or she could fail and suggestions what the student could do to increase the chance of succeeding based on what other students did. [12]

Another example is CourseMIRROR by Fan et al. which is an intelligent mobile learning system that reminds students to recap their reflections on a lecture and collects it while reviewing the quality of this feedback. With the help of Natural Language Processing (NLP), these reflections are summarized. This way, it can be of great advantage to the teachers as well as the students to keep them engaged. [8]

3.4 Prototyping techniques

As already mentioned, it is difficult for UX designers nowadays to design for AI/ML systems because they miss prototyping tools. An approach for a human-centered design prototyping technique for AI developments is the Wizard of Oz (WOz) method that was introduced in this context by J. T. Browne and enables the researchers to evaluate their designs. The WOz method is a prototyping technique where the function of the system is performed by the researcher to give the user the impression he or she is actually interacting with the system. [2]

One big benefit of this technique is that *„a low-fidelity method to prototype ML experiences would bridge the gap in getting designers involved earlier* *on in the process, helping designers understand the models more sincerely, and generating user feedback on the model design“* [2]. How is an earlier understanding during the process provided? As a low-fidelity prototype can only provide limited interaction, the designer has to have an understanding of the behavior and the limits in order to set this technique up. Additionally, this requires an early collaboration between UX designers and AI developers. Limitations of the WOz method are that usually large data sets are used by the system to „learn“ over time from the user's behavior. However, to make this method applicable, the researcher could receive information directly during the session and react to it immediately. [2]

3.5 Collaboration between developers and designers

The prototyping technique WOz is just one type of collaboration between UX designers and AI developers. In general, a collaboration could help ML researchers to understand human requirements, UX designers to comprehend AI algorithms better as well as determine appropriate human-centered goals from the beginning. Crucial factors for a successful cooperation are: showing results and examples to update each other on the current status of the project, communicate processes and timelines as well as setting timings and goals together. [15]

Other approaches are that UX designers learn about AI by reading and applying education materials, attending workshops and listening to data scientists demos [30]. An exemplary workflow could involve three stages: in the first phase, the design goals should be determined with the user in mind and probably also include telemetry data (data about the user's behavior of services currently in use). In stage two, both the UX designers as well as the developers begin with the implementation while in phase three the system is iteratively evaluated and improved. [30]

3.6 Avoiding biased outcome

One suggestion to avoid biased data that results in unfair results are third parties with the task of dealing with sensitive data of the client company to detect the origin of biased outcomes. Other possibilities involve linked databases of many organizations to benefit from the other's experiences and conducting an exploratory analysis to get ahead of features that cause a biased outcome. What has to be considered is that achieving perfect unbiased data in reality is almost impossible as there are many different definitions of how to define fairness and usually not all required data is present. [26]

4 Explanation user interfaces

Users often do not understand why and how complex AI algorithms work. Therefore, they do not trust the system. This is why the next section should serve the goal to provide some implementation examples for explanation user interfaces that include some of the previously mentioned opportunities and considerations.

4.1 Static explanations

A rather static example for offering the user an explanation for datasets is "Datasheets for Datasets" by Gebru et al. which has already been used by e.g. Microsoft, Google, IBM. Just as electronic appliances are received with additional descriptions, datasets could as well be accompanied by documentation to support transparency especially because of limited accessibility. [10]

To go one step further, "Model Cards" are introduced which are frameworks for describing the performance of machine learning models. Model Cards are intended to be only short documents that should make the ML models more intelligible, help developers to standardize explanation processes and clarify potential errors. The information included in Model cards are details about the model, the intended use, intended users and relevant factors such as groups of people and environments. Finally, the metrics, evaluation data, and training data, quantitative analyses and ethical considerations should be covered. [21]

However, a drawback that can not be neglected is the user's limited motivation for reading the documentation and the time required to set them up [10].

4.2 Interactive and dynamic explanation UIs

A more visual and interactive approach is offered by the Scheduling Assistant by Kocielnik et al. which automatically identifies meeting enquiries in text-based emails. After one of the highlighted text parts are clicked, the user can request that this appointment is added to the calendar via a pop-up window. For this purpose, an accuracy indicator, an example-based explanation, and a control slider were introduced to lead the user's expectation into a realistic direction. [16]

The goal of the accuracy indicator is that the user gets a realistic impression of the performance of the system. It is visualized as a gauge chart expressing 50 percent accuracy. The goal of the example-based explanation is to convey how the AI system actually works and to prepare the user for the potential inaccurate outcome or performance. Hence, it is clarified that each sentence is analyzed by presenting exemplary sentences. Subsequently, it is estimated how likely it is that a meeting request was mentioned. All these features can be observed in Figure 1. Eventually, the goal of the control slider is to increase the trust of the user by allowing him or her to determine the decision threshold of the system for whether a meeting request is included. With the help of the labels and the small visualizations, it is clarified that if the user moves the interactive slider to the left, less but more precise results will be presented while moving the slider to the right will lead to more but less accurate results. Kocielnik et al. could already proof in a user study that all three design ideas were able to influence the expectations of the users and thus prepare for the performance of the system as well as raise trust and satisfaction. [16]

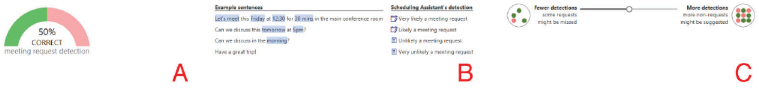


Figure 1 Accuracy indicator (A), example-based explanation (B) and control slider (C) designed by Kocielnik et al. [16].

Another interactive and dynamic explanation approach is Explanatory Debugging by Kulesza et al. which visually presents the reasons for the system's predictions. It is implemented in a prototype called EluciDebug which represents an email program that assigns emails to the respective folder. The user is able to correct the behavior and thus the predictions of the model. This can be done by adding or removing words via the explanation UI which is transferred to the respective model. Additionally, the importance of each word can be adapted by dragging the bars higher or lower presented in Figure 2. The objective of Explanatory Debugging is that the user establishes an accurate mental model of how the system works. Especially when only little training data exists in the beginning, Explanatory Debugging can be useful to push the system in the right direction. In a user study, Kulesza et al. were able to show that participants were able to manipulate the system's model in a ten percent more accurate way. Eventually, by providing interactive and dynamic explanation approaches that immediately includes the feedback of the user, the cost of actually dealing with these explanations seems to be lower. [17]

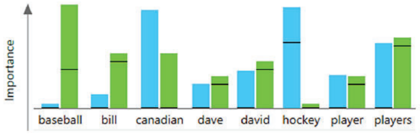


Figure 2 Explanatory Debugging as an interactive explanation UI which provides the opportunity for the user to add and remove words from the classifier and shift the importance by Kulesza et al. [17].

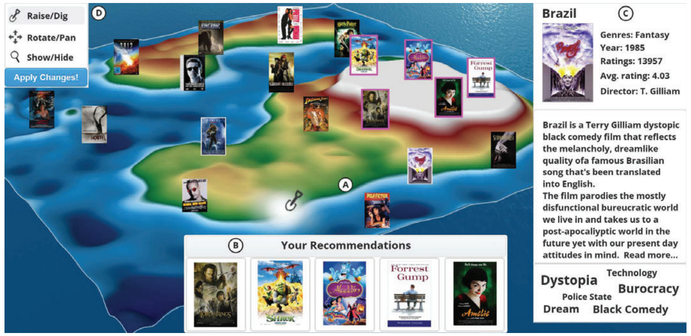
4.3 Recommender systems

There are several recommender systems that are already known to the public such as Amazon, Spotify or Netflix. The algorithms implemented in those systems are mostly based on machine learning and statistical techniques because of the huge data collected from the user's behavior. The services are able to offer personalized recommendations to their users which often results in a positive user experience. The three main algorithms are content-based, collaborative and context-based filtering. [11, 25]

Next, some explanation UI implementation examples associated with recommender systems are presented.

To help customers with their decision of which product to buy, the product characteristics and other customer's reviews should be made available to the user. The Preference-based Organization (Pref-ORG) views product proposals with a categorical structure and was extended by the explanation interface Senti-ORG which was introduced by Chen and Wang. It provides the product's feature description and extracts other customer's opinions about the product. A user experiment could, in fact, confirm that the customers gain better insights into the products and are supported in their decision making process. [4]

An explanation user interface for recommender systems that is mainly based on visual feedback is the 3D item space visualization by Kunkel et al. [18]. The goal of the explanation UI is to clarify why specific recommendations were made and to allow the user to manipulate how the system operates. The map is built up by using collaborative filtering which includes an item representation for the region of the map with the help of explicit or implicit (by tracking the user's behavior) feedback of general users of the system. High (high rates) or low (low rates) areas are created by the preferences of the current user. Finally, it is possible to individually customize the height of the areas based on the user's preferences. Such an approach showed that a usually tedious task e.g. rating movies can be overcome by interaction. The demonstrator is a web-based application that implements the 3D item space visualization shown in Figure 3. In an empirical user study, it was proved that the recommender system was perceived as very useful in situations where participants were uncertain about their movie choice. Additionally, the transparency was improved and the reasons why the system made those recommendations were comprehended. [18]



Brazil

Genres: Fantasy
Year: 1985
Ratings: 13957
Avg. rating: 4.03
Director: T. Gilliam

Brazil is a Terry Gilliam dystopic black comedy film that reflects the melancholy, dreamlike quality ofa famous Brasilian song that's been translated into English.
The film parodies the mostly disfunctional bureucratic world we live in and takes us to a post-apocaliyptic world in the future yet with our present day attitudes in mind. Read more...

Figure 3 Explanation UI of a 3D item space visualization including the working area (A), the recommended items (B), detail information (C) and interaction tools (D) by Kunkel et al. [18].

5 Conclusion and future work

For this paper, an extensive literature review of human-centered AI challenges and possibilities was conducted accompanied by implementation examples of explanation user interfaces to clarify the discussed theoretical proposals. The provided solutions support human-centered AI research and therefore serve the goal of offering explainable and usable intelligent systems. To enable this review, the most common challenges and their resolutions were filtered out and various details from different perspectives were added.

It was identified that the main challenge regarding AI/ML systems is the "black box" problem. It is not comprehensive for the user how and why the system operates the way it does. Consequences could be trust issues and eventually abandonment of the system. To avoid this potential outcome, collaborative decision making can be used to make the user better understand the way it works by using interaction and allowing intervention to adapt its behavior. Moreover, explanations should be provided to the user to ensure that transparency is present. Unfortunately, users are often not interested in dealing with explanations. However, motivation can be increased by collaborative decision making in order to tell the system what information people are interested in and to avoid errors by including the user's input. Hence, collaborative decision making seems to be successfully integrated into a lot of intelligent systems as the performance as well as user satisfaction benefit from it. Nevertheless, misaligned expectations can be caused which makes the user expect a better performance of the system than it can in fact offer. The goal is to align expectations by communicating the system's performance and potential errors beforehand. As these intelligent systems act independently, many users fear to lose control over AI systems. In that case, collaborative decision making is also an approach as they receive a better overview of what the system actually does and can therefore serve as reassurance.

Considering the perspective of the designers and developers, concrete design guidelines can approximate the "black box" problem which UX designers have to deal with too. This is because the proposals would include them earlier in the development process and therefore simplify the complexity of the algorithms. Other challenges where these concrete guidelines serve the same purpose is the lack of prototyping techniques as well as inter- and multidisciplinary research. The Wizard of Oz technique was mentioned as a practicable way for prototyping for dynamic systems as well as interdisciplinary workshops for collaborations with the goal of better understanding. This approach proved to be very successful and therefore should be followed up. However, prototyping for AI systems remains challenging because of the great data amount that is used in actual applications. This is why the WOz technique is a step in the right direction but not yet adequate. Suggestions for the insufficient data that can result in biased outcome includes third party companies for the analysis of the data or linked databases to benefit from other company's knowledge. These ideas are rather based on theoretical research and are probably difficult to realize because of competitive thinking and the complexity of analyzing huge datasets.

Eventually, every developer and UX designer has to assess which guidelines serve which purpose for which kind of AI system. Do they fit for what should be achieved, are other approaches more useful or do some have to be combined? The opportunities obviously have to be refined and adapted to their individual purpose but concrete guidelines can definitely promote and simplify the development of AI systems.

In the future, an attempt could be launched to design applicable intelligent ML systems that implement as much of the examined guidelines and ideas instead of only some for each device to achieve a proof of concept for instance. Furthermore, research on what aspects to focus on for specific types of intelligent system e.g. VUIs could be carried out.

References

1 Saleema Amershi, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, Eric Horvitz, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, and Paul N. Bennett. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, pages 1–13, Glasgow, Scotland Uk, 2019. ACM Press. URL: `http://dl.acm.org/citation.cfm?doid=3290605.3300233`, `doi:10.1145/3290605.3300233`.

2 Jacob T. Browne. Wizard of Oz Prototyping for Machine Learning Experiences. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems - CHI EA '19*, pages 1–6, Glasgow, Scotland Uk, 2019. ACM Press. URL: `http://dl.acm.org/citation.cfm?doid=3290607.3312877`, `doi:10.1145/3290607.3312877`.

3 Andrea Bunt, Matthew Lount, and Catherine Lauzon. Are Explanations Always Important? A Study of Deployed, Low-Cost Intelligent Interactive Systems. page 10, 2012.

4 Li Chen and Feng Wang. Explaining Recommendations Based on Feature Sentiments in Product Reviews. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces - IUI '17*, pages 17–28, Limassol, Cyprus, 2017. ACM Press. URL: `http://dl.acm.org/citation.cfm?doid=3025171.3025173`, `doi:10.1145/3025171.3025173`.

5 Maartje de Graaf, Somaya Ben Allouch, and Jan van Dijk. Why Do They Refuse to Use My Robot?: Reasons for Non-Use Derived from a Long-Term Home Study. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction - HRI '17*, pages 224–233, Vienna, Austria, 2017. ACM Press. URL: `http://dl.acm.org/citation.cfm?doid=2909824.3020236`, `doi:10.1145/2909824.3020236`.

6 Tilman Deuschel and Ted Scully. On the Importance of Spatial Perception for the Design of Adaptive User Interfaces. In *2016 IEEE 10th International Conference on Self-Adaptive and Self-Organizing Systems (SASO)*, pages 70–79, Augsburg, Germany, September 2016. IEEE. URL: `http://ieeexplore.ieee.org/document/7774388/`, `doi:10.1109/SASO.2016.13`.

7 Graham Dove, Kim Halskov, Jodi Forlizzi, and John Zimmerman. UX Design Innovation: Challenges for Working with Machine Learning as a Design Material. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, pages 278–288, Denver, Colorado, USA, 2017. ACM Press. URL: `http://dl.acm.org/citation.cfm?doid=3025453.3025739`, `doi:10.1145/3025453.3025739`.

8 Xiangmin Fan, Wencan Luo, Muhsin Menekse, Diane Litman, and Jingtao Wang. Scaling Reflection Prompts in Large Classrooms via Mobile Interfaces and Natural Language Processing. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces - IUI '17*, pages 363–374, Limassol, Cyprus, 2017. ACM Press. URL: `http://dl.acm.org/citation.cfm?doid=3025171.3025204`, `doi:10.1145/3025171.3025204`.

9 Umer Farooq and Jonathan Grudin. Human-computer integration. *interactions*, 23(6):26–32, October 2016. URL: `http://dl.acm.org/citation.cfm?doid=3012754.3001896`, `doi:10.1145/3001896`.

10 Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumeé III, and Kate Crawford. Datasheets for Datasets. *arXiv:1803.09010 [cs]*, April 2019. arXiv: 1803.09010. URL: `http://arxiv.org/abs/1803.09010`.

11 Carlos A. Gomez-Uribe and Neil Hunt. The Netflix Recommender System: Algorithms, Business Value, and Innovation. *ACM Transactions on Management Information Systems*, 6(4):1–19, December 2015. URL: `http://dl.acm.org/citation.cfm?doid=2869770.2843948`, `doi:10.1145/2843948`.

12 Jiazhen He, James Bailey, Benjamin I P Rubinstein, and Rui Zhang. Identifying At-Risk Students in Massive Open Online Courses. page 7.

13 Eric Horvitz. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems the CHI is the limit - CHI '99*, pages 159–166, Pittsburgh, Pennsylvania, United States, 1999. ACM Press. URL: `http://portal.acm.org/citation.cfm?doid=302979.303030`, `doi:10.1145/302979.303030`.

14 Kori Inkpen, Stevie Chancellor, Munmun De Choudhury, Michael Veale, and Eric P. S. Baumer. Where is the Human?: Bridging the Gap Between AI and HCI. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems - CHI EA '19*, pages 1–9, Glasgow, Scotland Uk, 2019. ACM Press. URL: `http://dl.acm.org/citation.cfm?doid=3290607.3299002`, `doi:10.1145/3290607.3299002`.

15 Claire Kayacik, Sherol Chen, Signe Noerly, Jess Holbrook, Adam Roberts, and Douglas Eck. Identifying the Intersections: User Experience + Research Scientist Collaboration in a Generative Machine Learning Interface. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems - CHI EA '19*, pages 1–8, Glasgow, Scotland Uk, 2019. ACM Press. URL: `http://dl.acm.org/citation.cfm?doid=3290607.3299059`, `doi:10.1145/3290607.3299059`.

16 Rafal Kocielnik, Saleema Amershi, and Paul N. Bennett. Will You Accept an Imperfect AI?: Exploring Designs for Adjusting End-user Expectations of AI Systems. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, pages 1–14, Glasgow, Scotland Uk, 2019. ACM Press. URL: `http://dl.acm.org/citation.cfm?doid=3290605.3300641`, `doi:10.1145/3290605.3300641`.

17 Todd Kulesza, Margaret Burnett, Weng-Keen Wong, and Simone Stumpf. Principles of Explanatory Debugging to Personalize Interactive Machine Learning. In *Proceedings of the 20th International Conference on Intelligent*

User Interfaces - IUI '15, pages 126–137, Atlanta, Georgia, USA, 2015. ACM Press. URL: `http://dl.acm.org/citation.cfm?doid=2678025.2701399`, `doi:10.1145/2678025.2701399`.

18 Johannes Kunkel, Benedikt Loepp, and Jürgen Ziegler. A 3d Item Space Visualization for Presenting and Manipulating User Preferences in Collaborative Filtering. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces - IUI '17*, pages 3–15, Limassol, Cyprus, 2017. ACM Press. URL: `http://dl.acm.org/citation.cfm?doid=3025171.3025189`, `doi:10.1145/3025171.3025189`.

19 Brian Y. Lim and Anind K. Dey. Assessing demand for intelligibility in context-aware applications. In *Proceedings of the 11th international conference on Ubiquitous computing - Ubicomp '09*, page 195, Orlando, Florida, USA, 2009. ACM Press. URL: `http://portal.acm.org/citation.cfm?doid=1620545.1620576`, `doi:10.1145/1620545.1620576`.

20 Jiaying Liu, Xiangjie Kong, Feng Xia, Xiaomei Bai, Lei Wang, Qing Qing, and Ivan Lee. Artificial Intelligence in the 21st Century. *IEEE Access*, 6:34403–34421, 2018. URL: `https://ieeexplore.ieee.org/document/8325446/`, `doi:10.1109/ACCESS.2018.2819688`.

21 Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. Model Cards for Model Reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency - FAT* '19*, pages 220–229, 2019. arXiv: 1810.03993. URL: `http://arxiv.org/abs/1810.03993`, `doi:10.1145/3287560.3287596`.

22 Emilee Rader, Kelley Cotter, and Janghee Cho. Explanations as Mechanisms for Supporting Algorithmic Transparency. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, pages 1–13, Montreal QC, Canada, 2018. ACM Press. URL: `http://dl.acm.org/citation.cfm?doid=3173574.3173677`, `doi:10.1145/3173574.3173677`.

23 Gonzalo Ramos, Jina Suh, Soroush Ghorashi, Christopher Meek, Richard Banks, Saleema Amershi, Rebecca Fiebrink, Alison Smith-Renner, and Gagan Bansal. Emerging Perspectives in Human-Centered Machine Learning. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems - CHI EA '19*, pages 1–8, Glasgow, Scotland Uk, 2019. ACM Press. URL: `http://dl.acm.org/citation.cfm?doid=3290607.3299014`, `doi:10.1145/3290607.3299014`.

24 Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, pages 1135–1144, San Francisco, California, USA, 2016. ACM Press. URL: `http://dl.acm.org/citation.cfm?doid=2939672.2939778`, `doi:10.1145/2939672.2939778`.

25 Von Jan-Hinrik Schmidt, Jannick Sørensen, Stephan Dreyer, and Uwe Hasebrink. Wie können Empfehlungssysteme zur Vielfalt von Medieninhalten beitragen? page 10.

26 Michael Veale and Reuben Binns. Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. *Big Data & Society*, 4(2):205395171774353, December 2017. URL: `http://journals.sagepub.com/doi/10.1177/2053951717743530`, `doi:10.1177/2053951717743530`.

27 Danding Wang, Qian Yang, Ashraf Abdul, and Brian Y. Lim. Designing Theory-Driven User-Centric Explainable AI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, pages 1–15, Glasgow, Scotland Uk, 2019. ACM Press. URL: `http://dl.acm.org/citation.cfm?doid=3290605.3300831`, `doi:10.1145/3290605.3300831`.

28 Daniel S. Weld and Gagan Bansal. The Challenge of Crafting Intelligible Intelligence. *arXiv:1803.04263 [cs]*, October 2018. arXiv: 1803.04263. URL: `http://arxiv.org/abs/1803.04263`.

29 Wei Xu. Toward human-centered AI: a perspective from human-computer interaction. *Interactions*, 26(4):42–46, June 2019. URL: `http://dl.acm.org/citation.cfm?doid=3342541.3328485`, `doi:10.1145/3328485`.

30 Qian Yang, Alex Scuito, John Zimmerman, Jodi Forlizzi, and Aaron Steinfeld. Investigating How Experienced UX Designers Effectively Work with Machine Learning. In *Proceedings of the 2018 on Designing Interactive Systems Conference 2018 - DIS '18*, pages 585–596, Hong Kong, China, 2018. ACM Press. URL: `http://dl.acm.org/citation.cfm?doid=3196709.3196730`, `doi:10.1145/3196709.3196730`.

Design challenges and opportunities for notifications in AR/VR environments

David Klein

Ludwig-Maximilians-Universität München, München, Deutschland
d.klein@campus.lmu.de

Zusammenfassung

This paper analyzes, summarizes and compares different approaches to designing notifications. These are then applied to challenges for notifications in AR and VR by discussing each match and considering the implications towards task interruption, as well as the differences between notifications for AR and VR.
In doing so the concepts of “context-awareness” for every aspect of the design and the use of “real life metaphors” for reliably conveying information seemed promising directions, which could be applied in proper notification design guidelines for AR and VR technologies.

2012 ACM Computing Classification Human-centered computing → Mixed / augmented reality

Keywords and phrases Augmented Reality; Virtual Reality; Notification Design; Task Interruption.

1 Introduction

Notifications of all kinds are part of our daily lives. From basic analog street signs to digital notifications on computers and phones, there are many kinds of notifications. These are necessary to inform people (or the user) of infrastructure, events and basically everything that may be of importance. To be quickly understood and noticed these notifications usually follow common design guidelines.
With the increasing use and spread of AR (Augmented Reality) and VR (Virtual Reality) technology, the need for adapting design guidelines to fit XR (Extended Reality) is obvious. Especially when considering what a lasting impact smart phones had and have on society, it is not hard to imagine what level of changes can result from widespread use of XR technologies.
Use cases for notifications in AR and VR are highly dependent on the environment. An additional person entering the room in which VR is used in, for example, is reason to notify the immersed user about that event. But not every

© David Klein;
licensed under Creative Commons License CC-BY
Cite as: David Klein. Design challenges and opportunities for notifications in AR/VR environments. In *5th Seminar on Ubiquitous Interaction (UBIACTION 2020)*. Editors: Florian Lang, Pascal Knierim, Jakob Karolus, Fiona Draxler, Ville Mäkelä, Francesco Chiossi, Luke Haliburton, Matthias Hoppe, Albrecht Schmidt. January 27, 2020. Munich, Germany. pp. 10:1–10:22.

notification warrants breaking immersion, especially for Augmented Reality as it may be necessary to not distract the user.

Distracting or breaking the users' immersion is the same as in real life, when on-going work tasks are interrupted and thereby need more time to accomplish. Therefore the same caution should be applied before doing so.

In this paper existing literature concerning notification design (with focus on AR and VR) shall be analyzed and compared to thereby identify common aspects, challenges and promising directions for further research.

2 Related Work

2.1 AR & VR

2.1.1 Definition

Augmented and Virtual Reality are generally viewed as being closely related. Yet they already differ greatly when comparing their individual definitions.

Mekni and Lemieux [6] stated that Augmented Reality was first discussed at Boing in 1990 (in comparison to 1965 for Virtual Reality) to improve factory worker guidance by showing them individual instructions for each aircraft using Head-Mounted Displays (HMD).

As of today most authors define AR in a way that requires a HMD, which would limit AR to only those technologies. Therefore Mekni and Lemieux [6] defined AR to not require any specific technology (disregarding the obvious, i.e. computing) by three characteristics:

- Combination of real and virtual.
- Interaction are done in real time.
- Uses the 3D environment.

In comparison to AR Virtual Reality was, as stated by Mandal [5], first proposed in 1965 by Ivan Sutherland as "make that (virtual) world in the window look real, sound real, feel real, and respond realistically to the viewer's actions" [6, p. 1].

After presenting Sutherland's initial proposal she continues to give an overlook the process VR underwent in the past decades and further summarizes the terminology surrounding VR. In this she characterizes immersive VR as follows:

- Viewing in three dimension virtual space uses the natural head movement.
- Relation and scaling in the virtual space are done in reference to human size.
- Increasing immersion can be achieved by additionally serving non-visual stimulus (haptic, auditory, etc.).

- The environment can be shared between users.

2.1.2 Challenges

Vi, Silva da Silva and Maurer [10] discussed several (usually not mentioned) challenges, as they are no challenge in designing notifications explicitly, but are challenges for designing for XR in general, they do still apply.

- XR ressources are mainly on VR due to the higher technological availability.
- User comfort is often emphasized due to the ergonomics of HMD's
- Most authors user metaphors, leaving few working examples
- Few empirical studies have been done, leaving many unvalidated concepts

Ghosh et. al. [3] examined different designs of notifications and further developed a prototype VR environment (NotifiVR) testing the previously gained insights. In their results they compared reaction time, task time and perception of each notification type and modality. Comparing the perception results, they identified four main challenges:

- Noticeability.
- Understandability.
- Urgency.
- Intrusiveness.

As ambient notifications are embedded in the environment and have to be aware of their surrounding, they share some design challenges with notifications in AR/VR.

In their paper about challenges for ambient notifications in smart homes Wier et. al. [11] discussed and identified three major challenges. They concluded that (1) the design of ambient notifications should be chosen at runtime in regards to the situational context. (2) Ambient notifications should convey their information calmly over a longer period of time. (3) Psychological constraints (attention-span, focus, mental load) have to be considered concerning information presentation and task interruption.

2.1.3 Opportunities

AR and VR provide different notification design opportunities due to their different immersion and information load characteristics. Ghosh et. al. [3] conducted a design exercise with several experienced VR designers, during which it was often mentioned that using known metaphors (phone notifications etc) is well received and offers a multitude of opportunities of varying quality. They also noted that urgency and intrusiveness show a positive correlation

(r=0.51), which needs to be taken in account as to not irritate the user (and decrease the notifications' effectiveness).
Zenner et. al. [13] proposed a framework for VR notifications in which notifications adapt to the setting and context of the virtual environment instead of breaking immersion by using generic pop-up messages.
Lucero and Vetek [4] did a study using interactive glasses to evaluate acceptability and usefulness of minimalistic mobile notifications. The glasses used were a simple prototype and therefore just represent a fraction of the capabilities of true AR devices. Participants in their study especially appreciated the minimalistic notification design, which did not distract them immediately from their task, and that their hands remained free for other things.

2.2 Interruptions

Czerwinksi, Cutrell and Horvitz [1] showed that "notifications reliably harm faster, stimulus-driven search tasks more than effortful, cognitively taxing search tasks" and noted that markers could help to return to a interrupted task.
Czerwinksi, Horvitz and Wilhite [2] in another study examined knowledge workers' daily routines, especially task interruptions and the time needed to return to a given task. They reported that the experienced difficulty for switching or returning to a previously interrupted task increased with task complexity, duration and type, as well as number of interruptions and their respective duration. Additionally they further explored tools for memory guidance and concluding that return time can be diminished by returning the system to the state before it was interrupted (positioning windows as before, showing the last performed steps, etc).
Ghosh et. al. [3] did a survey on typical types of interruptions for VR users and implemented them in NotifiVR. As there are obviously things that justify interrupting a given task, they noted that notifications should reflect the urgency of the interruption and therefore should be timed to not interrupt critical tasks or draw attention to another critical situation.

2.3 Notifications

2.3.1 In the real World

There are many known types of notifications in the real world. Car traffic alone offers a multitude of auditory (horn), haptic (paving) and visual (traffic lights) notifications. As technology has evolved, so have the notifications that are produced and needed to convey the right information at the right time.
Shirazi et. al. [7] found that users receive large numbers of notifications on their mobile phones every day and are accordingly assessing them by the perceived importance. Yet they found only a weak correlation between important notifications and user interaction, indicating that not every important notification requires immediate attention. As not all notifications can be important, there should be a clear distinction indicating importance. Further did they note, that users interact with notifications regarding communication and event apps (messenger, calendar) the most.
Schaub et. al. [9] assessed the state of privacy (and legal) software notifications and concluded that improvement is needed. Especially the well-known Terms-Of-Use before installing any software is a showcase for ineffective notification design. Mobile Notifications are in comparison better designed, yet often fail to convey all necessary information. Schaub et. al. [9] concluded that notifications have to be designed with the target audience in mind, have to provide some form of interaction and choice and have to be suited to the information they are intended to convey. In their design space they therefore proposed four characteristics for (privacy) notifications: Timing, Channel, Modality and Control.
Schaub, Balebako and Cranor [8] later further remarked that if complex information has to be conveyed, it should be broken down and delivered context-dependent to keep notifications as short as possible.

2.3.2 In AR & VR

As noted before (see 2.1.2) Vi, da Silva and Maurer [10] created guidelines for generally designing applications in XR. These share several characteristics with the previously mentioned design space by Schaub et. al. [9] (see 2.3.1) and the main challenges by Ghosh et. al. [3] (see 2.1.2), indicating a logical generalization hierarchy in which Vi's [10] guidelines are the most general, Schaub's [9] design space the most specialized and Ghosh's [3] characteristics lay in between.
Ghosh et. al [3] explored several types and modalities of notifications and noted that combined modalities (audio-visual, etc.) are generally more effective, as each (auditory, visual, haptic) have individual strengths. Especially the use of pure haptic notifications performs poorly if it is not paired with another modality. They further noted that audio performs best in drawing attention and visual best in conveying information, making audio-visual the overall preferred method. But they also added, that modalities should be added with increasing urgency. Ghosh et. al. [3] also referred to ambient displays (see 2.1.2) as visuals have to be distinct from the environment, proposing the use of fixed interaction places between virtual and real world (for example a virtual phone, which presents notifications like a real phone). This approach still leaves possibilities for harsher notification types if needed, as they noted that

different levels of urgency require different, basically context-aware, approaches in notifying the user.

Wiehr et. al. [11] noted for their challenges for designing ambient notifications that the notification itself should be context-aware (audio cues mixing in with background music) to minimize the mental disruption by drawn away attention. They followed that this approach needs to choose its notification design at run-time to be able to choose the proper design for each scenario. Additionally this does fit Ghosh et. al.'s instructions to design notifications in relation to the urgency of their respective information.

Lucero and Vetek [4] concluded from their study with their interactive glasses prototype, that users are able to handle more complexity than provided by simple animations. Even though distraction wasn't an major issue, it is mentioned as potential risk factor, leading them to propose more subtle motions to draw attention (a slightly shaking icon is better detected than a color or shaped coded one).

2.4 Notification design

As mentioned in the last section some of the proposed designs are either intended for a more general level (XR applications [10]), the same level (NotifiVR [3]) or a more specialized level (Privacy Notifications [9]). Additionally several authors examined different aspects relating to notification designs in AR and VR (Ambient notifications [11], Interruptions [3, 2], Immersion [13]). As the design space provided by Schaub et. al. [9] is the most detailed in breaking down notification design, the following subsections mirror their structure to then group the remaining guidelines in the same fashion.

2.4.1 Channel

Schaub et. al. [9] distinguish between primary (on-screen notification), secondary (flashing LEDs) and public (warning sign) channels, as not all devices have the technological capabilities to serve more than a single channel. As this design space is for privacy notification, we can at least compare the discussed channels to task interruption in VR, as each channel's intrusiveness should be quite distinct and therefore offer a valid design option for AR and VR environments as well.

Zenner et. al. [13] proposal includes adding a context function to the notification, so it can be embedded ambiently in each virtual environment. This context function could be used to dictate the intended channel to convey notifications as needed.

For ambient notifications Wiehr et. al. [11] advised to to embed notifications into the environment (sound cues into music, change street signs, etc) which would translate to using the public channel within the virtual world.

2.4.2 Modality

Schaub et. al. [9] chose to include four modalities (visual, auditory, haptic, machine-readable) in their approach comparing to the three (visual, auditory, haptic) mentioned by Ghosh et. al. [3]. Yet as the work of Ghosh is intended for VR, it may be argued that machine-readable notfications can be neglected (for now), as today's common vision of Virtual Reality is confined within a room. In contrast this argument cannot be made for AR, as its common vision is to be used mobile. The argument made by Schaub [9] for machine-readable notifications is the effortless transmission of privacy notices between systems (or as privacy beacon). As privacy concerns are a relevant topic in AR development, this should be reflected in design guidelines for Augmented Reality applications and notifications.

It has to be noted that the design space of Schaub et. al. [9] is focusing more heavily on conveying relevant information in comparison to Ghosh et. al. [3] whose focus is more towards drawing user attention. Both do not disregard the other aspect though.

Regarding the identical modalities of both proposals (auditory, visual, haptic) both authors [9, 3] agree on using visuals for conveying information and audio cues for drawing attention, while haptic remains at a supporting role (for now).

2.4.3 Timing

Schaub et. al. [9] did differentiate between 6 options (At setup, Just in time, Context-dependent, Periodic, Persistent, On-demand [access to the notification]) for timing notifications and indicated different uses for each. Ghosh et. al. [3] did use different timing for the different interruption scenarios, but did not consider it when designing their prototype.

As previously acknowledged there is a need for timing notifications to not interrupt (more important) tasks, which makes each notification in a sense context-dependent. Notifications in VR therefore seemingly have to be context-dependent and combined with another of the remaining options, especially again considering the usual space confinement of VR environments. For AR this basic need for context-dependent notifications is given as well, even with there being other use cases. But the multitude of use cases, as well as AR being generally context sensitive, could suggest using context-dependent approaches for more than just the timing aspect of notification design.

2.4.4 Control

Schaub et. al [9] discussed three forms of control mechanisms (blocking, non-blocking, decoupled) for notifications so the user always has a choice. These are characterized by their “level of interruption”, meaning whether the user is interrupted or distracted from the current task. In the case of Schaub et. al [9] it is because of privacy reasons empowering users by providing the option to decline (or circumvent) the collection of private data. Luckily this fits well to user wishes to be able to interact with notifications in both AR [4] and VR [3]. Especially in the case of (mobile) AR, privacy notifications seem a relevant topic in the future, therefore further justifying the inclusion of proper control mechanisms.

3 Existing notification design guidelines

3.1 In general

In their guidelines for warning designs Wogalter, Conzola and Smith-Jackson [12] defined several characteristics of which most can be found in the design guidelines for notifications as well.

- Salience
- Wording
- Layout/Placement
- Symbols
- Auditory Warnings
- Personal Factors (demographic variables)
- Familiarity

Still there are factors listed, which were not yet mentioned (at least in detail) by the previously presented guidelines of Ghosh and Schaub [3, 9]. First are demographic variables, meaning to consider needs of different age groups (especially older individuals) and to consider the changing meaning of notification elements (symbols, etc.) in different cultures. Second is Familiarity of which Wogalter et. al. [12] noted that familiarity in a domain (repeating warning signs) diminishes the attention towards them over time. Using changing designs (LED signs, HD displays, etc) may prevent this from occuring.
Overall those characteristics can be broken down to match the challenges for AR and VR (see 2.1.2). Layout, Symbols and Wording are each concerning understandability, while Salience and Auditory Warnings are each concerning noticeability (attention) and urgency.

3.2 For AR & VR

Vi, da Silva and Maurer [10] noted that there exists a wealth of knowledge about designing for Extended Reality in the community surrounding AR and VR, but this knowledge is not well represented within academic research and the scientific world as of now. Additionally they reported on the disproportionate amount of research done in VR compared to AR. Searching for design spaces and guidelines confirmed this notions, as there is no complete set of guidelines to be found generally and most design recommendations are intended for VR. Most papers concerning notifications in Augmented and Virtual Reality do not provide design spaces, comparable to the one provided by Schaub et. al. [9], but rather prefer to provide a set of guidelines, which therefore are formulated more loosely and do not cover every aspects of notification design.
As previously discussed (see 3.1) the challenges formulated by Ghosh et. al. [3] seem accurate, as other suggestions can be broken down into these four components (Noticeability, Understandability, Urgency, Intrusiveness).
Overall the work of Ghosh [3] seems to be the most well rounded approach, as it recognizes the need to consider interruptions in designing notifications for AR and VR environments. Further it provides design recommendations concluding from the initially explored challenges and the examination of different modalities and notification designs.

- Creating Distinction.
- Using Controllers efficiently.
- Reducing Visual Search.
- Avoid jump scares.
- Using Familiar metaphors.
- Switching context & details-on-demand.

Vi, Silva da Silva and Maurer [10] collected a set of 11 guidelines for designing applications in Mixed Reality (MR), which includes AR and VR. Additionally to the guidelines the authors gave explanation and reason for introducing each guideline:

- Efficent use of virtual and physical environment.
- Customization for individual needs.
- Prioritize virtual and physical safety and comfort.

- Prevent overwhelming or (unintentionally) distracting the user.
- Consider hardware limitation to prevent breaking immersion.
- Guide user attention, choice and information with cues (notifications).
- Increase immersion using additional non-visual channels.
- Utilize knowledge of the real world.
- Improve user understanding through consistent feedback.
- Empower the user by providing effective control mechanisms.
- Actions need to be reversible, as errors are expected.

Zenner et. al. [13] proposed a framework for context-aware notifications, but expressed its intention as a tool for the future development and derivation of principle design guidelines:
"[...] Immersive applications should implement notification events that fit the virtual world experienced by the user."

Shirazi et. al. [7] assessed mobile notifications and reported several insights on notifications:

- Nature of notifications is disruptive.
- Important notifications do not necessarily cause immediate attention.
- Notifications are for messaging.
- Important notifications are about people and events.
- Not all notifications are important.

These insights seem to correlate well with the previously proposed ideas of context-aware notifications, timed interruptions and the general call for individual customization.

Lucero and Vetek [4] concluded from their results that for very simple notifications user seem to be able to generally handle distraction well. Yet for moving elements this was not always the case, as movement has been shown to be one of the most effective visual ways of drawing attention. Further they reported on users expressing the wish to be able to interact with a given notification and on the display visibility varying due to changing light circumstances.

3.2.1 Review

In this section the previously presented guidelines shall be reviewed, summarized and categorized by the challenges for VR notifications (see 2.1.2) defined by Ghosh et. al. [3], as well as Schaub's [9] design space for privacy notifications.

Guideline	*Challenge*	*Design Space*
Ghosh: NotifyVR[3]		
1. Creating Distinction	NO, UN	CHA, MOD
2. Using Controllers efficiently	NO, UN, UR	CHA, MOD
3. Reducing Visual Search	NO, IN	CHA
4. Avoid Jump Scares	NO, UR, IN	MOD, TIM
5.Using Familiar metaphors	NO, UN, UR, IN	CHA, MOD, CON, TIM
6. Switching context & details-on-demand	UN, UR	CHA, CON, TIM
7. Efficent use of virtual and physical environment	NO, UR	CHA
8. Customization for individual needs	NO, UN, UR, IN	CHA, MOD, TIM
9. Prioritize virtual and physical safety and comfort	IN	CHA, MOD
10. Prevent overwhelming or (unintentionally) distracting the user	NO, UN	CHA, MOD, CON, TIM
11. Consider hardware limitation to prevent breaking immersion	UN	MOD
12. Guide user attention, choice and information with cues (notifications)	NO, UN, UR, IN	CHA, MOD, CON, TIM
13. Increase immersion using additional non-visual channels	UR, IN	CHA, MOD
14. Utilize knowledge of the real world	NO, UN, UR, IN	CHA, MOD, CON, TIM
15. Improve user understanding through consistent feedback	UN	MOD
16. Empower the user by providing effective control mechanisms	UN, IN	CON
17. Actions need to be reversible, as errors are expected	UN	CON
Zenner: Adaptive VR notif.[13]		
18. Immersive applications should implement notification events that fit the virtual world experienced	UR, IN	CON, TIM
Shirazi: Assessing Mobile notif.[7]		
19. Nature of notifications is disruptive	NO, IN	CHA, TIM

NO = Noticeability, UN = Understandability, UR = Urgency, IN = Intrusiveness;
CHA = Channel, MOD = Modality, CON = Control, TIM = Timing;

Guideline	*Challenge*	*Design Space*
Vi: HMD XR applications[10]		
20. Important notifications do not necessarily cause immediate attention	UN, UR	CON, TIM
21. Notifications are for messaging	UN, UR	
22. Important notifications are about people and events	UN	TIM
23. Not all notifications are important	UR, IN	CON, TIM
Lucero: NotifEye[4]		
24. Simplistic design is less distracting	NO, IN	MOD
25. Movement draws attention	NO, UR, IN	CHA, MOD, TIM
26. Provide interaction	UN, UR	CON, TIM
27. Expect visibility issues from changing light	NO, UN	CHA, MOD

NO = Noticeability, UN = Understandability, UR = Urgency, IN = Intrusiveness;
CHA = Channel, MOD = Modality, CON = Control, TIM = Timing;

Explanations:
1. External notifications must be distinct from the virtual environment and its notifications.
2. Controllers are real objects, but also represented within the virtual world. As the relation to the real world already exists, it is plausible to use them as plattform for presenting notfications. Vibrations (haptic) may be helpful to draw attention.
3. As visuals are not always noticed (poorly positioned, out of view, etc.) it is suggested to use a dedicated place for displaying real world notifications.
4. Visual and auditory notfications may cause jump scares, if spawned too close to the camera (aka. the users' body).
5. Familiar metaphors are more easy understood and accepted and therefore increase immersion and easy-of-use.
6. Notification should be interactable (dismissal, reaction) and provide further details if needed. It should be possible at all times to switch context to handling notifications.
7. Create virtual environments carefully to no overwhelm users. Do this by considering visual (virtual) and physical (real) restrictions and grouping of similar objects with each other.
8. Enable users to form their own environment by letting them customize as much as possible of their workflow and comfort functions. Design interactions with every level of experience in mind.

9. Design for XR by prioritizing physical (physical strain), physiological (motion sickness) and environmental (claustrophobia, etc) user comfort. Also be aware of discomfort (intrusiveness) triggered by spawning objects too close to the body or head of the user (personal space).
10. The amount of information presented should be carefully selected to neither over- or underwhelm the user. Keep secondary objects (tools, additional information) ready without being distracting (minimalistic design, objects are move- and hideable).
11. Be aware of individually differing hardware specs and limitations. Design for the strengths of your system.
12. Guide users to prevent frustration and help utilize the system. Do this by directing their attention to where it needs to be and simplify choices as much as possible. Further to not overwhelm the user you should only include necessary notifications, provide helping cues to tutorial functions and always provide information when an action may result in an error.
13. Provide a complete and appealing experience by enhancing visuals with audio (and haptic) and making the environment as comprehensible as possible.
14. Using familiar metaphors can help users understand an object's functionalities, but also may instill expectations for it to have additional real-life like uses. Also consider that virtual objects may be explored in different (zoom, “stepping inside”, details-on-demand) ways than in reality.
15. All interactions should have distinct visual states (with distinct characteristics each), which are consistent for the whole virtual environment. Feedback should also help the user learn to interact the intended way by identifying and assisting in preventing erroneous states. This should also encourage the user to explore and trust the provided feedback.
16. The user should always feel in control, which is achieved by always leaving control of the camera (and actions in general) to the user. Also exits should be provided at every state so it can be left or be undone.
17. Every action should be easily reversible, so the user is never “trapped” in an unwanted state. Reversibility also improves trust and therefore independent exploration of the application.
18. Provide a set of notifications (different types and modalities) explicitly designed to fit the context of the application the user is experiencing. This could also be used to react accordingly to the urgency of the external notification.
19. All notifications are disruptive, as time between seeing a visual cue and clicking on it has been shown to be less than 30 seconds.
20. Important notification tend to be interacted with faster, yet as this correlation is only weak it suggests that only because a notification is deemed important, it does not always necessitates a immediate reaction and interaction.

21. Notifications for messaging services are rarely being blocked and always associated with a high importance. This is mainly true for social media messenger types, but not necessarily for other types (e-mail etc).
22. Messaging and calendar applications are rarely blacklisted, as they are providing details and context to the communication and context (events, appointments, etc) of the user.
23. Beside the obvious truth of this statement, notifications should be reviewed if they are actually needed, before implementing them. As large numbers of notifications are not appreciated by users, notifications should only be used if they are actually useful.
24. Notifications with limited complexity have been shown to be handled well in terms of distraction from important events. Even though it is suggested that users are able to handle more complex notifications, simplicity seems the plausible route.
25. It has been shown that moving (i.e. shaking) objects are noticed quickly and are more effective in drawing attention to a notification than other modalities (color, shape). This suggests that movement has to be used carefully as to not immediately draw attention to sudden movements.
26. In a user tests with simple notifications participants express the wish to be able to effectively interact with the notifications.
27. Especially for AR, but in some cases for VR as well, visibility can be an issue, as the user is changing context (outdoors, etc) the quality and intensity of the overall illumination is constantly changing.

3.2.2 Categorization

In this section the previously collected and explained statements are grouped by similarity and then summarized into a single statement.
This can easily be achieved by categorizing each guideline by the challenges [3] their are relating to (as done in 3.2.1). Additionally we can further categorize them by assigning them to the related elements of the design space for privacy notifications [9].
To save space only the correlating numbers to the list above will be noted.

- **Noticeability:**
- Associated with guideline numbers: *1, 2, 3, 4, 5, 7, 8, 10, 12, 14, 19, 24, 25, 27*, Total: 14
- Sorting them further:
- Guidelines tagged for Channel: *1, 2, 3, 5, 7, 8, 10, 12, 14, 19, 25, 27*, Total: 12

- Guidelines tagged for Modality: *1, 2, 3, 5, 8, 10, 12, 14, 24, 25, 27*, Total: 11
- Guidelines tagged for Control: *5, 10, 12, 14*, Total: 4
- Guidelines tagged for Timing: *3, 5, 8, 10, 12, 14, 19, 25*, Total: 8

- **Understandability:**
- Associated with guideline numbers: *1, 2, 5, 6, 8, 10, 11, 12, 14, 15, 16, 17, 19, 20, 21, 22, 26, 27*, Total: 18
- Sorting them further:
- Guidelines tagged for Channel: *1, 2, 5, 8, 10, 12, 14, 19, 27*, Total: 9
- Guidelines tagged for Modality: *1, 2, 5, 8, 10, 11, 12, 14, 15, 27*, Total: 10
- Guidelines tagged for Control: *5, 10, 12, 14, 16, 17, 20, 26*, Total: 8
- Guidelines tagged for Timing: *5, 8, 10, 12, 14, 19, 20, 22, 26*, Total: 9

- **Urgency:**
- Associated with guideline numbers: *2, 4, 5, 6, 7, 8, 12, 13, 14, 18, 20, 21, 23, 25, 26*, Total: 15
- Sorting them further:
- Guidelines tagged for Channel: *2, 5, 6, 7, 8, 12, 13, 14, 25*, Total: 9
- Guidelines tagged for Modality: *2, 4, 5, 8, 12, 13, 14, 25*, Total: 8
- Guidelines tagged for Control: *5, 6, 12, 14, 18, 20, 23, 26*, Total: 8
- Guidelines tagged for Timing: *4, 5, 6, 8, 12, 14, 18, 20, 23, 25, 26*, Total: 11

- **Intrusiveness:**
- Associated with guideline numbers: *3, 4, 5, 8, 9, 12, 13, 14, 16, 18, 19, 23, 24, 25*, Total: 14
- Sorting them further:
- Guidelines tagged for Channel: *3, 5, 8, 9, 12, 13, 14, 19, 25*, Total: 9
- Guidelines tagged for Modality: *3, 4, 5, 8, 9, 12, 13, 14, 24, 25*, Total: 10
- Guidelines tagged for Control: *5, 12, 14, 16, 18, 23*, Total: 6
- Guidelines tagged for Timing: *4, 5, 8, 12, 14, 18, 19, 23, 25*, Total: 9

3.2.3 Summary

As the scope of this paper is a literature research, this categorization is subjective and by thought only and therefore is not to be viewed as verified.
A first glance at the number of entries per (sub-) category suggest most guidelines were concerning understandability which seems intuitive as conveying information is the first priority of simple notifications. Further the numbers suggest that thoughts of control mechanisms barely influence noticeability which seem intuitive as well, as drawing attention preceeds interaction opportunities.
At a deeper look entries concerning known metaphors and context dependence are appearing so regular to make it sensible calling them a basic design principle of notifications.
To further stick to the categorization the summaries of the guidelines will be done similarly.

Noticeability:
The decision which **channel** should be used should regard *user expectations* (phone ringing etc.), but also be *distinctly visible* as the lighting situation is likely to change. Additionally *available space* and *usable objects/surfaces* have to be considered.
Choose *multiple channels* to increase chances of drawing attention, but increase levels of perceived urgency and intrusiveness as well.
Choose **modalities** to imitate *familiar concepts* (phone vibration patterns, knocking sounds, etc), thereby increasing ease-of-use and trust (immersion). Also choose the notifications *complexity* to not overwhelm the user by it, yet do not understate the importance of the notification.
Choose *multiple modalities* to increase chances of drawing attention, but increase levels of perceived urgency and intrusiveness as well.
Noticeability does not require **control** mechanisms.
Choose the notifications **timing** depending on the choices for channel and modality, as different *metaphors* use different timings (phone ringing using periodic timing, mail notifications being either persistent or just-in-time, etc). Allow *customization* especially for "planned"notifications (alarm, appointment, etc).
Choice of timing depends on the level of *urgency* as well, as repeated (persistent, periodic) cues are obviously more noticeable. Additional context dependent timing should be considered for situations when distraction is deemed unsafe or too disruptive.

Understandability:
As notifications always carry some information, the choice of **channel** should

allow for the *encoding of all needed information.*
familiarities to the real world should be taken in account to decide which channel would be *expected* by the user to be used.
Consider the needed attention, mental effort and caused interruption for the notification to be understood and choose the *least disruptive* (encoding a weather forecast into haptic morse code will not be understood).

The choice of **modality** in terms of being understood should refer to the *type* of notification and its *real world association.*
Initially convey *just as much* information as needed, provide *details on demand* (a messenger notification usually only shows name, time and a content preview).
To retain immersion as much as possible a *context-dependent* choice of modalities is advised. Yet *user safety* and notification *urgency* must always be the first priority.

Control mechanisms should be designed to resemble the notifications *real life association* to be quickly understood. A simplistic design offering *details on demand* should suffice most times.

Again **timing** should relate to the *real life equivalent*, but more importantly consider *context* of the presently performed task. Therefore notifications should be timed by *user safety, task progression* (identify good state(s) for interruptions?) and notification *urgency.*

Urgency:
Increase number of **channels** with increasing urgency. As channels are not hierarchically in terms of urgency or intrusiveness, choose channel(s) fitting to the chosen *modality, context, and available space.*
Interrupt an ongoing task only if the notifications content deemed more important.

Choose **modality** type fitting to *notification content.* Achieve increased urgency by using *multiple modalities.*
Using moving objects or placing objects *close* to the users *head and/or body* will draw in attention, but is also considered to be very intrusive and can even scare users.
As *sudden changes* can irritate users and quickly draw attention (high urgency, high intrusiveness) it could be a better (less urgent, less intrusive) way to let modalities *fade-in.*

Provide (better) **control** with *increasing urgency*, as reactions get more probable or urgent. Even for low urgency notifications there should always be a control mechanism to *discard or delay* the notification. *Details-on-demand* seems a tried and tested (mobile notifications) method for minimizing notifications.
To not endanger users or interrupt important tasks it may be necessary to guide users to a *safe state* before starting the notification.

Urgent notifications should be relayed as soon as possible. The **timing** should therefore be appropriate for the *assigned urgency*, but still adhere to *real life equivalents* of the used concept, as well as to the *context* of application running.
Customization options should be provided, as previously planned notifications (appointments, alarm, etc) are handled different by each individual.

Intrusiveness:
As intrusiveness is not desirable choose as *few* **channels** as possible, *without understating* the importance of the notification.
Choose channel by *context* of the application to keep familiarity up and intrusiveness down.

Intrusiveness increases with the number and intensity of the **modalities** used. Increasing *intensity and number* of modalities over time is perceived less intrusive than a (or multiple) sudden change in the environment. Therefore *refrain* from suddenly spawning objects, especially in close proximity to the users' body or head.

Increase *trust* by providing *effective and immediate* **control** mechanisms to *empower user decisions* (instead of forcing them a certain way).
To avoid breaking immersion or "harshly"changing context, provide controls fitting for the *context* of the application.

As before **timing** should be chosen according to *urgency*, but also to where and when it would be *expected* by users.
Collect or delay notifications and provide them at a *safe time and place.*

4 Notification Opportunities

4.1 In AR

As the term "context-aware"was used quite heavily in the last section, it seems necessary to examine it closer. In the design space of Schaub et. al. [9] the term "context-aware"can be found as one of the options for timing, but remember that this design space is for privacy notifications. Therefore its main concern is to inform correctly and timely about external data collection practices, whereas users of AR and VR need to get not only the known (mobile, etc.) notifications, but also notifications on events happening around them (person entering room, someone speaking, etc) but may be auditory or visually obscured. Comparing context-awareness for privacy to the same AR (and regarding the last sections) it seems necessary to assign context-awareness to more than just timing, but almost everything.
As previously stated it would be necessary to estimate a notifications urgency to accurately determine whether to interrupt a users workflow and immersion, especially if the user is performing critical tasks. So it follows that context for the notifications urgency has to be provided, maybe in form of an “urgency scale” and an “interruptibility scale”.
Furthermore context-awareness is also suggested for the choice of a notifications' channel, modality and their respective placements. For Augmented Reality this means that it is key to know the level and form of augmentation happening, to be able to determine fitting context. For a minimal augmentation (for example an HMD doing nothing but replacing italian words with their english translation) a great variety of notifications, or even notifications in general, could be irritating. Whereas a heavily augmented environment (for example transforming the environment (buildings, people, etc) to a game world) would probably generate a fair amount of own notifications and could already have a dedicated space for notifications.
Other, more concrete, opportunities for notifying in AR can be found in the environment. As example we assume an augmented reality navigation application, the location of a city street corner and the task to navigate the user to an address. At this point the system has to choose the design of the notification telling the user to turn right.
Here it uses at least one of the primary (the AR display itself), secondary (LED's or other output methods, usually associated with display-less devices; here: virtual representation of real object (i.e. controller)) or public (street signs, etc) channels. On the one hand are all notifications using the primary channel (as everything is "displayed") but on the other hand it intuitively seems right to think about the primary channel as concerning the momentary active application experienced by the user as the display to notify on (frontal

car screen). As for the secondary channel, notifications may be designed using the virtual representation of the controller or even using the application itself (controller vibrating on the right, blinking street sign). Regarding the public channel its mention may be questionable, as it would only concern shared augmented environments with shared notifications, which seems to be unreasonable. Yet on a general note it would be using a public space notifying anyone interacting with it (street signs).

Another decision to be made is which modality to use and when to notify. Both of these relate in AR heavily to the context of the user. To not interrupt or distract it is necessary to choose a non-distracting modality or preferably determine the user's context (maybe using gaze detection). The same can be said for the timing as it changes by context ("turn right"notification is not needed at the corner but earlier). For conveying the information correctly it is necessary to provide distinction, meaning that visual and audio cues have to be well discernible from the surrounding.

4.2 In VR

Notifications in VR could be done exactly as suggested for AR, but using the virtual world for it. But this would not only disregard opportunities but even endanger the user, as VR devices have to consider the physical limitations of the surrounding space. At least we can assume that context-awareness can be thought of in the same way for purely the virtual world. Additionally less user context determination has to be done via observation, but can be inferred from data generated by the (virtual) environment as every process takes place within it. Adding to the suggestion of Zenner et. al. [13] for adaptive (context-aware) notifications to preserve immersion, they should also regard the context of user interruptibility and notification urgency.

As previously mentioned is the real physical environment another context which must be considered to prevent the user from self-harm and to notify about events happening within the room. The work of Ghosh et. al. [3] presents a sound analysis on conveying those notifications effectively.

5 Conclusion

This paper summarized and compared different approaches to designing notifications in the context of Augmented and Virtual Reality. Both of these technologies have different needs and opportunities for notifications, but seem to share the same preferred design choices by calling for the use of real life metaphors and context-awareness. While metaphors prevail in direct interaction, the latter is considering the user's context most.

Especially for AR this is a important topic, as it probably will be used in all aspects of daily life and therefore needs to prioritize not endangering its user in any way.

VR on the other hand has to consider its users safety as well, but is commonly used in fixed location. Additionally it has to relay real life information to the user, while trying to not brake immersion.

For both technologies context-awareness will probably be the leading concern in designing notifications and interactions and therefore should always be considered. Timing, Channel, Modality and Control should all be chosen considering the information importance and cost of interruption, real life events and user safety, immersion by fitting the environmental context.

Further research and empiric evidence will be needed to create and verify an appropriate design space for notifications in each form of Extended Reality.

Literatur

1 Edward B. Cutrell, Mary Czerwinski, Mary Czerwinski, and Eric Horvitz. Effects of instant messaging interruptions on computing tasks. In *CHI '00 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '00, pages 99–100, New York, NY, USA, 2000. ACM. URL: `http://doi.acm.org/10.1145/633292.633351`, `doi:10.1145/633292.633351`.

2 Mary Czerwinski, Mary Czerwinski, Eric Horvitz, and Susan Wilhite. A diary study of task switching and interruptions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pages 175–182, New York, NY, USA, 2004. ACM. URL: `http://doi.acm.org/10.1145/985692.985715`, `doi:10.1145/985692.985715`.

3 S. Ghosh, L. Winston, N. Panchal, P. Kimura-Thollander, J. Hotnog, D. Cheong, G. Reyes, and G. D. Abowd. Notifivr: Exploring interruptions and notifications in virtual reality. *IEEE Transactions on Visualization & Computer Graphics*, 24(04):1447–1456, apr 2018. `doi:10.1109/TVCG.2018.2793698`.

4 Andrés Lucero and Akos Vetek. Notifeye: Using interactive glasses to deal with notifications while walking in public. In *Proceedings of the 11th Conference on Advances in Computer Entertainment Technology*, ACE '14, pages 17:1–17:10, New York, NY, USA, 2014. ACM. URL: `http://doi.acm.org/10.1145/2663806.2663824`, `doi:10.1145/2663806.2663824`.

5 Sharmistha Mandal. Brief introduction of virtual reality & its challenges. 2013.

6 Mehdi Mekni and André Lemieux. Augmented reality : Applications , challenges and future trends. 2014.

7 Alireza Sahami Shirazi, Niels Henze, Tilman Dingler, Martin Pielot, Dominik Weber, and Albrecht Schmidt. Large-scale assessment of mobile notifications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 3055–3064, New York, NY, USA, 2014. ACM. URL: `http://doi.acm.org/10.1145/2556288.2557189`, `doi:10.1145/2556288.2557189`.

8 Florian Schaub, Rebecca Balebako, and Lorrie Faith Cranor. Designing effective privacy notices and controls. *IEEE Internet Computing*, 21(3):70–77, May 2017. `doi:10.1109/MIC.2017.75`.

9 Florian Schaub, Rebecca Balebako, Adam L. Durity, and Lorrie Faith Cranor. A design space for effective privacy notices. In *Proceedings of the Eleventh USENIX Conference on Usable Privacy and Security*, SOUPS '15, pages 1–17, Berkeley, CA, USA, 2015. USENIX Association. URL: `http://dl.acm.org/citation.cfm?id=3235866.3235868`.

10 Steven Vi, Tiago Silva da Silva, and Frank Maurer. User experience guidelines for designing hmd extended reality applications. In *Human-Computer Interaction - INTERACT 2019 - 17th IFIP TC 13 International Conference, Paphos, Cyprus, September 2-6, 2019, Proceedings, Part IV*, volume 11749 of *Lecture Notes in Computer Science*, pages 319 – 337. Springer, 2019. `doi:10.1007/978-3-030-29390-1`.

11 Frederik Wiehr, Alexandra Voit, Dominik Weber, Sven Gehring, Christoph Witte, Daniel Kärcher, Niels Henze, and Antonio Krüger. Challenges in designing and implementing adaptive ambient notification environments. pages 1578–1583, 09 2016. `doi:10.1145/2968219.2968552`.

12 Michael S. Wogalter, V. C. Conzola, and Tonya L. Smith-Jackson. Research-based guidelines for warning design and evaluation. *Applied ergonomics*, 33 3:219–30, 2002.

13 André Zenner, Marco Speicher, Sören Klingner, Donald Degraen, Florian Daiber, and Antonio Krüger. Immersive notification framework: Adaptive & plausible notifications in virtual reality. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI EA '18, pages LBW609:1–LBW609:6, New York, NY, USA, 2018. ACM. URL: `http://doi.acm.org/10.1145/3170427.3188505`, `doi:10.1145/3170427.3188505`.

Die Reise ins (virtuelle) Ich - Einflussfaktoren auf das Gefühl von Embodiment in Virtual Reality

Anke Pellhammer

Ludwig-Maximilians-Universität München, München, Deutschland
`Anke.Pellhammer@campus.lmu.de`

Zusammenfassung

Wie fühlt es sich an in einen anderen Körper einzutauchen und diesen kontrollieren zu können? Mit dem Anstieg an Möglichkeiten der Erfahrbarkeit dieses vielfältigen Erlebnisses durch Mixed Reality stellt sich die Frage nach dem Gefühl des Embodiments (GvE). Dieses setzt sich zusammen aus dem Gefühl des Körperbesitzes, der Körperkontrolle und der Lokalisierung der eigenen Person in dem Körper bzw. dem Volumen, welches dieser einnimmt. Diese Ausarbeitung beschäftigt sich näher damit, inwieweit ein virtueller Körper sein reales Pendant abbilden kann, wie die Wahrnehmung des Nutzers beeinflusst werden kann und welche Bedeutung die Perspektive zur Darstellung des digitalen Ichs einnimmt. Grundlegend für die Erfahrung von Embodiment ist das Zusammenspiel aus Eindrücken über die Sinnesorgane und der Imaginationskraft, wobei auch das äußerliche Erscheinungsbild des virtuellen Egos ebenso wie die gewählte Perspektive eine zentrale Rolle spielt. Mithilfe der First Person Perspektive blickt der Nutzer aus dem Körper. Bei der Third Person Perspektive wird der Charakter visuell in die Umgebung eingebettet. In beiden Fällen kann eine immersive Wirkung unter bestimmten Umständen bis zu einem gewissen Grad ermöglicht werden. Im Vergleich zu anderen Medientypen wie Büchern, Bildern, Filmen oder Computerspielen bietet VR eine sehr hohe immersive Wirkung. Die realitätsnahe Illusion von VR wird durch die verwendete Dreidimensionalität, der Synchronität des tatsächlichen und virtuellen Körpers sowie interaktiver Steuerelemente erzeugt.

2012 ACM Computing Classification Human-centered Computing – Human Computer Interaction (HCI) – Interaction Paradigms – Virtual Reality

Keywords and phrases Embodiment; Virtual Body Illusion; Body Transfer; Body Ownership Illusion; Perspective.

© Anke Pellhammer;
licensed under Creative Commons License CC-BY

Cite as: Anke Pellhammer. Die Reise ins (virtuelle) Ich - Einflussfaktoren auf das Gefühl von Embodiment in Virtual Reality. In *5th Seminar on Ubiquitous Interaction (UBIACTION 2020)*. Editors: Florian Lang, Pascal Knierim, Jakob Karolus, Fiona Draxler, Ville Mäkelä, Francesco Chiossi, Luke Haliburton, Matthias Hoppe, Albrecht Schmidt. January 27, 2020. Munich, Germany. pp. 11:1–11:18.

1 Motivation: Der Begriff Embodiment, die Identifikation mit dem virtuellen Körper und das Präsenzgefühl in VR

Stellen Sie sich vor, Sie sollen in eine Rolle schlüpfen und sich so für eine bestimmte Zeit in eine andere Figur verwandeln. Für Schauspieler ist dieses sogenannte Method Acting ein Training, um eine andere Person zu verkörpern. Ein ähnliches Konzept kommt auch bei der Nutzung von Virtual Reality zum Einsatz. Einen fremden Körper zu steuern, sich mit diesem zu identifizieren und eine andere Rolle einnehmen sind essentielle Komponenten in virtuellen Umgebungen. Inwiefern kann jedoch das Setting einer virtuellen Technologie die Erfahrung des Einnehmens eines anderen Körpers erleichtern?

Die Psychologie widmet sich zunehmend mit der Fragestellung, welche Rolle der körperlichen Interaktion mit der Umwelt für unsere Gedanken, Gefühle und Verhaltensweisen zukommt und inwiefern sich dies auf die subjektiv wahrgenommene Realität auswirkt. Eine Übertragung von äußeren Einflüssen auf den eigenen Körper kann beispielsweise über Herstellung einer metaphorischen Verknüpfung von physischen mit moralischen Zuständen erfolgen [21]. So wird das Händewaschen einerseits mit Sauberkeit des Körpers andererseits auch als Reinwaschen von Schuld interpretiert [34].

Der biologische Körper dient auch in virtuellen Umgebungen als Ausgangspunkt für ein Erlebnis. Ob und bis zu welchem Grad ein Nutzer die gleichen Empfindungen mit einem virtuellen Körper (sog. Avatar) erlebt wie mit seinem echten Körper, wird in der Wissenschaft vielseitig behandelt [13]. Unter dem Begriff *Embodiment* (dt. Verkörperung, Personifikation) wird ein Empfindungszustand zusammengefasst. Nach Kilteni et al. [12] beschreibt er die subjektive Erfahrung einen Körper zu besitzen und zu steuern. Hierbei setzt sich der Begriff aus drei Teilaspekten zusammen. Das *Gefühl der Selbstlokalisierung* (engl. *Sence of Self-localization*) bezeichnet das Volumen im Raum, wo sich eine Person lokalisiert. Meist entspricht die Selbstlokalisierung dem physischen Körper, außer es kommt zu einer außerkörperlichen Erfahrung [16]. Die zweite Komponente des *Gefühls von Embodiment* (GvE) stellt das *Gefühl des Körperbesitzes* (engl. *Sence of Body Ownership*) dar, welche den Körper als Quelle der erlebten Empfindungen ansieht. Drittens spielt außerdem das *Gefühl der Körperkontrolle* (engl. *Sence of Agency*) eine Rolle, da es die Steuerung über aktive Bewegungen ermöglicht. Neben einem hohen GvE gilt auch das Gefühl von Präsenz als essentieller Bestandteil für die Nutzererfahrung in Virtual Reality (VR). Der Terminus beschreibt das Gefühl des „Dort Seins“ in Bezug auf eine räumliche Umgebung und bezieht sich somit auf die Illusion an einem bestimmten Ort (engl. *Place Illusion*) zu sein [28]. Auf eine ähnliche Wechselwirkung spielt das Gefühl der Selbstlokalisierung an, welches aber nicht auf die Beziehung zwischen Selbst und räumlichen Umfeld, sondern auf die Beziehung des Selbst und des eigenen Körpers anknüpft [13].

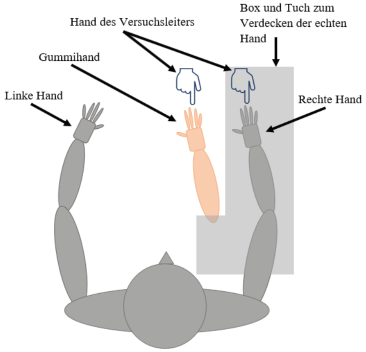
Sowohl in der Unterhaltungsindustrie, als auch in anderen Bereichen, wie dem medizinischem Sektor oder für militärische Zwecke, können Erkenntnisse über das Gefühl von Embodiment in VR den Fortschritt fördern und neue Möglichkeiten schaffen. Ziel dieser Arbeit ist es, einen Überblick zu vermitteln, welche Möglichkeiten existieren, das GvE im virtuellen Raum systematisch zu beeinflussen und wozu dies dienlich sein kann. Zu Beginn erfolgen dazu wissenschaftliche Beobachtungen zu allgemeinen und daraufhin zu spezifischeren Faktoren. Bei der Ausarbeitung liegt der Schwerpunkt auf der Bedeutung der Perspektive, da es sich hierbei um ein essentielles Mittel in VR zur Darstellung des Selbst handelt.

2 Einflussfaktoren und Limitationen der Identifikation mit dem virtuellen Ich

Dieses Kapitel erarbeitet schrittweise unterschiedliche Teilbereiche, welche bei dem Identifikationsprozess mit einem virtuellen Körper zum Tragen kommen. Dieses Wissen ist notwendig, um zu verstehen, wie sowohl die Wahrnehmung als auch das Verhalten eines VR-Nutzers beeinflusst werden kann. Dafür behandelt der erste Teil grundlegende Prozesse und Parameter, welche sich aus dem Zusammenspiel äußerer Stimuli und deren Interpretation auf den eigenen Körper ergeben. Des Weiteren wird im zweiten Part auf die Differenzen in der Identifikation mit dem virtuellen Charakter abhängig von der gewählten Perspektive eingegangen, da diese grundlegende Unterschiede aufweist, wie das Ich wahrgenommen wird.

2.1 Wahrnehmung externer Reize und Übertragung derer auf den eigenen Körper

Unter bestimmten Umständen können Stimuli aus der Umgebung missinterpretiert und auf den eigenen Körper bezogen werden. Dies verdeutlicht die Rubber Hand Illusion von Botvinick und Cohen [3] (siehe Abbildung 1), bei der Probanden anstelle ihrer echten Hand eine Gummihand sahen. Sowohl die echte Hand, als auch die Gummihand wurden zeitgleich sensorisch vom Versuchsleiter gereizt, wobei die Versuchspersonen dem Gefühl erlagen, die Gummihand sei ihre eigene. Eine zentrale Rolle für das Funktionieren der Illusion erhält die Synchronität der visuellen und taktilen Informationen, die an den Empfänger übermittelt wird. Petkova und Ehrsson [24] beschäftigten sich weiterführend mit der Frage, inwiefern diese Illusion auf den gesamten Körper übertragbar ist. Bei dem Body Swapping-Experiment (siehe Abbildung 2) standen die Testpersonen neben einer Schaufensterpuppe, welche mit Kameras am Kopf mit Blick hinab auf den Körper ausgestattet war. Während der Proband über ein Head-Mounted Display aus der Perpektive der Puppe an deren Körper hinuntersah, wurde die Versuchsperson sowie die Puppe am Bauch mit einer Bürste gestreichelt. Bei synchronen Berührungen übertrug sich der egozentrische Referenzrahmen der meisten Probanden auf den neuen Körper. Diese Beispiele zeigen, dass eine systematische perzeptuelle Manipulation zu einer immersiven Wirkung in andere Körper bzw. Körperteile führen kann.



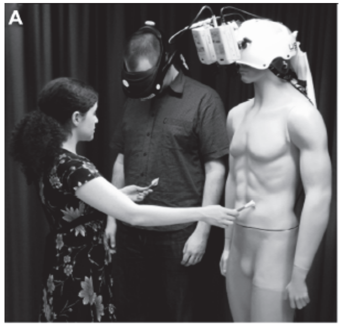


Abbildung 1 Der experimentelle Aufbau der Rubber Hand Illusion mit der Gummihand (mittig). (Eigene Darstellung in Anlehnung an Paton et al. [22])

Abbildung 2 Body-Swapping Experiment mit Berührung am Bauch als Stimuli. (Quelle: Petkova und Ehrsson [24])

Das erfolgreiche Hineinfühlen in einen Körper oder eine Situation erfordert laut Lombard und Jones [19] neben Sinneseindrücken zudem die geistige Vorstellungskraft der Person. Das Gefühl, sich in einem Leib wiederzufinden, basiert demnach auf dem Zusammenspiel zwischen sensorisch aufgenommenen Reizen, deren Ursprung aus externen Quellen der echten Welt stammen, und internen geistigen Prozessen, welche insbesondere im Gehirn verarbeitet werden. Mentale Vorgänge können dabei lebhafte Erinnerungen, Träume oder Empfindungen verursachen. Ein Beispiel dafür ist das Auftreten von Phantomschmerzen in amputierten Gliedmaßen. Obwohl kein sensorischer Input über die nicht mehr vorhandenen Nerven aufkommen kann, empfindet der Patient Schmerzen an diesen Stellen. Bei der therapeutischen Behandlung werden Imaginationsübungen appliziert, um das fehlinterpretierte Schmerzempfinden im Gehirn zu korrigieren [32]. Eine zentrale Rolle trägt die Vorstellungskraft auch bei dem gegenteiligen Effekt des GvE. Es handelt sich hierbei um sogenannte außerkörperliche Erfahrungen (engl. *Out-Of-Body Experience*), welche im medizinischen Bereich z.B. bei der Behandlung von Angststörungen, Anwendung findet. Durch die Illusion, den Körper gezielt zu verlassen und auf ihn zu blicken, erhält die Person eine andere Sichtweise auf ihren Körper und erlangt das Potential, die Situation sowie das Selbst zu abstrahieren [4].

Ein weiterer Schlüsselfaktor, der zu unterschiedlichen Identifikationsgraden führen kann, ist die Darstellung des virtuellen Selbst. Insbesondere die virtuelle Realität bietet zahlreiche Möglichkeiten eine andere Repräsentation des Selbst zu schaffen, denn morphologische Einschränkungen (dick, dünn, groß, blond etc.), pathologische Limitationen (Schmerz, Erschöpfung) und emotionale (Trauer, Freude) sowie soziale (Höflichkeit, Adäquates Verhalten) Grenzen können umgangen werden. Es spielt demnach eine grundlegende Rolle, welche Ausprägungen das Aussehen des alternativen Ichs aufweist. Aber verändert sich auch unsere Selbstwahrnehmung gegenüber dem virtuellen Körper, wenn wir diesen modifizieren? Laut Wauck et al. [31] ist die Übereinstimmung des Aussehens, Kleidungsstils, etc. von virtuellem und realem Ich hierbei nur bis zu einem gewissen Grad förderlich bzw. sinnvoll. Wie dies genau auf die Wahrnehmung wirkt, wird in Kapitel 2.2 in Abhängigkeit der Perspektive ausführlicher erläutert. Ein allgemeineres Merkmal hierbei spielt aber unter anderem das Geschlecht. Die alternative Realität bietet des Weiteren die Möglichkeit ein anderes Geschlecht anzunehmen. Die Übereinstimmung des virtuellen mit dem realen Geschlecht kann zu einer Verbesserung oder Verschlechterung bei der Identifikation mit dem gezeigten Avatar führen. Bei den Betrachtungen von Wauck et al. [31] war dies allerdings nicht der Fall, denn hier zeigten sich keine Abweichungen. Hingegen fiel in dem Experiment von Aymerich-Franch und Baileson [2], in dem Teilnehmer ihren virtuellen Doppelgänger dabei beobachteten, wie dieser vor einem Publikum eine Rede hielt, auf, dass Männer weniger Furcht vor der Situation hatten und sich eine bessere sprachliche Kompetenz zuschrieben als Frauen. Schwind et al. [27] bestätigten ein unterschiedlich wahrgenommenes Präsenzgefühl in Abhängigkeit der visuell präsentierten Hände. Frauen lokalisierten sich weniger in der Situation, wenn sie auf männliche Avatarhände blickten. Indessen unterstützten sowohl männliche als auch weibliche realistisch aussehende Hände das Präsenzgefühl von Männern. Es sei angemerkt, dass es sich bei der Darstellung eines männlichen oder weiblichen Avatars in VR nur um eine von vielen Optionen handelt, da auch geschlechtslose Repräsentationen eingesetzt werden können, beispielsweise durch die Weglassung von Geschlechtsmerkmalen oder durch die Darstellung von animalischen Charakteren bzw. roboterartigen Figuren.

Aber nicht nur die Wahrnehmung, sondern auch das Verhalten des Nutzers wird so aktiv je nach Situation beeinflusst. Aus äußerlichen Merkmalen der virtuellen Gestalt kann sich unterbewusst eine Erwartungshaltung entwickeln, welche den Nutzer gemäß den damit assoziierten Verhaltensmuster agieren lässt. Beispielsweise teilten Kilteni et al. [11] in ihrer Studie die Versuchsteilnehmer in zwei Gruppen ein, wobei alle Teilnehmer eine bestimmte Zeit frei zu einer Musik trommeln sollten. Vorab erhielten die Versuchspersonen kulturelle Hintergrundinformationen über die Djembe Trommel mittels eines Videos, worin auch afrikanische Trommler beim traditionellen Spielen zu sehen waren. Anfangs wurde jeder Teilnehmer durch zwei konturenfreie neutrale Hände repräsentiert. Im zweiten Schritt erhielt eine Hälfte der Probanden einen Avatar mit zwangloser Kleidung und einer dunkleren Hautfarbe, die andere Gruppe wurde in Business Outfit und mit hellem Hautton an dem virtuellen Körper repräsentiert. Der Proband blickte hierbei in der First Person Perspektive auf seine Hände, konnte aber im Spiegel den ganzen Körper des Avatar sehen (siehe Abbildung 3). Das Spielverhalten von Personen aus der ersten Gruppe, welche den legere gekleideten dunkelhäutigen Mann verkörperte, wies mehr Variationen beim Musizieren mit der Trommel auf als die zweite Hälfte. Die Wissenschaftler führten dies darauf zurück, dass die Probanden unbewusst mit dem äußerlichen Erscheinungsbild ein bestimmtes Verhalten assoziierten und sich gemäß des Stereotyps einfacher in die Rolle fanden. Dieses Phänomen wird auch Proteus-Effekt genannt [33].

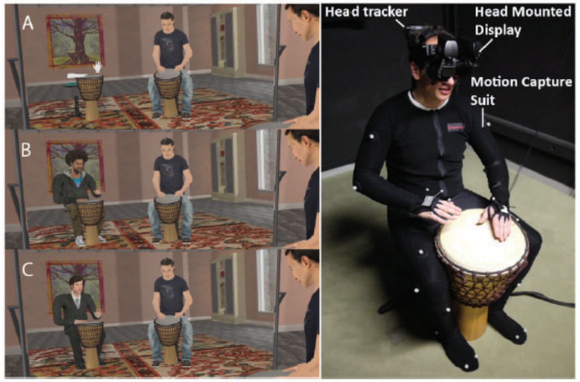


Abbildung 3 Links: Blick des Probanden in den Spiegel mit der Repräsentation des Selbst durch (A) neutrale Hände (B) einen dunkelhäutigen legere gekleideten Avatar (C) einen hellhäutigen formal angezogenen Avatar. Rechts: Versuchsaufbau des sitzenden Versuchsteilnehmers mit dem am Kopf befestigten Display und den am Körper angebrachten Sensoren während dem Trommeln. (Quelle: Kilteni et al. [11])

Zusammenfassend lässt sich festhalten, dass sich das Gefühl von Embodiment aus zwei Komponenten zusammensetzt. Einerseits spielt die Wahrnehmung und Verarbeitung von Sinneseindrücken eine zentrale Rolle. Andererseits ist die mentale Vorstellungskraft ein ebenso wichtiger Bestandteil für das gelungene Eintauchen in einen fremden Körper. Außerdem kommt es zu Unterschieden des GvEs je nachdem, ob sich die Person durch den gezeigten Zielkörper, beispielsweise einen Avatar, verhältnismäßig korrekt repräsentiert fühlt. Das Besitz- und Kontrollgefühl über einen bestimmten Körper kann bis zu einem gewissen Grad gezielt manipuliert werden, wobei auch die gegenteilige Wirkung erzeugbar ist. Sowohl bei der Illusion des „Hineinschlüpfens“ als auch beim Verlassen des Körpers wird der Blickwinkel als elementares Mittel zur Erzeugung des gewünschten Effekts eingesetzt. Das folgende Kapitel widmet sich daher intensiver mit der Bedeutung der Perspektive.

2.2 Intensitätsunterschiede des Identifikationsgefühls bei verschiedenen Perspektiven

Eine zentrale Bedeutung bei der Darstellung des virtuellen Selbst erhält die gewählte Perspektive. Denn sowohl auf das Gefühl der Selbstlokalisierung, als auch des Körperbesitzes und der -kontrolle kann laut Galvan Debarba et al. [6] und Maselli und Slater [20] mithilfe der Perspektive Einfluss genommen werden. Daher werden im Folgenden charakteristische Merkmale jeweils in Bezug auf die First Person (1PP; siehe Abbildung 4 links) sowie die Third Person Perspektive (3PP; siehe Abbildung 4 rechts) hervorgehoben. Je nach Perspektive werden unterschiedliche Teile des Körpers sichtbar, welche für den Nutzer im Fokus des Bildausschnitts stehen. Daher wird zudem jeweilig auf die Darstellungsformen näher eingegangen, welche hinsichtlich des Betrachtungswinkels relevant sind. Diese Vorgehensweise dient der Verdeutlichung, wie der Blickwinkel je nach Situation variiert werden kann, um gezielt das Verkörperungsgefühl zu manipulieren.

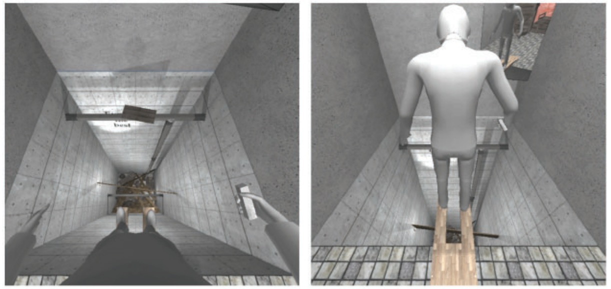


Abbildung 4 Sicht des Nutzers aus der First Person Perspektive aus dem Körper (links) und Third Person Perspektive auf den Körper (rechts). (Quelle: Galvan Debarba et al. [6])

First Person Perspektive

Die First Person Perspektive zeigt die Gliedmaßen des Avatar als würde der Nutzer tatsächlich in dem Körper sein (siehe Abbildung 4 links). Mit der Bewegung des Kopfes kann der User sich wie in der echten Welt im virtuellen Raum umsehen. Die Ego Perspektive erzeugt in VR [7] als auch unter realen Umständen [25] einen maximalen Grad an Körperbesitz und erleichtert ein Wiederfinden im ganzen Körper. Forschungsergebnisse zeigen die Tendenz, dass kompliziertere oder Fingerspitzengefühl erfordernde Interaktionen aus der 1PP im Vergleich zur 3PP bessere Ergebnisse erzielen. Ebenso schätzen Nutzer die Navigationsdauer in der 1PP realistischer ein [7]. Wie bereits angedeutet, charakterisiert sich die 1PP durch den Blick aus dem Körper heraus auf die Hände, welche hierbei stets zu sehen sind. Es liegt demnach nahe anzunehmen, dass die gezeigten Hände ein Schlüsselelement für das GvE sind. Mit der Frage, inwiefern das Aussehen der Hände Auswirkungen auf das GvE hat, beschäftigten sich Lin und Jörg [18]. In Abbildung 5 sind die getesteten von ihnen Darstellungsformen gezeigt. Diese wurden auf die virtuelle Handillusion untersucht, welche ähnlich wie die RHI die Wirkung des Gefühls des Körperbesitzes mit dem Unterschied der Verwendung einer virtuellen Hand auslösen kann. Die durchgeführte Studie deutete auf große individuelle Unterschiede hin, jedoch führten menschenähnliche Hände zu einer stärkeren Illusion des Körperbesitzes, wobei das realistische Handmodell den größten Effekt erzielte. Das Holzstück erzeugte lediglich bei manchen Probanden diese Wirkung. Bei allen Modellen wurde ein ähnlicher Grad an dem empfundenen Gefühl der Körperkontrolle festgestellt, sodass das Aussehen in diesem Aspekt keine Veränderung mit sich brachte. Auch das Experiment von Argelaguet et al. [1] bestätigt die Ergebnisse, dass das Gefühl des Körperbesitzes bei realistisch aussehenden Händen höher ausfällt. In Bezug auf das Gefühl der Körperkontrolle ergab diese Studie allerdings höhere Werte bei der Verwendung einer abstrakten Hand. Überdies kann bereits eine Variation der Fingeranzahl (siehe Abbildung 6) laut Schwind et al. [26] unterschiedliche Wahrnehmungen hervorrufen. Je weniger Finger die Probanden bei der realistischen Hand erkennen konnten, desto mehr sank das Präsenzgefühl. Im Vergleich dazu verhielt es sich bei der abstrakten Darstellung stets sehr gleichmäßig. An dieser Stelle wäre von hohem Interesse gewesen, inwiefern die einzelnen Komponenten des GvEs hierdurch beeinflusst werden und insbesondere ob durch die Reduktion der Hände ein niedrigeres Gefühl der Körperkontrolle wahrgenommen wird. Dass die Versuchsteilnehmer sich mit den Händen identifizieren konnten, demonstrierte hingegen das Auftreten von Phantomschmerzen bei nicht-ammputierten Personen, woraus ein gewisser Grad an GvE abgeleitet werden kann. Ein weiteres manipulierbares Mittel zur Erhöhung des Gefühls von Körperbesitz ist laut Perez-Marcos et al. [23] die Verbundenheit der Hände zum virtuellen Körper mittels der Darstellung des Armes. Wird die Hand ohne Bezug zum restlichen Avatar gezeigt, fühlte sich der Proband vergleichsweise weniger als Herr dieser Gliedmaßen.

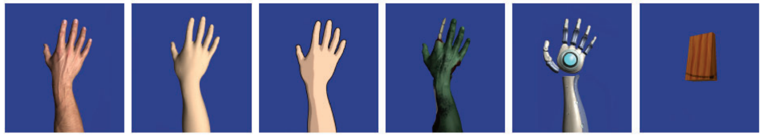


Abbildung 5 Getestete unterschiedlich aussehende Handmodelle in Lin und Jörgs Studie von links nach rechts: Realistische Hand, vereinfachte realistische Hand, weiter vereinfachte realisitsche Hand, Zombiehand, Roboterhand und Holzstück. (Quelle: Lin und Jörg [18])

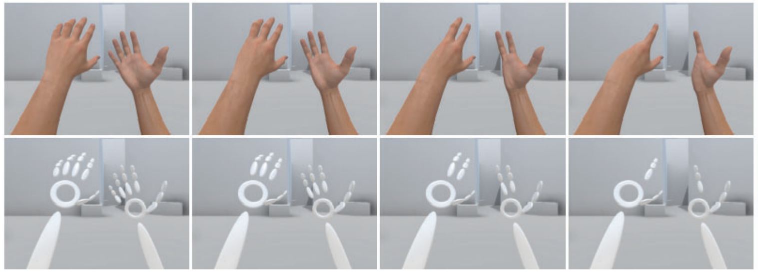


Abbildung 6 Getestete realistische (oben) und abstrakte (unten) Handmodelle mit abnehmender Anzahl an Fingern. (Quelle: Schwind et al. [26])

Third Person Perspektive

Bei der Third Person Perspektive besitzt der Nutzer nicht mehr den für ihn gewohnten Blickwinkel aus seinem Körper heraus, sondern er sieht auf den Körper der virtuellen Repräsentation seines Ichs (siehe Abbildung 4 rechts). Der Nutzer kann den Betrachtungswinkel auf den Körper kontrollieren, indem er den Kopf bewegt. Im Fokus des Blickfeldes steht daher stets der Avatarkörper. Ob und inwieweit unter diesen Umständen ein Gefühl von Embodiment auch in der Third Person Perspektive möglich ist, welche Auswirkungen die Perspektive auf das Erlebnis haben kann und wie diese gezielt genutzt werden können, soll in diesem Kapitel näher beleuchtet werden.

Beginnend mit dem Gefühl des Körperbesitzes lassen sich bezüglich des GvEs verschiedene Studienergebnisse finden. Laut Maselli und Slater [20] ist die Illusion, den virtuellen Körper ganz zu besitzen, von der 1PP abhängig. Dass durch die Verwendung der 3PP das Possessionsgefühl zumindest gehemmt wird, zeigen auch die Forschungsergebnisse von Galvan Debarba et al. [6]. Trotzdem ist es durchaus möglich das Gefühl von Körperbesitz über einen Avatar hervorzurufen, welches der 1PP ähnelt. Ein deutlich stärkeres Gefühl des Körperbesitzes kann in der 3PP hervorgerufen werden, wenn visuell-taktile oder visuell-motorische Reize mit dem gezeigten Körper übereinstimmen [6]. Dennoch hat die Perspektive einen größeren Einfluss als die Synchronität visuell-taktiler Reize [30]. Das Vorhandensein einer visuell-taktilen Kongruenz unterstützt nicht nur das Gefühl von Körperbesitz, sondern auch das der Körperkontrolle sowie der Selbstlokalisierung [20].

Ein Blick auf weitere Faktoren des GvEs zeigt jedoch, dass VR-Nutzer sich durchaus mit dem virtuellen Körper identifizieren können [16] [17]. Eine Lokalisierung des Ichs im Avatar war ebenfalls erkennbar, allerdings nicht so ausgeprägt wie in 1PP [6]. In Bezug auf die empfundene Präsenz in der virtuellen Umgebung konnte die 3PP im Vergleich zur 1PP höhere Werte erzielen [14].

Durch die 3PP wird laut Gorisse et al. [7] der virtuelle Raum bewusster wahrgenommen und Veränderungen können schneller erfasst werden, da der Körper im Kontext der umgebenden Welt aufgefasst wird. Zudem wird bei der Verwendung von 3PP die Situation generell als sicherer wahrgenommen, was insbesondere in Panik verursachenden Momenten ein erwünschter Effekt ist. Dies könnte beispielsweise für therapeutische Zwecke eine gute Möglichkeit sein, den Patienten nicht zu überfordern bzw. ihm eine Möglichkeit zu bieten, bei einer wahrgenommenen Bedrohung des Körpers von der 1PP in die 3PP zu wechseln. Dies kann insbesondere der Behandlung von Angststörungen zu Gute kommen, zu dessen Zweck außerkörperliche Erfahrungen in einem kontrolliertem Rahmen nachgeahmt werden. Zur Durchführung wurde beispielsweise bei Bourdins Studie [4] mithilfe der Verwendung der 3PP und der Nichtanwendung visuell-taktiler Synchronität bewusst ein Gefühl hervorgerufen, welches impliziert, den Körper nicht in seiner natürlichen Form zu besitzen, wobei durchschnittlich ein niedrigeres Level an Angst vor dem Tod bemerkbar war. Außerdem kann bei der Wahl der 3PP das Risiko einer auftretenden Cyberkrankheit im Vergleich zur 1PP verringert werden [14]. Dies stellt einen großen Vorteil für Personen dar, die durch den Gebrauch von VR-Equipment unter Kopfschmerzen oder Schwindel leiden. Warum die 3PP aber diese positive Eigenschaft besitzt, wird in der Untersuchung nicht weiter erläutert.

Dass die äußerliche Form Auswirkungen sowohl auf die Selbstwahrnehmung als auch das Verhalten des Nutzers haben kann, wurde bereits für die 1PP gezeigt. Wie verhält sich dies jedoch, wenn der Nutzer nicht nur die Hände, sondern den ganzen Körper des Avatars sieht? Die Proteus-Effekt Theorie [33] behandelt Implikationen des Aussehens des virtuellen Ichs in der 3PP in einer immersiven Umgebung. Attraktivere Avatare veranlassten Nutzer eher mit einem Fremden zu interagieren und durch Repräsentationen mit einer gesteigerten Körpergröße verhielten sich die Teilnehmer selbstbewusster bzw. waren eher bereit sich auf Diskussionen einzulassen. Gorisse et al. [8] beschäftigten sich in ihrer Studie mit dem Umstand, inwiefern der Faktor Ähnlichkeit bzw. Wiedergabetreue des virtuellen Körpers mit dem Realen des Nutzers in 3PP Einfluss auf das GvE hat. Getestet wurde dies anhand von drei verschiedenen Avataren mit steigendem Realitätsgrad (Roboter, Anzug und Doppelgänger), abgebildet in Grafik 7, welche verschiedene spielerische Aufgaben absolvieren sollten. Aus der qualitativen Analyse folgten Aussagen, welche auf ein sehr hohes GvE schließen lassen, da Probanden beispielsweise den Eindruck hatten, nicht nur einen Charakter, sondern auch sich selbst zu steuern. Eine Person dachte auf den ersten Blick sogar, dass es ihr echter Körper sei. Demgegenüber äußerten sich Personen aber auch gegenteilig, sodass sie den Doppelgänger als gruselig empfanden oder eine ungenaue Kopie von sich Selbst gegenüber sahen. Einige Teilnehmer fühlten sich mit dem Roboter stärker und selbstbewusster, da sie mit einem Maschinenkörper gefühlsmäßig weniger anfällig für Verletzungen seien. Die Forscher schlossen daraus, dass die Repräsentation des Ichs auch Einfluss auf das subjektive Spielerlebnis haben kann. Die statistische Auswertung zeigte für den Doppelgänger einen höheren Wert für das Gefühl des Körperbesitzes als bei den anderen beiden Körpermodellen. Die Komponenten Selbstlokalisierung sowie Körperkontrolle fielen bei allen drei Variationen ohne signifikante Unterschiede aus. Dass das Körpermodell aber nicht zwingend dem realen Körper des Nutzers entsprechen muss, zeigten Krekhof et al. [15].

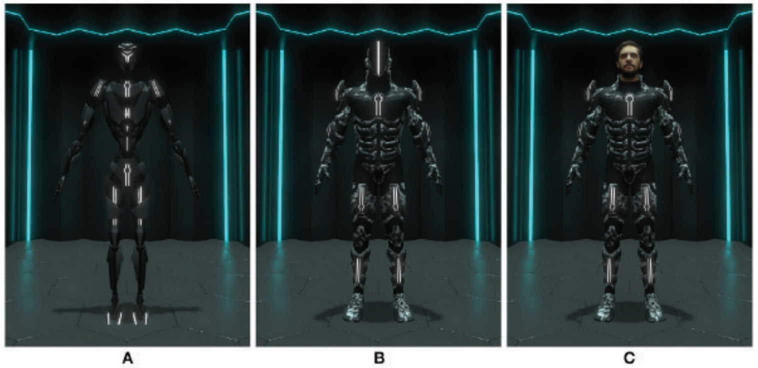


Abbildung 7 Getestete Körpermodelle mit Steigerung der äußerlichen Ähnlichkeiten zum Nutzer: (A) Roboter, (B) Anzug und (C) Doppelgänger. (Quelle: Gorisse et al. [8])

Hierbei wurden die Versuchsteilnehmer mittels VR nicht in humanoide, sondern in einen animalischen Avatar hineinversetzt. Blieb der Blickwinkel stets von hinten (wie in Abbildung 8) auf dem Spinnenkörper gerichtet und bewegte sich die tierische Repräsentation synchron zu dem Proband, so fiel das Gefühl von Körperbesitz in 3PP fiel im Vergleich zur 1PP nicht signifikant niedriger aus.

Abbildung 8 3PP auf den Spinnenkörper und Reflektion des Avatars im Spiegel (Quelle: Krekhof et al. [15])

Resümierend bietet die Third Person Perspektive zahlreiche Möglichkeiten passend zum Anwendungsfall dem Nutzer das Gefühl zu suggerieren sich mit einem Körper verbunden zu fühlen, aber auch falls nötig eine gewisse Distanz aufzubauen. Eine Erhöhung des GvE kann einerseits durch die Synchronität visuell-taktiler und visuell-motorischer Reize gefördert werden. Andererseits kann die äußerliche Erscheinungsform des Avatars das Eintauchen in den fremden Körper begünstigen. Eine Anleitung zum Erstellen des perfekten Avatars für das optimale GvE gibt es hier allerdings nicht. Da sich der Nutzer jedoch selbst damit identifizieren soll, liegt die Option nahe, ihn möglichst frei wählen zu lassen, wie dieser aussieht.

3 Verhältnis des Gefühls von Embodiment von Virtual Reality zu anderen immersiven Medien

Das Eintauchen in einen (fremden) Körper oder eine fremde Situation lässt sich nicht nur bei VR, sondern auch in weiteren Kommunikationsträgern finden. Aber welche Unterschiede und Gemeinsamkeiten bestehen zwischen VR und anderen Medien? Zur Beantwortung dieser Frage wurden verschiedene Kommunikationstäger betrachtet, deren Hauptmerkmale in Bezug auf das GvE hervorgehoben und mithilfe einer subjektiven Evaluierung die Unterschiede erarbeitet.

Zu Beginn wird hierzu die Differenzierung der Begriffe Präsenz und Immersion vorgenommen, da diese in den unterschiedlichen Disziplinen eine wichtige Rolle spielen. Das Präsenzgefühl beschreibt auf der psychologischen Ebene die subjektive Empfindung des Nutzers sich in einer Welt tatsächlich wiederzufinden und diese somit als „echte“ Erfahrung zu werten. Außerdem bedeutet präsent sein, mit der Umgebung zu interagieren. Im Gegensatz dazu zielt der Begriff Immersion auf den technischen Aspekt des Systems ab. Die Wahrnehmung des Nutzers wird überlistet, indem beispielsweise durch eine höhere Displayauflösung oder echt klingende Geräusche versucht wird ein höherer Realitätsgrad zu erreichen [29]. Ein direkter Bezug zu den körperlichen Umständen des Mediums besteht hierbei nicht, jedoch impliziert das Eintauchen in eine Geschichte oder eine Situation häufig auch das Einnehmen einer Figur. Als Ausgangspunkt sind ebenso für andere Medien zur Entstehung von dem GvE, wie in Kapitel 2.1 erarbeitet, sowohl der zeitlich und räumlich übereinstimmende sensorische Input essentiell, als auch die individuelle Imaginationskraft. In allen folgend angeführten Beispielen besteht stets eine Beziehung zwischen dem Informationen sendenden Medium zum Empfänger, welcher im Computerspiel ein Nutzer, bei Filmen aber ein Zuschauer ist [10].

Das Eintauchen in einen anderen Charakter, seinen Körper, sogar dessen Leben und die passende Umgebung beginnt bereits mit der Übermittlung von sprachlichen Konstrukten. Das narrative Mittel eines Ich-Erzählers taucht in verschiedenen Medien auf. Hierzu gehören Texte, Bücher, Hörbücher oder Podcasts. Der Leser oder Zuhörer erhält die Möglichkeit eine andere Person zu sein und deren Gefühle oder Handlungen nachzuempfinden. Ein Eingreifen in die Situation oder Entscheidungen ist dem Rezipient im Regelfall nicht möglich. Die Besonderheit liegt darin, dass sie rein mit der Vorstellungskraft arbeitet. Das gesprochene bzw. geschriebene Wort liefert keine zusätzlichen Eindrücke über die Sinne, da die Vorstellung selbst der genauesten Beschreibung von einer Sache, Gefühlen oder sonstigen Umständen letztendlich der Fantasie des Individuums überlassen bleibt. Beispielsweise Rollenspiele wie Pen-&-Paper zeigen, dass eine (hohe) Immersion nicht zwingend vorhanden sein muss, um sich als eine andere Person zu fühlen.

Anders verhält es sich hiermit bei statischen oder bewegten Bildern. Durch die zusätzliche Darbietung visueller Reize werden laut Robin Curtis [5] die sinnlichen Möglichkeiten und körperlichen Erfahrungen erweitert. Dem Betrachter wird so die Ausgestaltung der Situation abgenommen und er kann sich mit wenig mentaler Anstrengung auf das Geschehen einlassen. Eine aktive Interaktion der Nutzers ist nicht erforderlich, wodurch sich eine zweidimensionale Welt vor den Zuschaueraugen zu einer neuen Realität transformiert [10].

Der nächste Schritt in eine noch realistischere Verkörperlichung folgt beispielsweise in Videospielen mit der Dreidimensionalität, mit deren Hilfe die Abbildung der Realität täuschend echt simuliert werden kann. Das empfangende Subjekt ist nun nicht mehr nur Betrachter, sondern kann als Nutzer auch in das Geschehen eingreifen. Mit abnehmender Passivität und mehr interaktiven Möglichkeiten kann der Spieler selbst bestimmen, wie er in das Geschehen eingreift. Die Abbildung der Realität im virtuellen Raum wird insbesondere bei der Betrachtung von Konsolen wie Wii Sports oder Eyetoy deutlich. Bewegungen des Körpers werden mittels Controllern oder Sensoren gemessen und auf den virtuellen Körper projiziert.

Gleichermaßen erkennen Virtual oder Argumented Reality Brillen die Bewegungen des Kopfes und passen die digitale Repräsentation an. Im Vergleich zu den klassischen statischen Monitoren kann hierbei mehr auf den Nutzer eingegangen werden. Auch auditive Stimuli können die Interaktivität und die Präsenz des Users in seiner Umgebung weiter erhöhen, indem Geräusche synchron zu den ausgeführten Bewegungen und dem gezeigten Bild abgespielt werden [9].

Die nähere Betrachtung zeigt die Entwicklung zu realitätsabbildenden Technologien, welche primär die Sinne bedienen. Verglichen mit anderen Medien bietet VR durch Dreidimensionalität der Animationen, Synchronität des tatsächlichen und virtuellen Körpers sowie interaktiver Steuerelemente realitätsnahe Optionen, welche dem Rezipienten das Gefühl der Einnahme einer alternativen Realität erwecken können, welches nicht nur auf der Imagination beruht.

4 Zusammenfassung und Ausblick

Zu Beginn der Ausarbeitung wurde die Frage gestellt, wie es sich anfühlt, in einen anderen Körper einzutauchen und diesen kontrollieren zu können. Mithilfe von Virtual Reality wurde eine Möglichkeit geschaffen diese Empfindung bis zu einem gewissen Grad nachzuahmen und sich so nicht nur in fremde Welten, sondern auch andersartige Körper hinein zu versetzen. Die Illusion, einen fremden Körper als den eigenen wahrzunehmen, kann entstehen und durch bestimmte Bedingungen gefördert oder gehemmt werden. Der Begriff Embodiment bezeichnet dabei, inwieweit die Person das Gefühl des Besitzes, der Kontrolle und der Selbstlokalisierung in bzw. zu dem Körper empfindet. Abhängig sind diese Eindrücke von sensorischem Input sowie der individuellen Imagination. In VR gibt es verschiedene Möglichkeiten das Gefühl von Embodiment gezielt zu erzeugen und einzusetzen. Von großer Bedeutung erweist sich hierbei einerseits das äußerliche Erscheinungsbild des virtuellen Ichs, welches nicht nur den Identifikationsprozess, sondern auch das Verhalten des Nutzers

über die mit dem Aussehen verbundene Erwartungshaltung maßgeblich beeinflusst. Andererseits bestimmt auch die Wahl der Perspektive, aus welcher der Nutzer auf das Geschehen blickt, inwiefern ein Verbundenheitsgefühl zwischen virtuellem und realem Ich entsteht. Sowohl in der First Person Perspektive, als auch in der Third Person Perspektive kann ein starkes GvE hervorgerufen werden, wobei erstere den Zustand der Realität näher kommt und somit Wiederfinden in dem (virtuellen) Körper erleichtert. In der 3PP nimmt der User allerdings den virtuellen Raum bewusster wahr und Veränderungen sind leichter zu erkennen, was möglicherweise an der wahrgenommenen Einbettung des virtuelle Ichs in seiner Umgebung liegt. Vor allem bei therapeutischen Zwecken kann der Einsatz der 3PP einen positiven Nutzen erzielen, da der Patient dies häufiger als sicherer empfindet.

Für die Forschung über Wahrnehmungsunterschiede bei variierenden Perspektiven in Virtual Reality bzw. bei weiteren Einflussfaktoren in Mixed Reality Umgebungen besteht weiterhin eine hohe Relevanz. In vielen Anwendungsbereichen wie z.B. dem medizinischem Sektor bei der Durchführung von Operationen, bei der Vorbereitung realistischer Trainings für militärische Zwecke oder in der Unterhaltungsindustrie, könnte der Einsatz von Virtual Reality in Kombination mit deren Erkenntnissen die Arbeit vereinfachen oder sogar verbessern. Es fehlen dazu bisher allerdings Forschungen, welche Perspektive für die konkreten Anwendungsfälle beispielsweise bei Operationen geeigneter wäre, um ein besseres Ergebnis zu erzielen. Generell weist die Sparte der Third Person Perspektive bisher wenige Ergebnisse auf, welche Einblicke in den Identifikationsprozess der eigenen Person mit einem Avatar geben. Als ein möglicher Grund dafür könnte die hohe Verbreitung der First Person Perspektive im Unterhaltungsbereich für VR sein. Interessant für die zukünftige Wissenschaft könnte sein, Erkenntnisse über die Wirkung der Perspektive für das GvE bei der Interaktion mit anderen Personen zu sammeln, da hierzu wenig Forschungsergebnisse auffindbar sind. Des Weiteren liegen zwar einige Ergebnisse vor, wie das Gefühl von Embodiment durch die Stimulation einzelner Sinneskanäle beeinflusst werden kann, jedoch ergibt sich ferner die Frage, inwiefern sich die Kombination aus zeitgleich ausgeübten Reizen auf möglichst viele Sinnesorgane positive oder negative Wirkungen aufweist. Schließlich gilt es während des ganzen Prozesses stets abzuwägen, inwieweit die Nachahmung der Realität für den gegebenen Anwendungsfall sinnvoll ist oder ob der gegenteilige Effekt gegebenenfalls bevorzugt werden sollte.

Literatur

1 Ferran Argelaguet, Ludovic Hoyet, Michael Trico, and Anatole Lecuyer. The role of interaction in virtual embodiment: Effects of the virtual hand representation. In *2016 IEEE Virtual Reality (VR)*, pages 3–10, [S.l.], 3/19/2016 - 3/23/2016. IEEE. `doi:10.1109/VR.2016.7504682`.

2 Laura Aymerich-Franch and Jeremy Baileson. The use of doppelgangers in virtual reality to treat public speaking anxiety: a gender comparison. *Proceedings of the International Society for Presence Research 2014*, pages 173–186, 2014. URL: `https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.718.2915&rep=rep1&type=pdf`.

3 M. Botvinick and J. Cohen. Rubber hands 'feel' touch that eyes see. *Nature*, 391(6669):756, 1998. `doi:10.1038/35784`.

4 Pierre Bourdin, Itxaso Barberia, Ramon Oliva, and Mel Slater. A virtual out-of-body experience reduces fear of death. *PloS one*, 12(1):e0169343, 2017. `doi:10.1371/journal.pone.0169343`.

5 Robin Curtis, editor. *Synchronisierung der Künste*. Fink, Paderborn, 2013.

6 Henrique Galvan Debarba, Sidney Bovet, Roy Salomon, Olaf Blanke, Bruno Herbelin, and Ronan Boulic. Characterizing first and third person viewpoints and their alternation for embodied interaction in virtual reality. *PloS one*, 12(12):e0190109, 2017. `doi:10.1371/journal.pone.0190109`.

7 Geoffrey Gorisse, Olivier Christmann, Etienne Armand Amato, and Simon Richir. First- and third-person perspectives in immersive virtual environments: Presence and performance analysis of embodied users. *Frontiers in Robotics and AI*, 4:3, 2017. `doi:10.3389/frobt.2017.00033`.

8 Geoffrey Gorisse, Olivier Christmann, Samory Houzangbe, and Simon Richir. From robot to virtual doppelganger: Impact of visual fidelity of avatars controlled in third-person perspective on embodiment and behavior in immersive virtual environments. *Frontiers in Robotics and AI*, 6:3, 2019. `doi:10.3389/frobt.2019.00008`.

9 Matthias Hoppe, Jakob Karolus, Felix Dietz, Paweł W. Wozniak, Albrecht Schmidt, and Tonja-Katrin Machulla. Vrsneaky. In Stephen Brewster, Geraldine Fitzpatrick, Anna Cox, Vassilis Kostakos, and Anna L. Cox, editors, *CHI 2019*, pages 1–9, New York, New York, 2019. The Association for Computing Machinery. `doi:10.1145/3290605.3300776`.

10 Dawid Kasprowicz. *Der Körper auf Tauchstation*. Nomos Verlagsgesellschaft mbH & Co. KG, 2019. `doi:10.5771/9783845298689`.

11 Konstantina Kilteni, Ilias Bergstrom, and Mel Slater. Drumming in immersive virtual reality: the body shapes the way we play. *IEEE transactions on visualization and computer graphics*, 19(4):597–605, 2013. `doi:10.1109/TVCG.2013.29`.

12 Konstantina Kilteni, Raphaela Groten, and Mel Slater. The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments*, 21(4):373–387, 2012.

13 Konstantina Kilteni, Jean-Marie Normand, Maria V. Sanchez-Vives, and Mel Slater. Extending body space in immersive virtual reality: a very long arm illusion. *PloS one*, 7(7):e40867, 2012. `doi:10.1371/journal.pone.0040867`.

14 Mingyu Kim, Jiwon Lee, Changhun Kim, and Jinmo Kim. Tpvr: User interaction of third person virtual reality for new presence and experience. *Symmetry*, 10(4):109, 2018. `doi:10.3390/sym10040109`.

15 Andrey Krekhov, Sebastian Cmentowski, and Jens Kruger. The illusion of animal body ownership and its potential for virtual reality games. In *IEEE Conference on Games 2019*, pages 1–8, [Piscataway, New Jersey], 2019. IEEE. `doi:10.1109/CIG.2019.8848005`.

16 Bigna Lenggenhager, Michael Mouthon, and Olaf Blanke. Spatial aspects of bodily self-consciousness. *Consciousness and cognition*, 18(1):110–117, 2009. `doi:10.1016/j.concog.2008.11.003`.

17 Bigna Lenggenhager, Tej Tadi, Thomas Metzinger, and Olaf Blanke. Video ergo sum: manipulating bodily self-consciousness. *Science (New York, N.Y.)*, 317(5841):1096–1099, 2007. `doi:10.1126/science.1143439`.

18 Lorraine Lin and Sophie Jörg. Need a hand? In Eakta Jain and Sophie Joerg, editors, *Proceedings of the ACM Symposium on Applied Perception*, pages 69–76, New York, NY, USA, 2016? ACM. `doi:10.1145/2931002.2931006`.

19 Matthew Lombard, Frank Biocca, Jonathan Freeman, Wijnand IJsselsteijn, and Rachel J. Schaevitz, editors. *Immersed in media: Telepresence theory, measurement et technology*. Springer International Publishing, Cham, 2015. `doi:10.1007/978-3-319-10190-3`.

20 Antonella Maselli and Mel Slater. The building blocks of the full body ownership illusion. *Frontiers in human neuroscience*, 7:83, 2013. `doi:10.3389/fnhum.2013.00083`.

21 Brian P. Meier, Simone Schnall, Norbert Schwarz, and John A. Bargh. Embodiment in social psychology. *Topics in cognitive science*, 4(4):705–716, 2012. `doi:10.1111/j.1756-8765.2012.01212.x`.

22 Bryan Paton, Jakob Hohwy, and Peter G. Enticott. The rubber hand illusion reveals proprioceptive and sensorimotor differences in autism spectrum disorders. *Journal of autism and developmental disorders*, 42(9):1870–1883, 2012. `doi:10.1007/s10803-011-1430-7`.

23 Daniel Perez-Marcos, Maria V. Sanchez-Vives, and Mel Slater. Is my hand connected to my body? the impact of body continuity and arm alignment on the virtual hand illusion. *Cognitive neurodynamics*, 6(4):295–305, 2012. `doi:10.1007/s11571-011-9178-5`.

24 Valeria I. Petkova and H. Henrik Ehrsson. If i were you: perceptual illusion of body swapping. *PloS one*, 3(12):e3832, 2008. `doi:10.1371/journal.pone.0003832`.

25 Valeria I. Petkova, Mehrnoush Khoshnevis, and H. Henrik Ehrsson. The perspective matters! multisensory integration in ego-centric reference frames determines full-body ownership. *Frontiers in psychology*, 2:35, 2011. `doi:10.3389/fpsyg.2011.00035`.

26 Valentin Schwind, Pascal Knierim, Lewis Chuang, and Niels Henze. "where's pinky?". In Ben Schouten, Panos Markopoulos, Zachary Toups, Paul Cairns, and Tilde Bekker, editors, *Proceedings of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '17*, pages 507–515, New York, New York, USA, 2017. ACM Press. `doi:10.1145/3116595.3116596`.

27 Valentin Schwind, Pascal Knierim, Cagri Tasci, Patrick Franczak, Nico Haas, and Niels Henze. "these are not my hands!". In Gloria Mark, Susan Fussell, Cliff Lampe, m.c. schraefel, Juan Pablo Hourcade, Caroline Appert, and Daniel Wigdor, editors, *CHI '17*, pages 1577–1582, New York, NY, 2017. The Association for Computing Machinery. `doi:10.1145/3025453.3025602`.

28 Mel Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 364(1535):3549–3557, 2009. `doi:10.1098/rstb.2009.0138`.

29 Mel Slater, Vasilis Linakis, Martin Usoh, and Rob Kooper. Immersion, presence and performance in virtual environments. In Mark Green, Kim Fairchild, and Michael Zyda, editors, *Virtual reality software and technology*, pages 163–172, New York, NY, 1996. ACM Inc. `doi:10.1145/3304181.3304216`.

30 Mel Slater, Bernhard Spanlang, Maria V. Sanchez-Vives, and Olaf Blanke. First person experience of body transfer in virtual reality. *PloS one*, 5(5):e10564, 2010. `doi:10.1371/journal.pone.0010564`.

31 Helen Wauck, Gale Lucas, Ari Shapiro, Andrew Feng, Jill Boberg, and Jonathan Gratch. Analyzing the effect of avatar self-similarity on men and women in a search and rescue game. In CHI, editor, *CHI 2018*, pages 1–12, New York, NY, 2018. ACM. `doi:10.1145/3173574.3174059`.

32 Sharon R. Weeks, Victoria C. Anderson-Barnes, and Jack W. Tsao. Phantom limb pain: theories and therapies. *The neurologist*, 16(5):277–286, 2010. `doi:10.1097/NRL.0b013e3181edf128`.

33 Nick Yee and Jeremy Bailenson. The proteus effect: The effect of transformed self-representation on behavior. *Human Communication Research*, 33(3):271–290, 2007. `doi:10.1111/j.1468-2958.2007.00299.x`.

34 Chen-Bo Zhong and Katie Liljenquist. Washing away your sins: threatened morality and physical cleansing. *Science (New York, N.Y.)*, 313(5792):1451–1452, 2006. `doi:10.1126/science.1130726`.