

UBIACTION 2019

Edited by Pascal Knierim, Jakob Karolus, Fiona Draxler, Ville Mäkelä, Tonja Machulla, Florian Lang, Thomas Kosch, Albrecht Schmidt

UBIACTION 2019

4th Seminar on Ubiquitous Interaction
July 16, 2019, Munich, Germany

Edited by

Pascal Knierim

Florian Lang

Albrecht Schmidt

Editors

Pascal Knierim, Jakob Karolus, Fiona Draxler, Ville Mäkelä,
Tonja Machulla, Florian Lang, Thomas Kosch, Albrecht Schmidt

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

ACM Classification 1998 - H.5 INFORMATION INTERFACES AND PRESENTATION

ISBN-13: 978-1082354762

Publication date

16. July 2019

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.

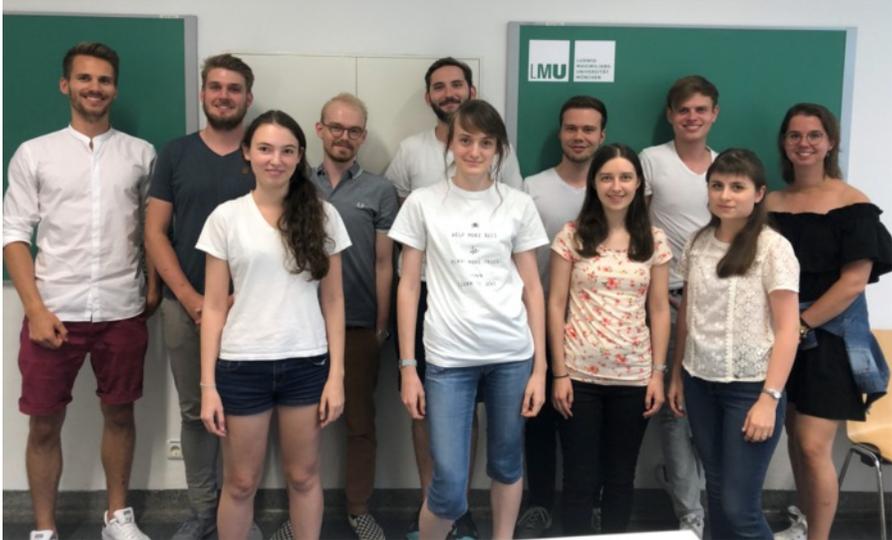
■ Vorwort

Die Digitalisierung hat massive Auswirkungen auf die verschiedenen Seiten unseres Lebens und Arbeitens. Zentrale Aspekte in unserer Gesellschaft, wie zum Beispiel Kommunikation, Mobilität oder Sicherheit, verändern sich durch ubiquitäre Computersysteme maßgeblich. Unsere Interaktion mit unserer Umgebung wird zunehmend zur Mensch-Computer-Interaktion, und Kommunikation mit anderen Menschen wird zunehmend durch digitale Technologien vermittelt. Die verwendeten Interaktionstechnologien und das Interaktionsdesign haben somit einen zentralen Einfluss auf die Erfahrungen, die wir in der Welt machen. Um einen Einblick in zukünftige Möglichkeiten im Bereich der Digitalisierung zu erlangen, wurde die Veranstaltung „Ubiaction“ vom Lehrstuhl „Human-Centered Ubiquitous Media“ der Ludwig-Maximilians-Universität München ins Leben gerufen. Bereits zum vierten Mal präsentierten die Studierenden des Seminars „Menschzentrierte Interaktion in ubiquitären Computersystemen“ am 16.07.2019 ihre Vorträge zu verschiedenen Themengebieten. In diesem Jahr wurden außerdem die Ergebnisse der Veranstaltung „Fortgeschrittene Themen der Mensch-Maschine-Interaktion“ sowie des Praktikums „Entwicklung von Mediensystemen“ im Rahmen einer Ausstellung präsentiert. Die schriftlichen Ausarbeitungen der Studierenden aus dem Sommersemester 2019 wurden in diesem Buch als vierter Band der Serie „Ubiaction“ zusammengeführt.

■ Inhaltsverzeichnis

Interaktionsmöglichkeiten in Projected Reality <i>Alexander Georg</i>	1:1–1:18
Augmented Reality-Interaktion im Fremdsprachenunterricht <i>David Gasser</i>	2:1–2:17
Learning in Mixed Realities <i>Michaela Schmiel</i>	3:1–3:15
Motivation im Zusammenhang mit der automatischen Generierung von Inhalten für das Erlernen von Sprachen <i>Anna-Carina Gehlisch</i>	4:1–4:18
Implizite Interaktionen durch physiologische Messungen <i>Maximilian B. Lamm</i>	5:1–5:20
Verdeckte Interaktion durch Elektromyografie <i>Evelyn Müller</i>	6:1–6:22
Implizite Interaktion mittels Eye-Tracking <i>Johannes Sylupp</i>	7:1–7:21
Augenbasierte Activity Recognition <i>Margaretha Lucha</i>	8:1–8:21
„Hey Siri, wo war ich stehen geblieben?“ - Mobile Wiederaufnahme unterbrochener Aufgaben mit Attention-aware Interfaces <i>Lars Reisig</i>	9:1–9:27
Wenn Displays und Augmented Reality Technologien verschmelzen - Herausforderungen und Möglichkeiten <i>Helena Bayerl</i>	10:1–10:17
UI-Design für ältere Menschen: Physische und kognitive Probleme <i>Oliver Hein</i>	11:1–11:13

■ Teilnehmer



Alexander Georg

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

Anna-Carina Gehlich

a.gehlich@campus.lmu.de
Medieninformatik

David Gasser

david.gasser@campus.lmu.de
Mensch-Computer-Interaktion

Maximilian B. Lammel

Maximilian.Lammel@campus.lmu.de
Mensch-Computer-Interaktion

Michaela Schmierl

M.Schmierl@campus.lmu.de
Medieninformatik

Evelyn Müller

Eve.Mueller@campus.lmu.de
Medieninformatik

Johannes Sylupp

sylupp.johannes@campus.lmu.de
Medieninformatik

Helena Bayerl

h.bayerl@campus.lmu.de
Mensch-Computer-Interaktion

Margaretha Lucha

M.Lucha@campus.lmu.de
Mensch-Computer-Interaktion

Oliver Hein

oliver.hein@campus.lmu.de
Mensch-Computer-Interaktion

Lars Reisig

lars.reisig@campus.lmu.de
Mensch-Computer-Interaktion

Interaktionsmöglichkeiten in Projected Reality

Alexander Georg

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

Zusammenfassung

In dieser Arbeit soll ein Überblick über die verschiedenen Interaktionsmöglichkeiten in Projected Reality gegeben werden. Dabei wird zunächst der Begriff definiert und von den anderen Formen der Realitäten abgegrenzt. Es werden verschiedene Anwendungsfälle vorgestellt, die demonstrieren, in welcher Form eine Interaktion in Projected Reality geschehen kann. Dabei wird zwischen einer Interaktion mit Menschen und einer Interaktion mit Objekten unterschieden. Anschließend wird diskutiert, welche Vorteile sich daraus im Vergleich zu Augmented und Virtual Reality ergeben. Zum Abschluss wird ein Ausblick auf technische Entwicklungen im Bereich Projected Reality gegeben.

2012 ACM Computing Classification Human-centered computing → Interaction techniques

Keywords and phrases Projected Reality; Interaktion

1 Einführung

Die Idee, mittels Projektionen eine neue oder erweiterte Realität zu erschaffen, ist bereits alt. Im 18. Jahrhundert waren in Europa sogenannte Phantasmagoria weit verbreitet. In diesen Horrorschaustücken wurden unter anderem handgehaltene Projektoren (Laterna Magica) genutzt, die Bilder von Skeletten und Dämonen projizierten. Durch eine Bewegung des Projektors wurde dabei die Illusion erzeugt, die Figuren würden sich durch den Raum bewegen. Wurden mehrere dieser Projektoren genutzt, konnten die Figuren ähnlich wie beim Puppentheater auch interagieren und eine Geschichte erzählt werden [4]. Damals noch begrenzt in den Möglichkeiten, werden heutzutage bereits Systeme wie PoMoCo getestet, bei denen die projizierten Figuren automatisch miteinander agieren. Der Anwendungszweck gilt dabei nicht mehr nur noch der Unterhaltung sondern dem Socialising von Filesharing und Inventory Management [23]. Interaktive Projektoren finden demnach vermehrt auch Anwendung



© Alexander Georg;

licensed under Creative Commons License CC-BY

Cite as: Alexander Georg. Interaktionsmöglichkeiten in Projected Reality. In *3rd Seminar on Ubiquitous Interaction (UBIACTION 2019)*. Editors: Pascal Knierim, Jakob Karolus, Fiona Draxler, Ville Mäkelä, Tonja Machulla, Florian Lang, Thomas Kosch, Albrecht Schmidt. July 16, 2019. Munich, Germany. pp. 1:1–1:18.

1:2 Interaktionsmöglichkeiten in Projected Reality

in produktiven Bereichen wie beispielsweise der Medizin und im Design.

Dabei können nicht mehr nur noch flache Oberflächen als Projektionsfläche dienen, sondern auch Projektionen auf bewegliche 3D-Objekte gemappt werden. Dadurch ergibt sich die Möglichkeit, neue Arten von haptischen Benutzerschnittstellen zu gestalten. Die Projektoren müssen dabei auch nicht mehr zwingend in der Hand gehalten oder stationär befestigt werden, sondern können an Helmen [3] oder sogar Drohnen [13] eingesetzt werden.

Ziel dieser Arbeit ist es, einen Überblick über die verschiedenen Interaktionsmöglichkeiten zu geben, die sich in Projected Reality eröffnen. Außerdem sollen Chancen im Vergleich zu Virtual und Augmented Reality diskutiert werden.

2 Einordnung und Definition von Projected Reality

Um Projected Reality richtig klassifizieren und definieren zu können, ist es sinnvoll, vorab die verschiedenen Formen von Realitäten zu betrachten.

Virtual Reality (VR) beschreibt das komplette Eintauchen in eine virtuelle Welt, in welcher der Nutzer interagieren kann. Die dargestellten Inhalte können dabei entweder in Echtzeit von einem Computer generiert werden oder vorab gerendert werden, wie beispielsweise 360°-Videos [16].

Augmented Reality (AR) beschreibt das Überlagern von digitalen Inhalten und der echten Welt mit Hilfe von technischen Geräten [16].

Mixed Reality (MR) stellt eine Mischung zwischen VR und AR dar. MR beinhaltet sowohl die Interaktionsmöglichkeiten, die in VR gegeben sind als auch die Vermischung von virtuellen und digitalen Inhalten [16].

Teilweise fällt es schwer, Anwendungen genau einer dieser Realitäten zuzuordnen, da besonders zwischen MR und AR bzw. MR und VR nur eine geringe Trennschärfe besteht.

Projected Reality (PR) In verschiedenen Arbeiten werden teils unterschiedliche Begrifflichkeiten dafür genutzt, um zu beschreiben, dass digitale Inhalte in die reale Welt projiziert werden. Häufig genutzte Begriffe sind beispielsweise Spatial Augmented Reality [15], Tangible Augmented Reality [25] und Projected Augmented Reality [7]. Die Verwendung des Begriffs AR in

Kombination mit den Begriffen Spatial, Tangible und Projected deutet zwar darauf hin, dass auch bei diesen Formen der Realität eine Vermischung von virtuellen und realen Inhalten entsteht, lässt allerdings die physischen Interaktionsmöglichkeiten außer Acht, die bei AR so nicht gegeben sind. Jochen Ehnes und Michitaka Hirose fassten dahingehend den Begriff Projected Augmented Reality zusammen und benutzen in ihren Arbeiten fortgehend den Begriff Projected Reality [7, 8].

3 Technische Umsetzung

Ein minimales Setup für eine PR-Anwendung muss mindestens aus einem Gerät zum Tracking, beispielsweise einer Kinect und einem Projektor bestehen. Technische Aspekte, die bei der Umsetzung beachtet werden müssen, sind nach der Arbeit von Joan Sol Roo und Martin Hachet wie folgt [20]:

- Kalibrierung: Die Position der Kamera und des Projektors in Relation zueinander
- Geometrie: Oberflächeninformationen der realen Objekte auf die Inhalte projiziert werden sollen
- Position: Position der Objekte in Relation zu Kamera und Projektor
- Material: Oberflächengegebenheiten der Projektionsfläche wie beispielsweise der Reflexionsgrad
- Licht: Umgebungsbeleuchtung und Helligkeit von Objekten
- Schatten: Verdeckung von Objekten
- Dynamik: Veränderungen der Szene hinsichtlich der vorangestellten Punkte

Das RoomAlive Toolkit von Microsoft adressiert einige der aufgeführten Herausforderungen und unterstützt bei der Implementierung von eigenen PR-Anwendungen. Die zugrundeliegende Hardware besteht aus sogenannten Procam Einheiten, die aus einem Projektor, einer Tiefenkamera und einem Computer bestehen. Es können mehrere dieser Procams über ein Netzwerk miteinander verbunden werden. Die Kalibrierung der einzelnen Einheit wird dabei über eine Gray-Code-Sequenz umgesetzt. Dabei projiziert zunächst jeder Projektor ein Gray-Code-Muster, welches dann von den Tiefenkameras der anderen Einheiten gescannt wird. Dadurch können die Pixel-Koordinaten der Tiefenkameras auf die Pixel-Koordinaten der Projektoren abgebildet werden. Aus den Informationen von allen Kameras wird dann eine Szene für Unity erstellt, die um Scripte und virtuelle Objekte erweitert werden kann. Mit dem RoomAlive Plugin für Unity kann ein Frame-by-Frame Mapping von Objekten umgesetzt werden, wodurch auch mit der Szene interagiert werden kann. Die Projektionen der virtuellen Inhalte können sowohl kontext- als auch nutzergebunden sein. Beim Scannen des Raumes werden die erkannten Oberflächen

1:4 Interaktionsmöglichkeiten in Projected Reality

automatisch kategorisiert. Dadurch kann eine Projektion von Gras, beispielsweise dem Kontext entsprechend, nur auf der als Boden erkannten Oberfläche abgebildet werden. Nutzergebundene Projektionen werden speziell auf den Betrachter angepasst. Inhalte, die hologrammartig zwischen Betrachter und Projektionsoberfläche erscheinen sollen, können per Headtracking und einer Simulation des Parallax Effekts umgesetzt werden. Des Weiteren kompensiert das RoomAlive Plugin automatisch die unterschiedlichen Projektionsoberflächen durch eine Anpassung der Projektionshelligkeit [12].

Durch das dynamische Mapping und die Möglichkeit mehrere Geräte zu kalibrieren, deckt das ToolKit bereits einige der aufgeführten technischen Herausforderungen wie Kalibrierung, Geometrie, Position und Dynamik ab. Auch das Problem der Verdeckung von Objekten kann dadurch gelöst werden, dass mehrere Projektoren aus verschiedenen Positionen ein Objekt bestrahlen. Weitere Aspekte wie Licht und Material können schon bei der Raumgestaltung dahingehend optimiert werden, dass besonders helle Gegenstände mit einer flachen Oberfläche als Projektionsoberfläche gewählt werden [9].

Als weitere Software für die Umsetzung von PR-Anwendungen wäre PapARt zu nennen, welche allerdings in ihrem Anwendungsgebiet deutlich limitierter ist als RoomAlive [14].

Trotz der existierenden Software sind die Anwendungen von der Hardwareseite aus meist noch prototypisch, da sie in der Regel speziell für einen Anwendungsfall konzipiert werden. Die Komplexität der Geräte variiert dabei je nach Verwendungszweck. Simple Anwendungen können beispielsweise schon mit einfachen Wii Remotes als Tracking Sensoren und einem Mini Projektor realisiert werden [2]. Komplexere Anwendungen wie beispielsweise die Anwendungsbeispiele für Microsofts RoomAlive können aus bis zu sechs Procams bestehen, mit denen die Fläche eines gesamten Raumes getrackt und auch bestrahlt werden kann [5].

4 Interaktionsmöglichkeiten

Da in PR die digitalen Inhalte in der realen Welt dargestellt werden, ergeben sich verschiedene Möglichkeiten mit realen Objekten zu interagieren. Die Eigenschaften des jeweiligen Objektes bedingen dabei auf welche Art und Weise eine Interaktion stattfinden kann. Wenn zusätzliche Inhalte auf greifbare Objekte projiziert werden, kann durch eine Veränderung der Position oder Rotation des Objektes eine Interaktion stattfinden. Wird eine Projektion auf einem

verformbaren Objekt abgebildet, kann die physische Erscheinung des Objektes verändert werden und die Projektion passt sich an die neue Form an. Des Weiteren wird in PR eine Interaktion zwischen Menschen begünstigt, da die digitalen Inhalte in der geteilten Realität dargestellt werden.

Im Folgenden werden verschiedene Arbeiten und Anwendungen vorgestellt, welche die verschiedenen Arten der Interaktion veranschaulichen.

4.1 Interaktion mit Objekten

Einer der Vorteile, die PR im Gegensatz zu VR und AR bietet, ist, dass tatsächlich mit den physischen Objekten interagiert werden kann. Die Interaktion muss somit nicht durch einen Controller, wie es bei VR der Fall ist, abstrahiert werden. Wie sich dies in der Interaktion mit Objekten in PR äußert, wird im Folgenden erläutert.

4.1.1 Greifbare Objekte

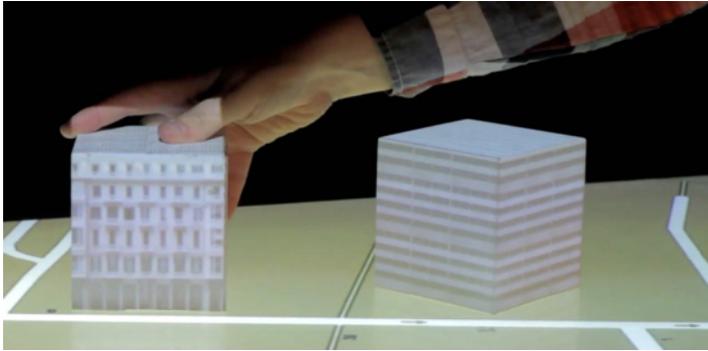
Nachgestellte Arbeiten thematisieren Anwendungen, bei denen sich die Projektion in Relation zu der Position und der Orientierung der getrackten 3D-Objekte ändert.

Tangible User Interfaces. Die Interaktion mit greifbaren Objekten lässt sich beispielsweise bei sogenannten Tangible User Interfaces (TUI) beobachten. Grundkonzept dieser TUIs ist es, auf herkömmliche Eingabegeräte wie Maus und Tastatur zu verzichten und physische Objekte (Tangibles) sowohl als Ein- als auch Ausgabegerät zu nutzen [11]. Die herkömmliche Interaktion mit Computern wird dadurch noch um haptische Reize erweitert.

Tangible 3D Tabletops. Ein Anwendungsbeispiel für TUIs im Bereich PR sind die im Paper von Peter Dalsgaard und Kim Halskov vorgestellten Tangible 3D Tabletops [6]. Dieses System kombiniert die Interaktionsmöglichkeiten von TUIs mit 3D-Projektionen, welche die Tangibles mit zusätzlichen Informationen überlagern. Einer der vorgestellten Use Cases ist die Visualisierung von 3D-Karten. Dabei werden sowohl auf den Tisch (Stadtplan) als auch auf die darauf befindlichen Objekte (Häuserfassaden) Inhalte projiziert, die sich je nach Anordnung der Objekte ändern (siehe Abbildung 1). Die Position und Lage der Tangibles wird dabei mit Hilfe von Bezugsmarken bestimmt, die sich unter jedem Objekt befinden und durch die halbtransparente Tischplatte von einer sich unter dem Tisch befindlichen Kamera getrackt werden können. Zwei oberhalb des Tisches befindliche Projektoren bilden die digitalen Inhalte

1:6 Interaktionsmöglichkeiten in Projected Reality

auf den Tangibles ab und ein weiterer Projektor unter dem Tisch projiziert zusätzliche Informationen auf die Tischplatte. Dieses System bietet sich vor allem für die Stadtplanung oder andere Bereiche an, in denen ein gutes räumliches Verständnis von Nutzen ist.



■ **Abbildung 1** Visualisierung einer 3D-Karte mit zusätzlichen Tangibles [6]

Strip’TIC. Ein weiteres Beispiel für den Einsatz von TUIs im Bereich PR ist der für die Flugsicherung entwickelte Prototyp Strip’TIC. Ziel war es, die bewährte Arbeitsweise mit physischen Kontrollstreifen beizubehalten, diese aber mit digitalen Informationen zu hinterlegen. Fluglotsen erhielten somit zusätzliche Informationen über den Status eines Fluges oder ihre Zuständigkeit. Von jedem physischen Kontrollstreifen existierte außerdem ein digitales Pendant im System, welches auch die handschriftlichen Notizen beinhaltete. Somit konnte ein digitales Feedback auf eine physische Eingabe gegeben werden und vorab gewarnt werden, wenn sich Flugrouten überschneiden [25]. Umgesetzt wird dies mit einem hybriden Stylus, der sowohl als analoger als auch digitaler Stift fungiert und es somit ermöglichte auf den physischen und digitalen Kontrollstreifen zu schreiben (siehe Abbildung 2). Der Stylus streamte dafür seine Position, um zu identifizieren, auf welchem Kontrollstreifen geschrieben wird. Mittels einer Infrarotkamera in der Spitze des Stiftes wird ein Prüfmuster auf dem physischen Kontrollstreifen getrackt, um die Schreibbewegungen nachzuvollziehen und das Geschriebene auf den digitalen Kontrollstreifen zu übernehmen. Das Tracking funktioniert mit einer Webcam, die Marker an der Rückseite der Kontrollstreifen scannt. Zum Darstellen der digitalen Inhalte werden zwei Projektoren benutzt, wobei einer oberhalb und der andere unterhalb des Stripboardes montiert ist [10]. Das digitale Feedback in Kombination mit der räumlichen Organisation der Kontrollstreifen fördert die Embodied Cognition und sorgt für einen besseren Workload der Fluglotsen als bei rein digitalen Systemen [25].



■ **Abbildung 2** Arbeit mit dem hybriden Stylus am Stripboard mit digitalen und physischen Kontrollstreifen [25]

Projektion zu nutzen, um reale Objekte mit zusätzlichen Informationen zu überlagern, bietet sich nicht nur dafür an, um TUIs zu erweitern, sondern auch um zusätzliche Inhalte in einen räumlichen Kontext zu setzen. Digitale Inhalte können somit leichter in Relation zu physischen Objekten gesehen werden. Vor allem Anwendungen aus dem medizinischen Bereich wie beispielsweise ProjectDR oder MoSART machen sich dies zum Nutzen [26, 3]. Beide Systeme erlauben eine Abbildung von MRT- und CT-Scans auf den Körper des Patienten, welches für ein besseres Verständnis der Anatomie sorgt. Ärzten soll dadurch bei der Diagnose geholfen werden (siehe Abbildung 3).

4.1.2 Verformbare Objekte

Neben der Interaktion mit greifbaren aber starren Objekten kann in Projected Reality auch eine Interaktion mit verformbaren Objekten stattfinden. Dabei kann entweder die Projektion auf die Verformung des Objektes reagieren oder das Objekt sich entsprechend der projizierten Inhalte verformen.

AR Sandbox. Ein Beispiel, bei dem die Projektion auf die Verformung des Objektes reagiert, ist das OpenSource Projekt AR Sandbox. Die Anwendung wurde als Lerntool für Geowissenschaften an Schulen und Universitäten konzipiert. Es bietet die Möglichkeit, die Topologie eines Geländes mit Sand zu modellieren und durch die Projektion in Echtzeit zu sehen, wie sich beispielsweise ein Wasserkreislauf in dem Gelände verhalten würde. Die räumliche

1:8 Interaktionsmöglichkeiten in Projected Reality

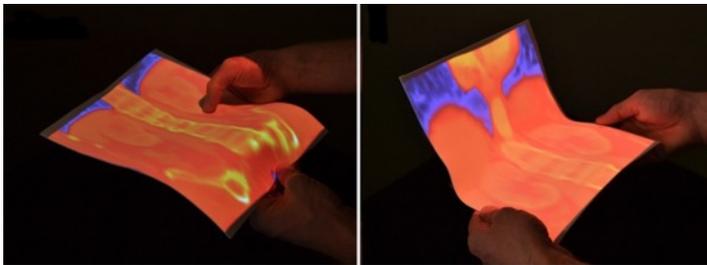


■ **Abbildung 3** Projektion eines volumetrischen Scans auf einen Körper mit ProjectDR [26]

Darstellung sorgt dabei für ein besseres Verständnis von komplexen Vorgängen wie Erosionen und Überflutungen [19, 27]. Um die topografische Karte für das modellierte Gelände zu erzeugen, wird die Oberfläche des Sandes von einer Kinect getrackt. Dafür sendet die Kinect Infrarot Impulse, die von dem Sand reflektiert und der Kamera aufgenommen werden. Dadurch lassen sich die Koordinaten für die Oberfläche ermitteln, aus denen die Geometrie für die Karte berechnet wird. Für die Projektion wird ein einzelner Projektor oberhalb des Sandkastens benötigt [21]. Die AR Sandbox wurde in verschiedenen Geographielaboren getestet und erhielt dabei ein durchweg positives Feedback der Probanden. Studenten fiel es nach eigener Aussage deutlich leichter, mit der AR Sandbox topografische Karten mitsamt der Sedimentierung zu verstehen [27].

Flexpad. Eine weitere Anwendung, bei der der digitale Inhalt der Verformung des Objektes folgt, ist das Flexpad [24]. Das Konzept ist es, durch die Projektion auf ein Blatt Papier oder ein anderes flexibles Material ein komplett deformierbares Display zu erzeugen. Ein vorgestelltes Anwendungsbeispiel ist das Analysieren von volumetrischen Datensätzen. Aus einem CT-Scan können mit dem Flexpad Querschnitte erzeugt werden, indem man das Display (Blattpapier) durch den in den Raum projizierten Scan bewegt (siehe Abbildung 4). Die Verformung des Papiers wird dabei nicht mittels von Markern bestimmt, sondern anhand von virtuell gerenderten Formen verglichen. Das zugrundeliegende Konzept dafür nennt sich Analyse durch Synthese. Dabei werden vorab

definierte Verformungen des Papiers mit dem von der Kinect gescannten Bild abgeglichen. Die gerenderte Form, welche der tatsächlichen Verformung am ähnlichsten ist, bestimmt dabei die Projektion. Um die Verformung möglichst genau bestimmen zu können, werden des Weiteren die Oberflächenbeschaffenheiten analysiert. Dies dient dazu, um zwischen den Händen des Nutzers und der Projektionsoberfläche zu differenzieren. Für die Abbildung des Schnittbildes wird lediglich ein Projektor auf Höhe der Kinect benötigt [24]. Der Vorteil, der sich durch das verformbare Display ergibt, ist, dass auch komplexe Formen wie die Aorta oder der Darm dargestellt werden können. Dies wäre mit einem planaren Display nicht möglich, da die schlauchförmigen Strukturen durch die ganzheitliche Darstellung in einem normalen Querschnitt abgeflacht und somit perspektivisch verzerrt werden. Eine akkurate Abbildung, die Rückschlüsse auf die tatsächliche Anatomie schließen lässt, ist somit nur mit einem verformbaren Display möglich [22]. Ziel ist es, das Flexpad in Zukunft auch noch mit einem Touch Input zu versehen, womit die Interaktionsmöglichkeiten nochmals expressiver werden [24].

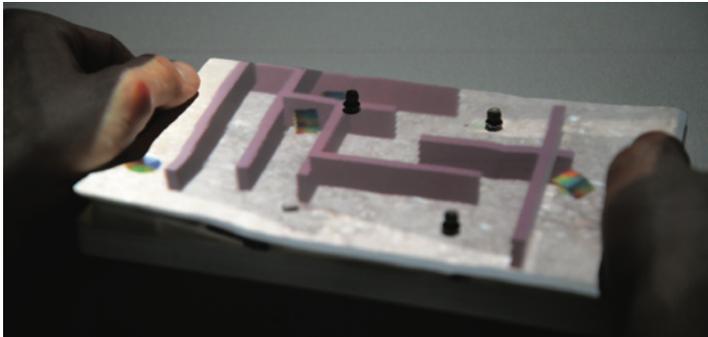


■ **Abbildung 4** Veränderung des Bildausschnitts durch verformen des Flexpads [24]

Shape-Changing Interface. Ein Beispiel, bei dem die Form des Objektes dem dargestellten Inhalt folgt, ist das von David Lindlbauer et al. vorgestellte Konzept, bei dem Spatial Augmented Reality mit einem Shape-Changing Interface kombiniert wird [15]. Ziel war es, die optische und physikalische Erscheinung eines Objektes dynamisch aneinander anpassen zu können, um ein virtuelles 3D-Objekt möglichst realistisch darstellen zu können. Umgesetzt wurde diese unter anderem dadurch, dass die Bump-Maps eines 3D-Objektes physikalisch auf dem Shape-Changing Interface dargestellt wurden. Dafür konnten mittels sechs Servomotoren einzelne Bereiche des Interfaces in der Höhe verstellt werden. Gesteuert werden konnten die einzelnen Motoren per Bluetooth mit einem verbundenen Computer. Das Projektionsmapping basiert, wie bei dem Flexpad, auf einem Abgleich zwischen virtuellen 3D-Objekten und der getrackten Form des Interfaces. Für das Tracking wurde dabei das OptiTrack System eingesetzt, welches mit dem Projektor oberhalb des Interfa-

1:10 Interaktionsmöglichkeiten in Projected Reality

ces positioniert war. Ein Showcase für das Konzept war ein Labyrinth-Spiel, bei dem sich das Interface immer an der Stelle erhöhte, an der sich der Ball befand (siehe Abbildung 5). Bei der Evaluation des Konzeptes gaben Probanden an, dass sie das Spiel mit dem zusätzlichen physischen Reiz sehr realistisch empfanden und wiesen dementsprechend eine hohe Immersion auf [15].



■ **Abbildung 5** Shape-Changing Interface mit zusätzlichen Projektionen [15]

Zwischenresümee. Durch die aufgeführten Arbeiten wird ersichtlich, dass die physischen Interaktionsmöglichkeiten mit Objekten dafür genutzt werden können, um intuitive und haptische Nutzerschnittstellen zu gestalten. Die Interaktion kann dabei durch die Positionierung der Objekte entstehen, wie es bei den Tangible 3D Tabletops der Fall ist, oder aber auch durch die Modellierung des Objektes wie beispielsweise bei der AR Sandbox [6, 19].

Des Weiteren können Projektionen auf Objekte auch dafür nützlich sein, um objektspezifische Informationen für ein leichteres Verständnis in Kontext zu dem Objekt zu setzen. Beispielhafte Anwendungen dafür sind ProjectDR und MoSART [26, 3].

4.2 Interaktion mit Menschen

Neben einer Interaktion mit Objekten kann in PR auch eine Interaktion zwischen Menschen entstehen. Folgende Arbeiten stellen dar, in welcher Form dies geschehen kann.

Mano-a-Mano. Das Mano-a-Mano ist ein System, das eine Interaktion zwischen zwei Personen in PR ermöglicht. Durch den Einsatz von mehreren Projektoren wird dabei jeweils eine Perspektive speziell für eine Person dargestellt. Interagiert eine Person mit einem virtuellen Objekt, kann dies

von seinem Gegenüber mitverfolgt werden, ohne dass dieser dafür eine Brille tragen muss. Ein gemeinsames Erleben von virtuellen Inhalten in der realen Welt wird dadurch möglich. Für die Anwendung wurde auf das bereits in Kapitel 3 vorgestellte RoomAlive Toolkit zurückgegriffen. Um den gesamten Raum tracken und Projektionen darzustellen zu können, wurden drei Projektor-Kinect-Kombinationen verwendet. Jeweils ein Projektor wurde für die Darstellung aus Sicht eines Nutzers benötigt. Die genaue Perspektive wurde dabei über das Headtracking der Kinect bestimmt. Die dritte Projektor-Kinect-Kombination wurde eingesetzt, um den Raum zwischen den Nutzern zu bestrahlen. Mit den Informationen aus allen Kinects wurde in Unity ein virtuelles Abbild des Raumes erzeugt. Dies diente dazu, um die digitalen Inhalte in der virtuellen Szene platzieren zu können. Die zusätzlichen virtuellen Objekte wurden dann über die drei Projektoren in dem realen Raum dargestellt. Da die Projektionen an kein reales Objekt gebunden waren, wurde die Interaktion durch Bodytracking der Nutzer realisiert. Das System ist dabei schon soweit entwickelt, dass sogar Ballfangspiele umgesetzt werden können. Ziel ist es, dass System in Zukunft auch noch in weiteren Bereichen zu testen und Anwendung zu finden [1].

Room2Room. PR kann auch dann Anwendung finden, wenn sich die an der Interaktion beteiligten Personen nicht in demselben Raum befinden. Room2Room ist ein System für Telepresence, das es schafft, durch eine lebensgroße Projektion des Gegenübers das Gefühl einer Face-to-Face Interaktion zu vermitteln. Bei der Evaluierung dieses Systems zeigte sich, dass durch die Projektion sowohl die Physical Presence, als auch das Verständnis für Mimik und Gestik im Vergleich zu normalen Teleconferencing Tools deutlich anstieg. Des Weiteren konnten die kollaborativen Aufgaben, die den Probanden in der Studie gestellt wurden, deutlich schneller gelöst werden. Das Feedback für das System fiel sehr positiv aus und die Probanden gaben Anregungen Room2Room auch in anderen Anwendungsbereichen, wie beispielsweise der Unterhaltungsbranche einzusetzen. Für die Implementierung wurde wieder auf das RoomAlive Toolkit zurückgegriffen, welches noch um die notwendige Netzwerkfunktionalität erweitert wurde. Auch bei diesem Setting wurden mehrere Projektoren und Kinects benutzt, um die Projektionen kontextabhängig abbilden zu können. Dies bedeutet, dass je nachdem ob der Gesprächspartner stand oder saß, die Projektion abhängig davon im Raum platziert wurden. Des Darüber hinaus wurde durch das Headtracking der Kinect ein Parallax Effekt simuliert, um den räumlichen Eindruck nochmals zu verbessern [17]

Zwischenresümee. Das Mano-a-Mano zeigt, dass sich der Einsatz von Projected Reality besonders dann anbietet, wenn ein gemeinsames Erleben

1:12 Interaktionsmöglichkeiten in Projected Reality

von digitalen Inhalten gefördert werden soll, ohne dabei auf zusätzliche Geräte wie beispielsweise VR-Brillen zurückgreifen zu müssen [1].

Des Weiteren kann PR auch bei der Kommunikation von Menschen eingesetzt werden, die sich an verschiedenen Orten befinden. Room2Room zeigt auf, dass sich auch im Bereich der Telekommunikation neue Möglichkeiten durch die Projektionen eröffnen [17].

5 Diskussion

Alle verschiedenen Formen der Realitäten (siehe Kapitel 2) bieten Vor- und auch Nachteile. Dementsprechend ist es anwendungsbedingt, welche Wahl am sinnvollsten ist. Im Folgenden wird diskutiert, welche Vorteile PR im Vergleich zu AR und VR bietet.

5.1 Vorteile

Keine externen Geräte. PR bietet im Vergleich zu VR und AR den großen Vorteil, dass keine zusätzlichen Geräte von den Nutzern getragen werden müssen. Durch die Projektion ergibt sich für den Nutzer auch ein deutlich größeres Sichtfeld und eine höhere Auflösung als es beispielsweise bei AR-Anwendungen mit der HoloLens der Fall ist [18].

Ein mögliches Einsatzgebiet, bei dem die erwähnten Vorteile von Nutzen sein können, sind medizinische Anwendungen. Mit ProjectedDR ist ein erster Prototyp für ein System evaluiert worden, welcher in diesem Gebiet Einsatz findet [26]. Eine PR-Anwendung eignet sich deutlich besser für den Einsatz im Operationssaal als eine AR-Anwendung, bei der auf ein zusätzliches Gerät zurückgegriffen werden muss, welches im Umfeld des Operationssaals eher hinderlich ist.

Kontext Visualisierung. Ein weiterer Vorteil, der sich vor allem auch bei medizinischen Anwendungen bemerkbar macht, ist, dass digitale Inhalte in Relation zu einem realen Objekt gesetzt werden können. Dies ermöglicht, wie am Beispiel von ProjectDR und MoSART bereits beschrieben (vgl. Kapitel 4.1.1), dass CT-Scans direkt auf dem Körper abgebildet werden können. Somit muss der Arzt seinen Blick nicht mehr zwischen Display und Patienten wechseln und gewinnt ein besseres Verständnis für die Anatomie.

Intuitive Interaktion. Außerdem kann eine direkte und intuitive Interaktion stattfinden, ohne dass die Bedienung mit einem Controller erlernt werden muss. Bei der AR Sandbox kann beispielsweise direkt an dem physischen Modell gearbeitet werden und im Falle von Strip’Tic kann sogar der gesamte Arbeitsprozess beibehalten werden, ohne diesen speziell für die PR-Anwendung zu adaptieren [19, 25].

Shared Experience. PR bietet außerdem den Vorzug der Shared Experience. In nahezu jeder der aufgeführten Arbeiten (siehe Kapitel 4) ist eine soziale Komponente enthalten. Auch wenn die Systeme nicht explizit auf eine Interaktion zwischen Menschen abzielen, ist dadurch, dass die digitalen Inhalte in den Kontext der realen Welt gesetzt werden, ein Ansatzpunkt für eine Interaktion gegeben. Das Konzept der AR Sandbox zielte beispielsweise auf eine intuitive Interaktion ab. Durch die Visualisierung in der realen Welt ist die Anwendung aber auch prädestiniert dafür, in Lernumgebungen eingesetzt zu werden [19].

Keine VR-Sickness. Dadurch, dass die tatsächliche Realität der Nutzer durch die Projektionen erweitert wird und sie nicht ein HMD tragen müssen, werden damit einhergehende Probleme wie Tracking und zu niedrige Frameraten umgangen, die zu VR-Sickness führen können.

5.2 Limitationen

Die Intuitivität kann nur auf Kosten der Flexibilität von PR-Anwendungen umgesetzt werden. Da mit realen Objekten interagiert werden kann, bestimmen diese letztlich auch die Anwendung. Das Hardware-Setup von der AR Sandbox ist beispielsweise nur für diese Anwendung ausgelegt und auch die Tracking Methode genau für diese Anwendung angepasst. Im Vergleich dazu können mit einer HTC Vive und einem Computer beliebig viele Anwendungen umgesetzt werden.

Eine weitere Limitation ist, dass die Projektionen, wenn sie nicht auf einer horizontalen Fläche abgebildet werden, immer an die Perspektive des Betrachters angepasst werden müssen. Dies ist besonders wichtig, wenn die Projektionen hologrammartig aussehen sollen. RoomAlive nimmt eine Anpassung über das Headtracking vor, welches allerdings nur für eine einzelne Person umgesetzt werden kann [12]. Dementsprechend führt es immer noch zu Problemen, wenn mehrere Personen von verschiedenen Standpunkten aus eine Projektion betrachten. Room2Room ist aus diesem Grund auch auf jeweils nur eine Person pro Raum limitiert [17].

5.3 Fazit

Zieht man Bilanz zwischen den Vorteilen und Nachteilen kommt man zu dem Schluss, dass sich der Einsatz von PR vor allem dann anbietet, wenn die Anwendungen ohne eine lange Einlernphase bedienbar sein müssen und speziell für einen Zweck konzipiert werden können. Die um Projektionen erweiterten TUIs, wie man sie beispielsweise bei Strip’Tic findet, sind demnach ein prädestiniertes Einsatzgebiet für PR, zumal die aufgeführten Probleme bei der Projektion weitestgehend eingegrenzt werden.

6 Ausblick

Im Vergleich zu VR und AR existieren für PR noch verhältnismäßig wenige kommerzielle Produkte. Ein Beispiel für eines der ersten kommerziellen Geräte ist das Lightform¹. Es bietet neben einer Hardwarelösung, bestehend aus Projektor und Kamera, auch bereits die dazugehörige Software. Das Einsatzgebiet liegt dabei eher in kreativen Bereichen, sodass es bereits Anwendung bei Kunstprojekten und im Interiordesign gefunden hat.

Für Anwendungen in dem Bereich Telepresence könnte des Weiteren das holografische Display von ASKA3D² von Bedeutung sein. Die spezielle holografische Scheibe erlaubt es, in Kombination mit einem herkömmlichen Display Hologramme zu erzeugen. Auf dem asiatischen Markt sind bereits erste Produkte angekündigt, die das holografischen Display nutzen.

Ein Beispiel, für ein noch in der Konzeptionsphase befindliches Gerät, ist das Lumen³. Dieser handgehaltene Projektor soll unter anderem als Zusatz zu technischen Geräten fungieren und beispielsweise Inhalte wie das Plattencover einer gerade abgespielten CD auf die Stereoanlage projizieren. Eine weitere Anwendung sehen die Entwickler in einem Storytelling Tool für Museen, mit dem Statuen durch Animationen zum Leben erweckt werden sollen.

Durch Entwicklungen in der Technologie wie beispielsweise Picoprojektoren werden sich in Zukunft Möglichkeiten zur Herstellung von kompakteren und billigeren Projektoren ergeben, welche die Produktion von neuen Geräten im Bereich PR begünstigen können.

¹ <https://lightform.com>

² <https://aska3d.com/en/>

³ <http://arvindsanjeev.com/lumen.html>

7 Zusammenfassung

Ziel der Arbeit war es, die verschiedenen Interaktionsmöglichkeiten in Projected Reality vorzustellen und zu erläutern. Es wurde ersichtlich, dass sowohl bei der Interaktion mit Menschen als auch bei der Interaktion mit Objekten davon profitiert werden kann, dass virtuelle Inhalte in den Kontext der realen Welt gesetzt werden. Dadurch können sowohl intuitivere Benutzerschnittstellen gestaltet, als auch das Verständnis von digitalen Inhalten verbessert werden.

Als eines der prädestinierten Anwendungsgebiete für PR haben sich dadurch die TUIs ergeben, bei denen sich alle der in Kapitel 5 erläuterten Vorteile bemerkbar machen. Des Weiteren werden die in Abschnitt 5.2 besprochenen technischen Limitationen auf ein Minimum reduziert.

Durch den Ausblick auf die technischen Entwicklungen wird ersichtlich, dass sich nun auch mehr kommerzielle Geräte in Produktion befinden, durch deren Markteinführung die Anzahl an PR-Anwendungen nochmals steigen wird.

Als Fazit lässt sich feststellen, dass PR besonders dann in Betracht gezogen werden sollte, wenn die Interaktion im Mittelpunkt der Anwendung steht. Damit bildet PR mit seinen spezifischen Vorzügen gegenüber AR und VR eine gute Ergänzung zu den anderen Formen der Realitäten.

Literatur

- 1 Hrvoje Benko, Andrew D. Wilson, and Federico Zannier. Dyadic projected spatial augmented reality. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14, pages 645–655, New York, NY, USA, 2014. ACM. doi:10.1145/2642918.2647402.
- 2 Siam Charoenseang and Nemin Suksen. Active location tracking for projected reality using wiimotes. In Randall Shumaker, editor, *Virtual and Mixed Reality - Systems and Applications*, pages 309–317, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.
- 3 Guillaume Cortes, Eric Marchand, Guillaume Brincin, and Anatole Lécuyer. Mosart: Mobile spatial augmented reality for 3d interaction with tangible objects. *Frontiers in Robotics and AI*, 5:93, 2018. doi:10.3389/frobt.2018.00093.
- 4 Karl D. D. Willis. A pre-history of handheld projector-based interaction. *Personal and Ubiquitous Computing*, 16:5–15, 2012. doi:10.1007/s00779-011-0373-5.
- 5 Andrew D. Wilson and Hrvoje Benko. Holograms without headsets: Projected augmented reality with the roomalive toolkit. In *Proceedings of the*

1:16 Interaktionsmöglichkeiten in Projected Reality

- 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '17, pages 425–428. ACM, 2017. doi:10.1145/3027063.3050433.
- 6 Peter Dalsgaard and Kim Halskov. Tangible 3d tabletops: Combining tangible tabletop interaction and 3d projection. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*, NordiCHI '12, pages 109–118, New York, NY, USA, 2012. ACM. doi:10.1145/2399016.2399033.
 - 7 Jochen Ehnes and Michitaka Hirose. Projected reality - content delivery right onto objects of daily life. In Zhigeng Pan, Adrian Cheok, Michael Haller, Rynson W. H. Lau, Hideo Saito, and Ronghua Liang, editors, *Advances in Artificial Reality and Tele-Existence*, pages 262–271, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.
 - 8 Jochen Ehnes and Michitaka Hirose. Projected reality - enhancing projected augmentations by dynamically choosing the best among several projection systems. *IEEE Virtual Reality Conference (VR 2006)*, pages 283–284, 2006. doi:10.1109/VR.2006.111.
 - 9 Markus Funk, Thomas Kosch, Katrin Wolf, Pascal Knierim, Sven Mayer, and Albrecht Schmidt. Automatic projection positioning based on surface suitability. In *Proceedings of the 5th ACM International Symposium on Pervasive Displays*, PerDis '16, pages 75–79, New York, NY, USA, 2016. ACM. doi:10.1145/2914920.2915014.
 - 10 Christophe Hurter, Rémi Lesbordes, Catherine Letondal, Jean-Luc Vinot, and Stéphane Conversy. Strip'tic: Exploring augmented paper strips for air traffic controllers. *Proceedings of the Workshop on Advanced Visual Interfaces AVI*, 2012. doi:10.1145/2254556.2254598.
 - 11 Hiroshi Ishii and Brygg Ullmer. Tangible bits: Towards seamless interfaces between people, bits and atoms. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, CHI '97, pages 234–241, New York, NY, USA, 1997. ACM. doi:10.1145/258549.258715.
 - 12 Brett Jones, Rajinder Sodhi, Michael Murdock, Ravish Mehra, Hrvoje Benko, Andy Wilson, Eyal Ofek, Blair MacIntyre, Nikunj Raghuvanshi, and Lior Shapira. Roomalive: Magical experiences enabled by scalable, adaptive projector-camera units. In *UIST '14 Proceedings of the 27th annual ACM symposium on User interface software and technology*, pages 637–644. ACM, 2014. doi:10.1145/2642918.2647383.
 - 13 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*, CHI '18, pages 433:1–433:6, New York, NY, USA, 2018. ACM. doi:10.1145/3173574.3174007.
 - 14 Jeremy Laviolle, Lauren Thevin, Jeremy Albouys-Perrois, and Anke Brock. Nectar: Multi-user spatial augmented reality for everyone: Three live de-

- monstrations of educative applications. In *Proceedings of the Virtual Reality International Conference - Laval Virtual, VRIC '18*, pages 27:1–27:6, New York, NY, USA, 2018. ACM. doi:10.1145/3234253.3234317.
- 15 David Lindlbauer, Jens Emil Grønbaek, Morten Birk, Kim Halskov, Marc Alexa, and Jörg Müller. Combining shape-changing interfaces and spatial augmented reality enables extended object appearance. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16*, pages 791–802, New York, NY, USA, 2016. ACM. doi:10.1145/2858036.2858457.
 - 16 Steve Mann, Tom Furness, Yu Yuan, Jay Iorio, and Zixin Wang. All reality: Virtual, augmented, mixed (x), mediated (x, y), and multimEDIATED reality. *CoRR*, abs/1804.08386, 2018. <https://arxiv.org/pdf/1804.08386.pdf>.
 - 17 Tomislav Pejša, Julian Kantor, Hrvoje Benko, Eyal Ofek, and Andrew Wilson. Room2room: Enabling life-size telepresence in a projected augmented reality environment. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing, CSCW '16*, pages 1716–1725, New York, NY, USA, 2016. ACM. doi:10.1145/2818048.2819965.
 - 18 Ramesh Raskar, Greg Welch, and Henry Fuchs. Spatially augmented reality. In *Proceedings of the International Workshop on Augmented Reality : Placing Artificial Objects in Real Scenes: Placing Artificial Objects in Real Scenes, IWAR '98*, pages 63–72, Natick, MA, USA, 1999. A. K. Peters, Ltd. URL: <http://dl.acm.org/citation.cfm?id=322690.322696>.
 - 19 Sarah Reed, Sherry Hsi, Oliver Kreylos, M. Burak Yikilmaz, Louise H. Kellogg, S. Geoffrey Schladow, Heather Segale, and Lindsay Chan. An interactive, augmented reality sandbox for advancing earth science education, 2016. <https://doi.org/10.1029/2016E0056135>.
 - 20 Joan Sol Roo and Martin Hachet. Interacting with spatial augmented reality, 2016. <https://hal.archives-ouvertes.fr/hal-01284005/document>.
 - 21 Sergio Álvarez Sánchez, Laura Delgado Martín, Miguel Ángel Gimeno-González, Teresa Martín-García, Fernando Almaraz-Menéndez, and Camilo Ruiz. Augmented reality sandbox: A platform for educative experiences. In *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality, TEEM '16*, pages 599–602, New York, NY, USA, 2016. ACM. doi:10.1145/3012430.3012580.
 - 22 Laurent Saroul, Sebastian Gerlach, and Roger D. Hersch. Exploring curved anatomic structures with surface sections. In *Proceedings of the 14th IEEE Visualization 2003 (VIS'03), VIS '03*, pages 27–34, Washington, DC, USA, 2003. IEEE Computer Society. doi:10.1109/VISUAL.2003.1250351.
 - 23 Roy Shilkrot, Seth Hunter, and Patricia Maes. Pocomo: Projected collaboration using mobile devices. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Ser-*

- vices, MobileHCI '11, pages 333–336, New York, NY, USA, 2011. ACM. doi:10.1145/2037373.2037424.
- 24 Jürgen Steimle, Andreas Jordt, and Pattie Maes. Flexpad: Highly flexible bending interactions for projected handheld displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '13*, pages 237–246, New York, NY, USA, 2013. ACM. doi:10.1145/2470654.2470688.
- 25 Jean-Luc Vinot, Catherine Letondal, Rémi Lesbordes, Stéphane Chatty, Stéphane Conversy, and Christophe Hurter. Tangible augmented reality for air traffic control. *Interactions*, 21:54–57, 2014. doi:10.1145/2627598.
- 26 Ian Watts, Pierre Boulanger, and Greg Kawchuk. Projectdr: Augmented reality system for displaying medical images directly onto a patient. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology, VRST '17*, pages 70:1–70:2, New York, NY, USA, 2017. ACM. URL: <http://doi.acm.org/10.1145/3139131.3141198>, doi:10.1145/3139131.3141198.
- 27 Terri L. Woods, Sarah Reed, Sherry Hsi, John A. Woods, and Michael R. Woods. Pilot study using the augmented reality sandbox to teach topographic maps and surficial processes in introductory geology labs. *Journal of Geoscience Education*, 64(3):199–214, 2016. doi:10.5408/15-135.1.

Augmented Reality-Interaktion im Fremdsprachenunterricht

David Gasser

Ludwig-Maximilians-Universität München, München, Deutschland
david.gasser@campus.lmu.de

Zusammenfassung

Augmented Reality bietet für den Unterricht in der Schule viele neue Möglichkeiten, das Lernen effektiver und spannender zu gestalten. Allerdings wird sie bisher nur selten im Klassenzimmer eingesetzt. Vor allem für das Lernen einer Fremdsprache wird ein lernförderlicher Effekt von Multimedia-Anwendungen erwartet, da multimedial dargebotene Inhalte (z. B. Bilder zu begleitenden Texten) besser im Langzeitgedächtnis behalten werden können. Auf Basis von Theorien des multimedialen Lernens und bereits erarbeiteten Designprinzipien für die Integration von Augmented Reality in den Unterricht wird eine Anwendung für den Fremdsprachenunterricht konzipiert. Mit der Anwendung soll sowohl das Lernen und Üben von Vokabeln als auch deren Einsatz im alltäglichen Sprachgebrauch gefördert werden. Ein weiteres Ziel der Anwendung ist, die Schüler für das Lernen zu motivieren und den Unterricht spannender und abwechslungsreicher zu gestalten. Durch diese Maßnahmen kann ein positiver Effekt auf den Lernerfolg der Schüler angenommen werden. Weitere Möglichkeiten von Augmented Reality für den Unterricht werden im Ausblick diskutiert.

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

Keywords and phrases Augmented Reality; Ubiquitous Computing; Language Learning; Classroom Interaction; Multimedia Learning.

1 Einführung

In der Schule werden Fremdsprachen meist mit Hilfe eines Textbuchs gelehrt. Die Vokabeln werden direkt aus dem Buch oder mit Karteikarten auswendig gelernt. Problematisch dabei ist, dass das eigentlich sehr praktische Wissen, die Sprache tatsächlich zu sprechen, überwiegend theoretisch gelernt und nicht direkt angewandt wird. Die Integration von Multimedia-Anwendungen in den Unterricht bietet an dieser Stelle viele Möglichkeiten, den Unterricht ansprechender und interaktiver zu gestalten. So kann der Einsatz von Multimedia



© David Gasser;

licensed under Creative Commons License CC-BY

Cite as: David Gasser. Augmented Reality-Interaktion im Fremdsprachenunterricht. In *3rd Seminar on Ubiquitous Interaction (UBIACTION 2019)*. Editors: Pascal Knierim, Jakob Karolus, Fiona Draxler, Ville Mäkelä, Tonja Machulla, Florian Lang, Thomas Kosch, Albrecht Schmidt. July 16, 2019. Munich, Germany. pp. 2:1–2:18.

dazu beitragen, die Schüler für das Lernen zu motivieren und entsprechend den Lernerfolg zu steigern [4, 11]. Anwendungen für das Lernen mit moderner Multimedia haben bereits in verschiedenen Formen ihren Weg in die Schule gefunden. So kommen im Unterricht beispielsweise Smartboards zum Einsatz oder Schüler lernen mit Software auf ihrem Tablet [3]. Im Folgenden wird es um ein konkretes Beispiel gehen, wie Multimedialernen im Unterricht gestaltet werden kann: die Augmented Reality (AR). AR bedeutet, dass die reale Welt durch virtuelle Aspekte erweitert wird. Das Lernen einer Fremdsprache kann durch AR noch stärker gefördert werden [9]. Obwohl AR bereits in vielen Bereichen, auch im Unterricht, eingesetzt wird, findet sie im Fremdsprachenunterricht bisher wenig Beachtung. Der Einsatz einer AR-Anwendung im Fremdsprachenunterricht verspricht, dass die Sprache noch effektiver gelernt wird, da sie in einer (teils) virtuellen Realität konkret angewendet werden kann.

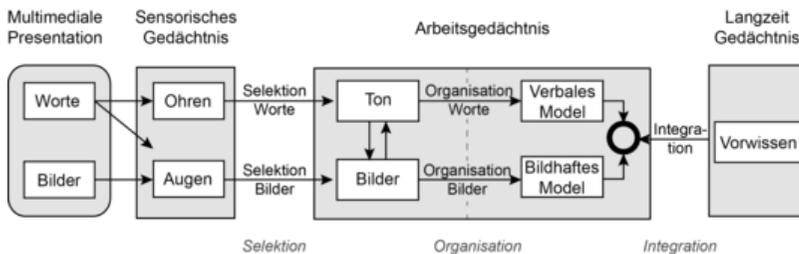
Der vorliegende Beitrag beschäftigt sich daher mit der Frage, wie AR im Fremdsprachenunterricht lernförderlich eingesetzt werden kann. Hierfür werden in einem ersten Schritt die kognitiven Prozesse beim multimedialen Lernen anhand der Theorie von Mayer [13] erläutert und deren Anwendung im Unterricht diskutiert. Im nächsten Teil wird aufgezeigt, welche Möglichkeiten und Vorteile AR-Anwendungen bieten, wie deren Inhalt erstellt wird und wie durch Gamification-Elemente die Motivation des Nutzers gesteigert werden kann. Der letzte Teil der Arbeit befasst sich mit der Integration von AR-Anwendungen im Unterricht. Anhand von Designprinzipien wird abschließend dargestellt, wie eine Anwendung für den Fremdsprachenunterricht konzipiert werden kann.

2 Multimedialernen

Als Multimedia werden verschiedene Objekte, technische Geräte oder Konfigurationen bezeichnet, mit denen sich Informationen kommunizieren lassen [25]. Das trifft beispielsweise auf ein Buch, die Tafel, den Overheadprojektor oder Beamer zu. Multimedialernen bedeutet, dass dem Nutzer für den zu lernenden Inhalt mehrere Darstellungsformate zur Verfügung gestellt werden, die interaktiv genutzt werden können. Die kognitive Theorie des multimedialen Lernens nach Mayer [13] beschäftigt sich damit, welche Sinne von Schülern angesprochen werden sollten, um Lernen durch das Medienangebot zu optimieren. Ein Lernmehrwert ist dieser zufolge dann zu erwarten, wenn das Medienangebot mehrere Sinnesmodalitäten der Schüler gleichzeitig anspricht. Hierzu untersuchten Plass et al. in ihrer Studie das Lernen von Deutsch als Fremdsprache in einer Multimedia-Umgebung [16]. Durch das Anzeigen von Bildern zu Schlagwörtern in einem Text konnten sich die Lernenden besser

an die Übersetzungen erinnern, als in dem Fall, bei dem nur eine der Darstellungsformen gewählt wurde. Im Weiteren wird nun genauer auf die kognitiven Prozesse beim Lernen mit Multimedia eingegangen.

In der kognitiven Theorie des multimedialen Lernens nach Mayer [13] wird zwischen drei Gedächtnisspeichern des Menschen differenziert: der sensorische Speicher, das Arbeitsgedächtnis und das Langzeitgedächtnis (siehe Abb. 1). Informationen der Außenwelt werden in Form von Bildern und geschriebenen oder gesprochenen Wörtern über die Sinneskanäle, die Augen und die Ohren, im sensorischen Speicher wahrgenommen. Im Arbeitsgedächtnis erfolgt die erste Verarbeitung der aufgenommenen Informationen. Hier werden die neuen Inhalte kanalspezifisch zu visuellen bzw. sprachlichen Repräsentationen organisiert und in die bereits bestehende Wissensstruktur (Langzeitgedächtnis) integriert. Diese kognitiven Prozesse werden in der Abbildung 1 grafisch dargestellt.



■ **Abbildung 1** Mayers Theorie zu multimedialem Lernen [1]

Da die Kapazität des Arbeitsgedächtnis begrenzt ist (vgl. Cognitive Load Theory [23]), muss bei der Konzeption von Multimedia-Anwendungen für den Unterricht darauf geachtet werden, dass eine kognitive Überlastung des Nutzers ausgeschlossen ist, z. B. indem auf redundante Informationen und auffällige Gestaltung des Mediums verzichtet wird.

Beim Lernen einer Fremdsprache liegt die visuelle Repräsentation eines Wortes häufig bereits vor und kann mit der neuen Vokabel in der Fremdsprache verknüpft werden. Die von den Sinneskanälen aufgenommenen Informationen (z. B. die neue Vokabel) werden im Arbeitsgedächtnis in das Vorwissen integriert. Das neue Wort wird also dem bereits bekannten Wort in der Muttersprache oder einer visuellen Repräsentation zugeordnet.

Durch „Spaced Repetition“, also die verteilte Wiederholung, wie sie von Ebbinghaus beschrieben wird, werden die gelernten Informationen besser im Gedächtnis behalten [7].

Aufgrund des sogenannten „Spacing Effect“, also dem Lernen von Inhalten über einen längeren Zeitraum, bleiben Inhalte besser im Gedächtnis hängen. Besonders beim Lernen von großen Mengen oder komplexen Themen hilft

2:4 AR im Fremdsprachenunterricht

diese Methode, die Information lange im Gedächtnis zu behalten. Wiederholt man Vokabeln regelmäßig, so können sie beim Gebrauchen einer Sprache eher angewandt werden. Durch den Einsatz von Multimedia-Anwendungen für das Lernen von Vokabeln kann dieser Effekt genutzt werden, indem der Lernende bereits gelernte Themen in bestimmten Abständen wiederholt.

Durch Multimediasysteme und -anwendungen werden diese Lerntheorien im Unterricht umgesetzt. Mit AR entstehen neue Möglichkeiten, das multimediale Lernen noch interaktiver und spannender zu gestalten.

3 Augmented Reality

In der AR wird die reale Welt durch virtuelle Informationen erweitert. Anders als bei der Virtual Reality (VR) sieht der Nutzer seine reale Umgebung. Die AR wird entweder auf dem Display eines Gerätes dargestellt, das durch eine Kamera die Umgebung aufnimmt [17], oder auf eine durchsichtige Oberfläche projiziert, durch die der Nutzer die reale Welt weiter betrachten kann. Auch das Smartphone, das die meisten ständig bei sich tragen, kann als Plattform für AR-Anwendungen dienen. Es existieren zum Beispiel Applikationen, wie SketchAR¹ oder IKEA Place², die AR auf dem Smartphone umsetzen.

AR ermöglicht es dem Nutzer, Informationen zu Objekten aus der Umgebung zu erhalten und mit diesen zu interagieren. Die Umgebung des Nutzers dient dabei als implizierter Input für die AR [17]. AR-Anwendungen werden bereits in vielen Bereichen, wie Medizin oder Ingenieurwesen, erfolgreich eingesetzt. Die Anwendungen erleichtern den Nutzern durch das Darstellen von zusätzlichen Informationen die Interaktion mit der realen Welt [2]. Durch die kognitive Entlastung kann der Nutzer sich besser auf seine primäre Aufgabe, bei dem ihn die AR-Anwendung unterstützt, konzentrieren. Zum Beispiel können AR-Anwendungen Ärzte bei schwierigen Operationen assistieren [22] oder Monteuren Reparaturanleitungen und Bauteile anzeigen [15]. Wie die Inhalte für AR-Anwendungen generiert werden und welche Möglichkeiten sich aus dem Einsatz von AR für das Lernen ergeben, wird im Folgenden erläutert.

3.1 Technische Grundlagen zur Generierung von AR-Inhalten

Die Inhalte von AR-Anwendungen werden auf unterschiedliche Arten erzeugt und dargestellt. Zum einen können marker- oder ortsbezogene Daten verwendet

¹ <https://sketchar.tech/>

² https://play.google.com/store/apps/details?id=com.inter_ikea.place&hl=de

werden, um die Informationen zu erhalten oder Objekte in der Umgebung werden automatisch erkannt und die Informationen dynamisch erstellt.

Die meisten AR-Anwendungen nutzen optische Sensoren, wie z. B. eine Kamera am Gerät, um die AR-Umgebung zu erstellen. Auf Objekten, in Büchern oder an Wänden werden Marker platziert. Diese Art von AR-Anwendung wird als markerbasierte AR bezeichnet. Die Marker, z. B. ein Barcode oder ein Funksignal, werden von den Sensoren erkannt und lösen eine Aktion aus, die in der Anwendung hinterlegt ist, wie z. B. das Darstellen von Informationen, das Abspielen von Sound oder das Zeigen eines Bildes oder Videos [8].

Im European Kitchen Projekt lernen Teilnehmer mit einer AR-Anwendung das Kochen in einer Fremdsprache [21]. Die Küchenutensilien, Zutaten und Küchengeräte sind mit Markern ausgestattet. Die Anwendung erkennt die Gegenstände, die die Teilnehmer verwenden und gibt über ein Tablet Feedback in der ausgewählten Fremdsprache. Auch das HELLO Projekt verwendet markerbasierte AR zum Fremdsprachenlernen [12]. Objekte in einer Schule, die mit einem QR-Code versehen sind, lösen Dialoge mit einem virtuellen Tutor aus, die die Schüler absolvieren müssen.

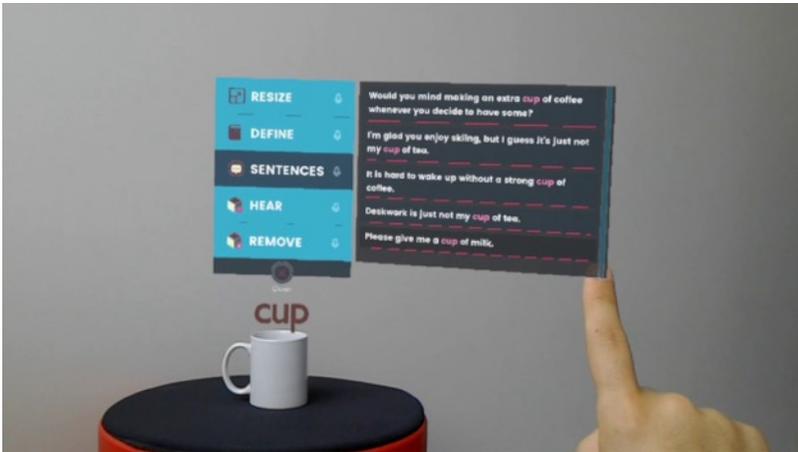
Neben markerbasierter AR können die Inhalte für eine AR-Umgebung auch ortsbasiert erstellt werden. Im Jahr 2016 gab es einen großen Hype um das mobile-AR Spiel Pokemon GO. In dem Spiel kann der User basierend auf seiner Position Aktionen ausführen. So muss sich der Nutzer beispielsweise für das Fangen von Pokemon an einem gewissen Ort aufhalten. Diese Art von AR nennt man ortsbasierte AR, da die Aktionen in der Anwendung über den Ort ausgelöst werden. Der Ort wird z. B. über die GPS-Position oder über Bewegungssensoren erkannt. Oft werden auf diese Art virtuelle Tour-Guides umgesetzt, in denen der Nutzer an unterschiedliche Orte der Welt reisen kann und Informationen erhält.

Ein weiteres Beispiel für eine ortsbasierte AR ist die Anwendung Mentira, ein AR-Spiel zum Spanisch lernen. In dem Spiel muss ein Mordfall auf Spanisch aufgeklärt werden. Dafür muss der Spieler sich an verschiedene Orte begeben und mit virtuellen Charakteren Dialoge führen und Texte lesen, um den Mord aufzuklären. Dabei sind die Aufgaben, die der Nutzer erfüllen muss, abhängig von dem Ort, an dem er sich gerade befindet. [20]

Für diese Art der AR-Anwendungen müssen alle Gegenstände, die verwendet werden, vorher definiert und markiert werden oder die genauen Positionen für Ereignisse hinterlegt werden. Dies begrenzt den Umfang der Anwendungen und das Erweitern ist mit dem Aufwand verbunden, neue Marker zu platzieren oder neue Orte in der Anwendung zu speichern. Die Informationen für die Anwendung können aber auch dynamisch erstellt werden, ohne dass die Gegenstände vorher definiert werden müssen. Ein Beispiel hierfür liefern Vazques et al. mit ihrer Anwendung WordSense [24], welche den Inhalt für die

2:6 AR im Fremdsprachenunterricht

Fremdsprachen-Anwendung dynamisch generiert. Die Applikation generiert neue Lernsituationen für das „zufällige Sprachenlernen in realen Alltagssituationen“ [24]. Dem Nutzer stehen für ein Objekt verschiedene Inhalte zur Verfügung (siehe Abbildung 2).



■ **Abbildung 2** In der AR-Anwendung WordSense von Vazquez werden dem Nutzer für Objekte in seiner Umgebung sowohl die Übersetzung als auch die Definition, Beispielsätze und Audioclips dargestellt [24].

Mithilfe der Google Vision API kann die Kamera des AR-Device Objekte in der Umgebung des Nutzers erkennen. Für das erkannte Objekt wird das Wort daneben in der Fremdsprache in der AR dargestellt. Die Applikation funktioniert auch in die entgegengesetzte Richtung: Wird ein geschriebenes Wort erkannt, so wird ein Bild für den Gegenstand eingeblendet. Für das Wort können dann die Definitionen und Beispielsätze angeschaut werden, die aus einer Datenbank für Beispielsätze stammen. Der Gebrauch und die Aussprache des Wortes können anhand eines kurzen Videoclips angeschaut bzw. über eine Audiodatei angehört werden. Diese Beispielclips und Audiospuren werden aus Filmen übernommen und oberhalb des Objektes eingeblendet. Diese Anwendung erlaubt dem Nutzer in seiner realen Umgebung zu lernen und dadurch die Assoziation zwischen Objekten oder Situation und dem Lernmaterial herzustellen. Zusätzlich ist diese Art von Applikation nicht durch die Marker begrenzt. Es können immer neue Links hergestellt werden und dadurch die Neugierde und die Aufmerksamkeit des Nutzers aufrecht erhalten werden.

3.2 Integration von Gamification-Elementen

Viele AR-Anwendungen machen Gebrauch von Gamification-Elementen aus dem Spieldesign. Durch deren Integration soll die Motivation des Nutzers gesteigert werden. Gerade bei monotonen Inhalten (z. B. Vokabeln lernen) oder sehr komplexen Themen können mit Gamification-Elementen die Lernfrequenz und die Begeisterung des Lernenden gesteigert werden. Khaleel et al. haben in ihrer Arbeit Gamification Elements für Lernapplikationen untersucht, welche Elemente eine positive Auswirkung auf das Lernen haben [11]:

- Der **Erfolg** wird z. B. durch ein Punktsystem oder die Vergabe von Sternen dargestellt. Der Nutzer sieht, wie gut er die Aufgaben gemeistert hat.
- Für das erfolgreiche Absolvieren einer Aufgabe erhält der Nutzer eine **virtuelle Belohnung** oder ein Abzeichen.
- Mittels einer **Rangliste** kann der Nutzer sehen, wie gut er im Vergleich zu anderen Nutzern abgeschlossen hat.
- Der Nutzer hat einen **sichtbaren Status** um zu sehen, wie weit er in seinen Aufgaben schon fortgeschritten ist.
- Durch **Echtzeit** wird der Nutzer verleitet, schnell Aufgaben zu lösen um ein neues Level oder die nächste Stufe im Spiel zu erreichen.
- Nutzer können **Profile** anlegen, in denen sie Informationen über ihre Person bereitstellen, Profilbilder teilen oder einen Avatar erstellen.

Diese Elemente können für das Lernen einer Fremdsprache einfach in eine AR-Anwendungen integriert werden. Die Schüler werden durch die einzelnen Elemente motiviert, regelmäßig die Anwendung zu nutzen und auch die Aufgaben möglichst gut abzuschließen.

3.3 Vorteile und Möglichkeiten von AR für das Lernen von Fremdsprachen

Es wurde bereits an einigen Beispielen gezeigt, wie AR-Anwendungen für das Lernen genutzt werden. Auch für das Lernen einer Fremdsprache bieten sich durch AR viele neue Möglichkeiten. Der Lernende bleibt in der AR in seinem realen und sozialen Umfeld. Die AR führt den Nutzer durch „partizipative und metakognitive Lernprozesse“ [14]. Das bedeutet, dass die Lernenden in der AR ihren Lernprozess selbst mitgestalten und ihre Ideen mit einbringen können (partizipativer Aspekt) und die Möglichkeit haben, ihr eigenes Lernen zu steuern und zu überwachen (metakognitiver Aspekt). Das AR-Interface ermöglicht dabei die „nahtlose Interaktion zwischen der realen und der virtuellen Welt“ [2].

Der kurze Weg zwischen realer und virtueller Welt begünstigen auch den Transfer der gelernten Inhalte auf die Praxis. In der Anwendung WordSense

(siehe Kapitel 3.1) beispielsweise werden beim Lernen der Vokabeln die Wörter in der Fremdsprache mit realen visuellen Darstellungen verknüpft. Dadurch wird den Lernenden bereits beim Lernen der Wörter ein direkter Bezug zur Anwendung in der Sprache gegeben.

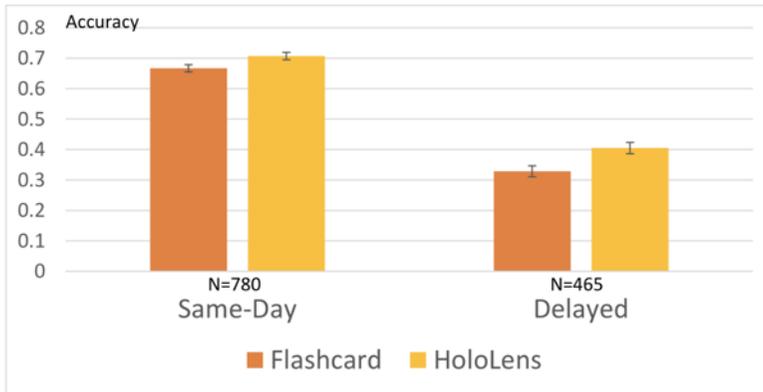
Die AR ermöglicht es zudem, in einem sicheren Rahmen neue Dinge auszuprobieren und Fehler ohne schwerwiegende Folgen zu machen. Es können sorglos neue und unterschiedliche Sachen ausprobiert werden, da jede Aktion rückgängig gemacht werden kann. Außerdem kann die AR-Umgebung dem Nutzer sofort persönliches Feedback geben, wie die zu lösende Aufgabe abgeschlossen wurde. Die Lernumgebung in AR kann bereits im Vorfeld so geplant werden, dass sie den Nutzer zum Ausprobieren ermutigt, mögliche Fehler berücksichtigt oder zu deren Vermeidung beiträgt.

Für das Lernen von Fremdsprachen ist das Anwenden der Sprache ein wichtiger Aspekt. In einer AR-Umgebung kann der Lernende die direkte Kommunikation üben und erkennen, wie die jeweilige Sprache funktioniert. Die Anwendung kann dem Nutzer sofort Feedback geben. Fehler im Sprachgebrauch bleiben ohne schwerwiegende Folgen, aber der Nutzer kann erkennen, wie der Konversationspartner reagiert. Auch wenn der Lernende Hilfe benötigt, zum Beispiel wenn ihm Wörter nicht einfallen, kann die AR diese sofort bereitstellen und unterstützen.

Ein wichtiger Aspekt beim Gestalten einer Lernumgebung ist das „Fördern von sozialer Interaktion von Nutzern, die sich im selben Raum befinden“ [18]. Durch den Einsatz von AR können mehrere Nutzer, die sich zusammen in einem realen Raum befinden, gemeinsam mit den selben virtuellen Objekten interagieren. Der Lernprozess wird durch die natürliche Interaktion, über Sprache und Gestik, gefördert [10].

Einige AR-Anwendungen für das Fremdsprachenlernen wurden bereits wissenschaftlich begleitet. Ibrahim et al. überprüfen beispielsweise in ihrer Studie den Effekt von Multimedia-Anwendungen zum Lernen von Vokabeln in einer Fremdsprache. Mit der AR-Anwendung „ARbis Pictus“ werden Objekte in der Umgebung mit den entsprechenden Wörtern markiert. Diese Lernmethode vergleichen sie mit dem herkömmlichen Lernen mit Karteikarten. Die Teilnehmer, die mit der AR lernten, erzielten ein signifikant besseres Ergebnis im Wiederholen der Wörter als die Teilnehmer, die auf traditionelle Art lernten. Beim Abfragen der Wörter wussten die AR-Teilnehmer am gleichen Tag 7% mehr der Wörter als die Karteikarten-Teilnehmer. Vier Tage später konnten die Teilnehmer sogar 21% mehr der Wörter wiederholen, wie in Abbildung 3 gezeigt wird. Außerdem beschreiben die Teilnehmer, dass die Erfahrung mit AR zu lernen sowohl effektiver als auch angenehmer war, als das Lernen mit Karteikarten.

Auch Santos et al. zeigen, dass durch den Einsatz von Multimedia neue



■ **Abbildung 3** Durch AR (heller Balken) erzielen Teilnehmer der Studie von Ibrahim et al. ein besseres Ergebnis beim Abfragen der Wörter, als mit Karteikarten (dunkler Balken). Der linke Teil zeigt die Auswirkung am selben Tag, der Rechte vier Tage später [9].

Lernerfahrungen gemacht werden können [19]. Sie zeigen anhand einer AR-Anwendung, dass sowohl die Motivation der Schüler zu Lernen und die Aufmerksamkeit für das Thema gesteigert werden. Zusätzlich können sie sich die gelernten Vokabeln über einen längeren Zeitraum merken.

4 Integration von AR in den Fremdsprachenunterricht

Die Integration von AR-Anwendungen im Unterricht kann den Lerneffekt und die Bereitschaft zum Lernen bei Schülern steigern. Diese positiven Effekte können durch die Integration von AR in den Unterricht zum Lernerfolg beitragen. Bisher wurden nur Anwendungen erklärt, die zwar zum Lernen von Fremdsprachen angewandt werden, nicht aber im Unterricht. Für den Einsatz von AR in der Schule müssen beim Design bestimmte Aspekte beachtet werden, um einen reibungslosen Ablauf zu garantieren und damit die Schüler den intendierten Lernstoff lernen.

4.1 Designprinzipien für die Integration von AR in den Unterricht

Damit eine AR-Anwendung in den Unterricht integriert werden kann, muss die Interaktion zwischen dem Nutzer und dem System reibungslos funktionieren. In der Mensch-Computer-Interaktion wird das als die Usability definiert. Die Usability einer Unterrichts-anwendung kann auf drei Ebenen gemessen

2:10 AR im Fremdsprachenunterricht

werden [5, 6]. Auf der ersten Ebene kann durch die Performanz, die Anzahl der Fehler oder den Cognitive Load gemessen werden, wie gut ein System funktioniert. Interagieren mehrere Nutzer mit dem System, so kann auf der zweiten Ebene die Qualität der Konversation zwischen den Nutzern oder wie häufig der Interagierende innerhalb der Gruppe wechselt, gemessen werden. Die dritte Ebene wird durch den Orchestration Load, dem Einsatz den ein Lehrkraft aufwenden muss, um verschiedene Aktivitäten und Lernprozesse im Unterricht zu koordinieren, definiert. Weiterhin definieren Cuedent et al. fünf Designprinzipien, um den Aufwand der Lehrkraft zu reduzieren, wenn eine AR-Interaktion im Unterricht eingesetzt wird:

- **Integration:** Die Lernumgebung wird in den Unterrichtsablauf integriert.
- **Empowerment:** Die Lehrkraft nimmt dann eine aktive Rolle im Unterrichtsgeschehen ein, wenn es erforderlich ist.
- **Awareness:** Die Lernumgebung informiert die Lehrkraft über den Status und Fortschritt der Lernenden.
- **Flexibilität:** Der Lehrkraft muss die Möglichkeit gegeben sein, flexibel auf unerwartete Ereignisse im Lernprozess zu reagieren und den Ablauf entsprechend anzupassen.
- **Minimalismus:** Die Lernumgebung stellt nur die Informationen und Funktionen zur Verfügung, die zu einer bestimmten Zeit benötigt werden.

Diese Designprinzipien wurden von Cuedent et al. in Form eines Tangible User Interface (Die Interaktion mit dem System erfolgt über physische Gegenstände) für den Unterricht umgesetzt. Die Anwendung TinkerLamp (siehe Abbildung 4) ist für Logistik-Auszubildende konzipiert. Sie wird dafür eingesetzt, Konzepte der Lagerung und des Transports von Waren zu lehren und anzuwenden. Das System besteht aus einer Kamera und einem Projektor, der das Szenario auf den Tisch projiziert. Mit kleinen Miniaturregalen interagieren die Auszubildenden mit den virtuellen Lagern. Der Lehrer bekommt so jederzeit einen Überblick, was die Lernenden machen, wie viele Regale sie platziert haben und wie weit sie mit ihren Aufgaben fortgeschritten sind, auch wenn er einige Meter entfernt steht. Dadurch wird die Awareness im System umgesetzt. Die Schüler bekommen die Aufgaben, die sie bearbeiten müssen, in Form von TinkerSheets, die unter dem Projektor auf dem Tisch platziert werden. Das System trackt das Arbeitsblatt in Echtzeit, setzt Parameter, wie die Größe des Raumes oder der verwendeten Maschinen, und gibt den Schülern Feedback. Die TinkerSheets setzen sowohl das Prinzip der Integration um, da die Arbeitsblätter dem Erscheinungsbild und Inhalt des Textbuchs entspricht, das für den Unterricht verwendet wird, als auch das Prinzip des Minimalismus, da das System keinen Login erfordert, sondern hierfür die Arbeitsblätter dienen. Der Lehrer kann den einzelnen Gruppen

kleine Papierkärtchen, die TinkerKeys, austeilen, falls er intervenieren möchte, wenn zum Beispiel eine Gruppe zu schnell ist oder er die Aufmerksamkeit der ganzen Klasse haben möchte. Die Karten machen das System flexibel, da der Lehrer die Gruppen unterschiedlich verfahren kann. Zum Beispiel kann er einer schnellen Gruppe eine Karte geben, um die Aufgaben komplexer zu machen, aber gleichzeitig eine langsamere Gruppe unterstützen. Zusätzlich gibt es eine Karte, bei der sämtliches Feedback in der Klasse ausgeblendet wird. Das entspricht dem Prinzip des Empowerment, da der Lehrer schnell die Aufmerksamkeit aller Auszubildenden erreicht. Eine weitere Komponente des Systems ist das TinkerBoard, ein Display im Klassenzimmer, das den Fortschritt der einzelnen Gruppen anzeigt. Zusätzlich kann der Lehrer den Bildschirm nutzen, um die Aufgaben nachzubesprechen. Durch das TinkerBoard wird sowohl die Awareness als auch die Flexibilität umgesetzt.



■ **Abbildung 4** Umsetzung der Designprinzipien anhand der TinkerLamp. Links das Setup mit Projektion des Lagerraums und Miniaturregalen, rechts ein Tinkersheet, das als Aufgabenblatt dient [5]

4.2 Konzeptualisierung einer AR-Anwendung für den Fremdsprachenunterricht

In diesem Beitrag wurde bisher das Lernen mit Multimedia und die Möglichkeiten von AR für den Unterricht behandelt. Auf Grundlage der bisherigen Aspekte zu diesen Themen soll ein Konzept für eine AR-Anwendung präsentiert werden, die im Fremdsprachenunterricht angewandt werden kann. Dieses Konzept orientiert sich an den gegebenen Beispielen WordSense und TinkerLamp (siehe Kapitel 3.1). Die Anwendung bezieht sich auf keine spezifische Sprache, sondern soll als allgemeiner Wegweiser für die Konzeption einer solchen Applikation für das Sprachenlernen im Allgemeinen dienen. Der Entwurf erfolgt in zwei Schritten. In einem ersten Schritt wird der Lerninhalt festgelegt,

2:12 AR im Fremdsprachenunterricht

der durch die Anwendung abgedeckt werden soll. Im zweiten Schritt werden die Komponenten der Anwendung und deren Funktionen herausgearbeitet. Anschließend wird anhand der Designprinzipien von Cuendet et al. überprüft, ob das Konzept der Anwendung im Unterricht eingesetzt werden kann.

4.2.1 Lerninhalt der AR-Anwendung

Fremdsprachenunterricht besteht aus mehreren Teilen, dem theoretischen Wissen und dem Anwendungswissen. In der Anwendung sollen das Vokabellernen und das Sprechen der Sprache umgesetzt werden. Die AR-Anwendung soll zusätzlich zum Lehrbuch eingesetzt werden. Es orientiert sich an den Kapiteln des Buches und beinhaltet den gleichen Lerninhalt. In jedem Kapitel werden neue Vokalen gelehrt. Den Schülern sollen nur die Vokabeln zur Verfügung stehen, die für das entsprechende Kapitel relevant sind. Um den Sprachgebrauch zu üben und die gelernten Vokabeln einzusetzen, können in der Anwendung Übungsaufgaben bearbeitet werden. Die Aufgaben können entweder aus dem Lehrbuch entnommen werden oder es wird weiteres Übungsmaterial von der Anwendung bereitgestellt. Zusätzlich sollen Themen und Vokabeln, die sonst in Aufgaben des Lehrbuchs vertieft werden, im Rahmen von interaktiven Dialogen und Spielen geübt werden.

4.2.2 Komponenten und Funktionen der AR-Anwendung

Die AR-Anwendung besteht aus zwei Teilen. Zum einen unterstützt sie die Schüler beim Lernen und Wiederholen von Vokabeln, zum anderen lehrt sie dem Schüler das Sprechen in der Sprache und den Sprachgebrauch. Im Folgenden wird auf wichtige Komponenten und Funktionen einer möglichen AR-Anwendungen für den Sprachunterricht eingegangen.

Die Anwendung kann entweder für das Smartphone oder ein Head-Mounted Display (HMD) entwickelt werden. Der Vorteil der mobilen AR ist, dass die meisten Schüler bereits in den unteren Jahrgangsstufen Smartphones besitzen und täglich nutzen. Die Schüler haben die Möglichkeit, die Anwendung auch außerhalb des Unterrichts weiter zu nutzen. Allerdings ist die Immersion auf dem HMD im Gegensatz zum Smartphone höher. Die Inhalte wirken realer in die Umgebung des Nutzers integriert und Lernende können besser in die AR-Szenarien eintauchen. Aktuelle AR-Brillen, wie die HoloLens, sind in der Anschaffung sehr teuer und benötigen in der Regel zusätzlich einen PC, was den Einsatz in größeren Gruppen limitiert. Aufgrund dieser Einschränkungen empfiehlt sich eine mobile AR-Applikation für das Smartphone.

Die Lehrkraft hat die Rolle, zu bearbeitende Arbeitsblätter zu dem Stoff aus dem aktuellen Kapitel bereitzustellen. Das AR-Device erkennt den Lerninhalt und erstellt daraus den Inhalt für die AR-Umgebung. Das Lernen von

Vokabeln erfolgt wie in der Anwendung WordSense (siehe Kapitel 3.1), die vor allem das Lernen neuer Vokabeln im Unterricht unterstützt. Auch bereits gelernte Vokabeln können mit Hilfe der Anwendung wiederholt oder wieder ins Gedächtnis gerufen werden.

Darüber hinaus kann die Anwendung auch als Vokabeltrainer außerhalb des Unterrichts eingesetzt werden. Für neue Vokabeln aus dem Buch werden visuelle Darstellungen bereitgestellt und das Wiederholen und Lernen von Vokabeln wird multimedial unterstützt.

In jedem Kapitel des Lehrbuchs sollen interaktive Übungsaufgaben und Dialoge eingebaut werden. Diese können entweder anhand von virtuellen Gesprächspartnern oder in Kollaboration mit Mitschülern stattfinden. Vergleichbar ist dies mit der Anwendung Mentira oder dem HELLO Projekt (siehe Kapitel 3.1). Anstatt die Aufgaben aus dem Lehrbuch auf dem Papier zu erledigen, können diese in Form von Spielen gelöst werden. Dies erfordert, dass die Aufgaben–insbesondere Aufgaben zu komplexen Sachverhalten–im Vorfeld definiert und in der AR gespeichert wurden. Die AR-Anwendung kann den Lernenden direktes Feedback auf die Bearbeitung der Aufgaben geben und weitere Inhalte für Dialoge generieren.

Zudem ermöglicht es die AR-Anwendung der Lehrkraft, sich einen Überblick zu verschaffen, wie weit die Schüler bereits fortgeschritten sind und wie gut sie ihre Aufgaben abgeschlossen haben. Auf einem Bildschirm, zum Beispiel einem Tablet oder einem Public Display im Raum, werden Status und Fortschritt der einzelnen Schüler angezeigt. So können auch die Schüler selbst sehen, wie weit sie im Vergleich zu den anderen bereits sind. In diesem Zusammenhang kann der Lehrer kann durch Aktionskarten, die von der AR-Anwendung erkannt werden, einschreiten. Ähnlich wie bei der TinkerLamp kann er so schnelle Schüler durch weitere Aufgaben beschäftigen, bestimmte Aufgaben erst nach dem Erreichen eines Lernziels erlauben oder die Aufmerksamkeit auf sich lenken, indem er alle AR Geräte sperrt und auf sich verweist.

4.2.3 Integration der AR-Anwendung in den Unterricht

Anhand der Designprinzipien von Cuendert et al. (siehe Kapitel 4.1) wird überprüft, ob sich das Konzept der Anwendung für die Integration in den Unterricht eignet und ob es die Theorien von Mayer und Ebbinghaus erfüllt.

- **Integration:** Die Anwendung richtet sich genau nach dem Lehrplan und dem Lehrbuch der Jahrgangsstufe. Außerdem werden zum Erstellen der Inhalte der AR Informationen aus dem Buch benötigt oder der Lehrer teilt Arbeitsblätter mit dem Lernmaterial aus. Die Integration der Anwendung in den Unterricht ist demnach einfach möglich.

2:14 AR im Fremdsprachenunterricht

- **Empowerment:** Der Lehrer kann zu jederzeit die Anwendung bei den Schülern sperren und die Aufmerksamkeit auf sich ziehen. Außerdem behält der Lehrer eine zentrale Rolle im Klassenzimmer, da er, wie bereits beschrieben, Aufgaben vorgeben oder intervenieren kann.
- **Awareness:** Der Lehrer hat jederzeit einen Überblick, wie weit ein Schüler im Prozess schon fortgeschritten ist. Dieses Prinzip wird aufgrund der Verwendung von Smartphones eingeschränkt, die für die Schüler ein hohes Ablenkungspotenzial bieten. Das Dashboard, das dem Lehrer zur Verfügung steht, muss also auch anzeigen können, ob ein Schüler sich in der Applikation befindet.
- **Flexibilität:** Die Anwendung ist flexibel, da der Lehrer den Ablauf der Anwendung steuern und schnellere Schüler durch Extraaufgaben individuell fördern kann, während er die langsameren unterstützt. Der Lerninhalt wird nicht von der Anwendung vorgegeben, sondern kann vom Lehrer bestimmt werden. So kann der Inhalt der Unterrichtsstunden am Lernprozess der Schüler ausgerichtet werden.
- **Minimalismus:** Die Schüler müssen sich nicht anmelden, um die Anwendung zu verwenden. Außerdem haben sie immer nur den Stoff zur Verfügung, den sie gerade benötigen. Das Abspielen der Applikation auf dem Smartphone würde den Einsatz von zusätzlicher Hardware vermeiden.

In dem Konzept der Anwendung für das Fremdsprachenlernen wird das multimediale Lernen nach Mayer (siehe Kapitel 2) umgesetzt. Damit Schüler die Vokabeln, die sie lernen, besser im Gedächtnis behalten, erhalten sie zusätzlich zum geschriebenen Wort eine visuelle Darstellung. Auch das Sprechen der Sprache wird nicht nur auf dem Papier gelernt sondern mit virtuellen Gesprächspartnern geübt. Die Anwendung unterstützt auch die Spaced Repetition nach Ebbinghaus. Der Schüler bekommt die Möglichkeit, Vokabeln in neuen Situationen zu wiederholen. Dafür muss er nicht das Buch oder die Karteikarten lernen, sondern kann im alltäglichen Umfeld bereits gelernte Worte zurück ins Gedächtnis holen. Durch Gamification-Elemente in der Anwendung, wie dem Fortschritt oder das Erhalten von Punkten, steigt die Motivation bei den Schülern, mit der Anwendung die Fremdsprache zu lernen.

5 Diskussion und Ausblick

In dieser Arbeit wurde ein Konzept für eine AR-Anwendung für den Fremdsprachenunterricht entworfen. Bisherige Ansätze zur Integration von AR-Anwendungen zeigen, dass AR viele neue Chancen für den Fremdsprachenunterricht bietet. Die Schüler lernen neue Inhalte multimedial, während sie in ihrem realen Umfeld bleiben. Durch die interaktive Anwendung können Schüler

das gelernte theoretische Wissen direkt in der Praxis anwenden. Darüber hinaus können mit AR kollaborative Unterrichtsmethoden realisiert werden. Das in diesem Beitrag vorgestellte Konzept für eine AR-Anwendung basiert auf den Design-Prinzipien von Cuendet. Allerdings ist die konkrete Umsetzung des Konzeptes auch mit Herausforderungen verknüpft. So brauchen die Lehrkräfte sowohl technische als auch (medien-)didaktische Fähigkeiten und Fertigkeiten, um die AR-Anwendung lernförderlich ins Unterrichtsgeschehen einzubinden. Weiterhin hängt die Umsetzung von der technischen Ausstattung der Schule ab. Die Anschaffung von AR-Brillen ist sehr teuer und in großen Mengen für viele Schulen nicht möglich. Auch bei einer Anwendung für das Smartphone muss mit großen Unterschieden gerechnet werden. Die Geräte unterscheiden sich sowohl in ihrem Betriebssystem als auch der Leistung, die sie erbringen können. Möglich ist auch, dass einige Schüler kein Smartphone besitzen. Außerdem kann das Smartphone als Plattform für eine AR-Anwendung zu einem Störfaktor im Unterricht werden, falls die Schüler sich mit anderweitigen Funktionen des Gerätes beschäftigen. Eine weitere Herausforderung für die erfolgreiche Integration von AR in den Unterricht ist die Vereinbarkeit mit dem Lehrplan und der Unterrichtsstruktur.

Zusammenfassend lässt sich sagen, dass das Lernen mit AR, als ein Ansatz für das multimediale Lernen im Unterricht, viel Potential birgt, den Lernerfolg bei Schülern zu steigern. In bereits vorgestellten Studien wurde gezeigt, dass das mit AR-Anwendungen erlernte Wissen nachhaltiger im Gedächtnis behalten werden konnte. Durch den zusätzlichen Einsatz von Gamification-Elementen kann zudem die Motivation bei Schülern gesteigert werden. In zukünftiger Forschung gilt zu untersuchen, ob diese Effekte beim Einsatz von AR im Unterricht sich auch langfristig zeigen. Außerdem muss überprüft werden, in welchen Jahrgangsstufen eine AR-Anwendung im Unterricht sinnvoll ist. Das umfasst sowohl, ob in unteren Klassen, also Schüler, die neu mit einer Fremdsprache starten, durch die AR gefördert werden, als auch in fortgeschrittenen Klassen, in denen der Fokus des Unterrichts häufig nicht mehr auf dem Lernen der Sprache liegt, sondern dem Lesen und Interpretieren von Texten.

AR hat für den Fremdsprachenunterricht das große Potential, Fremdsprachen erfolgreicher und nachhaltiger zu lernen. In Zukunft wird AR viele weitere Möglichkeiten für den Einsatz und das Lernen in der Schule eröffnen und den Unterricht effektiver und spannender gestalten.

Literatur

- 1 Theorie zum multimedialen lernen nach mayer. https://www.didaktik.physik.uni-muenchen.de/multimedia/lernen_mit_multimedia/psycho_theo/multimedia_mayer/index.html. Accessed: 2019-08-12.
- 2 Mark Billinghurst. Augmented reality in education. 2002.

- 3 Wilfried Bos, Ramona Lorenz, Manuela Endberg, Birgit Eickelmann, Rudolf Kammerl, and Stefan Welling. *Schule digital - der Länderindikator 2016: Kompetenzen von Lehrpersonen der Sekundarstufe I im Umgang mit digitalen Medien im Bundesländervergleich*. 2016.
- 4 Dorothy M. Chun and Jan L. Plass. Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal*, 80(2):183–198, 1996.
- 5 Sébastien Cuendet, Quentin Bonnard, Son Do-Lenh, and Pierre Dillenbourg. Designing augmented reality for the classroom. *Computers & Education*, 68:557–569, 2013.
- 6 Pierre Dillenbourg, Guillaume Zufferey, Hamed Alavi, Patrick Jermann, Son Do-Lenh, Quentin Bonnard, Sébastien Cuendet, Frédéric Kaplan, Pierre Dillenbourg, Patrick Jermann, and Sebastien Cuendet. Classroom orchestration: The third circle of usability. 2011.
- 7 Hermann Ebbinghaus. *Our knowledge concerning memory.*, pages 1–6. Teachers College Press, 1913.
- 8 Robert Godwin-Jones. AUGMENTED REALITY AND LANGUAGE LEARNING: From annotated vocabulary to place-based mobile games. *Language Learning*, 2016.
- 9 Adam Ibrahim, Brandon Huynh, Jonathan Downey, Tobias Höllerer, Dorothy Chun, and John O'Donovan. ARbis pictus: A study of language learning with augmented reality. *arXiv:1711.11243 [cs]*, 2017. [arXiv:1711.11243](https://arxiv.org/abs/1711.11243).
- 10 Hannes Kaufmann. Collaborative augmented reality in education. 2003.
- 11 Firas Layth Khaleel, Noraidah Sahari@Ashaari, Tengku Siti Meriam Tengku Wook, and Amirah Ismail. Gamification elements for learning applications. *International Journal on Advanced Science, Engineering and Information Technology*, 6(6), 2016.
- 12 Yang Liu, Daniel Holden, and Dongping Zheng. Analyzing students' language learning experience in an augmented reality mobile game: An exploration of an emergent learning environment. *Procedia - Social and Behavioral Sciences*, 228:369–374, 2016.
- 13 R. Mayer and R.E. Mayer. *The Cambridge Handbook of Multimedia Learning*. Cambridge Handbooks in Psychology. Cambridge University Press, 2005.
- 14 Sophio Morališvili. Augmented reality in foreign language learning.
- 15 Petar Pejić, Taško Rizov, Sonja Krasić, and Bojana Stajić. Augmented reality applications in engineering. 2014.
- 16 Jan L Plass, Dorothy M Chun, Richard E Mayer, and Detlev Leutner. Supporting visual and verbal learning preferences in a second-language multimedia learning environment. 1998.
- 17 Jun Rekimoto and Katashi Nagao. The world through the computer: computer augmented interaction with real world environments. In *Proceedings of the 8th annual ACM symposium on User interface and software technology - UIST '95*, pages 29–36. ACM Press, 1995.

- 18 A.A. Rizzo, J.G. Buckwalter, U. Neumann, C. Kesselman, M. Thiebaut, P. Larson, and A. van Rooyen. The virtual reality mental rotation spatial skills project. *Virtual Reality*, pages 113–120, 1998.
- 19 Marc Ericson C. Santos, Arno in Wolde Lübke, Takafumi Taketomi, Goshiro Yamamoto, Ma. Mercedes T. Rodrigo, Christian Sandor, and Hirokazu Kato. Augmented reality as multimedia: the case for situated vocabulary learning. *Research and Practice in Technology Enhanced Learning*, 11(1), 2016.
- 20 Olga Scivner, Julie Madewell, Cameron Buckley, and Nitocris Perez. Augmented reality digital technologies (ARDT) for foreign language teaching and learning. In *2016 Future Technologies Conference (FTC)*, pages 395–398. IEEE, 2016.
- 21 Paul Seedhouse, Anne Preston, Patrick Olivier, Dan Jackson, Philip Heslop, Madeline Balaam, Ashur Rafiev, and Matthew Kipling. The european digital kitchen project. *Bellaterra Journal of Teaching & Learning Language & Literature*, 7(1), 2014.
- 22 Jeffrey H Shuhaiber. Augmented reality in surgery. *ARCH SURG*, 139, 2004.
- 23 John Sweller. Chapter two - cognitive load theory. volume 55 of *Psychology of Learning and Motivation*, pages 37 – 76. Academic Press, 2011.
- 24 Christian David Vazquez, Afika Ayanda Nyati, Alexander Luh, Megan Fu, Takako Aikawa, and Pattie Maes. Serendipitous language learning in mixed reality. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '17*, pages 2172–2179. ACM Press, 2017.
- 25 B Weidenmann. Lernen mit medien. In *Pädagogische Psychologie*, volume 4, pages 415–465. Weinheim: Beltz Psychologie Verlags Union., 2001.

Learning in Mixed Realities

Michaela Schmierl

Ludwig-Maximilians-Universität München, München, Deutschland
M.Schmierl@campus.lmu.de

Zusammenfassung

In dieser Arbeit wird untersucht, wie Mixed Reality beim Lernen, besonders im naturwissenschaftlichen Bereich, helfen kann. Dafür wird zunächst geklärt, was unter Mixed Reality zu verstehen ist und verschiedene technische Umsetzungen von Mixed Reality Systemen erläutert, sowie verschiedene Lerntheorien herangezogen. Anhand ausgewählter Studien werden daraufhin konkrete Beispiele für die Realisierung von Mixed Reality Systemen für Lehrzwecke näher beleuchtet. Deren Ergebnisse lassen den Schluss zu, dass Mixed Reality einen positiven Einfluss auf den Lernerfolg haben kann.

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

Keywords and phrases Lernen; Mixed Reality; MR.

1 Naturwissenschaften erleben

Die Menschheit versucht seit jeher die Vorgänge in der Natur zu verstehen und zu erklären. Oftmals bestehen viele komplexe Zusammenhänge und diese zu vermitteln ist eine Herausforderung, die in jeder Zeitepoche anders angegangen wird. Von der reinen mündlichen Überlieferung über bildliche Darstellungen bis hin zu Modellen und Simulationen.

Ein Vorteil von Simulationen ist, dass naturwissenschaftliche Phänomene dargestellt werden können, die für einen Menschen für gewöhnlich nicht direkt erfahrbar sind. Dies gilt sowohl im Großen, beispielsweise dem Sonnensystem, als auch im Kleinen, wie der atomaren Ebene.

Seit einiger Zeit wird auch Mixed Reality erforscht, womit Informationen dem Nutzer ebenfalls näher gebracht werden können. Die Zusammensetzung von realen, physisch manipulierbaren und digitalen Objekten [5] kann ebenfalls zur Vermittlung von Lerninhalten genutzt werden. Derzeit ist Mixed Reality noch ein sehr junges Forschungsgebiet und die Möglichkeiten dieser Systeme sind noch nicht vollständig erfasst.

In dieser Arbeit wird deshalb erkundet, wie Mixed Reality für das Lernen genutzt werden kann. Zunächst wird festgestellt, was man unter Mixed Reality



© Michaela Schmierl;

licensed under Creative Commons License CC-BY

Cite as: Michaela Schmierl. Learning in Mixed Realities. In *3rd Seminar on Ubiquitous Interaction (UBIACTION 2019)*. Editors: Pascal Knierim, Jakob Karolus, Fiona Draxler, Ville Mäkelä, Tonja Machulla, Florian Lang, Thomas Kosch, Albrecht Schmidt. July 16, 2019. Munich, Germany. pp. 3:1–3:16.

3:2 Learning in Mixed Realities

versteht und einige Lerntheorien werden behandelt. Daraufhin werden ausgewählte Forschungsarbeiten zu Lernen mit Mixed Reality Systemen präsentiert, wobei der Fokus auf Forschungen im naturwissenschaftlichen Bereich liegt.

2 Mixed Reality

In diesem Kapitel wird zunächst dargestellt, was unter Mixed Reality zu verstehen ist und einige bereits vorhandene technische Umsetzungen werden erläutert.

2.1 Definition

Zunächst ist zu betonen, dass keine eindeutige und allgemeingültige Definition von Mixed Reality (MR) existiert. In der Forschung wird derzeit über verschiedene Auslegungen diskutiert, doch eine gemeinsame Begriffsbestimmung zu finden, ist schwer, zumal es sich bei MR um ein breites Feld handelt, wie im Folgenden noch ersichtlich wird. [19]

Eine populäre Sichtweise auf MR bieten Milgram et al. [18] mit ihrem *Realität-Virtualität-Kontinuum*. Eine vereinfachte Darstellung von Hein [5] ist in Abbildung 1 zu sehen.



■ **Abbildung 1** Vereinfachte Darstellung des Realität-Virtualität-Kontinuums von Hein [5], basierend auf der Arbeit von Milgram et al. [18]

An der linken Seite der Grafik steht die *Echte Umgebung*. Damit wird die Umgebung beschrieben, die ohne technische Veränderung wahrgenommen werden kann. Auf der rechten Seite ist die *Virtuelle Umgebung* zu sehen, die rein aus virtuellen Bestandteilen besteht. Dazwischen befinden sich der Bereich der *Mixed Reality*. Dieser umfasst sämtliche Mischformen aus *Echter* und *Virtueller Umgebung*, schließt jedoch nicht die Extrema mit ein. Werden auf einem Display sowohl reale, als auch virtuelle Objekte angezeigt, zählt dies folglich zu *Mixed Reality*. [18]

Durch Abbildung 1 ist ersichtlich, dass im Bereich der *Mixed Reality* die Menge an echten, physischen als auch virtuellen Bestandteilen variieren kann: Überwiegt der Anteil an echten Objekten, so wird dies als *Augmented Reality*

(AR) bezeichnet, überwiegt wiederum der Anteil virtueller Objekte, so fällt dies unter *Augmented Virtuality (AV)* [5].

Außerdem weist Le [12] darauf hin, dass die Objekte „in Echtzeit veränderlich sind“ [12].

Speicher et al. [19] merken allerdings an, dass diese Definition sehr auf die visuellen Aspekte reduziert ist und nicht alle Sinne umfasst, mit denen eine Umgebung wahrgenommen werden kann. So werden beispielsweise das Hören, Fühlen und Riechen nicht berücksichtigt.

Nach Speicher et al. [19] ist es durchaus möglich, dass verschiedene Definitionen von MR legitim sind, jedoch ist wichtig, den jeweiligen Kontext klarzustellen, in dessen Rahmen MR verstanden wird. Als Hilfestellung haben Speicher et al. sieben Dimensionen festgelegt. Darunter ist beispielsweise die Anzahl an Nutzern und der Level der Immersion. Auch ist der Level der Interaktion wichtig. Während manche Definitionen von MR lediglich implizite Interaktion voraussetzen, bedingen andere Definitionen auch explizite Interaktion. [19]

2.2 Existierende Mixed Reality Systeme

Wie bereits in dem vorherigen Abschnitt erwähnt, können Mixed Reality Systeme unterschiedlichster Natur sein. Je nach Anwendungszweck ist hierbei zu entscheiden, welche Art System den größten Nutzen bringt.

Eine bekannte Komponente sind sogenannte *Head-Mounted Displays (HMDs)*. Diese sind Displays, die am Kopf getragen werden. [10]. AR Brillen sind HMDs, mit denen noch immer die reale Umgebung wahrgenommen werden kann. Virtuelle Einblendungen können auf dem Display zusätzlich eingeblendet werden. Mit VR Brillen hingegen sind für gewöhnlich nur virtuelle Objekte zu sehen. Dennoch ist es möglich, reale Objekte einzublenden, wie McGill et al. [17] gezeigt haben, indem sie eine reale Tastatur mit einblenden, um Texteingaben zu erleichtern, während eine VR Brille getragen wird.

Handheld Devices sind kleine Computer mit einem Display, die in der Hand gehalten werden können [10]. Smartphones und Tablets sind Beispiele hierfür.

Auch Projektoren können genutzt werden. So wurde in einer Studie von Lindgren et. al. [13] ein digitales Weltall auf den Boden projiziert. In der Studie von Ferdous et. al. [4] wurde ein digitales Modell eines Menschen, dessen Muskeln zu sehen sind, auf einen realen Menschen projiziert, um im Physiotherapie-Unterricht beim Lernen zu helfen.

Zur Interaktion mit dem Mixed Reality System können ebenfalls viele verschiedene Techniken benutzt werden.

So können Kameras und Bilderkennungsoftware genutzt werden, um Objekte aus der realen Welt zu erkennen. Häufig benutzt werden Marker, die

3:4 Learning in Mixed Realities

dann mit virtuellen Objekten überblendet werden wie in der Studie von Liu et al. [14], in der auf einem Marker eine virtuelle Pflanze angezeigt wird.

Mit speziellen Handschuhe können Gesten ausgeführt werden. Su et al. [20] nutzten beispielsweise Handschuhe mit Beschleunigungssensoren.

Elektromuskelstimulation (EMS) ist eine Möglichkeit, Feedback an den Nutzer zu geben. Dabei werden Elektroden beispielsweise auf den Arm des Nutzers platziert. Lopes et al. [15] nutzen dies, um beim Bewegen virtueller Objekte Reibung auf dem Boden und den Widerstand an einer Wand zu simulieren.

3 Lernen mit Mixed Reality

In folgenden Abschnitten werden zunächst verschiedene Aspekte des Lernens beleuchtet, wie die Theorie des Konstruktivismus und kollaboratives Lernen. Auch wird die Nützlichkeit von Serious Games untersucht. Anschließend werden ausgewählte Forschungsarbeiten aus dem Bereich der Naturwissenschaften und anderer Gebiete vorgestellt, die Mixed Reality Systeme für Lehrzwecke nutzen. Daraufhin werden Herausforderungen betrachtet, die sich bei der Nutzung von Mixed Reality für das Lernen ergeben.

3.1 Theorien und Konzepte

Dieses Kapitel befasst sich mit der Theorie des Konstruktivismus, sowie dem Nutzen, den das Experimentieren für Lernende haben kann. Wie körperliche Aktivität beim Lernen helfen kann, wird im Kapitel Körperliche Interaktion untersucht. Auch soziale Interaktion ist für den Lernprozess von Bedeutung, was ebenfalls beleuchtet wird.

3.1.1 Konstruktivismus

Eine wichtige Lerntheorie ist die Theorie des Konstruktivismus. Laut Hein [6] besagt diese, dass der Lernende sämtliches Wissen selbst konstruiert. Dies ist ein aktiver Prozess. Da laut dem Konstruktivismus Bedeutungskonstruktion die einzige Form des Lernens ist, hat dies zwei Konsequenzen: Wenn über das Lernen nachgedacht wird, muss man den Blick auf den Lernenden und nicht auf die zu lernende Information richten. Außerdem besteht Wissen nicht außerhalb seines Kontextes (der Erfahrung des Lernenden), in dem es konstruiert wurde, beziehungsweise entstanden ist. [6]

Hein [6] zeigt folgende Prinzipien auf, die für das konstruktivistische Lernen grundlegend sind:

1. Lernen ist ein aktiver Prozess, was bedeutet, dass der Lernende sich aktiv mit der Welt um sich herum befassen muss, um Wissen zu erlangen.

2. Während der Mensch lernt, findet er gleichzeitig heraus, wie die Systeme strukturiert sind, in denen die Informationen gespeichert sind. Dieses Wissen kann er bei ähnlich strukturierten Informationen wieder anwenden.
3. Die Bedeutungskonstruktion findet im Geist statt. Körperliche Aktivität alleine reicht nicht aus.
4. Die Sprache des Lernenden beeinflusst das Lernen.
5. „Lernen ist eine soziale Aktivität“ [6]: Der Kontakt und der Austausch mit anderen Lernenden fördert das Lernen.
6. Wissen und dessen Kontext, in dem es entstanden ist, sind untrennbar miteinander verwoben. Das neu erworbene Wissen besteht in Beziehung mit bereits vorhandenem Wissen und den Erfahrungen des Lernenden.
7. Vorwissen ist nötig, um neues Wissen hinzuzugewinnen.
8. Lernen geschieht nicht sofort. Der Mensch braucht Zeit, um sich mit neuen Informationen und Ideen auseinanderzusetzen. Er muss aktiv darüber nachdenken, sie wiederholen und damit experimentieren, um ein tieferes Verständnis zu erlangen.
9. Motivation ist eine wichtige Voraussetzung, dass effektiv gelernt werden kann. Wenn für den Lernenden nicht nachvollziehbar ist, welchen Nutzen ihm das Wissen bringt, ist die Motivation nicht sehr hoch, sich eingehender damit zu beschäftigen.

3.1.2 Experimentieren

Wie bereits hervorgehoben, muss der Lernende eine aktive Rolle einnehmen, um Wissen zu konstruieren. Um den Lernenden zu ermutigen, sich mit neuen Informationen zu beschäftigen, kann ihm angeboten werden, mit den Informationen zu experimentieren. [22]

Das experimentelle Lernen kann grob in vier Phasen unterteilt werden. Zunächst hat der Lernende ein abstraktes Konzept, wie etwas funktioniert. Eine Idee besteht, wie Dinge zusammenhängen könnten. Dann wird dieses Konzept mit einem Experiment erprobt und eine neue Erfahrung bildet sich. Diese wird reflektiert. Ideen werden entweder angepasst, oder verworfen und neue Konzepte aufgestellt. [9]

Wojciechowski und Cellary [22] verweisen zudem auf das *learning-by-doing* Paradigma. Wenn gelernt werden soll, wie etwas zu tun ist, dann ist die beste und natürlichste Weise, es auszuprobieren. [22]

3.1.3 Körperliche Interaktion

In Mixed Reality sind sowohl digitale, als auch physische, greifbare Objekte vorhanden. Physische Objekte können beispielsweise benutzt werden, um digitale Objekte zu manipulieren. Diese Technologie kann für das physische

3:6 Learning in Mixed Realities

Experimentieren mit Ideen genutzt werden. Diese Verbindung aus physischer und digitaler Komponente sorgt dafür, dass bei Mixed Reality nicht mehr zwischen reinem, physischen Experiment und reiner digitalen Simulation, die z. B. nur auf Computerbildschirmen stattfindet, unterschieden werden kann. [2]

Su et al. [20] merken an, dass körperliche Interaktion helfen kann, eine enge Verbindung aus physischer Aktion und mentaler Vorgänge herzustellen und die Fähigkeit, Probleme zu lösen, verbessern kann. [20]

Die körperliche Interaktion mit der physischen Umgebung beeinflusst, wie Menschen über die Welt denken und die Theorien, die sie darüber haben. [13] Studien haben gezeigt, dass vorgeschriebene Bewegungen einen positiven Einfluss auf den Lernerfolg in Bereichen wie Algebra hatten. Jedoch muss bedacht werden, dass die Bewegungen mit dem zu lernenden Konzept übereinstimmen. Ist dies jedoch der Fall, dann kann die eigene Bewegung und das Bewegen anderer Objekte zum Lernen von Konzepten wie beispielsweise Bewegung führen. [13]

Die physische Komponente von Mixed Reality kann für das Lernen mit körperlicher Interaktion gut genutzt werden. Ein Beispiel hierfür ist eine Studie von Lindgren et al. [13], in der die Lernenden die Laufbahn eines Asteroiden vorhersagen, indem sie dessen Laufbahn in einem auf den Boden projizierten Weltall selbst abschreiten. [13]

Eine genauere Beschreibung der Studie kann in Kapitel 3.3 gefunden werden.

3.1.4 Kollaboratives Lernen

Lernsysteme, in denen eine Zusammenarbeit von mehreren Lernenden möglich ist, fördern nicht nur das Lernen des Einzelnen, sondern auch dessen soziale Kompetenzen.

Müssen die Gruppenmitglieder beispielsweise zusammen ein Problem lösen, müssen sich die Mitglieder gegenseitig beraten und für ihre Ansätze rechtfertigen können, sowie mit den anderen Mitgliedern verhandeln. Zum Schluss sollten sich die Lernenden einem gemeinsamen Verstehen angenähert haben. [2]

Die soziale Interaktion führt auch dazu, dass die Mitglieder über ihr eigenes Wissen und das der anderen nachdenken müssen, was zu einem Aufbau von gemeinsamem Wissen führt [14]. Das Besondere dabei ist, dass die Gruppenmitglieder erlangtes eigenes Wissen und Erfahrungen leicht miteinander austauschen können [22]. Die Qualität des gemeinsam konstruierten Wissens könnte sogar die Qualität von Wissen übertreffen, das durch einen Lernenden ohne Gruppe erarbeitet wurde [2].

Ein Beispiel für ein Lernsystem, das gemeinsames Lernen unterstützt, ist das von Johnson-Glenberg et al. [2] genutzte *SMALLab*. Mehrere Motion-

Tracking Kameras und eine Projektion auf den Boden spannen einen aktiven Raum auf, in dem mithilfe eines physischen, greifbaren Objekts im 3D Raum interagiert werden kann. Bis zu vier Lernende können gleichzeitig in dem aktiven Raum interagieren. Außerhalb des aktiven Raums können weitere Lernende sitzen und mitarbeiten, indem sie Ratschläge geben und auf einem Whiteboard schreiben können. [2]

3.2 Serious Games

Zunächst sollten die Begriffe *Game-Based Learning* und *Serious Games* unterschieden werden.

Game-Based Learning bezeichnet das „Erlernen von Wissen auf Basis eines Spiels“ [8, Seite 98] und umfasst dadurch viele Arten von Spielen.

Serious Games sind eine Variante davon. Hierbei steht im Gegensatz zu *Game-Based Learning* die „Problemlösung und der Lerneffekt im Vordergrund“ [8, Seite 99]. Serious Games werden also bereits mit der Absicht entwickelt, dem Spieler etwas beizubringen. „Teilnehmer können mit ihrer Hilfe auf effektive Weise einerseits mit einer neuen Materie vertraut gemacht werden und andererseits kann bereits erlangtes Wissen vertieft und trainiert werden.“ [8, Seite 100]

Es hat sich herausgestellt, dass Spiele die Motivation und die Bereitschaft, etwas zu lernen, erhöhen können [7, 8, Seite 98]. Wojciechowski und Cellary [22] sind der Meinung, dass die jüngeren Generationen beispielsweise durch Computerspiele bereits mit deren Technologien vertraut sind und diese auch in der Bildung erwarten. Die Nutzung von AR Technologie für spielerisches Lernen können laut Wojciechowski und Cellary [22] dabei helfen, die Motivation zum Lernen zu erhöhen.

Allerdings sei angemerkt, dass Engagement und Lerneffekt negativ beeinträchtigt werden können, wenn die Herausforderungen den Fähigkeiten des Spielers nicht entsprechen [7].

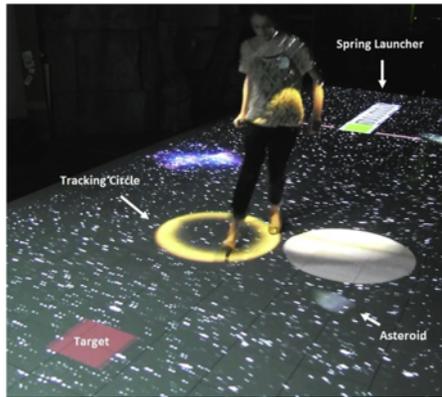
In den folgenden zwei Kapiteln werden einige Serious Games im Bereich Mixed Reality vorgestellt.

3.3 Forschungsarbeiten im Bereich der Naturwissenschaften

Die folgenden Studien zeigen, wie mithilfe von Mixed Reality Lehrinhalte aus dem naturwissenschaftlichen Bereich vermittelt werden können.

Lindgren et al. [13] untersuchten *MEteor*, eine Mixed Reality Lernumgebung, die Schülern die Gravitation und Planetenbewegung durch körperliche Interaktion vermittelt.

3:8 Learning in Mixed Realities



■ **Abbildung 2** *MEteor* Simulation von Lindgren et al. [13]

Die Lernumgebung, die in Abbildung 2 zu sehen ist, besteht aus einer Projektion auf den Boden, die Planeten im Weltall zeigt. Der Lernende hält sich innerhalb der Projektionsfläche auf, seine Bewegungen werden mittels einer Laser Scanning Technologie verfolgt. Aufgabe ist beispielsweise, einen digitalen Asteroiden zu einem bestimmten Ziel zu führen. Dafür wird zunächst Position, Winkel und Geschwindigkeit für den Start des Asteroiden festgelegt, indem der Lernende eine virtuelle Startschleuder platziert. Beginnend bei der Schleuder läuft der Lernende nun durch die projizierte Umgebung, wie er die Laufbahn des Asteroiden vermutet. Gleichzeitig wird auch ein digitaler Asteroid eingblendet, dessen Laufbahn durch physische Gesetzmäßigkeiten bestimmt ist, wie beispielsweise Geschwindigkeit und die Anziehungskraft anderer Planeten. Der Lernende kann dadurch seinen Kurs korrigieren, sollte er davon abweichen. Das Spiel ist erst erfolgreich beendet, wenn der Pfad des Lernenden und der tatsächliche Pfad nicht stark voneinander abweichen. [13]

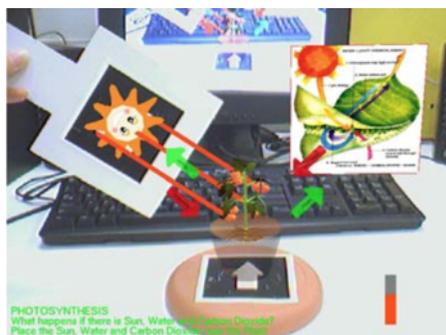
Die Mixed Reality Situation wurde in der Studie mit einer Desktop Version verglichen, in der lediglich mit der Maus interagiert werden konnte.

Untersucht wurde, ob sich der Lernerfolg und die Präsenz in den zwei Versionen unterschieden. Tatsächlich konnte die Mixed Reality Situation einen höheren Lernerfolg und höheres Engagement verzeichnen. Dies führen die Autoren darauf zurück, dass die Bewegung mit der Maus weniger körperliche Bewegung benötigte. Die Bewegung mit dem gesamten Körper rief eine höhere Präsenz hervor, wodurch die Lernenden näher an dem Lerninhalt sein dürften. [13]

Eine weitere Studie wurde von Liu et al. [14] mit Grundschulern durchgeführt. In ihrem System sollte spielerisch grundlegendes Wissen über Pflanzen vermittelt werden. Erforscht wurde die Einstellung der Nutzer gegenüber dem

Mixed Reality System und dem spielerischen Ansatz. [14]

Das System besteht aus mehreren physische Markern, die gefilmt werden. Auf einem Bildschirm ist das Videobild mitsamt digitaler Einblendungen für die gesamte Klasse sichtbar. [14]



■ **Abbildung 3** Veranschaulichung der Photosynthese [14]

Auf einem Marker wird ein virtueller Topf angezeigt, in dem eine virtuelle Pflanze zunächst gepflanzt wird, die daraufhin wächst. In Abbildung 3 ist zu sehen, wie mithilfe eines weiteren Markers Sonnenlicht simuliert und die Photosynthese veranschaulicht wird. Weitere Funktionen wie Bestäubung und Samenbildung sind ebenfalls möglich. [14]

Das Feedback der Nutzer war überwiegend positiv. Festgestellt wurde, dass die wahrgenommene Nützlichkeit und wahrgenommene Freude einen Einfluss auf das Interesse und Engagement an dem Mixed Reality System hatte. Die Qualität des System hatte einen indirekten Einfluss auf das Interesse, da sie die wahrgenommene Bedienbarkeit beeinflusste. [14]

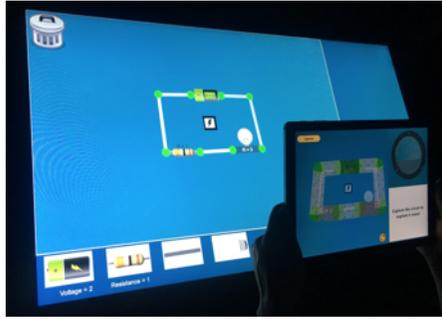
Diverse Forschungsarbeiten haben sich mit der Vermittlung von Kenntnissen über elektrische Stromkreisläufe durch Mixed Reality Systeme gewidmet.

Beheshti et al. [1] untersuchten beispielsweise *Spark*, ein interaktives Ausstellungsstück für Museen, mit dem die Besucher virtuelle elektrische Schaltungen bauen können. Auf einem multi-touch Display stellen die Besucher diese zusammen. Dabei können sie eine Simulation des Elektronenflusses auf demselben Display mitverfolgen, oder mit einem Tablet, das auf die Schaltung gerichtet ist, wie es in Abbildung 4 zu sehen ist. [1]

Grund für diese Art der Umsetzung war, dass die Besucher eine Verbindung zwischen den zwei verschiedenen Ansichten herstellen sollten, um einen größeren Lerneffekt zu erzielen. [1]

In ihrer Studie verglichen die Autoren die Wirkung der Darstellung des Elektronenflusses auf demselben Display (Single-Display Situation) mit der Darstellung auf dem Tablet (AR Situation). In einer Kontrollsituation wurde

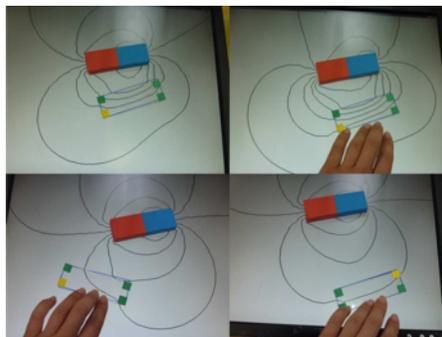
3:10 Learning in Mixed Realities



■ **Abbildung 4** Ein Museumsbesucher betrachtet die Schaltung mithilfe eines Tablets, auf dem der Elektronenfluss zu sehen ist. [1]

kein Elektronenfluss gezeigt. Die Studie wurde mit Kindern in Begleitung ihrer Eltern durchgeführt. Es stellte sich heraus, dass die AR Situation den höchsten Lernerfolg verzeichnete. [1]

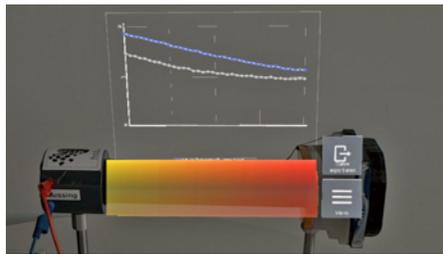
Beobachtet wurde zudem, dass das Engagement der Eltern je nach Situation variierte. So übernahmen die Eltern in der Single-Display Situation häufiger die Rolle eines Lehrers, der unter anderem dem Kind erklärte, wie Schaltkreise funktionieren. In der AR Situation hingegen nahmen die Eltern häufiger die Rolle eines Lernenden an. Die Autoren entwickelten zwei Theorien, um dieses Verhalten zu erklären. Zum einen könnte das Verhalten an der Neuheit der Technologie in der AR Situation liegen, sodass die Eltern sich in ihrer Rolle als Lehrer unwohl fühlten. Eine andere Theorie ist, dass das Kind, wenn es das Tablet in der Hand hält, mehr Kontrolle über das Erforschen und das Ausstellungsstück hat. [1]



■ **Abbildung 5** Magnetfeldlinien um Stabmagneten und transparente Metallstäbe [16]

Matsutomo et al. [16] entwarfen einen Simulator für Magnetfelder mit Augmented Reality.

Auf einem horizontal liegenden Bildschirm liegen Mock-ups für Stabmagnete und Metallstäbe. Eine Kamera ist über dem Bildschirm platziert, sodass sie darauf liegenden Objekte im Blickfeld hat. Mithilfe der Positionsdaten der Magnete und Metallstäbe, sowie der entsprechenden Berechnungen am PC, werden auf dem Bildschirm Magnetfeldlinien gezeichnet. Da dies in Echtzeit geschieht, können die Magnete verschoben werden und die Magnetfeldlinien werden entsprechend aktualisiert. Einer der Metallstab-Repräsentanten ist durchsichtig, sodass die Feldlinien innerhalb des Objekts beobachtet werden können. Beispiele dafür können in Abbildung 5 gesehen werden. [16]



■ **Abbildung 6** Sicht auf den erhitzten Metallstab durch eine AR-Brille. Die Temperatur kann anhand der Farbe und des Diagramms abgelesen werden. [11]

Knierim, Kiss, und Schmidt [11] führten ein Experiment zum Wärmefluss durch. Die Autoren erhoffen sich durch die visuelle Darstellung von physikalischen Größen wie der Wärme, bei Studenten ein besseres Verständnis für physikalische Phänomene zu schaffen. [11]

Für das Experiment wird ein Metallstab an einem Ende erhitzt und zur gleichen Zeit am anderen Ende mit einem Ventilator abgekühlt. Eine Wärmebildkamera filmt den Metallstab und schickt das Bild an eine AR-Brille. In Abbildung 6 ist der Metallstab aus Sicht des Betrachters mit der AR-Brille zu sehen: Die Temperatur des Metallstabs ist durch Farben sichtbar und zudem wird die Temperatur in einem Diagramm angezeigt. Temperaturveränderungen können so in Echtzeit beobachtet werden. [11]

3.4 Limitationen

Trotz der vielen Vorteile sollten die Herausforderungen, die Simulationen und Mixed Reality Systeme mit sich bringen, nicht unterschätzt werden.

Ein von Beheshti et al. [1] und Falloon [3] genanntes Problem ist, dass missverständlichen Interpretationen entstehen können. Lehrer und Schüler sollten eng zusammenarbeiten, um falsche Interpretationen zu erkennen und

aufzuklären [1].

Auch sind die Produktionskosten eines Mixed Reality Systems für Lehrzwecke nicht zu missachten. Spiele in Augmented Reality, als auch Augmented Virtuality sind laut Ternier et al. [21] komplexe Systeme, die sowohl flexibel, als auch simpel sein sollen. Zur Erstellung des Systems werden zudem Fachkräfte aus technischen Bereichen benötigt. Genauso sind Fachkräfte aus dem Bildungsbereich notwendig, um die Mixed Reality Systeme pädagogisch wertvoll zu machen und Lerninhalte sinnvoll zu vermitteln.

4 Fazit und Ausblick

In dieser Arbeit wurde erläutert, dass noch keine einheitliche Definition von Mixed Reality existiert und dass sehr viel mehr zu Mixed Reality zählen kann, als reine visuelle Inhalte. Daraufhin wurden diverse technische Hilfen beschrieben, mit denen Mixed Reality Systeme aufgebaut werden können. Nachdem einige Lerntheorien beleuchtet wurden, die auf Mixed Reality Systeme zutreffen können, wurden Studien beschrieben, die Mixed Reality im Bereich Naturwissenschaften untersuchten. Die Ergebnisse von beispielsweise Beheshti et al. [1] lassen den Schluss zu, dass Mixed Reality durchaus einen positiven Effekt auf den Lernerfolg haben können.

Viele der hier beschriebenen Mixed Reality Systeme setzen ihren Fokus auf visuell wahrnehmbare Inhalte. Jedoch wäre ebenfalls interessant, wie andere Sinne zum Lernen in naturwissenschaftlichen Gebieten verstärkt genutzt werden könnten. Die Möglichkeiten von Mixed Reality sind noch lange nicht ausgeschöpft und die Entwicklung, sowie Forschung im Bereich Mixed Reality schreiten schnell voran. Da Mixed Reality Systeme beim Lernen helfen können, besteht die Aussicht auf viele Mixed Reality Anwendungen in naher Zukunft.

Literatur

- 1 Elham Beheshti, David Kim, Gabrielle Ecanow, and Michael Horn. Looking Inside the Wires: Understanding Museum Visitor Learning with an Augmented Circuit Exhibit. pages 1583–1594, 2017. doi:10.1145/3025453.3025479.
- 2 Mina C. Johnson-Glenberg, David A. Birchfield, Lisa Tolentino, and Tatya-na Koziupa. Collaborative Embodied Learning in Mixed Reality Motion-Capture Environments: Two Science Studies. *Journal of Educational Psychology*, 106:86, 2014. doi:10.1037/a0034008.
- 3 Garry Falloon. Using simulations to teach young students science concepts: An experiential learning theoretical analysis. *Computers & Education*, 135:138–159, 2019. URL: <http://www.sciencedirect>.

- com/science/article/pii/S036013151930051X, doi:<https://doi.org/10.1016/j.compedu.2019.03.001>.
- 4 Hasan Shahid Ferdous, Thuong Hoang, Zaher Joukhadar, Martin N. Reinoso, Frank Vetere, David Kelly, and Louisa Remedios. "what's happening at that hip?": Evaluating an on-body projection based augmented reality system for physiotherapy classroom. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 234:1–234:12, New York, NY, USA, 2019. ACM. URL: <http://doi.acm.org.emedien.ub.uni-muenchen.de/10.1145/3290605.3300464>, doi:10.1145/3290605.3300464.
 - 5 Dimitri Hein. Bildung Augmentieren: Anwendung und Herausforderungen beim Lernen mit Augmented Reality. In *2nd Seminar on Ubiquitous Interaction (UBIACTION 2018-2)*. Editors: Fiona Draxler, Matthias Hoppe, Jakob Karolus, Thomas Kosch, Pascal Knierim, Albrecht Schmidt. June 28, 2018. Munich, Germany. pp. 13:1–13:18.
 - 6 George E. Hein. The significance of constructivism for museum education. In *The Museum and the Needs of People*, Haifa: Israel, 1992. National ICOM Committee. URL: <https://www.exploratorium.edu/education/ifi/constructivist-learning>.
 - 7 Azita Iliya Abdul Jabbar and Patrick Felicia. Gameplay engagement and learning in game-based learning: A systematic review. *Review of Educational Research*, 85(4):740–779, 2015. URL: <https://doi.org/10.3102/0034654315577210>, arXiv:<https://doi.org/10.3102/0034654315577210>, doi:10.3102/0034654315577210.
 - 8 Axel Jacob and Frank Teuteberg. *Game-Based Learning, Serious Games, Business Games und Gamification –Lernförderliche Anwendungsszenarien, gewonnene Erkenntnisse und Handlungsempfehlungen*, pages 97–112. Springer Fachmedien Wiesbaden, Wiesbaden, 2017. URL: https://doi.org/10.1007/978-3-658-16742-4_8, doi:10.1007/978-3-658-16742-4_8.
 - 9 Mmaki Jantjies, Trevor Moodley, and Ronel Maart. Experiential learning through virtual and augmented reality in higher education. In *Proceedings of the 2018 International Conference on Education Technology Management*, ICETM 2018, pages 42–45, New York, NY, USA, 2018. ACM. URL: <http://doi.acm.org.emedien.ub.uni-muenchen.de/10.1145/3300942.3300956>, doi:10.1145/3300942.3300956.
 - 10 Mehmet Kesim and Yasin Ozarslan. Augmented reality in education: Current technologies and the potential for education. *Procedia - Social and Behavioral Sciences*, 47:297 – 302, 2012. Cyprus International Conference on Educational Research (CY-ICER-2012)North Cyprus, US08-10 February, 2012. URL: <http://www.sciencedirect.com/science/article/pii/S1877042812023907>, doi:<https://doi.org/10.1016/j.sbspro.2012.06.654>.

- 11 P. Knierim, F. Kiss, and A. Schmidt. Look inside: Understanding thermal flux through augmented reality. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 170–171, Oct 2018. doi:10.1109/ISMAR-Adjunct.2018.00059.
- 12 Duc Anh Le. Freihand basierte Interaktion in Augmented Reality. In *2nd Seminar on Ubiquitous Interaction (UBIACTION 2018-2)*. Editors: Fiona Draxler, Matthias Hoppe, Jakob Karolus, Thomas Kosch, Pascal Knierim, Albrecht Schmidt. June 28, 2018. Munich, Germany. pp. 12:1–12:18.
- 13 Robb Lindgren, Michael Tscholl, Shuai Wang, and Emily Johnson. Enhancing learning and engagement through embodied interaction within a mixed reality simulation. *Computers and Education*, 95:174–187, 2016. URL: <http://dx.doi.org/10.1016/j.compedu.2016.01.001>, doi: 10.1016/j.compedu.2016.01.001.
- 14 Wei Liu, Adrian David Cheok, Charissa Lim Mei-Ling, and Yin-Leng Theng. Mixed Reality Classroom: Learning from Entertainment. In *Proceedings of the 2Nd International Conference on Digital Interactive Media in Entertainment and Arts, DIMEA '07*, pages 65–72, New York, NY, USA, 2007. ACM. URL: <http://doi.acm.org/10.1145/1306813.1306833>, doi:10.1145/1306813.1306833.
- 15 Pedro Lopes, Sijing You, Alexandra Ion, and Patrick Baudisch. Adding force feedback to mixed reality experiences and games using electrical muscle stimulation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18*, pages 446:1–446:13, New York, NY, USA, 2018. ACM. URL: <http://doi.acm.org.emedien.ub.uni-muenchen.de/10.1145/3173574.3174020>, doi:10.1145/3173574.3174020.
- 16 S. Matsutomo, K. Mitsufuji, Y. Hiasa, and S. Noguchi. Real time simulation method of magnetic field for visualization system with augmented reality technology. *IEEE Transactions on Magnetics*, 49(5):1665–1668, May 2013. doi:10.1109/TMAG.2013.2240672.
- 17 Mark McGill, Daniel Boland, Roderick Murray-Smith, and Stephen Brewster. A dose of reality: Overcoming usability challenges in vr head-mounted displays. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15*, pages 2143–2152, New York, NY, USA, 2015. ACM. URL: <http://doi.acm.org.emedien.ub.uni-muenchen.de/10.1145/2702123.2702382>, doi:10.1145/2702123.2702382.
- 18 Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. Augmented reality: A class of displays on the reality-virtuality continuum. *Telemanipulator and Telepresence Technologies*, 2351:282–293, 01 1994. doi: 10.1117/12.197321.
- 19 Maximilian Speicher, Brian D. Hall, and Michael Nebeling. What is mixed reality? In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19*, pages 537:1–537:15, New York, NY, USA,

2019. ACM. URL: <http://doi.acm.org/10.1145/3290605.3300767>, doi: 10.1145/3290605.3300767.
- 20 Mu-Chun Su, Gwo-Dong Chen, Yi-Shan Tsai, Ren-Hao Yao, Chung-Kuang Chou, Yohannes Budiono Jinawi, De-Yuan Huang, Yi-Zeng Hsieh, and Shih-Chieh Lin. Design of an Interactive Table for Mixed-Reality Learning Environments. In Maiga Chang, Rita Kuo, Kinshuk, Gwo-Dong Chen, and Michitaka Hirose, editors, *Learning by Playing. Game-based Education System Design and Development*, pages 489–494, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.
- 21 Stefaan Ternier, Roland Klemke, Marco Kalz, Patricia Van Ulzen, and Marcus Specht. AR Learn: Augmented reality meets augmented virtuality. *Journal of Cheminformatics - J Cheminf*, 18, 2012.
- 22 Rafal Wojciechowski and Wojciech Cellary. Evaluation of learners' attitude toward learning in aries augmented reality environments. *Computers & Education*, 68:570 – 585, 2013. URL: <http://www.sciencedirect.com/science/article/pii/S0360131513000535>, doi:<https://doi.org/10.1016/j.compedu.2013.02.014>.

Automatische Inhaltsgenerierung und Motivation beim Lernen von Sprachen

Anna-Carina Gehlich

Ludwig-Maximilians-Universität München, München, Deutschland
a.gehlich@campus.lmu.de

Zusammenfassung

Ohne Motivation ist es schwierig, Aufgaben anzufangen und sie zu Ende zu bringen. Dabei unterscheidet man zwischen intrinsischer Motivation ohne Einflüsse der Umwelt und extrinsischer, die durch äußere Einflüsse geweckt wird. Beim Erlernen einer Fremdsprache ist die Motivation am Anfang oft besonders hoch; den Spracherwerb jedoch kontinuierlich weiterzuverfolgen, fällt vielen Menschen schwer. Um Lernende zu unterstützen, bietet sich Microlearning an, das versucht, kleine Lektionen in den Alltag des Nutzers zu integrieren, um so bisher ungenutzte Zeit zur Verbesserung von Sprachkenntnissen zu nutzen. Zusammen mit der automatischen Generierung von Inhalten, die individuell auf Nutzende abgestimmt sind, kann es gelingen, die Motivation langfristig aufrecht zu erhalten.

2012 ACM Computing Classification Applied Computing, Education

Keywords and phrases Spracherwerb; Automatisch generierter Inhalt; Motivation; Microlearning.

1 Einführung

Mehrere Sprachen zu sprechen, ist das Tor zur Welt. Um sich in anderen Ländern zu verständigen, sind Fremdsprachenkenntnisse von großem Vorteil. Wenn man jedoch nicht mehr zur Schule geht und täglich Sprachunterricht besucht, steht man vor der Herausforderung, die neue Sprache zusätzlich zu alltäglichen Abläufen zu erlernen. Möglichkeiten, eine Sprache zu lernen, gibt es viele: Ob man einen Kurs besucht, sich Filme in der Fremdsprache ansieht, Bücher liest oder ins Ausland fährt, für jeden findet sich eine Möglichkeit. Oft stellt sich das Problem der Motivation. Anfangs ist die Motivation sehr hoch, man möchte möglichst schnell Fortschritte machen. Doch bald greift der Alltag um sich und die Motivation sinkt, die neuen Sprachkenntnisse geraten schnell wieder in Vergessenheit.

Abhilfe kann hier Microlearning schaffen. Microlearning versucht, kleine Lektionen in den Alltag zu integrieren, indem es an Stellen ansetzt, die man



© Anna-Carina Gehlich;

licensed under Creative Commons License CC-BY

Cite as: Anna-Carina Gehlich. Automatische Inhaltsgenerierung und Motivation beim Lernen von Sprachen. In *3rd Seminar on Ubiquitous Interaction (UBIACTION 2019)*. Editors: Pascal Knierim, Jakob Karolus, Fiona Draxler, Ville Mäkelä, Tonja Machulla, Florian Lang, Thomas Kosch, Albrecht Schmidt. July 16, 2019. Munich, Germany. pp. 4:1–4:18.

bereits täglich nutzt. Da sich in der heutigen Zeit immer mehr auf dem Smartphone und im Internet abspielt, steigt auch die Nachfrage nach Applikationen wie „Duolingo“. Duolingo ist eine Sprach-Lernapp, bei der Nutzende sich über kleine Einheiten Vokabeln, Grammatik und Hörverstehen aneignen können. Ein Belohnungssystem und ein Wettbewerb mit anderen Nutzenden soll Zielpersonen motivieren, jeden Tag eine weitere Lektion zu absolvieren. Um auch hier die Motivation hoch zu halten, stellt sich die Frage, ob es möglich ist, mithilfe automatischer Inhaltsgenerierung die Kurse besser an den einzelnen Nutzenden anzupassen: Automatische Inhaltsgenerierung versucht individuelle Inhalte für Nutzende zu erstellen, sodass die Motivation dauerhaft hoch bleibt und das Lernen nicht im Alltag untergeht.

In der folgenden Arbeit wird zuerst der Begriff der Motivation erläutert, anhand verschiedener Motivationstheorien und im Zusammenhang mit dem Erwerb von Sprachkenntnissen. Danach wird automatisch generierter Inhalt am Beispiel von Gaming erklärt und welche Möglichkeiten es gibt, Microlearning beim Sprachenlernen einzusetzen. Diese beiden Punkte, die Motivation und automatisch generierter Inhalt, werden im darauf folgenden Kapitel im Bezug auf Sprachen verknüpft und mit Beispielen illustriert. Erkenntnisse, Schwachpunkte und weitere Aufgabenfelder werden diskutiert und die Ergebnisse abschließend zusammengefasst.

2 Motivation

Motivation steuert das menschliche Verhalten, sie ist der Antrieb, den wir brauchen. Ohne sie ist es nahezu unmöglich, seine persönlichen Ziele zu erreichen. Dabei ist es irrelevant ob das Ziel ist, morgens aufzustehen oder den Mount Everest zu erklimmen. Motivation hat verschiedene Ursachen. Auf wie viele Arten die Frage „Warum gehen Menschen morgens zur Arbeit“ beantwortet werden kann, wird mit den folgenden Beispielen beschrieben:

- Sarah muss Geldverdienen, um sich und ihre Familie zu ernähren.
- Lorenz macht seine Arbeit Spaß.
- Marvin ist langweilig und er weiß nicht, was er sonst tun soll.
- Antonia strebt eine Karriere an und muss bei ihren Vorgesetzten Eindruck schinden.
- Emma freut sich darauf, ihre Kollegen und Kolleginnen zu treffen und sich mit ihnen auszutauschen.
- Jan hat in diesem Quartal bereits zu oft gefehlt und ihm wird sonst gekündigt.
- Michael kann sich in seinem Job selbst verwirklichen.
- Theo findet Lisa aus dem Büro gegenüber toll und will sie heute auf einen Kaffee einladen.

- Freitags gibt es immer Weißwürste im Büro.

Obwohl das Thema Motivation bereits seit langer Zeit Teil der Forschung ist, gibt es keine allgemein gültige Formel, mit der beschrieben werden kann, was einen Menschen antreibt. Die Ursachen unterscheiden sich von Mensch zu Mensch und von Zeitpunkt zu Zeitpunkt. Manche Gründe ähneln sich und können in Kategorien zusammengefasst werden. Die folgenden Theorien sind bereits etabliert und sollen im nachstehenden Kapitel kurz vorgestellt werden.

2.1 Motivation in der Psychologie

Warum verhält ein Mensch sich so, wie er es tut? In der Psychologie stellt man die Frage, aus welchen Gründen ein Mensch handelt. Diese Gründe werden je nach Theorie unterschiedlich gewichtet und führen so zu den verschiedenen Ansätzen:

Maslowsche Bedürfnispyramide: Motivation zur Erfüllung von Bedürfnissen ist die Basis der Theorie von Abraham Maslow. Er bringt die Bedürfnisse in eine Reihenfolge. Diese Reihenfolge sagt aus, aus welchen Motiven ein Mensch handelt, da er ausgehend von der Stufe, auf der er sich gerade befindet, unterschiedliche Bedürfnisse stillen möchte. Die Stufen bauen aufeinander auf, das heißt, um in eine höhere Stufe vorzudringen, müssen erst alle niedrigeren Bedürfnisse erfüllt werden. Die unterste Stufe besteht aus physiologischen Grundbedürfnissen, wie Hunger, Durst oder Schlaf. Als nächstes beschreibt Maslow das Bedürfnis nach Sicherheit, also den Wunsch nach Schutz und Angstfreiheit. Als drittes Bedürfnis wird Liebe und Zugehörigkeit genannt, gefolgt von dem Bedürfnis nach Wertschätzung und Anerkennung von Anderen. An der Spitze der Pyramide steht die Selbstverwirklichung. Nach Maslow kann diese Stufe erst erreicht werden, wenn alle anderen Bedürfnisse gestillt wurden. Danach, diese Stufe zu erreichen, strebt der Mensch sein Leben lang. Entsprechend versucht er, seine Talente und Fähigkeiten zu entwickeln und auszuüben. Die Pyramide ist dargestellt in Abbildung 1. [1, 9]

Henry Murray: Auch Murray sieht Motivation als eine Folge von Bedürfnissen. Er unterscheidet zwischen psychogenen Bedürfnissen und viszerogenen (biologischen) Bedürfnissen, wobei die viszerogenen mit den physiologischen Grundbedürfnissen von Maslow zu vergleichen sind. Psychogene Bedürfnisse hingegen beschreibt er ausführlicher hinsichtlich des angestrebten Ziels, der dafür nötigen Handlungen und der auftretenden Emotionen [2]. Diese Bedürfnisse beschreibt er als „need“, während äußere Einflüsse bzw. bedürfnisbezogene Situationsmerkmale als „press“ bezeichnet werden. Ebenso wie Maslow ordnet auch er Bedürfnisse hierarchisch an, für jeden Menschen ist



■ **Abbildung 1** Die Bedürfnisse eines Menschen nach Abraham Maslow [9]

diese Hierarchie jedoch individuell. Allerdings sind die viszerogenen Bedürfnisse die wichtigeren Bedürfnisse, da durch sie das Überleben des Menschen gesichert werden. Durch die individuelle Entwicklung von Motiven führt er Motivation auch als Persönlichkeitsmerkmal an.

Murrays bedeutendste Leistung war die Mitentwicklung des Thematischen Apperzeptionstests (TAT), mithilfe dessen es möglich ist, die Motive von Menschen herauszufinden. Dazu werden den Probanden Bilder vorgelegt, zu denen Sie eine Geschichte erfinden sollen. Mithilfe einer Analyse der Erzählung können dann die Motive der Person offenbart werden, da diese in die Geschichte hinein projiziert wurden. [1, 10].

Zwei-Faktoren-Modell nach Herzberg: Herzberg untersuchte die Arbeitszufriedenheit von Angestellten. In den Ergebnissen seiner Arbeit unterscheidet er zwischen Motivatoren und Hygienefaktoren. Motivatoren beeinflussen demnach die Arbeitszufriedenheit, während Hygienefaktoren die Arbeitsunzufriedenheit bestimmen. An einem Arbeitsplatz ist es sowohl nötig, Motivatoren zu schaffen, als auch die Hygienefaktoren zu beachten, um Mitarbeitende langfristig zu motivieren. Die beiden Faktoren sind unabhängig voneinander, sie müssen also gleichzeitig erreicht werden, bevor ein gutes Arbeitsklima entsteht. Unter Motivatoren versteht er zum Beispiel Anerkennung, Verantwortung, Aufstieg und Selbstverwirklichung. Hygienefaktoren

können das Gehalt, Arbeitsbedingungen, oder Sicherheit sein. [16]

Diese Theorien beschreiben die beiden wesentlichen Ursachen von Motivation: Motivation, die aus inneren Bedürfnissen hervorgeht, und Motivation, die durch äußere Einflüsse entsteht. Die folgenden beiden Unterkapitel beschäftigen sich ausführlicher mit den Begriffen der intrinsischen und extrinsischen Motivation.

2.1.1 Intrinsische Motivation

Sowohl Murray als auch Herzberg unterscheiden bereits zwischen inneren und äußeren Faktoren, die sich auf die Motivation auswirken. Unter intrinsischer Motivation werden all jene Faktoren aufgelistet, die aus innerem Antrieb geschehen, die also keinen äußeren Einfluss aufweisen. Richard Ryan definiert intrinsische Motivation als “doing of an activity for its inherent satisfactions rather than for some separable consequence” [19].

Alles, was aus intrinsischer Motivation heraus geschieht, geschieht, ohne dass Strafe oder Belohnung von außen erwartet wird, nur um der Aktivität willen. In dem Beispiel vom Beginn des Kapitels wären das Michael, der sich in seinem Job verwirklichen kann, und Lorenz, dem seine Arbeit Spaß macht. Alle anderen aufgelisteten Personen werden von äußeren Einflüssen motiviert, wie zum Beispiel durch Geld, Karriere oder soziale Bedürfnisse.

2.1.2 Extrinsische Motivation

Äußere Einflüsse sind deutlich häufiger anzutreffen als eine Motivation durch innere Triebe des Menschen. Jede Form der Motivation, die durch einen äußeren Einfluss entsteht, wird unter extrinsischer Motivation zusammengefasst. Wenn also erwartet wird, dass eine bestimmte Situation eintritt, nachdem man eine Aufgabe erledigt hat, ist das eine Form des äußeren Einflusses. Hierbei spielt es keine Rolle, ob die Erwartung positiv oder negativ konnotiert wird. Als Beispiel seien hier zwei Schüler genannt, die auf eine Klassenarbeit lernen. Die Eltern des ersten Schülers versprechen ihm bei einer Eins in der Klassenarbeit ein Eis. Die Erwartung ist positiv verknüpft, der Schüler wird durch die Belohnung motiviert. Die Eltern des anderen Schülers drohen ihm damit, dass er eine Woche nicht Fernsehen darf, wenn er etwas Schlechteres als eine Drei mit nach Hause bringt. Auch dieser Schüler ist extrinsisch motiviert, allerdings durch die Vermeidung einer Strafe.

2.2 Motivationsfaktoren beim Sprachenlernen

Jugendliche müssen während ihrer Ausbildung in der Schule eine Vielzahl von Fächern belegen, darunter auch Fremdsprachen. Im Schuljahr 2017/2018 war

Englisch die meist gelernte Sprache in allgemeinbildenden Schulen, gefolgt von Französisch, Latein und Spanisch [3]. Um eine Sprache zu lernen, ist es vorteilhaft, sich außer mit Vokabeln und Grammatik auch mit der Kultur des Landes auseinanderzusetzen, in dem diese gesprochen wird. Dies hilft sowohl bei einem späteren Auslandsaufenthalt, als auch um sprachliche Eigenheiten zu verstehen. Ein Verständnis der Kultur kann durch einen Auslandsaufenthalt, aber auch durch das Ansehen von Sendungen in der Fremdsprache oder durch das Lesen eines Buches aus dem fremden Land aufgebracht werden.

Tulving und Thomson haben bereits 1973 festgestellt, dass zwischen dem Kontext, in dem etwas gelernt, und dem Kontext, in dem etwas abgerufen wird, ein Zusammenhang besteht. Ein Wort wird öfter richtig abgerufen, wenn der Kontext in dem es gelernt wurde mit dem, in dem es abgerufen wird, übereinstimmt. Die Art und Weise, wie ein Wort wiedergefunden werden kann, hängt davon ab, wie es beim Lernen verknüpft wurde. Dieses Prinzip heißt „encoding specificity“ und sollte beim Einrichten von Lernapplikationen bedacht werden [21].

Auch die Auswirkungen der Vergessenskurve sind relevant. Neu Gelerntes wird weniger schnell vergessen, in Abhängigkeit davon, wie oft es bereits wiederholt wurde. „Spaced repetition“ setzt an dieser Stelle an. Es bedeutet, dass etwas Gelerntes wahrscheinlicher ins Langzeitgedächtnis gelangt, wenn es über längere Zeitabstände hinweg richtig abgerufen werden konnte, als wenn es nur kurz nacheinander wiederholt wurde. Ein gutes Beispiel dafür sind Karteikästen. Im ersten Durchgang werden die Vokabeln täglich wiederholt; sobald sie richtig übersetzt wurden landen sie ein Fach weiter. Im zweiten Fach werden sie bereits weniger oft wiederholt. Wird die Vokabel erneut richtig übersetzt, gelangt sie noch ein Fach weiter; wird sie hingegen nicht gewusst, landet sie wieder in Fach 1. Das Prinzip wiederholt sich so lange, bis die Vokabel im Langzeitgedächtnis verankert ist [13].

Diese Prinzipien sind in Applikationen zu verankern, um ein dauerhaftes Erlernen von Sprachen zu erreichen. Zusätzlich sollte nun auch die Motivation gestärkt werden. Dafür spielen weitere Faktoren eine Rolle. Gardner unterscheidet zwischen zwei Verhaltensaspekten beim Sprachenlernen: 1. „attitudes toward learning the language“ und 2. „attitudes toward the other-language community“. Dabei ist der erste Aspekt didaktisch relevant, während der zweite den sozialen Kontext abdeckt [11].

Eine Aufteilung nach inneren und äußeren Faktoren haben William und Burden 1997 aufgestellt, siehe Abbildung 2. Sie analysieren die Motivationsfaktoren eines Schulkindes in seiner Unterrichtsumgebung. Zu den inneren Faktoren zählen: Das intrinsische Interesse an der Aktivität, der wahrgenommene Wert der Aktivität, der Sinn für die Tätigkeit, das Können, die Selbsteinschätzung, die Einstellung zur Zielsprache und -kultur, andere affektive Zustände, das

Entwicklungsalter und die Entwicklungsstufe sowie das Geschlecht. Äußere Faktoren auf der anderen Seite sind signifikante Personen (Eltern, Lehrer, Mitschüler), sowie die Art der Interaktion mit ihnen, die Lernumgebung und der weitere Kontext. Diese Faktoren bestimmen die Stärke der Motivation des Lernenden. Einige von diesen Faktoren können beim Entwickeln von Applikationen schwierig eingebunden werden, wie das Vermeiden von gelernter Hilflosigkeit oder die Beurteilung von Erfolg und Misserfolg. Andere Faktoren lassen sich aber sehr gut beeinflussen und sollten daher mit bedacht werden. Zum Beispiel kann durch ein optimales Maß an Herausforderung das intrinsische Interesse des Lernenden an der Aktivität geweckt werden. Ein individuelles Anpassen der Schwierigkeit an das Können des Lernenden kann hier zu Erfolg führen. Auch das Näherbringen der Kultur des Landes, in dem die Fremdsprache gesprochen wird, kann zu höherer Motivation führen. Zusätzlich können Eltern und Mitlernende eingebunden werden, und angemessene Belohnungen an die Leistung des Lernenden geknüpft werden. [23]

Innere Faktoren	Äußere Faktoren
Intrinsisches Interesse an der Aktivität - Erwecken von Neugier - Optimales Maß an Herausforderung	Signifikante Personen - Eltern - Lehrer - Mitschüler
Wahrgenommener Wert der Aktivität - Persönliche Relevanz - Erwarteter Wert von Ergebnissen - Intrinsischer Wert, der der Aktivität beigemessen wird	Die Art der Interaktion mit den signifikanten Personen - Vermittelte Lernerfahrungen - Art und Häufigkeit von Feedback - Belohnungen - Art und Häufigkeit von angemessenem Lob - Strafen, Sanktionen
Sinn für die Tätigkeit - Selbstwirksamkeit (locus of causality) - Kontrollüberzeugung: Prozess und Ergebnisse - Fähigkeit angemessene Ziele zu setzen	Die Lernumgebung - Behaglichkeit - Ressourcen - Tageszeit, Woche, Jahr - Größe der Klasse und Schule - Klassen- bzw. Schulethos
Das Können - Kompetenzgefühl - Bewusstsein für das Entwickeln von Fertigkeiten und Können in einem selbstgewählten Gebiet - Selbstwirksamkeit	Der weitere Kontext - Breiteres Familiennetzwerk - Lokales Bildungssystem - Widersprüchliche Interessen - Kulturelle Normen - Soziale Erwartungen und Haltungen
Selbsteinschätzung - Realistisches Bewusstsein von persönlichen Stärken und Schwächen in den erforderlichen Kompetenzen - Persönliche Definitionen und Beurteilungen von Erfolg und Misserfolg - Selbstbezogene Bedenken - Gelernte Hilflosigkeit	
Einstellung - Zur Zielsprachengemeinschaft und Kultur	
Andere affektive Zustände - Selbstvertrauen - Beklemmung, Angst	
Entwicklungsalter und Entwicklungsstufe	
Geschlecht	

■ **Abbildung 2** Aufteilung von William und Burden der Motivation nach inneren und äußeren Faktoren [23]

3 Anwendungsgebiete für automatisch generierten Inhalt

Inhalt automatisch zu generieren und an Nutzende individuell anzupassen ist eine noch sehr junge Disziplin. 2002 meldete Benjamin Hosken ein Patent an, das automatischen Inhalt und zusammenarbeitsbasierte Systeme und Methoden für das Bestimmen und das Bereitstellen von Inhaltsempfehlungen verwendet [15]. Dabei ist das Ziel die Unterhaltungsbranche, mit Medien wie Musik oder Videos. Empfehlungen werden auf Grundlage der Profile der Nutzenden und einer Datenbank getroffen. Je mehr Ähnlichkeiten ein Profil mit dem von Anderen hat, desto besser können Empfehlungen ausgesprochen werden, die den Wünschen der Nutzenden entsprechen. Solche und ähnliche Algorithmen werden heutzutage in Diensten wie Spotify oder Youtube verwendet, aber auch in sozialen Netzwerken. Hier wird auch der Freundeskreis einbezogen: Eine enge Bezugsperson oder ein Familienmitglied kann so einen größeren Einfluss auf das, was die Zielperson sieht, haben, als eine fremde Person [8].

Im folgenden Kapitel wird automatisch generierter Inhalt in zwei Anwendungsfeldern betrachtet: Zunächst bei Spielen und dann im Anwendungsgebiet des Sprachenlernens.

3.1 Gaming

Spiele bieten Nutzenden vor allem Unterhaltung. Um diese möglichst spannend zu gestalten und Nutzende nicht zu verlieren, ist es wichtig, dass sie sich nicht langweilen. Um Abwechslung ins Spiel zu bringen, ist man hier schon früh auf die Idee gekommen, Level, Umgebung, Gegenstände oder andere Objekte anzupassen. Mithilfe automatischer Inhaltsgenerierung werden diese fortwährend verändert, sodass Nutzende während des Spielens immer neuen Gegenständen begegnen. Bisherige Methoden, wie die Umgebung mithilfe zufälliger Variablen zu generieren, haben hier nur begrenzt den gewollten Effekt, da sie zu unerwünschten Objekten führen können und sich außerdem nicht an individuelle Nutzende anpassen lassen. Das Videospiel „Galactic Arms Race“ von 2010 integrierte diese Technik. Es analysiert Nutzerverhalten im Spiel, um so den Inhalt anzupassen und das Spielerlebnis individuell zu verbessern. Dazu wird ein Algorithmus verwendet, der das Spielverhalten in Echtzeit analysiert und auf Basis des Ergebnisses den Inhalt anpasst [14].

Shaker, Yannakakis und Togelius haben gezeigt, dass es möglich ist, auch die einzelnen Level in einem Spiel individuell an Nutzende anzupassen. Als Basis dient eine abgewandelte Version des Spiels „Super Mario Bros“, einem Jump-‘n’-Run-Spiel. Daten von 327 Nutzenden wurden ausgewertet und später auf das Spiel angewandt. Als Kontrolle diente die Vorhersage der Emotionen der Nutzenden auf das Level, welche sehr gut vorausgesagt werden konnten [20].

3.2 Erlernen einer Fremdsprache

Onlinespiele werden heutzutage vermehrt eingesetzt, um Fremdsprachen zu lernen. Carrie Cal hat 2013 eine Version von Tetris adaptiert, in der zusammen mit den Blöcken regelmäßig Vokabeln erscheinen. Durch das Spiel werden Funktionen wie Herausforderung, Fantasie, Wettbewerb und Wiedererkennung aktiviert. Dies sind Faktoren, welche die intrinsische Motivation des Spielenden anregen und ihn so motivieren, weiterzumachen. In der Version des Spiels werden neue Vokabeln zunächst mit einem Block, einem Bild und dem Wort eingeführt. Später tauchen nur noch der Block und das Bild auf, und bevor Nutzende den Block rotieren können, muss er das Wort korrekt ausgesprochen haben. In der Vorstudie wurden bisher noch keine individuellen Vokabeln für die Nutzenden gewählt; dies ist aber bei späteren Versionen zu empfehlen, um sich an den Fortschritt der Nutzenden anzupassen [4].



■ **Abbildung 3** Links: Intelligente Untertitel die den Nutzer dabei unterstützen sollen, beim Ansehen von Filmen die Fremdsprache besser zu verstehen und zu lernen [17]; Rechts: Das Arcade Spiel Tetris so adaptiert, dass während des Spielens Vokabeln gelernt und wiederholt werden können [4]

Viele Menschen greifen auf Filme und Serien zurück, um sich in ihrer Freizeit auf eine Fremdsprache einzulassen und Sprachkenntnisse zu vertiefen. Unterstützung bieten bilinguale Untertitel in der Fremdsprache und in der Muttersprache, um der Handlung besser folgen zu können und sich mit Schreibweise und Bedeutung von Vokabeln auseinanderzusetzen. Kovacs und Miller haben intelligente Untertitel entwickelt, die nicht nur eine sinngemäße Übersetzung bieten, sondern Wort für Wort der gesprochenen Sprache folgen. Der Lernende wird so deutlich besser unterstützt. Außerdem bieten sie zusätzlich

Übersetzungen, Definitionen und eine bessere Steuerung durch den Film, mit erneuter lautmalerischer Wiedergabe von nicht verstandenen Textpassagen. Die Untertitel werden automatisch erzeugt und sind aktuell für die Sprachen Chinesisch, Japanisch, Französisch, Deutsch und Spanisch verfügbar. In der Vorstudie schafften es die Probanden, sich mit den intelligenten Untertiteln doppelt so häufig an neue Worte zu erinnern, als mit reinen bilingualen Untertiteln. Dabei wich die Dauer des Filmkonsums kaum voneinander ab und die Probanden sagten selbst aus, dass ihre Freude am Film nicht getrübt wurde. Sowohl die Wort-für-Wort Übersetzungen als auch die Navigation durch die Dialoge wurden intensiv genutzt [18].

4 Verstärkung der Motivation durch automatisch generierten Inhalt

Bisher wurde beleuchtet, welche Faktoren bei der Motivation zum Fremdsprachenerwerb eine besonders große Rolle spielen und wie Nutzenden mithilfe von automatisch generiertem Inhalt individuelle Lektionen geboten werden können. Der Fremdsprachenerwerb erfordert eine kontinuierliche Hingabe von Lernenden über mehrere Monate. Während dieser Zeit sollte man gleichmäßig motiviert bleiben um den Lernerfolg zu maximieren. Ob Motivation beim Sprachenlernen durch automatisch generierten Inhalt besonders lange erhalten werden kann, ist eine spannende Fragestellung. Durch passende Algorithmen können Inhalte individuell an den Nutzer angepasst werden, sodass weder Unter- noch Überforderung auftreten, die ihn demotivieren könnten. Auch Microlearning bietet sich an, da hier wertvolle bisher ungenutzte Zeit dem Lernen gewidmet werden kann. Im folgenden Kapitel werden einige Beispiele vorgestellt, wie sich mithilfe von Microlearning das Sprachenlernen in den Alltag integrieren lässt, um das Interesse des Nutzers möglichst lange zu erhalten. Dies soll zusätzlich durch automatische Inhaltsgenerierung unterstützt werden.

4.1 Lernschoner

Eine Möglichkeit, Microlearning in den heutigen Alltag zu integrieren, ist, es mit an den Bildschirmschoner zu knüpfen. Gassler, Hug und Glahn stellten 2004 einen Prototypen vor, der sich an den Bildschirmschoner anschließt. Dieser war gedacht, um bestehende Sprachkenntnisse aufzufrischen, indem Nutzende kleine Sätze übersetzen müssen. Wenn Nutzende den Computer erneut verwenden wollen, startet erst das Lernprogramm. Die Nutzenden evaluieren dabei ihre Antworten selbst. Aufgrund dieser Antworten werden zukünftige Sätze ausgewählt. Gassler, Hug und Glahn berichten selbst von Abbrecherquoten von nur 25%, während sie andere e-Learning Inhalte beim Sprachkenntniserwerb

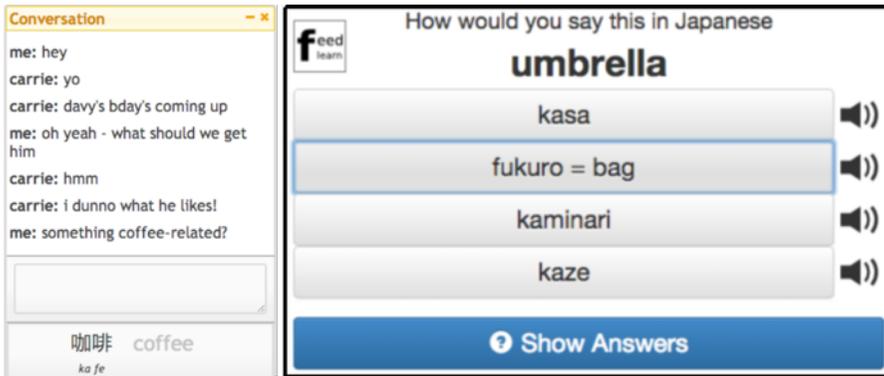
mit bis zu 85% Abbrecherquoten beziffern. Ihre Methode, Nutzende zu motivieren, bezieht sich auf eine technische Unterstützung der Initiation des Lernens. Dadurch müssen Nutzende nicht mehr aktiv das Lernprogramm starten, um sich ihren Lektionen zu widmen, sondern sie werden von technischer Seite darauf hingewiesen, ihre Lektionen abzuarbeiten, selbst wenn sie nicht daran gedacht oder etwas anderes vor hatten. Durch die kurzen Lektionen sind sie außerdem nicht stundenlang beschäftigt, sondern können sich gleich nach dem Lösen einer Aufgabe wieder anderen Tätigkeitsfeldern zuwenden. [12]

4.2 WaitChatter

Nachdem heutzutage das Smartphone ein ständiger Begleiter geworden ist und das Telefonieren immer häufiger von Chatdiensten abgelöst wird, bietet es sich an, ein System zu schaffen, die Wartezeit auf das Gegenüber zwischen den Nachrichten sinnvoll zu nutzen. „WaitChatter“ setzt an genau diesem Punkt an. Die Konversation wird analysiert, um so aktuell relevante Vokabeln in der Fremdsprache anzuzeigen, selbst wenn die Konversation in der Muttersprache geführt wird. Zwei Modi sind derzeit verfügbar: Ein Modus, in dem die Vokabel zusammen mit der Übersetzung angezeigt wird, und ein „Quiz-Modus“, indem die Vokabeln abgefragt werden. Die Vorstudie empfiehlt, unmittelbar nachdem eine Nachricht gesendet wurde Inhalte zum Spracherwerb anzuzeigen, da zu diesem Zeitpunkt der Nutzer zwar noch das Chatfenster fokussiert, jedoch offen für andere Inhalte ist. Die Motivation für Anwendende erschließt sich in diesem Fall aus dem sinnvollen Nutzen von kurzen Zeitabständen während des Wartens auf das Gegenüber. So kann diese Zeit genutzt werden um Vokabeln zu lernen, ohne dass das Gespräch beendet werden muss. Dadurch, dass die Vokabeln in Abhängigkeit zu der aktuell geführten Konversation angezeigt und gelernt werden, wird außerdem der Faktor „encoding specificity“ angeregt – das Lernen von Vokabeln in einer Umgebung, in der sie vermutlich auch später benutzt werden. [5]

4.3 MicroMandarin

Die Anwendung „MicroMandarin“ versucht ebenfalls, möglichst viel Wartezeit für den Erwerb von Sprachkenntnissen zu verwenden. Dabei werden vier Funktionen in den Vordergrund gerückt: Das Lernen abhängig davon, wo der Nutzer sich aktuell befindet, und das Verwenden der Sprache; das Durchsuchen aller bisher gelernten Vokabeln und die Fortschrittsanzeige. Durch die ersten beiden Funktionen kann auch hier Nutzenden eine individuelle Lernerfahrung geboten werden. Die App versucht, Nutzenden in einem bestimmten Kontext vorrangig Vokabeln zu bieten, in dem sie vermutlich auch bald verwendet werden. Anhand von Geodaten können abhängig von der Position der Nutzen-



■ **Abbildung 4** Links: WaitChatter, eine Anwendung in der Vokabeln während des wartens auf den Chatpartner gelernt werden können [5]; Rechts: Lektion die in den Facebook Feed integriert werden kann [17]

den relevante Vokabeln angezeigt werden. Durch einen Button „Used today“ können Nutzende angeben, ob sie die Vokabel an diesem Tag benutzt haben, also ob sie relevant für die Verwendenden war. Im Ausblick werden weitere Studien vorgeschlagen, um Nutzende zu motivieren, die Sprache auch aktiv zu sprechen. Dies kann eventuell durch einen höheren Spielreiz angeregt werden. Durch das Lernen in Abhängigkeit vom Ort soll eine höhere intrinsische Motivation angeregt werden, da die Relevanz der Vokabeln steigt. Allerdings ist fraglich ob die Relevanz alleine die Motivation hoch hält oder ob weitere äußere Faktoren eingebaut werden sollten um Nutzende langfristig zu motivieren. Auch wird von höherer Frustration bei Lernenden berichtet, bei denen die Karriere nicht abhängig vom Fremdsprachenerwerb ist. Diese Nutzende stoßen in realen Konversationen an Grenzen, die zu affektiven Filtern führen. Danach sind sie weniger gewillt, die Fremdsprache anzuwenden. Dies deckt sich mit der Theorie des Fremdsprachenerwerbs von William und Burden, die ebenfalls affektive Zustände als inneren Faktor der Motivation beim Fremdsprachenerwerb angeben. Lob durch Konversationsbeteiligte könnte hier entgegenwirken. [7, 23]

4.4 FeedLearn

Soziale Medien spielen im heutigen Leben eine große Rolle und werden von einem Großteil der Bevölkerung regelmäßig besucht. In der Arbeit von Kovacs wird der Facebook Feed zusätzlich dafür genutzt, Vokabeln zu lernen und zu wiederholen. FeedLearn integriert kleine Quizze in den Facebook Feed. Nutzende können dann zwischen verschiedenen Antwortmöglichkeiten die richtige Übersetzung auswählen. Das Quiz ist in beide Sprachrichtungen verfügbar.

In der Vorstudie wurde als Fremdsprache Japanisch gewählt. Alle sichtbaren japanischen Wörter können außerdem durch einen Klick auf einen nebenstehenden Lautsprecher laut vorgelesen werden (siehe Abbildung 4). Nach dem Beantworten erscheint an der gleichen Stelle das nächste Quiz. Es konnte festgestellt werden, dass die direkte Integration in den Feed im Gegensatz zu dem Klicken auf einen Link die Hürde senkt, mit dem Lernen zu beginnen. Auch hier wird durch eine technische Unterstützung die Hürde des Beginns einer Lektion für Nutzende minimiert. In eine Aktivität, die bereits regelmäßig ausgeführt wird–hier das Besuchen des Facebook Feeds–wird die Lernlektion integriert. Nutzende können selbst entscheiden, ob sie darauf eingehen, oder darüber hinwegscrollen. Auch hier wird das Prinzip der „spaced repetition“ angewandt, da Nutzende häufige Wiederholungen angeboten bekommen. [17]

4.5 Memreflex

Karteikarten sind ein beliebtes Mittel, um Vokabeln möglichst oft zu wiederholen. „Memreflex“-Karteikarten sind eine Applikation für das Smartphone. Sie funktionieren wie herkömmliche Karteikarten, bieten aber zusätzliche Informationen: zum Beispiel die Anzahl der noch abzuarbeitenden Karten; eine Anzeige über die Chronik der aktuellen Vokabel; eine Hochrechnung der Wahrscheinlichkeit, die aktuelle Vokabel wiederzuerkennen; eine Hochrechnung über die Anzahl der Vokabeln, die im Gedächtnis behalten werden in Zeitabständen von einer Minute, einer Stunde, einem Tag, einer Woche, einem Monat und einem Jahr. Damit wird der Fortschritt beim Lernen auch für die Nutzenden sichtbar. Edge et al. beziehen sich bei der Motivation auf die Hilfe von adaptiven Systemen, die dazu führen können, dass die Motivation sowohl kurzfristig als auch über einen langen Zeitraum gesteigert wird. Das Feedback, das Nutzende zusätzlich bekommen, in diesem Fall die Statistiken über gelernte Wörter, soll sie zusätzlich motivieren. Dies deckt sich auch mit den Erkenntnissen von William und Burden, die Feedback als äußeren Faktor für die Motivation angeben. [6, 23]

5 Diskussion

Motivation beim Lernen über einen langen Zeitraum hinweg aufrecht zu erhalten erweist sich oft als schwierig. Die vorgestellten Arbeiten bieten Lösungen an, wie kleine Lektionen in den Alltag integriert werden können, um so die Hürde des Beginns beim Lernen zu senken. Allerdings sind die Anwendungen bisher noch in keinen Langzeitstudien untersucht worden; ihre langfristige Tauglichkeit ist also noch in Frage zu stellen.

Sicherlich bietet Microlearning den Vorteil, dass die Aktivität des Lernens nicht separat begonnen werden muss, sondern direkt aus der vorher

ausgeführten Aktion in das Lernen übergeht. Beispiele wie „FeedLearn“ und „WaitChatter“ gehen an dieser Stelle mit sehr gutem Beispiel voran. Auch der „Lernschoner“ von Gassler [12] bietet die Möglichkeit kurz vor der eigentlichen Aktivität eine Sprachlektion zu absolvieren. Ob die Anbindung an den Bildschirmschoner jedoch zielführend ist, ist fraglich, nachdem Nutzende den Computer schließlich gerade nicht benutzen, wenn der Bildschirmschoner aktiviert wird. Womöglich fährt sich der Computer eher herunter, bevor Nutzende zurückkehren. Eine andere Möglichkeit, diese Lektion einzubinden, wäre zum Beispiel, sie an das Schließen von Anwendungen zu knüpfen. Dies geschieht häufiger als das Erreichen des Ruhezustands. Ob dies dann jedoch schneller nervt, muss in weiteren Studien exploriert werden.

Das Lernen an Spiele zu knüpfen, ist ebenfalls keine neue Idee. Im Beispiel des abgewandelten Tetris [4] ist ein positives Ergebnis feststellbar, allerdings ist die Zielgruppe nicht klar definiert. Ein Tetris-Nerd wird eine andere Applikation bevorzugen, bei der das Spielerlebnis nicht durch das Nichtwissen von Vokabeln beendet werden kann. Auf der anderen Seite ist es für weniger besessene Nutzer ein Anreiz dafür, mehr Vokabeln zu lernen, um nicht zu scheitern. Das Spiel Tetris bietet allerdings nur eine begrenzte Anzahl an Blöcken. Sollten also mehr Worte gelernt werden als Blockkombinationen vorhanden sind, werden Blöcke zwangsläufig doppelt besetzt. Ob dies beim Lernen hinderlich ist, wäre in weiterführenden Studien zu explorieren. Außerdem fehlt hier der Kontext zur Sprache. Ideal wäre ein Spiel mit möglichst simplem Spielprinzip, in dem Gegenstände mit passendem Kontext zur Sprache regelmäßig dargeboten werden.

Je besser das System individuell an den Nutzer angepasst ist, desto mehr Daten müssen vorher verarbeitet und ausgewertet worden sein. In Zeiten, in denen Datenschutzgesetze eine große Rolle spielen, wird es immer schwieriger solche Inhalte zu anonymisieren und vor fremden Zugriffen zu schützen. Außerdem ist für jede Anwendung eine große Community nötig, um Inhalte besonders gut mit anderen Daten abzugleichen und dann bessere Vorschläge für die Zielperson anbieten zu können.

Die App „Duolingo“ will in Zukunft besser auf Nutzende eingehen, die aufgehört haben, die Sprache zu lernen, aber später wieder einsteigen wollen [22]. Sie haben andere Bedürfnisse als neu einsteigende Nutzer.

Auch diese Nutzergruppe sollte in Zukunft berücksichtigt werden. Die Motivation konstant hoch zu halten sollte zwar das endgültige Ziel sein, für viele Nutzende ist es aber nahezu unmöglich, entsprechende Anwendungen konsequent in den Alltag einzubinden. Eigene Funktionen für diese Nutzergruppe einzubauen ist wichtig, um ihnen eine Möglichkeit zu bieten, wieder in das Lernen der Sprache einzusteigen, ohne dass sie komplett von vorne beginnen müssen.

Die Frage, ob Microlearning und automatische Generierung von Inhalten nachhaltig die Motivation steigern können, ist noch nicht klar beantwortet. Zu wenige Anwendungen wurden bisher in großem Umfang getestet und auf ihre Langzeitwirkung hin untersucht. Die Relevanz des Themas ist gegeben und ihr Potential sollte weiter erforscht werden. Ob mit solchen Anwendungen in Zukunft keine Langeweile mehr aufkommt, kann noch nicht belegt werden.

6 Zusammenfassung

Motivation hat verschiedene Ursachen. Beim Erlernen von Sprachen spielen sowohl intrinsische als auch extrinsische Faktoren eine große Rolle. Durch Spiele können intrinsische Faktoren angeregt werden, die dem Nutzer helfen, die Motivation zum Lernen einer Sprache aufrechtzuerhalten, aber auch extrinsische Faktoren spielen eine Rolle. Microlearning unterstützt den Nutzer zusätzlich, indem es kleine Aufgaben in den Alltag des Nutzers integriert, sodass er bei seinen täglichen Routinen lernen kann, ohne explizit eine Anwendung zu starten. Zusätzlich kann Microlearning mithilfe von automatischer Generierung von Inhalt die Lektionen individuell an den Nutzer anpassen, was eine weitere Steigerung der Motivation bewirken kann. Ob und wie lange diese Applikationen die Motivation aufrecht erhalten, muss in weiteren Studien exploriert werden. Dass dies aber ein Anwendungsfeld mit großem Potential ist, haben bisherige Arbeiten bereits gezeigt.

Literatur

- 1 Lyle E. Bourne and Bruce R. Ekstrand. *Einführung in die Psychologie*, chapter Motivation und Emotion. Verlag Dietmar Klotz, 1997.
- 2 V. Brandstätter, J. Schüler, R.M. Puca, and L. Lozo. *Motivation und Emotion*, chapter Klassische psychologische Ansätze als Vorläufer der modernen Motivationsforschung, pages 11–30. Springer Berlin Heidelberg, 2018. doi:https://doi.org/10.1007/978-3-662-56685-5_2.
- 3 Statistisches Bundesamt. Schüler/-innen mit fremdsprachlichem Unterricht. <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Schulen/Tabellen/allgemeinbildende-beruflicheschulen-fremdsprachl-unterricht.html>.
- 4 Carrie J Cai. Adapting arcade games for learning. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pages 2665–2670. ACM, 2013.
- 5 Carrie J Cai, Philip J Guo, James Glass, and Robert C Miller. Wait-learning: leveraging conversational dead time for second language education. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, pages 2239–2244. ACM, 2014.

- 6 Darren Edge, Stephen Fitchett, Michael Whitney, and James Landay. Mem-reflex: adaptive flashcards for mobile microlearning. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*, pages 431–440. ACM, 2012.
- 7 Darren Edge, Elly Searle, Kevin Chiu, Jing Zhao, and James A Landay. Micromandarin: mobile language learning in context. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3169–3178. ACM, 2011.
- 8 Donald Fischer. Automatic generation of content recommendations weighted by social network context, January 3 2012. US Patent 8,091,032.
- 9 Hermann-Josef Fisseni. *Persönlichkeitspsychologie*, chapter 22, pages 207–215. Hogrefe, 1998.
- 10 Hermann-Josef Fisseni. *Persönlichkeitspsychologie*, chapter 11, pages 96–106. Hogrefe, 1998.
- 11 Robert C Gardner. *Social psychology and second language learning: The role of attitudes and motivation*, chapter 3. Arnold, 1985.
- 12 Gerhard Gassler, Theo Hug, and Christian Glahn. Integrated micro learning—an outline of the basic method and first results. *Interactive Computer Aided Learning*, 4:1–7, 2004.
- 13 Robert Godwin-Jones. Emerging technologies from memory palaces to spacing algorithms: approaches to secondlanguage vocabulary learning. *Language, Learning & Technology*, 14(2):4, 2010.
- 14 Erin Jonathan Hastings, Ratan K Guha, and Kenneth O Stanley. Automatic content generation in the galactic arms race video game. *IEEE Transactions on Computational Intelligence and AI in Games*, 1(4):245–263, 2009.
- 15 Benjamin E Hosken. Automated content and collaboration-based system and methods for determining and providing content recommendations, August 20 2002. US Patent 6,438,579.
- 16 Simone Kauffeld and Carsten C. Schermuly. *Arbeitszufriedenheit und Arbeitsmotivation*, pages 237–259. Springer Berlin Heidelberg, Berlin, Heidelberg, 2019. doi:10.1007/978-3-662-56013-6_9.
- 17 Geza Kovacs. Feedlearn: Using facebook feeds for microlearning. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pages 1461–1466. ACM, 2015.
- 18 Geza Kovacs and Robert C Miller. Smart subtitles for vocabulary learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 853–862. ACM, 2014.
- 19 Richard M Ryan and Edward L Deci. Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology*, 25(1):54–67, 2000.
- 20 Noor Shaker, Georgios Yannakakis, and Julian Togelius. Towards automatic personalized content generation for platform games. In *Sixth Artificial Intelligence and Interactive Digital Entertainment Conference*, 2010.

- 21 Endel Tulving and Donald M Thomson. Encoding specificity and retrieval processes in episodic memory. *Psychological review*, 80(5):352, 1973.
- 22 Conor Walsh. Back from the brink: What duolingo learned about its resurrected users, 2017.
- 23 Marion Williams and Robert L Burden. *Psychology for Language Teachers: A Social Constructivist Approach*. ERIC, 1997.

Implizite Interaktionen durch physiologische Messungen

Maximilian B. Lammel

Ludwig-Maximilians-Universität München, München, Deutschland

Maximilian.Lammel@campus.lmu.de

Zusammenfassung

Digitale Produkte fordern heutzutage mehr und mehr Zeit und Aufmerksamkeit von uns Menschen. Das geht inzwischen soweit, dass Interaktionen mit den Geräten von der Kommunikation mit der Umgebung und den Mitmenschen ablenken. Dieses Paper stellt einen Weg vor, die Technologie wieder zurück in den Hintergrund zu bringen, wie es in Mark Weisers Vision von ubiquitous Computing vorhergesehen ist. Dazu wird der Ansatz der physiologischen Messungen genutzt, um implizite Interaktionen zu entwickeln. Verschiedene Methoden, wie der Körper ausgewertet werden kann, geben Aufschluss darüber, was machbar ist und welche Herausforderungen bei der Generierung von Daten gegeben sind. Anhand von bestehenden Beispielen werden anschließend Designprinzipien festgelegt, die eine erfolgreiche, und für den Nutzer sinnvolle, Umsetzung garantieren sollen. Dadurch wird der Fokus bei der Gestaltung von digitalen Produkten weg von der Interaktion mit dem Gerät hin zur natürlichen Interaktion mit der Umgebung geführt.

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

Keywords and phrases Ubiquitous Computing; Affective Computing; Implizite Interaktion; Physiologische Messungen; Biofeedback; NoUI.

1 Einleitung

Interaktion, die „Wechselbeziehung zwischen Handlungspartnern“ [5] ist synonym zur Kommunikation und begegnet uns Menschen eigentlich täglich. In der sozialen Interaktion ist es ein Zusammenspiel der verschiedenen Parteien und so erkennen Personen, die sich nahe stehen, oft bereits am Verhalten des Gegenübers, welche Gedanken diesen beschäftigt oder welche Bedürfnisse gerade bestehen. Das erschafft eine gewisse Vertrautheit und Nähe, wodurch man diesen Menschen mehr vertraut und zutraut, auch in dem Aspekt, dass



© Maximilian B. Lammel;

licensed under Creative Commons License CC-BY

Cite as: Maximilian B. Lammel. Implizite Interaktionen durch physiologische Messungen. In *3rd Seminar on Ubiquitous Interaction (UBIACTION 2019)*. Editors: Pascal Knierim, Jakob Karolus, Fiona Draxler, Ville Mäkelä, Tonja Machulla, Florian Lang, Thomas Kosch, Albrecht Schmidt. July 16, 2019. Munich, Germany. pp. 5:1–5:20.

5:2 Implizite Interaktion

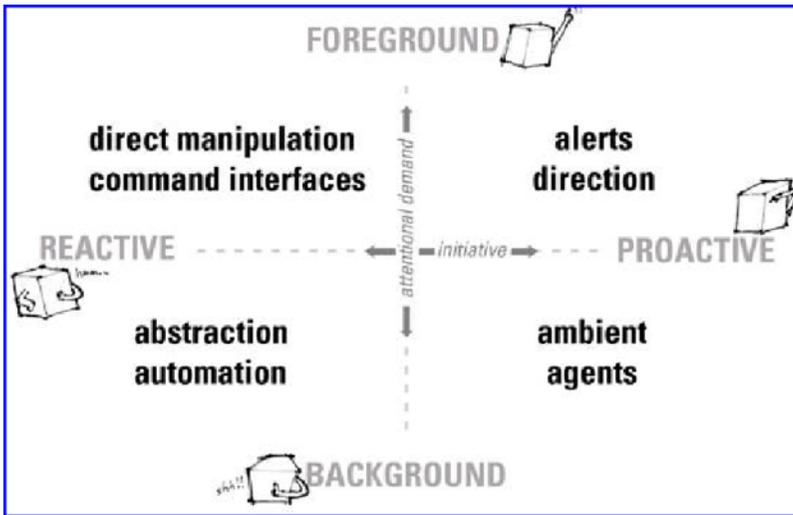
sie wissen, was man will, ohne, dass man es erst explizit zum Ausdruck bringen muss. Solche versteckten Interaktionen kann es ebenso zwischen gänzlich unbekanntem Leuten geben, beispielsweise wird von Servicemitarbeitern erwartet, dass sie sich der jeweiligen Situation anpassen. Zu welchem Zeitpunkt sollte man einem Gast im Restaurant das Glas nachschenken, wann die Rechnung bringen? Braucht ein Kunde im Modegeschäft Beratung oder sieht er sich nur um? Dabei hilft es, dass Menschen in der Interaktion meist auf die kleinen Hinweise bei der Aktion und Reaktion achten, die der Gegenüber durch sein Verhalten von sich gibt, und diese auswertet. So kann direkt auf die andere Person eingegangen werden ohne diese fragen zu müssen.

Von einem guten Service erwarten wir also, dass unsere Mitmenschen anhand des Kontexts oder der Körpersprache Situationen lesen und überflüssige Interaktionen bereits vorwegnehmen. Genauso haben wir uns inzwischen daran gewöhnt, dass technologische Geräte uns ohne Umstände weiterhelfen. Viele Türen öffnen sich automatisch, Lichter gehen von selbst an, Smart Homes und Mobiltelefone erkennen, wo der Nutzer sich befindet und passen ihre Funktionalität dementsprechend an. Dennoch ist es in der Interaktion mit Maschinen oft schwierig für die Anwendung zu erkennen, was eine Person tatsächlich vorhat und wie sie weiterhelfen kann. Dadurch kann es zu unerwarteten Situationen kommen, welche das Vertrauen in die Fähigkeiten einer Maschine zerstören. Denn wer will beispielsweise, dass das eigene Mobiltelefon während eines Meetings oder einer Präsentation klingelt [26]?

Es braucht also mehr als reine Automatisierung, um implizite Interaktion gelingen zu lassen, denn dahinter versteckt sich eine Absicht, selbst wenn diese nicht ausdrücklich kommuniziert wird [39]. Die Interaktion läuft in der Regel zwar automatisch ab und benötigt kaum Aufmerksamkeit oder Initiative, aber ist dennoch zielgerichtet. Demnach findet sich implizite Interaktion im Rahmen der Interaktion, wie in Abbildung 1 zu sehen, im Hintergrund wieder.

Explizite Interaktion hingegen befindet sich vor allem in der oberen, linken Ecke von Abbildung 1, ist also direkte Manipulation, die bewusst ausgeführt wird. Damit steht fest, dass sowohl Aufmerksamkeit benötigt wird und man die Interaktion aktiv gestalten muss, um an das gewünschte Ziel zu kommen. Heutzutage ist das meistens der schnellste Weg dorthin, bekanntermaßen sagt schon ein altes Sprichwort *wer etwas richtig gemacht haben will, der macht es selbst* und die Maschine führt in diesem Fall lediglich aus.

Denn das ist häufig der einfachste Weg - „Bist du krank? Es gibt eine App dafür! Willst du beten? Es gibt eine App dafür! Tot? Nunja, es gibt auch dafür eine App! [...] Es gibt einen besseren Weg“ (Übersetzt vom Autor) [28]. Wie Golden Krishna in seinem Buch „The best Interface is no interface“ beschreibt, geht die Tendenz noch immer dahin, Probleme mithilfe eines digitalen Interfaces zu lösen und tatsächlich wächst der Markt für Apps beständig [4]. Nicht



■ **Abbildung 1** Der Rahmen für implizite Interaktion. (Siehe [26], Seite 75)

nur durch das Smartphone, sondern im Allgemeinen sieht man mehr und mehr Displays, sei es im Auto, am Kühlschrank oder im öffentlichen Leben. Mark Weiser sprach schon vor über 20 Jahren davon, dass wir Menschen von Computern umgeben sein werden und diese ins tägliche Leben integriert sind [42]. Seine Vorstellung von Ubiquitous Computing beinhaltet jedoch mehr als die reine Nutzung von Bildschirmen, die an traditionellen Produkten befestigt werden, um sie dadurch zu *verbessern*. Denn wir Menschen passen uns immer mehr an Maschinen an und verbringen inzwischen viel Zeit damit, auf deren Anfragen einzugehen [26]. Stattdessen sollte zu jedem Zeitpunkt das richtige Gerät für die passende Situation bereit stehen, wodurch Technologie in den Hintergrund gerückt wird und immer mehr verschwindet [42].

Obwohl noch nicht alles ubiquitär ist gibt es Stand heute viele neue Möglichkeiten, die sich durch die Allgegenwärtigkeit von Computern ergeben [38] und die man ausnutzen kann. Passen sich diese Maschinen jedoch nicht an die Gegebenheiten an, so erscheinen sie relativ *dumm*, denn die immer steigenden Erwartungen des Nutzers werden nicht erfüllt. Um individuell auf den jeweiligen Nutzer eingehen zu können, kann man also nicht den Ansatz *One size fits all!* verfolgen, sondern muss sich mit der Person dahinter auseinandersetzen und deren Beweggründe, Absichten und Emotionen ermitteln.

Abhilfe schaffen können hier physiologische Sensoren, die den „psychologischen Zustand des Nutzers an ein adaptives System kommunizieren“ (Übersetzt vom Autor) [21]. „Man kann nicht nicht kommunizieren“ [41] lautet einer der

Grundsätze des Psychologen Paul Watzlawick und tatsächlich spiegelt unser Körper meist am besten, was in uns vorgeht. Analysen von Emotionen und Biodaten können Aufschluss geben und es einer Maschine dadurch ermöglichen, uns etwaige Interaktionsschritte abzunehmen [24].

In den folgenden beiden Kapiteln wird zuerst verstärkt auf die psychophysiologische Messung eingegangen, mit welchen Methoden Daten generiert werden und wie sich das auf die Technologie auswirkt. Anschließend werden Projekte aufgezeigt, die diese Technik ausnutzen, um verbesserte Erlebnisse zu schaffen, und welche Lehren daraus gezogen werden. Im letzten Kapitel wird diskutiert, welche Probleme damit gelöst werden und welche Herausforderungen aufkommen können, bevor eine Zusammenfassung die Arbeit abschließt.

2 Die Auswertung des menschlichen Körpers

Um ganz gezielte Anpassungen an den Menschen vorzunehmen ist es für Computer notwendig, Daten über diesen zu besitzen. Das sind auf der einen Seite die Daten, die fast zu jeder Zeit bei der Nutzung von technischen Geräten gesammelt werden, wie die Verweildauer auf Webseiten oder der Standort. Zusammen mit der Umgebung können diese bereits den Rahmen vorgeben, in welchem eine Interaktion stattfindet und sind deshalb von großem Vorteil bei der Automatisierung. Auf der anderen Seite können Daten interessant sein, welche direkt durch Betrachtung und Auswertung des Körpers erhoben werden. Diese Biodaten geben einen tieferen Einblick in das menschliche Innenleben, wodurch individuell auf die einzelne Person und Situation eingegangen werden kann. Ein bekanntes Beispiel hierfür sind Armbänder oder Smartwatches, wie Fitbit¹, welche als Fitnessstracker genutzt werden und dann Informationen über den eigenen Gesundheitszustand liefern. Selbst ohne Interaktion kann es spannend sein, mehr über sich selbst zu erfahren und gegebenenfalls Anpassungen im eigenen Leben durchzuführen. Doch Wearables sind lediglich eine Möglichkeit, wie Daten generiert werden können. Messungen der Psychophysiologie, einem „Teilgebiet der Psychologie, das sich mit den physiologischen und biochemischen Grundlagen psychischer Vorgänge befasst“ [7], können also Aufschluss über Körper und Geist eines Menschen geben.

2.1 Psychophysiologische Messungen

Der Mensch drückt sich nicht nur durch seine Aktionen aus, sondern ebenso durch seinen Körper. Denn dieser offenbart emotionale und körperliche Signale, die mithilfe von geeigneten Geräten messbar gemacht werden können [22]. Das

¹ <https://www.fitbit.com>

Feld der Psychophysiologie beschäftigt sich damit, ebene Signale durch Messungen zu analysieren und zu interpretieren, um daraus Ableitungen treffen zu können [13]. Fest steht, dass Körper und Geist eng miteinander verknüpft sind, weshalb die Auswertung des einen Teils auch einen Einblick in den Anderen ermöglicht.

Um herauszufinden, was im Körper eines Menschen vor sich geht, werden meist Sensoren benötigt, die am Körper festgemacht werden und kontinuierlich messen, was sich je nach Situation verändert. Biofeedback ist eine Technik, die die gemessenen Biodaten über einen Computer an den Nutzer weitergibt, um es diesem zu ermöglichen, eigene Anpassungen durchzuführen. Eine Untergruppe davon, welche sich speziell mit der Auswertung des menschlichen Gehirns auseinandersetzt, ist Neurofeedback. Beide Techniken ermöglichen es, Daten über den Menschen zu generieren, die sich weiterverarbeiten lassen, ganz gleich ob lediglich als Feedback für den Nutzer selbst oder als neuen Input für den Computer [24].

Im folgenden Abschnitt werden einige der Methoden vorgestellt, mit deren Hilfe biologische Daten erhoben werden können. In Abbildung 2 ist zu erkennen, wie die einzelnen Methoden aussehen können und welche Unterschiede bestehen. Vereint sind die Methoden allesamt in dem Sinne, dass sie nicht invasiv sind und dadurch eine höhere Chance besteht, dass sie sich in den Alltag integrieren lassen.



■ **Abbildung 2** Verschiedene Arten der psychophysiologischen Messgeräte: Smartwatch [6], Eyetracking [3], EMG [2], EEG [1](von links im Uhrzeigersinn)

2.1.1 Herzfrequenz und Blutdruck

Die Herzfrequenz und der Blutdruck sind im Alltag wohl die bekanntesten Biodaten, denn diese kann mithilfe von Fitnesstrackern oder Smartwatches bereits jeder an sich selbst messen und auswerten. Aus dem daraus gewonnenen Feedback kann jeder Nutzer selbst Schlüsse ziehen und Anpassungen für die eigene Gesundheit vornehmen. Obwohl daraus meist Aussagen über das Gesundheits- oder Fitnesslevel getroffen werden, ergeben sich ebenso Interaktionsmöglichkeiten. Werden Aufregung oder Stress gemessen so kann dies als Indikator genutzt werden, dass bei einer Tätigkeit Unterstützung notwendig ist und das ohne dass die betroffene Person selbiges äußern muss.

2.1.2 Elektroenzephalografie (EEG)

Es gibt verschiedene Wege um Gehirnaktivität zu messen, jedoch wird in dieser Arbeit der Fokus auf EEG gelegt, da diese Methode weder invasiv ist, noch übermäßig teuer und stationär im Vergleich zu Verfahren wie Magnetresonanztomographie (MRT). Zwar ist die Genauigkeit bei anderen Methoden oft besser und weniger störanfällig, allerdings ist EEG durch die zuvor genannten Eigenschaften am besten dazu geeignet in alltägliche Technologie eingebunden zu werden, zum Beispiel in Kombination mit einer Virtual Reality Brille [15].

Über Elektroden, die am Kopf befestigt werden, kann elektrische Aktivität im Gehirn gemessen werden und über einen längeren Zeitpunkt werden so die Gehirnströme abgebildet. Die zu sehenden Wellen können interpretiert werden, indem man Ausschläge mit dem Geschehenen vergleicht, um dadurch beispielsweise Erregung zu messen [10]. Zum Teil ist es selbst für geschultes Personal schwer, Muster in Gehirnaktivitäten zu entdecken und diese zu verstehen [24], weshalb meist ein größerer Zeitraum betrachtet werden muss, um die richtigen Schlüsse zu ziehen. Dann können allerdings zusätzlich Muster erkannt werden, über die sich der Nutzer selbst eventuell nicht im Klaren war.

2.1.3 Elektromyografie (EMG)

Mit Hilfe von EMG kann die Aktivität der Muskeln eines Menschen gemessen werden. Diese Methode nutzt im nicht invasiven Fall Elektroden, die auf der Haut über dem untersuchten Muskel befestigt werden. Anschließend kann der Zustand des Muskels ermittelt werden, indem der Ruhezustand bzw. die Kontraktion betrachtet wird. Je nach Muskelgruppe und Zustand können daraus Ableitungen getroffen werden, welche auf die generelle Gemütslage schließen lassen können. Dadurch

können digitale Produkte auf der einen Seite vom Nutzer durch propriozeptive Interaktion selbst gesteuert werden, indem er seine Haltung anpasst, um gewünschte Aktionen herbeizuführen [30]. Auf der anderen Seite können unbewusste Muskelbewegungen aufgedeckt werden, beispielsweise der Mimik [19], welche die eigentlichen Absichten der Person signalisieren. Demnach kann man, wie die Menschen es ebenfalls tun, von der Körperhaltung und Gestik viele Informationen ablesen.

2.1.4 Eye-Tracking

Dem Eye-Tracking wird in der Mensch Computer Interaktion großes Potenzial zugesprochen, sowohl in der direkten Manipulation, als auch in der impliziten Interaktion. Über die Augenbewegungen einer Person kann man die menschlichen Emotionen ablesen und Aussagen treffen, an was diese gerade interessiert ist [24] und was sie als nächstes vorhat. Das ist unter anderem an der Tatsache zu erkennen, dass der Blick Interesse signalisiert [34] und der Umsetzung von Aktionen vorausgehen kann [33]. Das hat zum Vorteil, dass man sofort auf die Augenbewegungen eingehen könnte und daraus Ableitungen ziehen kann, jedoch ist Eye-Tracking oft vage und benötigt deshalb zuerst eine Kalibrierung zwischen Augen und Tracker. Aber gerade im öffentlichen Umfeld ist es häufig so, dass sich die Menschen auf ihre Augen verlassen und indirekt tatsächlich durch sie kommunizieren, weshalb Eye-Tracking dort eine besondere Rolle einnehmen kann.

2.1.5 Elektrodermale Aktivität (EDA)

Informationen über Stress und kognitive Aktivität liefert die EDA, außerdem bekannt als Hautleitwert. Schweiß am Körper, bei der Messung vor allem der an der Hand, kann Klarheit geben, in welchem Zustand sich ein Mensch befindet, denn mit steigender Intensität oder Aufregung fängt man an, vermehrt zu schwitzen. Daran erkennt man beispielsweise, ob jemand lügt, aber es wird ebenso für die Bestimmung von Krankheitsbildern genutzt [24]. Denn der Schweiß lässt sich, da er ein elektrischer Leiter ist, gut durch Messgeräte aufzeichnen und daraufhin auswerten. Dabei reichen bereits kleine Änderungen, um gezielt Aussagen treffen zu können. Diese Methode ist zwar nicht speziell in Abbildung 2 aufgezeigt, jedoch eignen sich hier besonders Sensoren an Hand und Fingern, wie im linken Bild.

All die gesammelten Daten sind schon allein als Informationsquelle und Feedback interessant, um die betreffende Person evaluieren zu können. Mehr

über sich selbst zu erfahren, ist ein wichtiger Schritt um Verbesserungen vornehmen zu können und Kontrolle über den eigenen Körper zu gewinnen. Genauso können Experten davon profitieren, solche Biodaten auszuwerten, Muster zu erkennen und Ableitungen zu treffen. Dabei rangieren die Anwendungsfelder vom Leistungssport und der Verbesserung des Körpers bis hin zu Medizin und der Erkennung von Krankheiten.

Zusätzlich zur Bestimmung des Körpers können aus diesen Daten und durch weitere Methoden Aussagen über den emotionalen Zustand einer Person gemacht werden. Mit der Aufgabe, diese Emotionen in der Technologie nutzbar zu machen, beschäftigt sich das Feld von *Affective Computing*.

2.2 Emotion in der Technologie

Während manche Daten des Körpers klar darüber Aufschluss geben, was gerade vor sich geht, ist das bei der Betrachtung des Geistes oft nicht so deutlich. Verschiedene Faktoren befinden sich im Zusammenspiel, wodurch es zu Störfaktoren kommt, die die Auswertung beeinflussen können. Deshalb ist nicht gesagt, dass ein Wert in den Daten eins zu eins auf die dahinter liegende Stimmung abbildbar ist. Ein Blick auf ein Objekt etwa kann Interesse signalisieren oder im Tagtraum ohne Fokus geschehen. Ein Lächeln im Gesicht kann auf gute Laune hinweisen oder nur für eine Erinnerung an einen alten Witz stehen. Tatsächlich ist es schwer zu einem eindeutigen Ergebnis zu kommen, weshalb bei der Auswertung unter anderem auf künstliche Intelligenz zurückgegriffen wird. Rosalind Picard beschreibt schon 1997 in ihrem Buch „Affective Computing“, dass intelligente Maschinen unsere Emotionen erkennen und verstehen sollten, um letztendlich richtig mit uns Menschen zu interagieren [35]. Die Möglichkeiten, große Mengen an Daten, wie Tonaufnahmen oder Bilder und Videos von Gesichtern, auswerten zu können, zeigt zudem das bestehende Potenzial künstliche Intelligenz und Affective Computing zu verknüpfen. Denn obwohl Emotionen oft als untergeordneter Teil der Körper und Geist Beziehung betrachtet werden [23], sind sie doch ein großer Teil in der Betrachtung der gesamten Einheit. Diese Gesamtbetrachtung spielt eine Schlüsselrolle, um die wahren Intentionen einer Person zu erkennen und ist deshalb nötig, um natürliche Interaktionen zu schaffen, denn hierbei können schon kleine Fehler zu Unverständnis beim Nutzer führen.

2.3 Implizite Evaluation

„Durch die Nutzung von Hirn-, Körper-, Verhalten- und Umweltsensoren ist es nun möglich, subtile Veränderungen des kognitiven Zustands eines Nutzers in Echtzeit zu erfassen“ (Übersetzt vom Autor) [40]. Die Kombination von verschiedenen Auswertungen erhöht die Chancen, die richtigen Schlüsse zu

ziehen und dadurch die richtigen Aktionen einzuleiten. Implizite Interaktion wird also erst durch die passende Evaluation effektiv, weshalb sich die Frage stellt, wie man diese Untersuchungen durchführen kann ohne zu sehr in die Privatsphäre der untersuchten Person einzudringen. Nötig ist also eine implizite oder versteckte Evaluation, wodurch es dem Nutzer ermöglicht wird, sich natürlich zu verhalten, denn nur so kann gewährleistet werden, dass die Daten mit der Wirklichkeit übereinstimmen. Stand heute sind viele der oben genannten Methoden zwar nichtinvasiv, jedoch trotzdem nur schwer in den Alltag zu integrieren. Denn dafür müssten die Messungen mobil, schnell und ohne Beeinflussung durch die Umgebung durchführbar sein. Der einfachste Weg wäre, die Evaluation in bestehende Gegenstände zu integrieren, um möglichst wenig am Bestehenden zu verändern. Genutzt wird diese Technik bereits bei der Auswertung eines Nutzers vorm Bildschirm. Mausclicks und Tastatureingaben lassen beispielsweise auf Interesse schließen [18] und Webseiten lassen sich daraufhin implizit an den einzelnen Nutzer anpassen, abhängig von dessen Aktionen [12]. Vor dem PC können Eye-Tracking oder andere Messgeräte ebenfalls effektiv eingesetzt werden, denn hierbei ist ein stationäres Umfeld gegeben, welches klare Strukturen vorgibt. Weg vom Bildschirm ist es deutlich schwieriger an die geforderten Daten zu kommen, wodurch auf andere Mittel zurückgegriffen werden muss. Kameraaufnahmen können hier hilfreich sein, um ein authentisches Experiment zu schaffen. Mithilfe von versteckten Aufnahmen kann man an spontane und natürliche Gesichtsausdrücke gelangen [11]. Eine andere Möglichkeit ist eine Verknüpfung mit Wearables. Diese können zu jeder Zeit Messungen am Körper selbst durchführen und nützliche Daten gegebenenfalls an verbundene Geräte weiterleiten. Auf Abbildung 2 ist schnell zu erkennen, dass die Methoden bisher noch nicht alle im Alltag integriert sind, obwohl eine erfolgreiche Messung eigentlich voraussetzt, dass die Daten unbemerkt aufgenommen werden und damit keinen Einfluss auf das Geschehen nehmen. Die Datengenerierung ist ein offenes Problem, unter anderem wegen dem Datenschutz und der Privatsphäre jedes einzelnen. Es muss also zwischen Datenschutz und Nutzbarkeit abgewogen werden. Denn erst durch die richtige Art und Weise an Daten zu gelangen ist eine implizite Interaktion wirklich möglich, weil erst dann Natürlichkeit besteht.

3 Implizite Interaktion durch Biofeedback

Vernachlässigt man die Probleme der Datengenerierung vorerst, dann kann man feststellen, dass bereits viel Forschung im Bereich der physiologischen Messungen und anschließender Interaktion betrieben wurde. Denn zumindest in Laborumgebung kann man auf die Geräte zur Datenakquise zurückgreifen und demnach bereits testen, wie Menschen sich verhalten, wenn ihre Bioprofil

zur Interaktion genutzt werden. Denn für viele Nutzer ist dieser Eingriff und Interaktionsmethode gänzlich unbekannt.

Dieses Kapitel befasst sich mit Arbeiten, die zu dem Thema durchgeführt worden sind, beginnend mit Ideen aus der Computerspielbranche. Die Lehren, die daraus gezogen werden können, werden anschließend mit den Projekten, die alltäglichen Bezug haben, verglichen, um im letzten Unterpunkt Richtlinien für das Design von impliziten Interaktionen mit Hilfe von Biodaten aufzuzeigen.

3.1 Spielend lernen

Die Spieleindustrie lebt davon, der Konkurrenz einen Schritt voraus zu sein und innovative Produkte voranzutreiben und ist deshalb oft Vorreiter in der Forschung. In diesem hoch kompetitiven Umfeld kann eine Neuheit einen großen Vorteil mit sich bringen, weshalb ständig probiert wird, eine neuartige Technologien für sich zu nutzen. Kein Wunder also, dass bereits seit den 1980er Jahren unter anderem an Biofeedbacktechniken geforscht wird [31]. Die Branche profitiert dabei von dem kontrollierten Umfeld, das bei Computerspielen vorherrscht. Es ist vorgegeben, welche Controller und Geräte der Nutzer bedienen kann, genauso wie die Spielregeln, die vom Entwickler je nach Bedarf angepasst werden. In dieser *laborartigen* Umgebung im eigenen Zuhause sind verschiedene Szenarien und Tests durchführbar, wodurch Feedback von den Spielern selbst eingeholt werden kann. Da die Spieler bereits an Controller gewöhnt sind, ist es möglich, diese neben der Eingabefunktion zusätzlich als Sensoren zu benutzen. Aufgrund dessen kann man ohne große Änderung des Spielablaufs neue Tools ausprobieren und gleichzeitig ein erweitertes Spielerlebnis schaffen. Ein Beispiel dafür liefern Nacke et al., die mit Hilfe von direkten und impliziten physiologischen Eingaben das Spielgeschehen beeinflussen [31]. Während die expliziten Eingaben mehr Kontrolle beim Spielen zulassen, kann man durch die passive Sensoren intelligente Adaptionen durchführen, um das Spiel an den Zustand des Nutzers anzupassen. Beispielsweise kann sich die Welt im Hintergrund verändern oder aber die Schwierigkeit kann sich erhöhen, falls der Spieler unterfordert ist. Ansätze wie diesen verfolgt beispielsweise Valve Software², die in ihrem Spiel „Left4Dead“, einem Ego-Shooter gegen Zombies, die Erzeugung von Gegnern dynamisch an den Spielers anpasst [10]. Je nachdem wie erregt der Spieler ist kommen mehr oder weniger Gegner neu hinzu. Ein weiteres Projekt von Valve Software bezieht sich auf den auf Puzzle basierenden Ego-Shooter „Portal 2“, bei dem die normale Steuerung mit der Maus für Testzwecke durch Eye-Tracking ersetzt wurde. Physiologische Signale sind als Eingabe zwar einsetzbar, aufgrund der Kosten jedoch noch

² <https://www.valvesoftware.com>

nicht für den Konsumenten umzusetzen, so lautete letztendlich das Fazit [10]. Die Umstellung für den Spieler selbst ist also nicht das Problem, solange es eine natürliche Interaktion beinhaltet. Im Gegenteil kann durch den entstehenden *Novelty Effekt* ein Vorteil entstehen, da viele Spieler gerne etwas neues ausprobieren. Insgesamt bedeutet das, dass bei Sequenzen, in denen schnelle Reaktionen nötig sind, direkte Manipulation von Vorteil ist, eine implizite Auswertung der Signale aber durchaus zur Stimmung und damit zum Erfolg des Spieles beitragen kann [24]. Dennoch ist die implizite Nutzung von physiologischen Daten schwierig und muss eine Balance zwischen Schwierigkeit und Spaß für den Spieler erschaffen [29].

Die Ideen aus den Spielen lassen sich zwar nicht eins zu eins in die Wirklichkeit übertragen, da dort andere Regeln gelten, jedoch kann man aus den Lernerfahrungen Ableitungen für das *echte* Leben generieren. Die Nutzung von psychophysiologischen Signalen kann bei der Interaktion durchaus positiv gebraucht werden, aber insbesondere bei der impliziten Interaktion muss man darauf achten, den Nutzer nicht zu überfordern. Vorweggenommene Interaktion oder eine unterstützende Wirkung sind demnach besser als Funktionen, bei denen vom Nutzer gefordert wird, sich in einen bestimmten Zustand zu versetzen [31]. Obwohl sich also einige Punkte durchziehen, kommen in der Realität noch weitere Herausforderungen hinzu, welche untersucht werden müssen.

3.2 Implizite Interaktion im echten Leben

Obwohl die Gegebenheiten in der Realität nicht immer klar definiert sind, gibt es dennoch Situationen, in denen man die Rahmenbedingungen besser einschätzen kann. In diesen Szenarien, beispielsweise im Auto, sind Umfeld und Kontext gegeben, wodurch die Interaktionsmöglichkeiten beschränkt sind. Außerdem können aufgrund dessen die Messgeräte platziert werden, sodass sie möglichst im Hintergrund bleiben. Diese beiden Faktoren ermöglichen sowohl die Generierung von Daten, als auch die Verwendung derselbigen, wie an den folgenden Beispielen zu sehen ist.

- **Mobilität:** Wenngleich die Automatisierung des Autos vorangetrieben wird spielt trotzdem noch der Fahrer selbst die zentrale Rolle. Durch Unterstützungen, wie Spurwechselassistenten, Lichtautomatik oder Abstandswarnern, ist dieser jedoch schon lange nicht mehr auf sich allein gestellt. Dennoch liegt die Verantwortung laut dem Übereinkommen über den Straßenverkehr, welches 1968 in Wien abgeschlossen wurde, beim Fahrer, denn „jeder Führer muss dauernd sein Fahrzeug beherrschen oder seine Tiere führen können“ [8]. Hier kann sowohl Biofeedback ebenso wie implizite Interaktion ansetzen, um das Sicherheitsrisiko zu minimieren.

5:12 Implizite Interaktion

Die Auswertung des Fahrzeugführers gibt Aufschluss, ob und wie dieser unterstützt werden kann oder wann eine Weiterfahrt Risiko mit sich bringt. Riemer et al. konnten durch Messungen der Herzfrequenzvariabilität von Pendlern Stresssituationen auf die gefahrene Route abbilden, um bei Abweichungen von der Norm Warnungen auszusprechen [37]. Ein wichtiger Aspekt bezüglich der Sicherheit ist die Müdigkeit des Fahrers. Diese kann zum einen über Eye Tracking festgestellt werden [20] und zum anderen über die Körperhaltung [16]. Schnell wird deutlich, wie viele Möglichkeiten es gibt physiologische Messungen im Auto durchzuführen und mit dem Fahrer zu kommunizieren, sei es durch Kamera [20], Sitz [16] oder Lenkrad [14]. Neben Warnungen und Unterstützungen kann man direkt und individuell auf den Fahrer eingehen, indem man erkennt, wer das Fahrzeug steuert [36]. Denn in Zukunft wird die Herausforderung weniger die Kontrolle des Fahrzeugs sein und mehr die Kooperation zwischen Fahrer und Auto, insbesondere das Vorhersehen von Aktionen und die Bedeutung von Aufmerksamkeit [25].

- **Luftfahrt:** Ein ähnliches Szenario bietet die Luft- und Raumfahrt, bei der Piloten in einem Cockpit vor klare Anforderungen gestellt werden. Der Spielraum für Fehler ist gering, was mitunter zu hohen Stresssituationen führen kann, insbesondere wenn Passagiere an Bord sind. Deshalb bedarf es Unterstützung ebendiese Situation zu vermeiden. Analog zum Auto kann man die Messgeräte im Cockpit in die bestehenden Geräte einbinden ohne zusätzliche, kognitive Last für den Piloten zu erzeugen. Causse et al. haben durch Messungen nachgewiesen, dass Stress die Entscheidungen beim Fliegen nachhaltig beeinflussen kann [17]. Genau das sollte allerdings vermieden werden, weshalb es wichtig ist, diese Situationen zu erkennen und abzuwenden. Der entscheidende Punkt im Cockpit ist es, die Lage schnellstmöglich zu erfassen und die richtige Entscheidung treffen zu können [25]. Deshalb kann eine Automatisierung anhand der Bedürfnisse des Piloten einen großen Beitrag bei der Reduzierung der kognitiven Last leisten, um dadurch die Sicherheit aller Beteiligten zu erhöhen.
- **Smart Home:** Feste Strukturen sind ebenfalls im eigenen Haus zu finden. Gearbeitet wird am Schreibtisch, gekocht wird am Herd und geschlafen im Bett. Smart Home Geräte wie Tado ³ erkennen bereits ob jemand zuhause ist und passen sich dementsprechend an, im Falle von Tado die Raumtemperatur. Die Zeiten, dass man sich fragen musste, ob man den Herd ausgeschaltet hat, sind vorbei, denn durch diese Geräte kann sich der Nutzer auf die wichtigen Sachen fokussieren. Vergleichbare Anpassungen kann man ebenso durch Biofeedback steuern, wenn Aufgabe und Lage

³ <https://www.tado.com/>

vorgegeben sind. Ein Bett, das Atmung und Herzfrequenz aufnimmt [27], könnte durch Schlafüberwachung den richtigen Zeitpunkt zum Aufwachen finden, eventuell besser als es bereits von Apps⁴ gelöst wird. Aufgrund von Biodaten könnte auf fehlende Nährstoffe hingewiesen werden, welche über einen smarten Kühlschrank zudem gleich bestellt werden können. Feste Standorte der Geräte und klar verteilte Aufgaben ermöglichen es im Smart Home, Daten zu generieren und diese sinnvoll zu nutzen.

Die Möglichkeiten sind in diesen Bereichen bereits gegeben, weshalb es logisch ist, diese zu kombinieren und das Potenzial voll auszuschöpfen. Dadurch kann die Technologie weiter in den Hintergrund verschwinden, um Platz für das wesentliche zu schaffen.

In der *freien Wildbahn* liegen andere Gegebenheiten vor, weshalb dort eine weitere Herausforderung wartet. Neben den unzähligen Möglichkeiten zur Interaktion ist es schwierig, Daten zu beschaffen, denn dazu wird meist eine Nähe zum Nutzer benötigt, welche in der Öffentlichkeit nur selten gegeben ist. Fordert man diese Nähe ein, verliert man schnell den Vorteil der impliziten Interaktion und damit den Vorteil gegenüber einer simplen Abfrage des Nutzers, die dieser selbst angibt. Deshalb sollen einige Design Prinzipien helfen, implizite Interaktionen und den Umgang mit physiologischen Messungen richtig zu gestalten.

3.3 Design Prinzipien

Anhand der vorangegangenen Beispiele sollte schon deutlich geworden sein, dass es bei der Umsetzung von impliziter Interaktion mit, und selbst ohne, Biofeedback durchaus Stolpersteine geben kann und schnell Fehler geschehen. Im folgenden werden drei Prinzipien vorgestellt sind, die bei der Entwicklung von impliziten Interaktionen durch physiologische Messungen essentiell sind.

- **Datengenerierung im Hintergrund:** Implizit setzt voraus, dass die Daten ebenfalls im Hintergrund generiert werden sollten [9]. Denn kaum einer ist vollkommen entspannt, wenn Sensoren zur Auswertung direkt am Körper befestigt werden und aufgrund dessen sollten diese Sensoren immer mehr mit der Umwelt verschmelzen, sei es durch Wearables oder in den Interaktionsgeräten selbst. Denn obwohl die vorgestellten Methoden zur Messung nicht invasiv sind, nehmen sie im jetzigen Zustand doch Einfluss auf den ausgewerteten Menschen und damit auf die Ergebnisse.
- **Datenauswertung hinterfragen:** Sind die notwendigen Daten erst einmal gegeben gibt es dennoch die Herausforderung der richtigen Auswertung.

⁴ <https://www.sleepcycle.com/>

5:14 Implizite Interaktion

Fehler bei der Aufnahme und Beurteilung der Daten führen zu Fehlern bei der nachfolgenden Interaktion, was wiederum ein großes Maß an Frustration beim Nutzer mit sich bringt [25]. Deshalb muss abgeschätzt werden, ob direkt Einfluss auf die Situation genommen werden kann oder ob sich die Anpassungen eines Geräts erst auf zukünftige Interaktionen mit dem Nutzer fokussieren, wodurch eine genauere Auswertung der Daten ermöglicht wird [40].

- **Fokus auf die Akteure:** Einer der wichtigsten Punkte beim Design von impliziten Interaktionen ist das Zusammenspiel der Akteure [26]. Im Mittelpunkt steht letztendlich nicht die Interaktion, sondern das Erreichen der Ziele des Nutzers. Um diese zu erreichen müssen die beteiligten Parteien das gleiche wollen und das auch so kommunizieren. Dadurch wird gewährleistet, dass die Interaktion nicht von der einen Seite erzwungen wird und stattdessen ebenjenes Zusammenspiel gefördert wird.

Durch die Einhaltung dieser Prinzipien ist bereits ein großer Schritt hin zur Entwicklung sinnvoller Produkte getan. Deshalb eignen sich, wie in den Beispielen zu sehen, besonders Interaktionen, die ein klares Ziel verfolgen. Denn hierbei ziehen Nutzer und Produkt gemeinsam an einem Strang, wodurch selbst die Aufnahme von Biometrie nachvollzogen werden kann. Denn ob die Vorteile dieser Datenauswertung überwiegen oder doch die Privatsphäre Vorrang hat, ist ein offener Streitpunkt, welcher im nächsten Kapitel thematisiert werden soll.

4 Diskussion

Die Nutzer werden nur aufgrund eines gefühlten oder tatsächlichen Mehrwerts in Kauf nehmen, dass die Daten ihres Körpers ausgewertet werden. Obwohl die Aufzeichnung von Daten während der Nutzung eines Geräts inzwischen zum Standard geworden ist, wird der eigene Körper doch noch als privat angesehen. Wem gehören meine Daten, insbesondere die meines Körpers? Es muss also die Frage beantwortet werden, welchen Mehrwert die implizite Interaktion durch physiologische Messungen mit sich bringt und wann diese Technik sinnvoll eingesetzt werden kann. Um einen positiven Effekt zu erzielen dürfen die Anwendungen dem Nutzer weder physisch schaden, noch unbedarft mit dessen Daten umgehen. In den folgenden Abschnitten soll ein Ausblick darüber gegeben werden, wie sich das auf die Zukunft auswirken könnte, sowohl die positiven Aspekte als auch die negativen.

4.1 Interaktion ohne Bildschirm

Eingangs in Kapitel 1 wurde bereits von Weisers Vorstellung von ubiquitous Computing und der großen Anzahl von Computern und Bildschirmen gesprochen. Allerdings befinden sich diese nicht im Hintergrund der Welt, sondern kämpfen gegeneinander um unsere Aufmerksamkeit und unsere Zeit. Denn längst ist das zur wertvollsten Ware geworden, die ein Nutzer mit sich bringt. Tatsächlich ist es für digitale Produkte meist dann profitabel, wenn sich der Nutzer oft und lange auf der Plattform aufhält. Denn mit Zeit kommen Daten und Geld. Deshalb werden zum Teil Abhängigkeiten geschaffen, die vom eigentlichen Leben ablenken und den Fokus voll auf die Interaktion mit dem Gerät legen[32]. Das Smartphone und die verschiedenen Applikationen darauf sind inzwischen zum ständigen Begleiter geworden, welcher in jeder freien Minute um unsere Aufmerksamkeit buhlt. Doch nicht nur das eigene Smartphone ist schuld. Blinkende Displays rücken in unser Blickfeld und akustische Signale verdrehen uns den Kopf. Alles für eine erste Interaktion mit dem Menschen und potenziellen Kunden. Neben der digitalen Produktbranche spricht man von Nutzern lediglich im Zusammenhang mit Drogen, was zu denken geben sollte. Die Beispiele aus Kapitel 3 zeigen, dass das nicht nötig ist und Anwendungen, die sich auf ihre Hauptaufgabe fokussieren, einen großen Mehrwert mit sich bringen. Mit ihrer Hilfe kann Zeit gespart werden und gleichzeitig wird ein Erlebnis geschaffen, welches nicht die Technologie, sondern den Menschen in den Mittelpunkt setzt. Ebenso kann durch die gewonnene Zeit bei der impliziten Interaktionen Platz für das Wesentliche geschaffen werden und Zeit ist nicht nur für digitale Produkte wertvoll, sondern vor allem für den Menschen selbst.

4.2 Datenflut

Schon jetzt werden immens viele Daten über den Menschen gesammelt, doch mit der Akquirierung von Biodaten wird eine neue Stufe erreicht. Natürlich muss hinterfragt werden, wie man an diese Daten gelangt. Wird der Mensch ohne seines Wissens über Kameras oder andere periphere Aufnahmegeräte untersucht, dann ist das ein starkes Eindringen in die Privatsphäre. Oder man überlässt dem Nutzer selbst die Entscheidung, ob die Daten aufgenommen werden, beispielsweise mittels Wearables. So oder so stellt sich die Frage, wie diese Daten verwendet werden. Auf der einen Seite aus Gründen der Sicherheit, wie bei der Mobilität, kann argumentiert werden, dass es notwendig ist, den besten Ansatz zu verfolgen und jede Möglichkeit auszunutzen. Auf der anderen Seite wird Privatsphäre immer mehr zum Luxusgut, welches sich nur bestimmte Leute leisten können. Der Rest findet sich damit ab, dass die eigenen Daten zu allerlei Zwecken genutzt werden und oft ist es inzwischen sowieso komfortabler, wenn bereits Informationen preisgegeben worden sind,

5:16 Implizite Interaktion

beispielsweise wenn Vorlieben schon bekannt sind. Aber ob Komfort ein Grund für die Aufgabe seiner persönlichen Daten ist, ist ethisch zu hinterfragen, denn sind die Entscheidungen dann noch die eigenen? Lässt sich der Mensch überhaupt auf Variablen reduzieren oder ist er mehr, als man durch einzelne Messungen herausfinden kann? Eventuell sind die Gedanken und Emotionen doch ganz im Inneren versteckt, weshalb man Vorsicht walten lassen muss, um durch physiologische Messungen nicht einfach eine weitere Möglichkeit zu schaffen, Daten für Werbung zu generieren.

4.3 Ausblick

In Zukunft wird sich zeigen, ob die möglichen Vorteile Wirklichkeit werden können, aber dafür muss sich jeder, der daran arbeitet, für den richtigen Umgang mit den gewonnenen Daten einsetzen. Für die richtige Umsetzung wird ein interdisziplinäres Team aus Interaktionsdesignern, Entwicklern und Psychologen benötigt, die zusammen an der Lösung der aufgezählten Probleme arbeiten. Die oben aufgezeigten Beispiele sind Anhaltspunkte, in welchen Bereichen diese Technik Einsatz findet, aber ebenso wird es andere Anwendungsfelder geben, wie beispielsweise in der Medizin.

Zum Abschluss werden die wichtigsten Punkte der Arbeit zusammengefasst, zudem wird erneut auf das große Potenzial der Psychophysiologie hingewiesen und dem Appell, dieses voll auszuschöpfen, jedoch nur im Sinne der Nutzer.

5 Zusammenfassung

Die Abwägung zwischen nutzbaren Daten und Privatsphäre, genauso wie die letztendliche Verwendung der Daten sind die Schlüsselfaktoren ob psychophysiologische Messungen und die dadurch mögliche implizite Interaktion zum Erfolg werden können. Denn das Potenzial ist da, diese Technik positiv zu nutzen und dem Nutzer zu mehr Kontrolle zu verhelfen. Die Sicherheit kann erhöht werden, Menschen mit Krankheiten wie Autismus können bei der Emotionserkennung unterstützt werden oder es wird Arbeit reduziert. Gleichzeitig spielt aber der Kontrollverlust eine große Rolle, auf der einen Seite über die eigenen Daten und auf der anderen Seite über die Entscheidung bei Interaktionen.

Entscheidend ist, dass die psychophysiologische Messungen mit dem richtigen Ziel vor Augen angewendet werden, denn nur dann kann die Umsetzung gelingen.

Literatur

- 1 Eeg hospital. https://commons.wikimedia.org/wiki/File:EEG_Hospital.png. Accessed: 2019-08-11.
- 2 Electromyogram illustration. https://upload.wikimedia.org/wikipedia/commons/7/75/Electromyogram_Illustration.png. Accessed: 2019-06-07.
- 3 Eye tracking device. https://upload.wikimedia.org/wikipedia/commons/c/c1/Eye_Tracking_Device_003.jpg. Accessed: 2019-08-11.
- 4 Global app revenue reached \$19.5 billion last quarter, up 17% year-over-year. <https://sensortower.com/blog/app-revenue-and-downloads-q1-2019>. Accessed: 2019-06-07.
- 5 Interaktion, die. <https://www.duden.de/rechtschreibung/Interaktion>. Accessed: 2019-06-22.
- 6 Photo by kamil s on unsplash. <https://unsplash.com/photos/9nZQfPe0TiU>. Accessed: 2019-08-11.
- 7 Psychophysiologie, die. <https://www.duden.de/rechtschreibung/Psychophysiologie>. Accessed: 2019-06-23.
- 8 Übereinkommen über den strassenverkehr. <https://www.admin.ch/opc/de/classified-compilation/19680244/index.html>. Accessed: 2019-06-26.
- 9 Daniel Afergan. Using brain-computer interfaces for implicit input. In *Proceedings of the adjunct publication of the 27th annual ACM symposium on User interface software and technology - UIST'14 Adjunct*, pages 13–16, Honolulu, Hawaii, USA, 2014. ACM Press. URL: <http://dl.acm.org/citation.cfm?doid=2658779.2661166>, doi:10.1145/2658779.2661166.
- 10 Mike Ambinder. Biofeedback in gameplay: How valve measures physiology to enhance gaming experience. <https://steamcdn-a.akamaihd.net/apps/valve/2011/ValveBiofeedback-Ambinder.pdf>, 2011. Accessed: 2019-06-23.
- 11 Ioannis Arapakis, Ioannis Konstas, and Joemon M. Jose. Using facial expressions and peripheral physiological signals as implicit indicators of topical relevance. In *Proceedings of the seventeen ACM international conference on Multimedia - MM '09*, page 461, Beijing, China, 2009. ACM Press. URL: <http://portal.acm.org/citation.cfm?doid=1631272.1631336>, doi:10.1145/1631272.1631336.
- 12 Richard Atterer, Monika Wnuk, and Albrecht Schmidt. Knowing the user's every move: user activity tracking for website usability evaluation and implicit interaction. In *Proceedings of the 15th international conference on World Wide Web - WWW '06*, page 203, Edinburgh, Scotland, 2006. ACM Press. URL: <http://portal.acm.org/citation.cfm?doid=1135777.1135811>, doi:10.1145/1135777.1135811.

- 13 Albert F. Ax. Editorial. *Psychophysiology*, 1(1):1–3, July 1964. URL: <http://doi.wiley.com/10.1111/j.1469-8986.1964.tb02613.x>, doi:10.1111/j.1469-8986.1964.tb02613.x.
- 14 Frank Beruscha, Klaus Augsburg, and Dietrich Manstetten. Haptic warning signals at the steering wheel: A literature survey regarding lane departure warning systems. 4(5):6.
- 15 Tomislav Bezmalinovic. vr-brille misst hirnstroeme und erkennt augenbewegungen. <https://mixed.de/ces-2018-vr-brille-misst-hirnstroeme-und-erkennt-augenbewegungen/>. Accessed: 2019-06-23.
- 16 Andreas Braun, Sebastian Frank, Martin Majewski, and Xiaofeng Wang. CapSeat: capacitive proximity sensing for automotive activity recognition. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '15*, pages 225–232, Nottingham, United Kingdom, 2015. ACM Press. URL: <http://dl.acm.org/citation.cfm?doid=2799250.2799263>, doi:10.1145/2799250.2799263.
- 17 Mickaël Causse, Frédéric Dehais, Patrice Péran, Umberto Sabatini, and Josette Pastor. The effects of emotion on pilot decision-making: A neuroergonomic approach to aviation safety. *Transportation Research Part C: Emerging Technologies*, 33:272–281, August 2013. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0968090X12000551>, doi:10.1016/j.trc.2012.04.005.
- 18 Mark Claypool, Phong Le, Makoto Wased, and David Brown. Implicit interest indicators. In *Proceedings of the 6th international conference on Intelligent user interfaces - IUI '01*, pages 33–40, Santa Fe, New Mexico, United States, 2001. ACM Press. URL: <http://portal.acm.org/citation.cfm?doid=359784.359836>, doi:10.1145/359784.359836.
- 19 Ulf Dimberg, Monika Thunberg, and Kurt Elmehed. Unconscious facial reactions to emotional facial expressions.
- 20 M. Eriksson and N.P. Papanikotopoulos. Eye-tracking for detection of driver fatigue. In *Proceedings of Conference on Intelligent Transportation Systems*, pages 314–319, Boston, MA, USA, 1997. IEEE. URL: <http://ieeexplore.ieee.org/document/660494/>, doi:10.1109/ITSC.1997.660494.
- 21 Stephen H. Fairclough. Fundamentals of physiological computing. *Interacting with Computers*, 21(1-2):133–145, January 2009. URL: <https://academic.oup.com/iwc/article-lookup/doi/10.1016/j.intcom.2008.10.011>, doi:10.1016/j.intcom.2008.10.011.
- 22 Mariam Hassib, Max Pfeiffer, Stefan Schneegass, Michael Rohs, and Florian Alt. Emotion Actuator: Embodied Emotional Feedback through Electroencephalography and Electrical Muscle Stimulation. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, pages 6133–6146, Denver, Colorado, USA, 2017. ACM

- Press. URL: <http://dl.acm.org/citation.cfm?doid=3025453.3025953>, doi:10.1145/3025453.3025953.
- 23 Kristina Hook. The encyclopedia of human-computer interaction, 2nd ed., affective computing. <https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/affective-computing#>. Accessed: 2019-06-24.
 - 24 Benjamin Jillich. Acquisition, Analysis and Visualization of Data from Physiological Sensors for Biofeedback Applications. page 114.
 - 25 Wendy Ju. The Design of Implicit Interactions. *Synthesis Lectures on Human-Centered Informatics*, 8(2):1–93, March 2015. URL: <http://www.morganclaypool.com/doi/10.2200/S00619ED1V01Y201412HCI028>, doi:10.2200/S00619ED1V01Y201412HCI028.
 - 26 Wendy Ju and Larry Leifer. The Design of Implicit Interactions: Making Interactive Systems Less Obnoxious. *Design Issues*, 24(3):72–84, July 2008. URL: <http://www.mitpressjournals.org/doi/10.1162/desi.2008.24.3.72>, doi:10.1162/desi.2008.24.3.72.
 - 27 Juha M Kortelainen, Mark van Gils, and Juha Pärkkä. Multichannel Bed Pressure Sensor for Sleep Monitoring. page 4.
 - 28 Golden Krishna. *The best interface is no Interface*. New Riders Press, 2015.
 - 29 Kai Kuikkaniemi, Toni Laitinen, Marko Turpeinen, Timo Saari, Ilkka Kosunen, and Niklas Ravaja. The influence of implicit and explicit biofeedback in first-person shooter games. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*, page 859, Atlanta, Georgia, USA, 2010. ACM Press. URL: <http://portal.acm.org/citation.cfm?doid=1753326.1753453>, doi:10.1145/1753326.1753453.
 - 30 Pedro Lopes, Alexandra Ion, Willi Mueller, Daniel Hoffmann, Patrik Jonell, and Patrick Baudisch. Proprioceptive Interaction. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, pages 939–948, Seoul, Republic of Korea, 2015. ACM Press. URL: <http://dl.acm.org/citation.cfm?doid=2702123.2702461>, doi:10.1145/2702123.2702461.
 - 31 Lennart Erik Nacke, Michael Kalyn, Calvin Lough, and Regan Lee Mandryk. Biofeedback game design: using direct and indirect physiological control to enhance game interaction. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*, page 103, Vancouver, BC, Canada, 2011. ACM Press. URL: <http://dl.acm.org/citation.cfm?doid=1978942.1978958>, doi:10.1145/1978942.1978958.
 - 32 Eyal Nir. *Hooked: How to build habit-forming products*. Penguin Group, 2014.
 - 33 Ken Pfeuffer, Jason Alexander, Ming Ki Chong, Yanxia Zhang, and Hans Gellersen. Gaze-shifting: direct-indirect input with pen and touch modulated by gaze. In *UIST '15 Proceedings of the 28th Annual ACM*

- Symposium on User Interface Software and Technology*, pages 373–383. ACM, 11 2015. ACM, 2015. This is the author’s version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in UIST ’15 Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology <http://dx.doi.org/10.1145/2807442.2807460>. doi:10.1145/2807442.2807460.
- 34 Ken Pfeuffer, Jason Alexander, and Hans Gellersen. Partially-indirect bi-manual input with gaze, pen, and touch for pan, zoom, and ink interaction. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI ’16, pages 2845–2856, New York, NY, USA, 2016. ACM. URL: <http://doi.acm.org.emedien.ub.uni-muenchen.de/10.1145/2858036.2858201>, doi:10.1145/2858036.2858201.
- 35 Rosalind Picard. *Affective Computing*. MIT Press, 1997.
- 36 Andreas Riemer and Alois Ferscha. Supporting Implicit Human-to-Vehicle Interaction: Driver Identification from Sitting Postures. In *Proceedings of the First Annual International Symposium on Vehicular Computing Systems*, Dublin, Ireland, 2008. ICST. URL: <http://eudl.eu/?id=3545>, doi:10.4108/ICST.ISVCS2008.3545.
- 37 Andreas Riemer, Alois Ferscha, and Mohamed Aly. Heart on the road: HRV analysis for monitoring a driver’s affective state. In *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI ’09*, page 99, Essen, Germany, 2009. ACM Press. URL: <http://portal.acm.org/citation.cfm?doid=1620509.1620529>, doi:10.1145/1620509.1620529.
- 38 A. Schmidt, B. Pfleging, F. Alt, A. Sahami, and G. Fitzpatrick. Interacting with 21st-century computers. *IEEE Pervasive Computing*, 11(1):22–31, 2012. URL: <http://ieeexplore.ieee.org/document/6081843/>, doi:10.1109/MPRV.2011.81.
- 39 Barış Serim and Giulio Jacucci. Explicating "Implicit Interaction": An Examination of the Concept and Challenges for Research. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI ’19*, pages 1–16, Glasgow, Scotland Uk, 2019. ACM Press. URL: <http://dl.acm.org/citation.cfm?doid=3290605.3300647>, doi:10.1145/3290605.3300647.
- 40 Erin Treacy Solovey, Daniel Afergan, Evan M Peck, Samuel W Hincks, and Robert J K Jacob. Designing Implicit Interfaces for Physiological Computing: Guidelines and Lessons Learned using fNIRS. page 31.
- 41 Paul Watzlawick. Die axiome von paul watzlawick. <https://www.paulwatzlawick.de/axiome.html>. Accessed: 2019-06-22.
- 42 Mark Weiser. The computer for the 21st century. In *The scientific American*, 1991.

Verdeckte Interaktion durch Elektromyografie

Evelyn Müller

Ludwig-Maximilians-Universität München, München, Deutschland
Eve.Mueller@campus.lmu.de

Zusammenfassung

Interaktionen mit mobilen Geräten können in vielen Situationen des Alltags andere Personen in ihren aktuellen Tätigkeiten unterbrechen oder ablenken. Um soziale Akzeptanz zu erreichen, sollten Interaktionen jedoch subtil stattfinden. Hierzu eignet sich verdeckte Interaktion, die mithilfe von Elektromyografie (EMG) erfasst werden kann. EMG erkennt subtile Gesten, sogar wenn dabei keine Bewegungen stattfinden, indem elektrische Muskelaktivitäten gemessen werden. Da kleine, kaum bis nicht bemerkbare und kurze Gesten eher sozial akzeptiert sind, bietet sich EMG als Eingabemöglichkeit an. Einige Forscherteams haben bereits Prototypen solcher Devices entwickelt oder das ehemals auf dem Markt erhältliche Myo Armband getestet. In dieser Arbeit werden verschiedene dieser Prototypen und Forschungsarbeiten vorgestellt, die sich mit explizitem Input von verdeckten Gesten durch EMG beschäftigt haben. Dabei wird besonders auf die Arbeiten von Costanza et al. (*Intimate Interfaces*) und Saponas et al. (*Muscle-Computer Interfaces*) eingegangen.

2012 ACM Computing Classification Human computer interaction (HCI) → Ubiquitous and mobile computing

Keywords and phrases EMG; Elektromyografie; covert interaction; verdeckte Interaktion; subtile Interaktion.

1 Motivation

Auf einmal durchbricht das Klingeln oder Summen eines Handys die Stille, stört die Konzentration oder unterbricht den Vortrag, worauf der Angerufene sein Handy hervorkramt, es ausschaltet oder kurz telefoniert. Eine Situation, die Vielen bekannt sein dürfte, genauso wie die kurze Verwirrung, sich angesprochen zu fühlen, wenn die andere Person in Wahrheit lediglich mithilfe ihrer Kopfhörer telefoniert. Mobile Geräte sind immer und überall dabei und können so auch häufig für Unterbrechungen sorgen, bemerkt Costanza bereits 2006 [5], noch bevor beispielsweise WhatsApp auf dem Markt war [22]. Pri-



© Evelyn Müller;

licensed under Creative Commons License CC-BY

Cite as: Evelyn Müller. Verdeckte Interaktion durch Elektromyografie. In *3rd Seminar on Ubiquitous Interaction (UBIACTION 2019)*. Editors: Pascal Knierim, Jakob Karolus, Fiona Draxler, Ville Mäkelä, Tonja Machulla, Florian Lang, Thomas Kosch, Albrecht Schmidt. July 16, 2019. Munich, Germany. pp. 6:1–6:22.

vate Kommunikation und Öffentlichkeit treffen aufeinander, wobei nicht nur die Privatsphäre einer einzelnen Person leidet, sondern auch außenstehende Personen mehr oder weniger an einer Interaktion teilhaben, für die sie sich gar nicht aktiv entschieden haben [7]. Ein wichtiger Punkt für das Design mobiler Geräte ist jedoch, wie gut sie (und die zugehörige Interaktion) sozial akzeptiert werden [6]. Es sind Interfaces nötig, die subtil sowie diskret sind und die Personen in der Umgebung weder unterbrechen noch ablenken [8, 7]. Weisers Vorstellung, dass Computer in Zukunft „unsichtbar“ werden, indem sie in den Hintergrund verschwinden [34], kann jetzt umgesetzt, wenn nicht sogar erweitert werden [23]: Auch die Interaktion mit den Geräten sollte möglichst unsichtbar sein [6], womit auch die entgegen Weisers Vision nicht in den Hintergrund gerückten Smartphones [29] unauffälliger werden können. Um das zu erreichen, können verdeckte, subtile Gesten verwendet werden, die von anderen Personen nicht bemerkt werden und diese somit weder ablenken noch deren aktuelle Tätigkeit unterbrechen. Einige dieser Gesten, wie das bloße Anspannen eines Muskels [7], sind sogar kaum bis gar nicht wahrnehmbar. Doch wie können derartige Gesten erfasst werden? Hierzu eignet sich die Technik der Elektromyografie (EMG), die mithilfe von Elektroden die Aktivität von Muskeln erkennen kann [31], selbst wenn keine Bewegung zu sehen ist [6]. Costanza et al. [8, 6, 5, 7] beispielsweise nutzen diese Technik für ihre *Intimate Interfaces* und präsentieren einen Prototyp, der so subtile Gesten erkennt. Auch Saponas et al. [24, 23, 25, 26] verwenden EMG für ihre *Muscle-Computer Interfaces*, welche Handgesten auch erkennen, wenn ein Gegenstand in der Hand gehalten wird.

Diese Arbeit beschäftigt sich mit verdeckter Interaktion, die durch EMG ermöglicht wird. Zunächst wird die verdeckte Interaktion selbst genauer betrachtet (2 Verdeckte Interaktion). Da einige Begriffe in der Literatur auf unterschiedliche Weise verwendet werden, wird dazu z. B. „subtil“ für diese Arbeit definiert (2.1 Verdeckt, implizit und subtil). Anschließend wird analysiert, welche Gesten als sozial akzeptiert und subtil gelten (2.2 Subtile Gesten). Nach einer genaueren Beschreibung von EMG (zu Beginn von 3 Elektromyografie als Input) werden verschiedene Forschungsarbeiten vorgestellt, wobei in Abschnitt 3.2 Verdeckte Gesten als Input auf einige genauer eingegangen wird. Wie sogar Texteingabe mittels EMG möglich ist, beschreibt 3.2.1 Texteingabe, und die Forschungen von Costanza et al. (3.2.2 *Intimate Interfaces* bei Costanza) und Saponas et al. (3.2.3 *Muscle-Computer Interfaces* bei Saponas), die sich genauer mit EMG für verdeckte Gesten beschäftigt haben, werden genauer betrachtet. Danach werden Anforderungen für und Herausforderungen an derartige Interfaces zusammengefasst (4 Anforderungen und Herausforderungen) und schließlich untersucht ob und wie sich EMG Interfaces für verdeckten Input in Zukunft durchsetzen könnten (5 Ausblick).

2 Verdeckte Interaktion

Begriffe wie „subtil“ und „implizit“ werden in vielen Artikeln unterschiedlich definiert, wie Pohl et al. [19] und Serim et al. [30] festgestellt haben. Deshalb wird im Folgenden beschrieben, auf welche Weise diese Bezeichnungen und der Begriff „verdeckt“ in dieser Arbeit verwendet werden. Anschließend wird untersucht, was subtile Gesten und soziale Akzeptanz auszeichnet, und was bereits in diese Richtung geforscht wurde.

2.1 Verdeckt, implizit und subtil

In dieser Arbeit geht es vor allem um verdeckte Interaktionen, also Interaktionen, die von anderen Personen möglichst nicht bemerkt werden und die somit sozial akzeptiert sind und mehr Privatsphäre bieten. Diese Interaktionen können auch als subtil bezeichnet werden.

Das Wort „subtil“ wird in der Literatur zu Mensch-Computer Interaktion mit verschiedenen Bedeutungen verwendet [19]. Anderson et. al [3] definieren subtile Interaktion als von Anderen unbemerkter Input bzw. Output. Sie messen Subtilität anhand davon, welcher Anteil der Interaktionen bemerkt wird. Pohl et al. [19] haben solche Definitionen in insgesamt 55 Veröffentlichungen aus dem Bereich Mensch-Computer Interaktion untersucht. Sie identifizieren vier verschiedene häufige Definitionen von „subtil“: (1) unaufdringlich, (2) versteckt bzw. täuschend, (3) unaufwändig und (4) beeinflussend. Unaufdringlich (1) bezieht sich hierbei auf die Feedback Methode des Systems, welche wenig Aufmerksamkeit verlangt und unauffälliger gestaltet ist. Mit versteckt bzw. täuschend (2) ist Input gemeint, der von Anderen nicht gesehen werden kann (versteckt) oder unbemerkt bleibt, da über ihn hinweg getäuscht wird (täuschend). Auch unaufwändig (3) bezeichnet den Input des Nutzers, der in diesem Fall aus kleinen, wenig anstrengenden Bewegungen besteht. Die vierte Definition (4) behandelt die sanfte Einflussnahme auf den Nutzer, indem beispielsweise dessen Blick gelenkt wird.

In dieser Arbeit wird „subtil“ im Sinne von 2 und 3 verwendet, da vor allem der Input der Nutzer im Vordergrund steht. Es geht also um kleine Bewegungen, die nicht bemerkt oder nicht einmal gesehen werden.

Der Fokus dieser Arbeit liegt dabei auf explizitem Input, während impliziter Input in Artikel 5 „Implizite Interaktionen durch physiologische Messungen“ von Maximilian Lammel behandelt wird. Allerdings werden auch „explizit“ bzw. „implizit“ in der Literatur nicht immer gleich verwendet. Serim et al. [30] untersuchen die verschiedenen Bedeutungen von implizit und beleuchten fünf verschiedene Facetten. Demnach wird in dieser Arbeit Input betrachtet, der vom Nutzer absichtlich, mit voller Aufmerksamkeit und vollständigem Wissen über die Folgen, bewusst und ohne implizierte Bedeutungen geliefert wird.

Insgesamt geht es in dieser Arbeit um verdeckte, subtile Gesten, die explizit vom Nutzer durchgeführt werden.

2.2 Subtile Gesten

Doch welche Gesten sind verdeckt und subtil genug, damit sie sozial akzeptiert werden? Interaktionen müssen neben effizient und effektiv auch so designt sein, dass sie ein gutes Nutzererlebnis bieten [19], wobei auch auf den sozialen Kontext geachtet werden muss [6]. Um sozial akzeptiert zu werden, sollten Gesten subtil sein, möglichst unauffällig und unaufdringlich, sowie natürlich, und auch die Devices sollten nicht auffallen oder andere soziale Aktivitäten behindern [17, 21]. Soziale Akzeptanz kann sich von Kultur zu Kultur unterscheiden und mit der Zeit sogar verändern [17]. Aktuell gibt es keine einheitliche Weise, zu messen, wie subtil eine Geste ist, auch wenn z. B. erfasst werden kann, wie oft sie bemerkt oder als störend empfunden wird [19]. Die meisten der in dieser Arbeit vorgestellten Devices und zugehörigen Gesten wurden tatsächlich gar nicht auf deren Subtilität und soziale Akzeptanz getestet. Stattdessen wird häufig lediglich untersucht, ob eine von den Autoren als subtil bezeichnete Geste zuverlässig von ihrem System erkannt wird. Im Gegensatz dazu haben beispielsweise Costanza et al. (3.2.2 *Intimate Interfaces* bei Costanza) [7] mithilfe eines Videos getestet, wie häufig ihre Geste bemerkt wurde, und auch die Studienteilnehmer von Anderson et al. [3] sahen Videos. Anderson et al. teilten die Probanden in drei Gruppen mit unterschiedlichem Wissen ein: Personen ohne Wissen, mit Wissen über die Geräte, und Personen, die sowohl über das Gerät als auch die Interaktionen damit Bescheid wussten. Alströhm et al. [1] beschränkten sich nicht auf solche Laborbedingungen, sondern testeten, wie ihr Device in der Öffentlichkeit wahrgenommen wurde. Auch Wozniak et al. [35] testeten ihr Gerät in alltäglichen Situationen: beim Fahrradfahren und im Café. Welche Arten von Gesten und Geräten als subtil und sozial akzeptiert angesehen werden können, haben Montero [17], Alströhm [1] und Anderson [3] mit ihren jeweiligen Teams genauer untersucht.

Anderson et al. [3] beschäftigten sich mit der Fragen, wann, warum und wie verdeckte Interaktion stattfindet, wie Vorgehensweisen aus der Magie verwendet werden können, um die Geräte selbst zu tarnen, und wie diese klassifiziert werden können. Dazu führten sie zunächst mithilfe eines Online-Fragebogens mit 227 Teilnehmern eine Studie durch. So fanden sie beispielsweise heraus, dass mobile Geräte am häufigsten für das Nachschauen der Uhrzeit, Lesen oder Schreiben von Nachrichten und Fotografieren genutzt werden. Als Gründe dafür, dass diese Interaktionen verdeckt stattfanden, gaben 35% gesellschaftliche Normen an, 25% Verbote durch Vorgesetzte und 18% Regeln des jeweiligen Ortes, beispielsweise eines Museums. Für eine subtile Interaktion trotz dieser Regeln verließen sie entweder den Bereich, in dem sie Gefahr liefen, entdeckt

zu werden, oder sie verbargen das Gerät (zum Beispiel unter dem Tisch) bzw. die Interaktion selbst, indem sie diese als eine andere Tätigkeit tarnten. Anschließend untersuchten Anderson et al. wie Methoden der Magie dafür genutzt werden können, Interaktionen bzw. Geräte subtil zu gestalten, wozu sie sowohl Magier befragten, als auch Literatur recherchierten. Als Ergebnis stellen sie fünf Methoden vor: Anpassung an das Aussehen des Nutzers, Variation der Interaktion, natürlicher Umgang mit einem Gegenstand (der in Wahrheit das getarnte Gerät ist), Zeitliche Trennung von Ursache und Wirkung sowie Training der Nutzer für unauffälligere Interaktion. Sie kategorisieren Devices nach Art des Geräts (Input, Output), Sichtbarkeit des Geräts (sichtbar, unsichtbar, getarnt), Sichtbarkeit der Interaktion (wahrnehmbar, nicht wahrnehmbar, getarnt) und Transparenz der Aktion (wie genau wissen Beobachter, was getan wird). Die vorliegende Arbeit fokussiert sich auf Geräte für Input, auch wenn für einige vorgestellte Prototypen ebenfalls deren Output beschrieben wird. Ein Großteil dieser vorgestellten Geräte wird als Armband getarnt oder unsichtbar unter der Kleidung getragen und die meisten Interaktionen mit diesen sind – soweit untersucht – kaum bis nicht wahrnehmbar.

Montero et al. [17] teilen soziale Akzeptanz in „soziale Akzeptanz des Nutzers“ und „soziale Akzeptanz des Beobachters“. Die soziale Akzeptanz des Nutzers wird davon beeinflusst, wie er sich bei dieser Interaktion fühlt. Die soziale Akzeptanz des Beobachters hängt davon ab, wie er eine von einem Anderen ausgeführte Interaktion wahrnimmt, und ob er versteht, was derjenige tut. Wie der Beobachter selbst auf eine Geste reagieren würde, wirkt sich auch auf dessen Akzeptanz als Nutzer aus. Für die soziale Akzeptanz des Beobachters spielen die Weise, auf welche eine Interaktion stattfindet, und die wahrnehmbaren Ergebnisse eine große Rolle. Montero et al. haben vier verschiedene Arten von Interaktionen untersucht, wobei sie diese nach Reeves et al. [20] klassifizieren. Hier wird danach aufgeteilt, wie sichtbar Manipulation und Ergebnis sind, was zu den vier Kategorien „ausdrucksstark“, „spannend“, „geheim“ und „magisch“ führt. Bei ausdrucksstarken Gesten sind sowohl die Manipulation durch den Nutzer als auch deren Folgen sichtbar, während bei geheimen Gesten weder die Geste selbst noch die Reaktion des Systems wahrnehmbar sind. Magische Gesten sind als solche nicht sichtbar, jedoch deren Resultate, und im Gegensatz dazu sind spannende Gesten sichtbar, lassen Beobachter jedoch über deren Folgen im Unklaren. Zu jeder Art wurden in einer Nutzerstudie jeweils zwei Gesten in einem Video präsentiert, wozu die Teilnehmer angeben sollten, wie sie diese Geste in der Öffentlichkeit bewerten würden. Nach einer Erklärung von Manipulation und Ergebnis sollten sie bewerten, wie sie sich selbst bei Ausführung der Gesten in verschiedenen Situationen fühlen würden. Montero et al. kamen zu dem Ergebnis, dass ausdrucksstarke, geheime und magische Gesten akzeptiert werden, spannende, bei denen die

6:6 Verdeckte Interaktion durch EMG

Auswirkung der Geste geheim bleibt, hingegen nicht.

Alströhm et al. [1] untersuchten Gesten, die um das technische Gerät herum ausgeführt werden, im Zusammenhang mit Größe, Länge und Position bzw. Entfernung von diesem Device in der Öffentlichkeit. Da die Gesten daraus bestanden, Zahlen in die Luft zu zeichnen, bleibt fraglich, wie subtil diese Gesten sein können. Trotzdem konnten sie feststellen, dass kleine und kurze Gesten bevorzugt werden, und dass ihren Probanden die Reaktion von anderen Personen wichtig war. Sie konnten außerdem keinen Unterschied zwischen introvertierten und extrovertierten Personen erkennen.

Zusammenfassend findet verdeckte Interaktion aufgrund von gesellschaftlichen Normen und verschiedenen Regeln bereits ohne auf Subtilität ausgelegte Geräte statt (beispielsweise durch Verstecken des Devices). Um soziale Akzeptanz zu erreichen, können entweder die Geräte oder die Interaktion verborgen bzw. angepasst werden. Kleine Gesten von kurzer Dauer eignen sich dabei gut, ebenso wie ausdrucksstarke, geheime und magische Gesten, wohingegen spannende Gesten nicht akzeptiert werden.

3 Elektromyografie als Input

Subtile bzw. verdeckte Gesten können mithilfe von Elektromyografie (EMG) auch dann erkannt werden, wenn äußerlich keine Bewegung erkennbar ist [6]. Dabei wird die elektrische Aktivität von Muskeln gemessen [36]. Eine Geste beginnt damit, dass das Gehirn ein elektrisches Signal an sogenannte Motoneuronen sendet, die zusammen mit ihren jeweiligen Muskelfasern motorische Einheiten bilden. Das Zusammenziehen der Fasern wird wiederum durch Signale der Motoneuronen erreicht. EMG erfasst dann die gesamte elektrische Aktivität der aktiven motorischen Einheiten des Muskels. Dazu werden Elektroden verwendet, die entweder im Muskel (invasives EMG) oder auf der Haut (Oberflächen-EMG) angebracht werden, wobei letztere Methode zwar weniger genau, aber für Interfaces besser geeignet ist. [24, 25] EMG wird in der Medizin häufig für Diagnosen und Rehabilitation, aber auch für die Steuerung von Prothesen eingesetzt [23]. Da EMG sogar isometrische Muskelkontraktionen, also Muskelaktivitäten ohne auffällige oder sichtbare Bewegung eines Körperteils, erkennen kann [6], eignet sie sich hervorragend für verdeckte Interaktion. Denn für eine soziale Akzeptanz sollten Gesten klein, subtil und möglichst unsichtbar sein.

Input für ein System kann dabei dadurch erzeugt werden, welche Muskeln an der Geste beteiligt sind, in welchem Muster oder wie stark sie angespannt werden, wobei letzteres allerdings für den Nutzer schwerer zu merken ist [9, 11].

Die gewählte Geste für ein Device wird in der Forschung meistens mithilfe von Machine Learning anhand von gesammelten Daten trainiert, worauf dann

getestet wird, wie gut das System neue Daten zuordnen kann. Diese Tests finden größtenteils unter Laborbedingungen statt, wo manchmal sogar beispielsweise die Haltung des Armes [38] vorgeschrieben wird. In manchen Fällen wird ein Einsatz im Alltag simuliert (z. B. durch das Ablaufen eines Slalom-Parkours [6, 7]) oder das System sogar in der Öffentlichkeit getestet [35]. Neben eigenen EMG-Geräten und Algorithmen werden häufig auch bereits erhältliche verwendet. Einige Forscher arbeiteten sogar mit einem fertigen Input Device: dem Myo Armband [13, 14, 35, 38]. Dieser inzwischen nicht mehr erhältliche Controller [27] von ehemals *Thalmic Labs* (jetzt *North* [28]) erfasst mithilfe von acht Sensoren fünf verschiedene Gesten [14]. Er unterstützt außerdem Bluetooth und haptisches Feedback [32]. Der verwendete Algorithmus liefert jedoch nur eine Genauigkeit von 68%, kann allerdings ersetzt werden [9].

Einige Forschungsergebnisse und Prototypen werden im Folgenden vorgestellt, wobei der Schwerpunkt auf die Forschung von Costanza et al. [8, 6, 5, 7] und Saponas et al. [24, 23, 25, 26] gelegt wird, die sich besonders ausführlich mit EMG Geräten für verdeckte Interaktion beschäftigt haben.

3.1 Möglichkeiten der Interaktion

Elektromyografie ist tatsächlich so präzise, dass komplizierte Fingerbewegungen erkannt werden können, ja sogar, welche Akkorde auf einer Gitarre gegriffen werden [12]. Diese Technik ist dabei keineswegs auf Bewegungen der Hände beschränkt. Es können beispielsweise auch Muskelaktivitäten im Gesicht und dadurch Emotionen erfasst werden [33]. Insgesamt wurden bereits viele verschiedene Arten des (hier behandelten expliziten) Inputs getestet, die Genauigkeit verbessert, Lösungsansätze für Probleme gefunden und die Möglichkeiten des Myo Armbands untersucht. Einige dieser Arbeiten werden im Folgenden vorgestellt.

Wie Zhang et al. [39] zeigen, ist es beispielsweise möglich, Zungengesten (links, rechts, oben, unten, vorne, entspannt) durch Elektroden an Kinn und Kehle zu erkennen. Zhang et al. erreichten hier eine Genauigkeit von 94.17%.

Amma et al. [2] verwendeten im Gegensatz zu anderen Arbeiten mehr Elektroden in einer höheren Dichte, insgesamt 192 Elektroden. Sie erreichten für 27 verschiedene Gesten, bestehend aus Tippen, Krümmen und Strecken eines oder mehrerer Finger, durchschnittlich eine Genauigkeit von 90%, wobei das Testen in der selben Sitzung wie die Kalibrierung für einzelne Nutzer stattfand. Um die Gestenerkennung bei von Sitzung zu Sitzung leicht unterschiedlich platzierten Elektroden zu verbessern, stellen Amma et al. noch dazu zwei Algorithmen vor, welche diese Verschiebung anhand von Eigenschaften der Anatomie berechnen. Mit deren Hilfe können sie die Erkennungsrate von 59% auf 75% verbessern. Außerdem kommen sie zu dem Ergebnis, dass eine größere Zahl an Elektroden die Leistung erhöht, wobei ihr Gerät die besten Resultate

ab 100 Elektroden erreicht.

Huang et al. [10] konzentrierten sich auf bestimmte Bewegungen des Dauerns wie Swipes, langes Halten und Tippen, und erreichen eine durchschnittliche Erkennungsrate von 82.9%. Des Weiteren implementierten sie zwei Anwendungsbeispiele: ein Tool für Präsentationen und einen Musikplayer.

Kerber et al. [13] widmeten sich dem Problem, dass Touch-Interaktionen zum Beispiel mit Smartwatches nur durch die Hand des anderen Armes (ohne Uhr) durchgeführt werden können. Ist diese Hand beschäftigt, ist Input schwierig. Deshalb stellen Kerber et al. ein EMG-System mit Myo-Armband vor, welches Input durch die Hand mit Uhr möglich macht. In ihrer Studie wurde jedoch Touch Input bevorzugt und ihr System brachte keinen Geschwindigkeitsvorteil. Diese Ergebnisse führen Duente et al. [9] auf die Hardware zurück, da das Myo Armband lediglich eine Genauigkeit von 68% bietet.

In einer weiteren Forschungsarbeit verwendeten Kerber et al. [14] erneut das Myo Armband, aber dieses Mal mit einem eigenen Algorithmus, der die Erkennungsrate auf 95% verbessern konnte. Mit dessen Hilfe waren sie in der Lage, nicht nur die fünf Standard-Gesten, sondern insgesamt 40 Gesten zu erkennen.

Wozniak et al. [35] nutzten ebenfalls das Myo Armband, mit dessen Standard-Gesten ihre Versuchspersonen einen Musikplayer in zwei verschiedenen Situationen bedienten: in einem Café und beim Fahrradfahren. Das Gerät stellte sich jedoch als zu sensibel heraus, da es hier viele alltägliche Bewegungen als vermeintlichen Input erkannte. Die Interaktion mit dem System wurde von den Versuchspersonen darüber hinaus als zu auffällig wahrgenommen.

Wiederum einen eigenen Algorithmus für das Myo Armband verwendeten Zadeh et al. [38], wodurch sie eine Erfolgsrate von 95% erreichten. Allerdings testeten sie das Device mit einer festgelegten Arm-Haltung.

Duente et al. [9] dagegen nutzten einen eigenen Prototyp für den Umgang mit Benachrichtigungen. Muskeln wurden hierbei nicht nur als Input Methode, sondern auch für die Benachrichtigungen selbst genutzt. Durch elektrische Muskelstimulation (EMS) werden über Elektroden elektrische Impulse an die Muskeln abgegeben, was sie mit EMG verzahnten, sodass der Output nicht den Input verändern konnte. In einer Nutzerstudie konnten die Probanden auf die EMS Benachrichtigungen entweder mit „Annehmen“ oder „Ablehnen“ reagieren, was auf einem Bildschirm vorgegeben wurde. Jeder Teilnehmer testete drei verschiedene Arten der Interaktion: Touch Input via Smartwatch und zwei verschiedene Eingaben durch EMG. Für eine Input Version wurde lediglich ein Muskel verwendet (musterbasiert), der die Hand in Richtung Handrücken (nach „oben“) bewegt. Hier galt einmal Anspannen als Annehmen und zweimal als Ablehnen. Für die zweite Methode (richtungsbasiert) wurde auch der Gegenspieler dieses Muskels verwendet, welcher die Hand in Richtung Handfläche

(nach „unten“) bewegt. Eine Bewegung nach „oben“ wurde als Annehmen gewertet, eine nach „unten“ als Ablehnen. Insgesamt stellte sich die richtungsbasierte Methode als deutlich schneller im Vergleich zur musterbasierten und zu Touch heraus, was in Kontrast zu den Ergebnissen von Kerber et al. [13] steht. Duinte et al. vermuten jedoch, wie bereits erwähnt, dass die Ergebnisse von Kerber et al. durch das Myo Armband beeinflusst wurden. Die Fehlerrate des Touch Inputs war insgesamt am geringsten mit 0.14%, gefolgt von der richtungsbasierten Methode mit 3.9%. Die musterbasierte schnitt deutlich schlechter mit 20.8% ab, was zeigt, dass es schwieriger ist, verschiedenen Input mit nur einem Muskel zu erzeugen. Diese Methode wurde von den Probanden auch am schlechtesten bewertet, wobei die richtungsbasierte Methode bei fast allen Favorit war.

3.2 Verdeckte Gesten als Input

Viele dieser bereits vorgestellten Prototypen nutzen Gesten, die als subtil bzw. verdeckt angesehen werden können. Auch wenn kaum eine Geste hinsichtlich sozialer Akzeptanz untersucht wurde, ist es denkbar, dass sich einige dazu eignen könnten, wie beispielsweise die Zungengesten bei Zhang et al. [39] oder eine Auswahl der Gesten von Amma et al. [2]. Costanza et al. [8, 6, 5, 7] (*Intimate Interfaces* bei Costanza 3.2.2) hingegen legten bei ihrer Geste viel Wert auf soziale Akzeptanz. Sie stellen *Intimate Interfaces* für verdeckte Interaktion vor und testen deren Subtilität. Saponas et al. [24, 23, 25, 26] (*Muscle-Computer Interfaces* bei Saponas 3.2.3) beschäftigten sich ebenfalls genauer mit EMG Input, wobei auch deren Gesten als subtil gewertet werden können. Sie untersuchten außerdem, wie Trainingsdaten von vorherigen Sitzungen trotz anderer Elektroden-Positionen genutzt werden können, und ob ihre Gesten auch erkannt werden, wenn die Hände verschiedene Gegenstände tragen. Die Forschung dieser Teams um Costanza und Saponas werden in den folgenden Abschnitten behandelt. Zunächst jedoch wird ein weiterer Forschungsbereich vorgestellt: Die Texteingabe.

3.2.1 Texteingabe

Die meisten der schon beschriebenen Arbeiten nutzen weniger als 10 verschiedene Gesten für ihre Systeme. Die größte Anzahl erkennen Amma et al. [2] mit 27 Gesten und Kerber et al. [14] mit 40. Eine Texteingabe durch diese Gesten wäre sicherlich möglich (beispielsweise indem jedem Buchstaben eine Geste zugeordnet wird), aber vergleichsweise langsam und ungewohnt. Mithilfe von EMG ist es darüber hinaus auch möglich, Sprache selbst zu erkennen. Spracherzeugung ist ein komplizierter Vorgang, an dem viele Muskeln beteiligt sind [11], deren Aktivität durch Elektroden in der unteren Gesichtshälfte und

6:10 Verdeckte Interaktion durch EMG

an der Kehle erfasst werden können [16, 15, 18, 11]. Dafür ist es nicht nötig, Buchstaben bzw. Wörter laut auszusprechen, stumme Artikulation reicht aus.

Manabe et al. [16] erkannten durch ihr Device fünf Vokale mit einer durchschnittlichen Genauigkeit von über 90%. Naik et al. [18] erreichten für ihre drei Vokale 83.33%. Lee et al. [15] beschränkten sich nicht nur auf Vokale. Ihr System erkannte 60 verschiedene Wörter mit einer Genauigkeit von 87.07%.

Selbst wenn die Wörter bzw. Buchstaben nicht laut ausgesprochen werden, ist Input mithilfe dieser Methoden noch gut sichtbar. Beim Lesen tritt hingegen eine Art Sprache auf, die nicht erkennbar ist: das sogenannte Subvokalisieren, eine innere Stimme, die auch abseits des Lesens genutzt werden kann [11]. Kapur et al. [11] stellen ein mobiles Gerät vor, AlterEgo, dessen festes „Gerüst“ die Position der Elektroden bestimmt, und das mit Knochenschall Kopfhörern ausgestattet ist. Sie testeten ihr Device anhand der Zahlen von 0 bis 9 und erreichten eine Genauigkeit von 92%, was der Erkennungsrate von aktuellen Spracherkennungssystemen entspricht. Kapur et al. schlagen vor, dass solche Systeme sowohl als Input genutzt werden könnten, als auch für die Kommunikation mit anderen Personen. Zusätzlich sei es möglich, beispielsweise komplizierte Rechenaufgaben an einen Computer auszulagern, und so eine Art Symbiose zwischen Mensch und Maschine zu erreichen.

Der Input dieser vorgestellten Geräte mag unhörbar oder sogar unerkannt stattfinden, dennoch sind die Devices sehr auffällig. Die soziale Akzeptanz wird hier vermutlich eher gering ausfallen.

3.2.2 *Intimate Interfaces* bei Costanza

Das mobile Input Gerät von Costanza et al. [8, 6, 5, 7] ist im Vergleich zu den Geräten zur Spracherkennung deutlich verborgener: Sie stellen das Konzept von *Intimate Interfaces* vor, unauffälligen Geräten, die subtile Gesten erkennen. Ihr Controller besteht aus am Oberarm getragenen Elektroden, die in Zukunft in ein Armband integriert werden können, welches noch dazu durch Kleidung verdeckt werden kann. Costanza et al. [6] untersuchten für die Entwicklung ihres *Intimate Interfaces* zunächst verschiedene Muskeln als Inputmöglichkeiten und entschieden sich für ein kurzes Anspannen des Bizeps. Dieser Muskel im Oberarm zeichnet sich auch bei weniger sportlichen Menschen deutlich ab und sein Signal wird von anderen Muskeln kaum beeinflusst. Weil die Geste mithilfe von isometrischer Muskelkontraktion erzeugt wird, handelt es sich dabei um eine sehr subtile „bewegungslose Geste“. Anschließend führten Costanza et al. eine zweite Studie durch, um dieses Anspannen des Muskels mit Signalen bei anderen Aktivitäten wie beispielsweise Gestikulieren oder Anheben von Objekten zu vergleichen. Das System wurde mithilfe der hier gesammelten Daten trainiert, getestet und so lange gestresst und verbessert, bis es die Geste bei jedem beliebigen Nutzer mit kaum False-Positives erkannte. Darauf folgte

die Hauptstudie mit 10 Teilnehmern. Die Probanden liefen mehrmals einen Slalom Parkour ab, wobei die erste Runde ohne Interaktion stattfand. Darauf folgten vier weitere mit Gesten, deren Länge in drei Runden vorgeschrieben waren (lang, kurz und beides in zufälliger Abfolge). Feedback bei erkannten Gesten und der Stimulus, auf den die Geste folgen sollte, wurden über Audio gegeben. In der generischen Runde ohne festgelegter Gestendauer erreichte das System eine Genauigkeit von 96%. Das kurze Anspannen erzielte 97%, das lange 94%, und das kombinierte 87%. Die Unterscheidung von lang und kurz in der kombinierten Runde funktionierte weniger gut. Lediglich 55% der kurzen und 47% der langen Gesten wurden jeweils richtig zugeordnet. Insgesamt wurde die Art der Geste nur bei 51% erkannt.

Costanza et al. verbanden ihren Controller in der nachfolgenden Forschungsarbeit [7] dann mit einem konkreten Interface für den Umgang mit Anrufen. Ein eingehender Anruf konnte (1) ohne Nachricht abgelehnt, (2) ignoriert, sowie mit der Nachricht (3) „Ich rufe später zurück“ oder (4) „In einer Besprechung; zurückrufen wenn dringend“ abgelehnt werden. Sie erweiterten ihr Armband aus Elektroden dazu durch einen Vibrationsmotor für taktilen Output, der signalisierte, wenn eine Geste erkannt wurde. Um das Gerät zu testen, führten sie eine Studie mit 12 Personen durch, welche aus drei Teilen bestand. Auch hier sollten die Versuchsteilnehmer wieder im Slalom laufen, während sie gar nicht (ohne Gerät), mithilfe von einem Arm oder mit zwei Armen (zwei Controller) interagierten. Wenn zwei Controller verwendet wurden, war das Feedback durch Vibration von Vorteil, weil so besser erkannt werden kann, auf welchen Arm es sich bezieht. Die vier Möglichkeiten, mit einem (simulierten) Anruf umzugehen, konnten dabei durch ein Audio Menü ausgewählt werden: das aktuelle Element und das Auswählen einer Option wurden von einer Computerstimme über kabellose Kopfhörer an den Nutzer übermittelt. Wenn *ein* Controller bzw. Arm für die Interaktion genutzt wurde, ging das Menü automatisch alle Elemente immer wieder der Reihe nach durch, bis eines durch die Geste ausgewählt wurde. Wenn *beide* Arme genutzt wurden, ging der Nutzer die Elemente selbst mithilfe des dominanten Armes durch und selektierte durch den anderen Arm. Insgesamt wurden durch das System 96.2% der Anrufe korrekt beantwortet, 97.6% durch die Zwei-Arm Version, 94.6% durch die mit einem Controller. Die Laufgeschwindigkeit der Nutzer wurde durch die Interaktion mit dem Gerät nicht beeinträchtigt, ein negativer Effekt durch das Interface konnte also nicht festgestellt werden. Von den beiden Methoden bevorzugten die Versuchspersonen die Version mit zwei Controllern. Anschließend untersuchten Costanza et al. in einer weiteren Studie, wie auffällig diese Interaktion ist. Dazu betrachteten Versuchspersonen ein stummes Video, auf dem ein (geschulter) Schauspieler (tonlos) mit einer nicht sichtbaren Person sprach und mehrmals die Geste ausführte. Das Video

war dreiteilig: einmal trug der Darsteller ein Oberteil mit langen, einmal mit kurzen Ärmeln und einmal wurde das Armband in einer Nahaufnahme gezeigt. Die Versuchspersonen sollten angeben, wenn sie eine Interaktion entdeckten und wie sicher sie sich dabei waren. Bei der Version mit langen Ärmeln wurden lediglich 13.9% der Gesten korrekt erkannt, bei kurzen Ärmeln 33.1% und bei der Nahaufnahme 75.9%. Die Interaktion wird also, selbst wenn explizit darauf geachtet wird, nicht so leicht bemerkt, besonders nicht, wenn das Gerät von einem langen Ärmel verdeckt wird.

Der Armband-Controller von Costanza et al. ist ein unauffälliges Eingabegerät, das speziell auf verdeckte Gesten ausgelegt ist und soziale Akzeptanz unterstützt. Dadurch, dass die Geste nicht mit der Hand durchgeführt wird, ist eine Interaktion auch dann möglich, wenn die Hände bereits beschäftigt sind. Das Gerät musste außerdem im Gegensatz vieler anderer Prototypen nicht speziell für den Nutzer trainiert werden.

3.2.3 *Muscle-Computer Interfaces* bei Saponas

Saponas et al. [24, 23, 25, 26] haben ebenfalls ein EMG Eingabegerät entwickelt, welches sie als *Muscle-Computer Interface* bezeichnen und das Fingergesten verwendet [24]. Anders als bei Costanza et al. ist dadurch bei beschäftigten Händen Input schwieriger. Deshalb untersuchen Saponas et al. ob Handgesten mithilfe von EMG auch dann möglich sind, wenn beispielsweise etwas getragen wird [25]. Während Costanza et al. ihr System so entwickelten, dass pro Nutzer kein zusätzliches Training nötig war, musste das Gerät von Saponas et al. jeweils vor seinem Einsatz kalibriert werden. Um dies zu verbessern, entwickelten sie ihren Prototyp weiter und testeten, wie dieser mit Kalibrierungen an vorangegangenen Tagen umgehen konnte [26].

Zuerst realisierten Saponas et al. [24] ein Gerät, dessen Elektroden ebenfalls in einem Band (zukünftig in einem Armband), jedoch um den Unterarm, angeordnet werden. Sie nutzten dafür vorgefertigte Machine Learning Methoden und ein einfaches Setup für den EMG Input. Insgesamt verwendeten sie 18 Gesten einer Hand: Ein Tippen oder Anheben jeweils eines Fingers, sowie festen bzw. leichten Druck, Krümmen oder Ausstrecken des Mittel- oder Zeigefingers. Die Hand lag dabei auf einem Tisch. Sie testeten mit der Hilfe von 12 Versuchspersonen, die auf einem Bildschirm dargestellte Gesten durchführen sollten. Von den auf diese Weise erhaltenen Daten wurden pro Nutzer 90% für das Training, die restlichen für das Testen verwendet. Hier wurde das System für jeden Nutzer einzeln kalibriert, um dessen Gesten zu erkennen. Bei den Gesten Strecken und Krümmen erreichen sie eine Genauigkeit von 78%, bei Druck-Gesten 84%, für Tippen 78% und Anheben 95%.

In einer späteren Arbeit [25] verwenden Saponas et al. wieder ein Armband aus Elektroden am Unterarm. Sie beschäftigten sich mit der Frage, wie

Handgesten noch genutzt werden können, wenn der Nutzer etwas trägt. Als Geste wählten sie das Drücken jeweils eines Fingers gegen den Daumen bzw. ein festeres Zugreifen. Diese Geste wurde jedoch nur erfasst, wenn das System eine einfache und robuste Aktivierungsgeste erkannte: ein Zusammendrücken der zur Faust geballten anderen (nicht-dominanten) Hand. Diese Technik soll gewährleisten, dass das System zuverlässig erkennt, was als Geste gemeint ist. Sie trainierten und testeten ihr System in einer dreiteiligen Studie mit 12 Personen. In Teil A wurde ohne Objekt interagiert und nebenbei erforscht, wie sich die Haltung des Armes auswirkt. Für Teil B hielten die Teilnehmer einen Reise-Becher oder eine Tasche mit ca. 1kg, wie auch in Teil C, wo sie jedoch mit einem konkreten Interface, einem Musikplayer, interagierten. In den ersten beiden Teilen wurden außerdem jeweils Trainingsdaten gesammelt. Welche Geste aktuell von den Probanden gefordert war und welche Geste das System erkannte, wurde auf einem Bildschirm als Hand mit markiertem Finger dargestellt. Die Teilnehmer sollten mithilfe der Gesten das Menü eines Musikplayers bedienen und z. B. einen bestimmten Song auswählen. In Teil A (ohne Gegenstand) erreichte das System durchschnittlich eine Genauigkeit von 77%, wenn die getestete Armhaltung der trainierten entsprach. Stammten die Trainingsdaten von einer leicht anderen Pose, lag die Erkennungsrate bei 72%, bei stark veränderten Haltungen bei 63%. Wenn alle Posen des Armes für das Training verwendet wurden, stieg die Genauigkeit auf 79%. Bei lediglich drei Gesten verbesserte sich die Erkennungsrate auf 85%. Hielten die Nutzer einen Reisebecher in der Hand (Teil B), lag die Genauigkeit bei 65% ohne Feedback, mit Feedback sogar bei 85%. Auch hier verbesserten sich die Werte, wenn lediglich mit drei Fingern gearbeitet wurde, auf 77% ohne Feedback, 86% mit. Die Interaktionsgeschwindigkeit nahm hingegen mit Feedback ab. Sofern die Nutzer eine Tasche in der Hand hielten, erreichte das System 88% mit Feedback (bzw. 90% bei drei Gesten) und 86% ohne Feedback (bzw. 91%)

In einer weiteren Arbeit [26] widmeten sich Saponas et al. der Frage, wie sich ein zeitlicher Abstand zwischen Kalibrierung und Verwendung des Systems auf dessen Leistung auswirkt. Sie verwendeten subtile Gesten aus den beiden oben beschriebenen vorherigen Arbeiten mit kaum oder nicht erkennbarer Bewegung des Zeige-, Mittel- oder Ringfingers: Kneifen (bzw. Drücken eines Fingers gegen den Daumen oder ein getragenes Objekt) und Drücken eines Fingers auf eine Oberfläche, auf der die Hand liegt. Letztere kann auf jeder beliebigen Oberfläche, wie dem eigenen Bein, ausgeführt werden. Auch hier wurde zudem wieder die Aktivierungsgeste verwendet. Ihre Studie mit acht Teilnehmern fand an zwei verschiedenen Tagen statt. Am ersten Tag trainierten die Nutzer das System und testeten es sofort im Anschluss. Nach einer kurzen Pause, in der sie das Armband abgenommen hatten, wurden erneut Trainingsdaten gesammelt. Die darauf folgenden Tests verwendeten jedoch

6:14 Verdeckte Interaktion durch EMG

erneut die Daten aus der ersten Kalibrierung. Auch bei der nächsten Sitzung zwei Tage später wurden zunächst Trainingsdaten erfasst und danach Tests durchgeführt, welche wiederum die zweite Kollektion des ersten Tages verwendeten. Saponas et al. untersuchten auf diese Weise wie sich die Leistung ihres Systems unterscheidet, wenn es direkt vor der Nutzung, am selben oder an einem anderen Tag trainiert wurde. Drücken eines Fingers auf eine Oberfläche wurde anhand von kurz zuvor erfassten Trainingsdaten mit einer Genauigkeit von durchschnittlich 76% erkannt, durch Daten des selben Tages mit einer Erkennungsrate von 73% und an einem anderen Tag lediglich 66%. Kneifen wurde durch das System besser erkannt. Es wurden 86% mit kurz zuvor und 87% mit am selben Tag gesammelten Daten sowie 86% durch Daten eines anderen Tages erreicht.

Das Device von Saponas et al. ist wie das von Costanza et al. ein unauffälliges Armband. Saponas et al. testeten zunächst verschiedene subtile Gesten, verwendeten aber anschließend vor allem ein Kneifen. Sie führten zusätzlich ein Zusammendrücken der anderen Hand als Aktivierungsgeste ein, damit ihr System nur gewollte Gesten erfasst. Das Kneifen wird auch dann noch erkannt, wenn die Person etwas in der Hand hält, oder wenn die Trainingsdaten von einem anderen Zeitpunkt des selben Tages oder sogar eines anderen Tages stammen.

4 Anforderungen und Herausforderungen

Viele der vorgestellten Prototypen wurden nicht explizit für verdeckte Interaktionen entwickelt, eignen sich dennoch sehr gut dafür. Ein Gerät für verdeckte Interaktion muss subtile Gesten unterstützen, Gesten die eher klein, kurz und unbemerkbar sind, und außerdem auch selbst als Device kaum auffallen [1, 17, 21, 3]. Wie beispielsweise Costanza et al. und Saponas et al. bewiesen haben, können subtile Gesten mithilfe von EMG erkannt werden, selbst dann wenn diese nach außen hin nicht sichtbar sind. Costanza et al. konnten außerdem zeigen, dass die Interaktion mit ihrem Controller durch Anspannen des Oberarms für Beobachter schlecht zu erkennen ist. Viele Forscherteams verwenden zudem Devices, die als Armband getarnt werden können, und somit, wie Anderson et al. [3] vorschlugen, an den Nutzer angepasst bzw. unauffällig sind. Einige Geräte für Texteingaben, wie beispielsweise AlterEgo von Kapur et al. [11] sind hingegen noch sehr auffällig. Dadurch, dass Input mithilfe von Elektromyografie getarnt wird, können „spannende“ Gesten vermieden werden, die im Gegensatz zu den Resultaten der Interaktion sichtbar sind und sozial nicht akzeptiert werden [17]. Soziale Akzeptanz des Nutzers nach Montero et al. [17] kann somit ebenfalls erreicht werden. Zusätzlich sind einige Gesten sogar mit beschäftigten Händen möglich [25]. Gesten an sich können selbst

dann noch verwendet werden, wenn der Nutzer nichts sieht (beispielsweise im Dunkeln) oder wenn er unterwegs ist (z. B. auch bei erschwerten Bedingungen in einem fahrenden Bus) [31]. EMG Controller selbst sollten bei mobilen Nutzern ebenfalls gut funktionieren (d. h. unabhängig von Posen sein und möglichst auch bei beschäftigten Händen Interaktion ermöglichen), schnell an den entsprechenden Muskeln befestigt werden können und sofort funktionieren, ohne dass bei jedem Anlegen kalibriert werden muss, sowie in Echtzeit arbeiten und nur als Geste erkennen, was als solche gemeint ist [25, 26]. Viele der vorgestellten Prototypen wurden lediglich unter Laborbedingungen und teilweise sogar in festgelegten Posen getestet. Sie müssen zudem häufig vor der Nutzung kalibriert werden und sind teilweise überempfindlich (wie beispielsweise das Myo Armband [35]). Diese Überempfindlichkeit wird Midas-Touch genannt und kann beispielsweise durch eine Aktivierungsgeste [25] behoben werden. Einige weitere Punkte werden von verschiedenen Prototypen erfüllt, wie beispielsweise die Nutzerunabhängigkeit bei Costanza et al. [8, 6, 5, 7] und die Verwendung der Geste mit beschäftigten Händen bei Saponas et al. [24, 23, 25, 26]. Eine weitere Herausforderung für EMG Interfaces ist Crosstalk, Signale von umliegenden Muskeln, welche besonders bei subtilen Gesten mit vergleichsweise schwachen Signalen deren Erkennung erschweren. Weiterer Noise, der gefiltert werden muss, entsteht durch das Durchqueren des Signals von Fett und Haut, durch Bewegungen der Haut und durch die Umgebung selbst [24]. Die Signale unterscheiden sich noch dazu von Person zu Person [18] und je nach Positionierung der Elektroden [11], beispielsweise wenn das Gerät neu angelegt wird oder sich der Nutzer bewegt. Gegen eine Positionsänderung der Elektroden kann beispielsweise ein festes Gitter, in dem die Elektroden angeordnet sind, helfen [11]. Einige Forscherteams wie z. B. Amma et al. [2] und Kerber et al. [14] haben wiederum Algorithmen entwickelt, die mit Verschiebungen umgehen können. Costanza et al. [8, 6, 5, 7] und Kerber et al. [14] trainieren ihre Systeme unabhängig von bestimmten Nutzern und ermöglichen deshalb, dass ihre Controller mit vielen verschiedenen Nutzern umgehen können. Saponas et al. [24, 23, 25, 26] zeigen, dass ihre Kneif-Geste auch dann noch relativ gut erkannt wird, wenn das Training an einem anderen Tag stattgefunden hat. Für Interfaces, die verdeckte Gesten mithilfe von EMG als Input erkennen, muss außerdem entschieden werden, wie das Feedback stattfindet. Dieses ist zum Beispiel visuell, als Audio, taktil oder durch Smartglasses möglich [7]. Je nach Art des Feedbacks wird weniger (visuelle) Aufmerksamkeit benötigt, eine begrenzte Ressource, die dann stattdessen für die Umgebung genutzt werden kann [37]. Beispielsweise die Systeme von Costanza et al. [8, 6, 5, 7] und von Saponas et al. [24, 23, 25, 26] gaben den Nutzern zudem Feedback, wann immer eine Geste erkannt wurde. Bei Costanza et al. geschah dies durch Audio bzw. haptisches Feedback, während Saponas et al. dies bildlich dar-

stellten. Die Gesten der Nutzer werden durch Feedback genauer [4, 25]. Wie Saponas et al. feststellten [25], verlangsamt diese Art von Feedback jedoch die Interaktion, was bei Systemen, bei denen Fehler sich gravierend auswirken, in Kauf genommen werden kann, um die Zahl der Fehler zu reduzieren.

Einerseits bringt verdeckte Interaktion auch mehr Privatsphäre, andererseits jedoch auch eine Frage der Ethik mit sich: mithilfe solcher Systeme wird darüber hinweg getäuscht, dass gerade interagiert wird [19]. Es ist also möglich, dass das Gegenüber zwar so aussieht, als wäre die Person geistig anwesend, sie sich in Wahrheit aber mit ihrem Gerät beschäftigt. Dabei stellt sich auch die Frage, ob durch solche Geräte mehr Interaktionen in Situationen stattfinden, in denen diese sozial nicht akzeptiert sind.

5 Ausblick

Kein Klingeln oder Summen eines Handy unterbricht die Stille, stört die Konzentration oder unterbricht den Vortrag und doch wird jemand angerufen, der ohne sein Handy hervor zu kramen, dieses ausschaltet oder sogar ohne Worte telefoniert. So könnte eine verdeckte Interaktion in Zukunft aussehen. Mithilfe von EMG werden subtilste Gesten erkannt, die für Andere kaum bemerkbar oder sogar mit keiner sichtbaren Bewegung verbunden sind. Einige Handgesten sind selbst dann noch möglich, wenn Gegenstände getragen werden, andere Gesten werden ohne Hände durchgeführt. EMG Geräte erfassen beispielsweise Input für den Umgang mit Anrufen oder die Interaktion mit Musikplayern, ohne dass der Nutzer z. B. sein Handy hervorholen muss. Die Devices selbst können dabei häufig als Armband getarnt werden. Diese verdeckten Interaktionen lassen das jeweilige Gerät wie in Weisers Vision [34] in den Hintergrund verschwinden. Sie werden von anderen Personen nicht bemerkt und stören diese folglich auch nicht. EMG Controller für verdeckte Interaktion können somit soziale Akzeptanz erreichen. Doch wenn diese Geräte eigentlich auf soziale Akzeptanz ausgelegt sein sollten, warum haben sie sich nicht längst im Alltag etabliert? Wenn in den Studien häufig hohe Erkennungsraten erreicht wurden und es zu vielen Herausforderungen Lösungsansätze gibt, warum nutzt die große Masse diese Geräte nicht längst? Wie bereits erwähnt, wurde kaum ein Prototyp tatsächlich außerhalb des Labors oder einer simulierten alltäglichen Situation getestet. Die Werte beziehen sich größtenteils auf Gesten, die lediglich bei einem unbewegten Nutzer, teilweise mit vorgeschriebenen Armhaltungen, erkannt wurden. Im täglichen Leben der Nutzer würde jedoch weitaus mehr Bewegung stattfinden, die Noise erzeugt, eine Positionsänderung der Elektroden verursacht oder fälschlicherweise als Input erkannt wird. Wozniak et al. [35], die das Myo Armband z. B. während dem Fahrradfahren oder in einem Café testeten, mussten feststellen, dass dieses zum einen zu auffällig war,

aber auch zu sensibel. Das Hören eines Podcasts wurde beispielsweise auf einer holprigen Straße unmöglich, da immer wieder vermeintlicher Input erkannt wurde. Selbst jede scharfe Linkskurve wurde als Input gewertet. Noch dazu liegt die Erkennungsrate des Myo Armbands lediglich bei 68% [9]. Dieser Wert ist relativ gering, vor allem wenn man beachtet, dass in alltäglichen Situationen häufig Midas-Touch auftritt. Selbst die besseren Werte der anderen Prototypen aus den Laborstudien, würden sich in alltäglichen Situationen verschlechtern. Zudem stellt sich auch die Frage, wie gut diese Geräte noch funktionieren, wenn der Nutzer stark schwitzt, in Regen, bei großer Hitze oder Kälte unterwegs ist, oder wie lange beispielsweise EMG Armbänder getragen werden können, bevor sie unangenehm werden. Zusätzlich gibt es wenige Arbeiten, die testen, wie sozial akzeptiert die gewählte Geste tatsächlich ist.

Insgesamt bietet verdeckte Interaktion durch EMG viele Vorteile und hat großes Potential. Sie erreicht soziale Akzeptanz und ermöglicht subtile Interaktion mit Geräten. In der Forschung muss jedoch mehr Wert auf tatsächlichen Einsatz der Geräte im alltäglichen Leben gelegt werden, um sie in sozialer Akzeptanz und Genauigkeit zu optimieren, bis diese Geräte tatsächlich häufig zum Einsatz kommen. Denn wer würde beispielsweise gerne ein Smartphone benutzen, dessen Touch nicht zuverlässig funktioniert?

Literatur

- 1 David Ahlström, Khalad Hasan, and Pourang Irani. Are you comfortable doing that?: Acceptance studies of around-device gestures in and for public settings. In *Proceedings of the 16th International Conference on Human-computer Interaction with Mobile Devices & Services*, MobileHCI '14, pages 193–202, New York, NY, USA, 2014. ACM. URL: <http://doi.acm.org/10.1145/2628363.2628381>, doi:10.1145/2628363.2628381.
- 2 Christoph Amma, Thomas Krings, Jonas Böer, and Tanja Schultz. Advancing muscle-computer interfaces with high-density electromyography. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 929–938, New York, NY, USA, 2015. ACM. URL: <http://doi.acm.org/10.1145/2702123.2702501>, doi:10.1145/2702123.2702501.
- 3 Fraser Anderson, Tovi Grossman, Daniel Wigdor, and George Fitzmaurice. Supporting subtlety with deceptive devices and illusory interactions. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 1489–1498, New York, NY, USA, 2015. ACM. URL: <http://doi.acm.org/10.1145/2702123.2702336>, doi:10.1145/2702123.2702336.
- 4 Stephen Brewster, Joanna Lumsden, Marek Bell, Malcolm Hall, and Stuart Tasker. Multimodal 'eyes-free' interaction techniques for wearable devices. In *Proceedings of the SIGCHI Conference on Human Factors in*

- Computing Systems*, CHI '03, pages 473–480, New York, NY, USA, 2003. ACM. URL: <http://doi.acm.org/10.1145/642611.642694>, doi:10.1145/642611.642694.
- 5 Enrico Costanza. *Subtle, Intimate Interfaces for Mobile Human Computer Interaction*. Master thesis, Massachusetts Institute of Technology, Massachusetts, 2006.
 - 6 Enrico Costanza, Samuel A. Inverso, and Rebecca Allen. Toward subtle intimate interfaces for mobile devices using an emg controller. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, pages 481–489, New York, NY, USA, 2005. ACM. URL: <http://doi.acm.org/10.1145/1054972.1055039>, doi:10.1145/1054972.1055039.
 - 7 Enrico Costanza, Samuel A. Inverso, Rebecca Allen, and Pattie Maes. Intimate interfaces in action: Assessing the usability and subtlety of emg-based motionless gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 819–828, New York, NY, USA, 2007. ACM. URL: <http://doi.acm.org/10.1145/1240624.1240747>, doi:10.1145/1240624.1240747.
 - 8 Enrico Costanza, Alberto Perdomo, Samuel A. Inverso, and Rebecca Allen. Emg as a subtle input interface for mobile computing. In *Mobile Human-Computer Interaction - MobileHCI 2004*, pages 426–430. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.
 - 9 Tim Duentz, Justin Schulte, Max Pfeiffer, and Michael Rohs. Muscleio: Muscle-based input and output for casual notifications. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(2):64:1–64:21, July 2018. URL: <http://doi.acm.org/10.1145/3214267>, doi:10.1145/3214267.
 - 10 Donny Huang, Xiaoyi Zhang, T. Scott Saponas, James Fogarty, and Shyamnath Gollakota. Leveraging dual-observable input for fine-grained thumb interaction using forearm emg. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, UIST '15, pages 523–528, New York, NY, USA, 2015. ACM. URL: <http://doi.acm.org/10.1145/2807442.2807506>, doi:10.1145/2807442.2807506.
 - 11 Arnav Kapur, Shreyas Kapur, and Pattie Maes. Alterego: A personalized wearable silent speech interface. In *23rd International Conference on Intelligent User Interfaces*, IUI '18, pages 43–53, New York, NY, USA, 2018. ACM. URL: <http://doi.acm.org/10.1145/3172944.3172977>, doi:10.1145/3172944.3172977.
 - 12 Jakob Karolus, Hendrik Schuff, Thomas Kosch, Paweł W. Wozniak, and Albrecht Schmidt. Emguitar: Assisting guitar playing with electromyography. In *Proceedings of the 2018 Designing Interactive Systems Conference*, DIS '18, pages 651–655, New York, NY, USA, 2018. ACM. URL: <http://doi.acm.org/10.1145/3196709.3196803>, doi:10.1145/3196709.3196803.
 - 13 Frederic Kerber, Pascal Lessel, and Antonio Krüger. Same-side hand interactions with arm-placed devices using emg. In *Proceedings of the 33rd*

- Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '15, pages 1367–1372, New York, NY, USA, 2015. ACM. URL: <http://doi.acm.org/10.1145/2702613.2732895>, doi: 10.1145/2702613.2732895.
- 14 Frederic Kerber, Michael Puhl, and Antonio Krüger. User-independent real-time hand gesture recognition based on surface electromyography. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '17, pages 36:1–36:7, New York, NY, USA, 2017. ACM. URL: <http://doi.acm.org/10.1145/3098279.3098553>, doi:10.1145/3098279.3098553.
 - 15 Ki-Seung Lee. Emg-based speech recognition using hidden markov models with global control variables. *IEEE transactions on bio-medical engineering*, 55(3):930–940, 2008. doi:10.1109/TBME.2008.915658.
 - 16 Hiroyuki Manabe, Akira Hiraiwa, and Toshiaki Sugimura. Unvoiced speech recognition using emg - mime speech recognition. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '03, pages 794–795, New York, NY, USA, 2003. ACM. URL: <http://doi.acm.org/10.1145/765891.765996>, doi:10.1145/765891.765996.
 - 17 Calkin S. Montero, Jason Alexander, Mark T. Marshall, and Sriram Subramanian. Would you do that?: Understanding social acceptance of gestural interfaces. In *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '10, pages 275–278, New York, NY, USA, 2010. ACM. URL: <http://doi.acm.org/10.1145/1851600.1851647>, doi:10.1145/1851600.1851647.
 - 18 Ganesh R. Naik, Dinesh K. Kumar, and Sridhar P. Arjunan. Multi modal gesture identification for hci using surface emg. In *Proceedings of the 12th International Conference on Entertainment and Media in the Ubiquitous Era*, MindTrek '08, pages 90–94, New York, NY, USA, 2008. ACM. URL: <http://doi.acm.org/10.1145/1457199.1457219>, doi:10.1145/1457199.1457219.
 - 19 Henning Pohl, Andreea Muresan, and Kasper Hornbæk. Charting subtle interaction in the hci literature. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 418:1–418:15, New York, NY, USA, 2019. ACM. URL: <http://doi.acm.org/10.1145/3290605.3300648>, doi:10.1145/3290605.3300648.
 - 20 Stuart Reeves, Steve Benford, Claire O'Malley, and Mike Fraser. Designing the spectator experience. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, pages 741–750, New York, NY, USA, 2005. ACM. URL: <http://doi.acm.org/10.1145/1054972.1055074>, doi:10.1145/1054972.1055074.
 - 21 J. Rekimoto. Proceedings fifth international symposium on wearable computers. In *Wearable Computers, 2001. Proceedings. Fifth International Symposium on*, pages 21–27, s.l., 2001. IEEE Comput. Soc. doi:10.1109/ISWC.2001.962092.

- 22 Sebastian Sander. Unnützes wissen: Seit wann gibt es eigentlich whatsapp? URL: https://praxistipps.chip.de/unnuetzes-wissen-seit-wann-gibt-es-eigentlich-whatsapp_101570.
- 23 T. Scott Saponas. Enabling always-available input: Through on-body interfaces. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '09, pages 3117–3120, New York, NY, USA, 2009. ACM. URL: <http://doi.acm.org/10.1145/1520340.1520441>, doi: 10.1145/1520340.1520441.
- 24 T Scott Saponas, Desney S. Tan, Dan Morris, and Ravin Balakrishnan. Demonstrating the feasibility of using forearm electromyography for muscle-computer interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 515–524, New York, NY, USA, 2008. ACM. URL: <http://doi.acm.org/10.1145/1357054.1357138>, doi:10.1145/1357054.1357138.
- 25 T. Scott Saponas, Desney S. Tan, Dan Morris, Ravin Balakrishnan, Jim Turner, and James A. Landay. Enabling always-available input with muscle-computer interfaces. In *Proceedings of the 22Nd Annual ACM Symposium on User Interface Software and Technology*, UIST '09, pages 167–176, New York, NY, USA, 2009. ACM. URL: <http://doi.acm.org/10.1145/1622176.1622208>, doi:10.1145/1622176.1622208.
- 26 T. Scott Saponas, Desney S. Tan, Dan Morris, Jim Turner, and James A. Landay. Making muscle-computer interfaces more practical. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 851–854, New York, NY, USA, 2010. ACM. URL: <http://doi.acm.org/10.1145/1753326.1753451>, doi:10.1145/1753326.1753451.
- 27 Paul Sawers. Amazon-backed wearables company thalnic labs kills its myo armband, teases new product, 2018. URL: <https://venturebeat.com/2018/10/12/amazon-backed-wearables-company-thalnic-labs-kills-its-myo-armband-teases-new-product/>.
- 28 Paul Sawers. Thalnic labs rebrands as north, launches \$999 alexa-powered holographic glasses, 2018. URL: <https://venturebeat.com/2018/10/23/thalnic-labs-rebrands-as-north-launches-999-alexa-powered-holographic-glasses/>.
- 29 A. Schmidt, B. Pfleging, F. Alt, A. Sahami, and G. Fitzpatrick. Interacting with 21st-century computers. *IEEE Pervasive Computing*, 11(1):22–31, 2012. doi:10.1109/MPRV.2011.81.
- 30 Barış Serim and Giulio Jacucci. Explicating „implicit interaction“: An examination of the concept and challenges for research. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 417:1–417:16, New York, NY, USA, 2019. ACM. URL: <http://doi.acm.org/10.1145/3290605.3300647>, doi:10.1145/3290605.3300647.

- 31 Desney Tan, Dan Morris, and T. Scott Saponas. Interfaces on the go. *XRDS*, 16(4):30–34, June 2010. URL: <http://doi.acm.org/10.1145/1764848.1764856>, doi:10.1145/1764848.1764856.
- 32 Visual Dimension Center Fellbach. Thalmic labs - myo gesture control armband. URL: <https://www.vdc-fellbach.de/leistungen/technikbewertung-var/thalmic-labs-myo-gesture-control-armband/>.
- 33 SCOTT R. VRANA. The psychophysiology of disgust: Differentiating negative emotional contexts with facial emg. *Psychophysiology*, 30(3):279–286, 1993. URL: <https://onlinelibrary.wiley.com/doi/full/10.1111/j.1469-8986.1993.tb03354.x>.
- 34 Mark Weiser. The computer for the 21st century. *Scientific American*, 265(3):94–104, 1991. doi:10.1038/scientificamerican0991-94.
- 35 Mikołaj Woźniak, Patryk Pomykalski, Dawid Sielski, Krzysztof Grudzień, Natalia Paluch, and Zbigniew Chaniecki. Exploring emg gesture recognition - interactive armband for audio playback control. In *Proceedings of the 2018 Federated Conference on Computer Science and Information Systems*, Annals of Computer Science and Information Systems, pages 919–923. IEEE, 2018. doi:10.15439/2018F175.
- 36 Qiang Yang, Yongpan Zou, Meng Zhao, Jiawei Lin, and Kaishun Wu. Armin: Explore the feasibility of designing a text-entry application using emg signals. In *Proceedings of the 15th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, MobiQuitous '18, pages 117–126, New York, NY, USA, 2018. ACM. URL: <http://doi.acm.org/10.1145/3286978.3287030>, doi:10.1145/3286978.3287030.
- 37 Bo Yi, Xiang Cao, Morten Fjeld, and Shengdong Zhao. Exploring user motivations for eyes-free interaction on mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 2789–2792, New York, NY, USA, 2012. ACM. URL: <http://doi.acm.org/10.1145/2207676.2208678>, doi:10.1145/2207676.2208678.
- 38 A. Saleh Zadeh, A. P. Calitz, and J. H. Greyling. Evaluating a biosensor-based interface to recognize hand-finger gestures using a myo armband. In *Proceedings of the Annual Conference of the South African Institute of Computer Scientists and Information Technologists*, SAICSIT '18, pages 229–238, New York, NY, USA, 2018. ACM. URL: <http://doi.acm.org/10.1145/3278681.3278709>, doi:10.1145/3278681.3278709.
- 39 Qiao Zhang, Shyamnath Gollakota, Ben Taskar, and Raj P.N. Rao. Non-intrusive tongue machine interface. In Matt Jones, Philippe Palanque, Albrecht Schmidt, and Tovi Grossman, editors, *CHI 2014, one of a CHIInd*, pages 2555–2558, New York, NY, 2014. Assoc. for Computing Machinery. doi:10.1145/2556288.2556981.

Implizite Interaktion mittels Eye-Tracking

Johannes Sylupp

Ludwig-Maximilians-Universität München, München, Deutschland
sylupp.johannes@campus.lmu.de

Zusammenfassung

Der Mensch gibt über seine Augen viele Informationen preis. Implizite Interaktion mittels Eye-Tracking nutzt diese Tatsache, dass sich viel aus der Beobachtung der Augen ableiten lässt, um das Zusammenspiel mit Geräten zu vereinfachen, zu erweitern und komfortabler zu machen. Bisher schien der Fokus eher auf expliziter Interaktion zu liegen, in Zukunft könnte die implizite Form oder Mischvarianten in diesem Bereich häufiger werden. Das Review gibt eine Übersicht über vielversprechende Anwendungsmöglichkeiten von impliziter Interaktion mittels Eye-Tracking aus dem aktuellen Stand der Forschung, untergliedert in Echtzeit- und verzögerte Interaktivsysteme sowie Diagnosesysteme.

2012 ACM Computing Classification Human-centered computing → Interaction techniques

Keywords and phrases Implizite Interaktionstechniken; Eye-Tracking; Augenbewegungen; Mensch-Computer-Interaktion.

1 Einleitung

”Die Augen sind die Fenster der Seele”, das sagte bereits Hildegard von Bingen vor etwa 1000 Jahren¹. Auch heute wissen wir, dass die Augen zum Teil unsere Absichten [4], unser Interesse [1], aber auch unser Sprachverständnis [23] und generell unsere Aufmerksamkeit [1, 45] verraten können. Diese Tatsache kann man sich für die implizite Interaktion von Menschen mit technischen Geräten mittels Eye-Tracking zu Nutze machen, indem die Systeme auf nicht bewusst getätigten Nutzer-Input angemessen reagieren und sich Anwendungen und graphische Oberflächen entsprechend anpassen. Das kann entweder live in Interaktivsystemen oder durch das Erstellen von Modellen u.a. im Rahmen der nachträglichen Analyse der Daten in Diagnosesystemen geschehen [1].

¹ <https://www.aphorismen.de/zitat/16932>, zuletzt geprüft am 12.08.2019



© Johannes Sylupp;

licensed under Creative Commons License CC-BY

Cite as: Johannes Sylupp. Implizite Interaktion mittels Eye-Tracking. In *3rd Seminar on Ubiquitous Interaction (UBIACTION 2019)*. Editors: Pascal Knierim, Jakob Karolus, Fiona Draxler, Ville Mäkelä, Tonja Machulla, Florian Lang, Thomas Kosch, Albrecht Schmidt. July 16, 2019. Munich, Germany. pp. 7:1–7:22.

7:2 Implizite Interaktion mittels Eye-Tracking

In dieser Seminararbeit soll eine Übersicht zum aktuellen Stand der Forschung in Bereich impliziter Interaktion mit Hilfe von Eye-Tracking gegeben werden. Schmidt [46] sieht einen allmählichen Übergang von expliziter zu impliziter Interaktion durch neue technische Möglichkeiten. Graphische Oberflächen könnten durch anpassbaren sensorischen Oberflächen abgelöst beziehungsweise stärker ergänzt werden [51]. In dem Review werde ich unterschiedliche Anwendungsszenarien von anpassungsfähigen Systemen aufzeigen.

Es sind zahlreiche Anwendungsgebiete zu impliziter Interaktion mittels Eye-Tracking zu finden. Es wird z. B. beschrieben, dass Bilder, nachdem sie betrachtet wurden, recht gut zugeschnitten werden können [45]. Oder, dass personalisierte Werbung durch bloße Augenbeobachtung angezeigt werden kann, ohne dass lange Nutzerprofile angelegt werden müssen [1]. Auch Filme können zum Beispiel bei Abwesenheit automatisch pausiert werden [48]. Ein spannendes Anwendungsfeld ist ebenso ein Roboter, der Spielzeug automatisch einem Menschen reicht, sobald dieser auf das Objekt blickt und somit implizit Interesse daran signalisiert [38].

In der Arbeit werden zuerst die Begriffe implizite und explizite Interaktion in Kapitel 2 definiert und Anwendungsbeispiele im Gebiet Eye-Tracking aufgezeigt. Eine kurze Einweisung in Eye-Tracking findet sich dann im Abschnitt 3. Im Anschluss wird der Leser in Adaptive und Attentive User Interfaces eingeführt (Kap. 4). Nach etwas Hintergrundwissen wird die vorhandene Literatur zu impliziter Interaktion mittels Eye-Tracking nach diesen drei Gesichtspunkten behandelt: Echtzeit-Interaktivsysteme (im Kapitel 5), verzögerte Interaktivsysteme (siehe Kap. 6) und Diagnosesysteme im Kapitel 7. Danach werden Einschränkungen und Bedenken thematisiert (siehe Kap. 8). Zum Abschluss wird in Kapitel 9 ein Fazit gezogen und Ausblicke vorgestellt.

2 Implizite Interaktion mittels Eye-Tracking

Zuerst will ich die Definition des Begriffs "Implizite Interaktion" betrachten.

2.1 Definition von impliziter Interaktion

Im Bereich der Mensch-Maschine-Interaktion (MMI) existieren unterschiedliche Definitionen über implizite Interaktion. Serim et al. [47] identifiziert u.a. diese in der Literatur vorhandenen Attribute: Unbeabsichtigt und unbewusst. In dieser Arbeit beziehe ich mich auf deren Begriffsverständnis, dass ein *System auf nicht absichtlich getätigten Nutzerinput reagiert*. Serim et al. [47] schreibt:

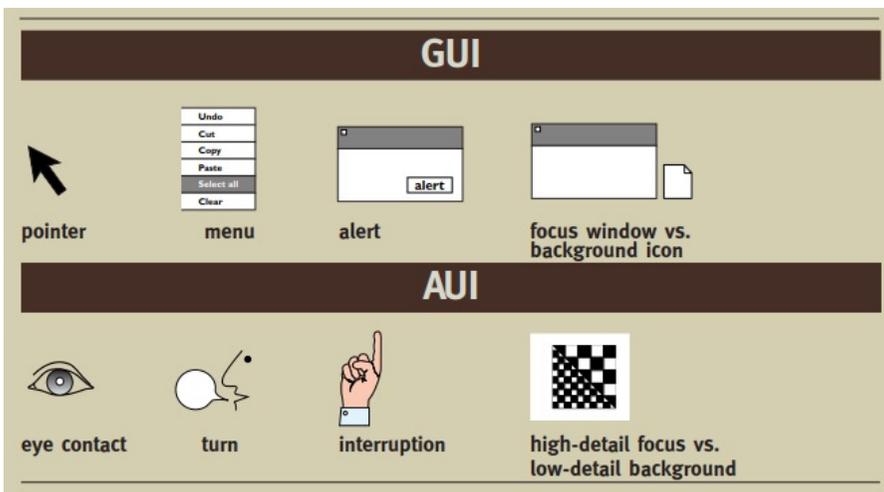
We define implicit interactions as interactions in which the appropriateness of a system response to the user input (i.e., an effect) does not rely on the user having conducted the input to intentionally achieve it.

2.2 Definition von expliziter Interaktion

Laut Atterer et al. [2] ist eine scharfe Abgrenzung von explizit und implizit nicht so einfach möglich, weil unklar ist, was genau als bewusst und was als unbewusst einzustufen sei [2]. Außerdem ergänzen sich oft beide Arten [46]. Ich definiere explizit hier als bewusste und beabsichtigte Interaktion. Mall sagt: "Von explizitem Input mittels Eye-Tracking spricht man, wenn Personen Augenbewegungen bewusst einsetzen, um einen Computer zu bedienen" [33].

2.3 Übergang von expliziter zu impliziter Interaktion

Schmidt [46] beschreibt einen allmählichen Übergang von expliziter zu impliziter Interaktion, der aus neuen sensorischen Fähigkeiten von technischen Geräten resultiert. Ein Wechsel sei gerade bei Smartphones notwendig, weil dort Aufgaben oft nur kurz und nebenbei getätigt werden [46]. Wo es früher einen Wechsel zu graphischen Oberflächen (GUIs) gab, liege nun vermutlich ein Übergang zu aufmerksamen und anpassbaren sensorischen Oberflächen (AUIs, siehe 4) vor uns [51]. Die Unterschiede werden in Abb. 1 ersichtlich.



■ **Abbildung 1** Gegenüberstellung von bisherigen graphischen Nutzeroberflächen (GUIs) und Attentive User Interfaces (AUIs). Ein Teil davon zeigt, dass explizite bzw. manuelle Interaktion durch implizite Interaktion bzw. Sprachbefehle ersetzt wird. Bildnachweis: "Attentive User Interfaces" von Vertegaal [51].

Explizite Interaktion wird aber vermutlich nicht komplett durch Implizite Interaktion ersetzt werden, weil je nach Intention des Herstellers und Anwendungsszenario eine andere Interaktionsart sinnvoll ist.

2.4 Anwendungsbeispiele von impliziter und expliziter Interaktion mittels Eye-Tracking

Explizite wie implizite Interaktion haben ihre Stärken und Schwächen, weswegen der Anwendungskontext relevant ist.

Es gibt Bereiche, wo die implizite Variante besser eingesetzt werden kann und sollte. Denn diese kann im Vergleich zur expliziten Interaktion für den Nutzer einfacher, weniger kognitiv anstrengend und schneller sein, u.a. weil keine extra Bürde wie beim Ausfüllen eines Feedback-Formular besteht [5, 9]. Eine klare Favorisierung davon ist in Situationen, wo auf Ablenkung oder Abwesenheit einer Person reagiert werden soll, wie im Abschnitt 5.6 beschrieben. So kann ein mobiles Gerät seinen Bildschirm automatisch beim Weglegen ausschalten [46] ohne dass umständlich ein Knopf gedrückt werden muss.

Andererseits ist manchmal der Gebrauch von expliziter Interaktion zu bevorzugen, z.B. wenn klare Aussagen vom Nutzer benötigt werden oder die volle Kontrolle gewährleistet werden soll. Ein Beispiel wären Bildschirmtastaturen für eingeschränkte Personen, die bewusst mit den Augen betätigt werden. Ebenso können Gesten verwendet werden. Diese haben den Vorteil, dass sie sehr sicher erkannt werden können, allerdings besteht die Gefahr, dass sie nur bedingt intuitiv und gewohnt für den Anwender sind [13].

Kombinationen beider Arten sind durchaus manchmal sinnvoll, wie z. B. bei der impliziten Objekt-Selektion von MAGIC Pointing (siehe 5.1) [56].

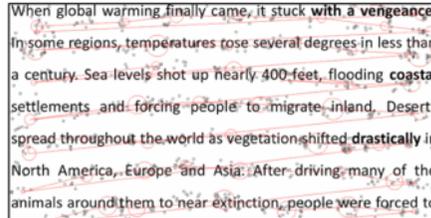
In vielen Fällen entscheidet letztlich der Fokus des Auftraggebers und das Anwendungsfeld darüber, welche Art besser passen könnte. Je nachdem ob eine subjektivere explizite Einschätzung, die auch Unter- oder Überschätzung einschließt, oder eine objektivere implizite Beobachtung gewünscht ist, sollte der Interaktionstyp anders gewählt werden. So kann auf der einen Seite z. B. eine Skalenbewertung zu Vorlieben und Interessen des Nutzers erbeten werden [10]. Jedoch werden explizite Bewertungen bzw. Interaktionen im Web-Kontext manchmal auch als störend empfunden [9, 10]. Auf der anderen Seite ist es möglich das Interesse an Produkten relativ genau durch implizite Interaktion mit Hilfe von Eye-Tracking herauszufinden [25, 30, 31].

2.5 Abgrenzung von Interaktiv- und Diagnosesystemen

Der Fokus der Arbeit liegt in der Beschreibung von Anwendungsbeispielen aus der Literatur. Dabei ist es hilfreich zwischen Interaktiv- und Diagnosesystemen zu unterscheiden. Während Echtzeit-Interaktivsysteme nach meiner Definition innerhalb weniger Sekunden auf den Nutzer reagieren, sammeln verzögerte Interaktivsysteme zuerst einige Sekunden bis Minuten Daten und passen dann den Inhalt benutzerspezifisch an. Diagnosesysteme werten die Daten erst im Nachhinein zwecks Nutzerfreundlichkeit oder Verhaltensanalyse aus [1].

3 Hintergrundinformationen zu Eye-Tracking

Die Augenbewegungen beim Lesen gleichen einer Sequenz aus Fixationen und Sakkaden [27, page 92] (siehe Abb. 2). Neben Eye-Tracking-Input werden für implizite Interaktion noch Tippgeschwindigkeit [2], Suchhistorie, Klickdaten, Scrolling, Lesezeichen oder Objekt-Anzeigedauer herangezogen [55].



■ **Abbildung 2** Beispielhafte Augenbewegung beim Lesen. Kleine Punkte: Rohdaten; roter Kreis: berechnete Fixation; rote Linie: Sakkade. Bildnachweis: [8, page 495].

3.1 Fixationen

Bei Fixationen schaut das Auge unbewegt auf einen Punkt [21, 28, 43], um Informationen zu extrahieren [28, 43]. Hierbei erfolgt hauptsächlich die Bildfassung und -verarbeitung [20]. Es ist für uns interessant, weil sich oft, wenn auch nicht immer [5], aus Fixationsmenge und -Dauer die Aufmerksamkeit [21, 45] und das Interesse des Nutzers ableiten lässt [9, 1]. Die Dauer einer einzelnen Fixation variiert in der Literatur von 100 - 500 ms [21, 27, 32, 43].

3.2 Sakkaden

Sakkaden sind sehr schnelle Sprünge des Auges zwischen Fixationen. Diese sind z. B. vorzufinden, wenn man auf neu in unsere Sicht gekommene Objekte schaut oder liest [28, 21, 43]. Die Sakkaden-Länge korreliert mit der Relevanz eines Textes für den Leser und ist deshalb für dessen Interessen-Ermittlung wichtig [5]. Eine Sakkade dauert i.d.R. zwischen 30 - 80 ms [21, 27].

3.3 Blinzeln

Das Blinzeln der Augen geschieht normalerweise unbewusst und unbemerkt [44]. Siegle et al. [49] beschreiben, dass Blinzeln wohl vermehrt vor und nach der Verarbeitung von Perzeption und Information geschieht [49]. Blinzeln-Erkennung kann zum expliziten Interagieren genutzt werden, z. B. zur bewussten Item-Selektion [34]. Es scheint als korreliere das Sprachverständnis eines Lesers mit längerer durchschnittlicher und totaler Blinzeldauer [23].

4 Adaptive User Interfaces

4.1 Definition

Laut Oppermann sind Adaptive User Interfaces Schnittstellen, die sich an den Nutzer mittels vorher über ihn gesammelten Daten anpassen [37]. Damit fallen alle Interaktivsysteme, die in Kapitel 5 und 6 erläutert werden, darunter.

Der Vorteil von den adaptiven Nutzerschnittstellen ist, dass sie nicht statischen Inhalt präsentieren, sondern diesen dynamisch auf den oder die aktuellen Anwender einstellen. Dadurch kann die Benutzerfreundlichkeit, der Komfort, die Zufriedenheit, aber auch Werbeeinnahmen durch zielgerichtetere und personalisiertere Anzeigen steigen. Ein Beispiel dafür wäre, dass die Schriftart einer graphischen Oberfläche graduell zunimmt, wenn erkannt wird, dass der Nutzer Schwierigkeiten beim Lesen hat.

4.2 Spezialfall: Attentive User Interfaces

Ein Spezialfall davon sind Attentive User Interfaces (AUIs). Diese erkennen und reagieren fokussiert auf die menschliche Aufmerksamkeit [51].

AUIs arbeiten ohne explizite Befehle von Maus oder Tastatur [1, 14] (siehe Abb. 1). Stattdessen wird die Aufmerksamkeit des Nutzers u.a. aus den Metriken Anwesenheit, Blickrichtung, Entfernung zum Gerät und mehrerer Augendaten berechnet [1, 45, 51]. Weil die kognitive Aufmerksamkeitsspanne des Menschen limitiert ist, versuchen AUIs diese nicht unnötig abzulenken, sondern schätzen die Nachrichtenpriorität zum Teil im Zusammenspiel mit anderen Geräten ein. Somit zeigen sie Informationen nur an, wenn sie die Aufmerksamkeit des Anwenders sowieso besitzen oder diese besonders wichtig sind. [1, 51] Mehrere Beispiele hierfür werden im Unterkapitel 5.6 beschrieben.

5 Echtzeit-Interaktivsysteme zur Impliziten Interaktion

Alt et al. [1] zeigt den Unterschied von Interaktiv- und Diagnosesystemen im Bereich Eye-Tracking in der Literatur auf [1]. Im Folgenden werde ich Beispiele für Echtzeit-Interaktivsysteme anführen, welche innerhalb weniger Sekunden auf den gesammelten impliziten Input eines Nutzers reagieren.

5.1 Implizite Selektion von Objekten

Es existieren verschiedene implizite Methoden zur Objekt-Selektion. Da Maus- und Augenposition bei PCs stark korrelieren [7], kann ein Klicken an der richtigen Stelle auf dem Bildschirm verschnellert und vereinfacht werden, indem der Mauszeiger automatisch den Augenbewegungen folgt [56].

Dies setzt MAGIC Pointing um, das ein schnelles Klicken durch Positionieren des Cursors auf dem betrachteten Bildschirmbereich ermöglicht [56].

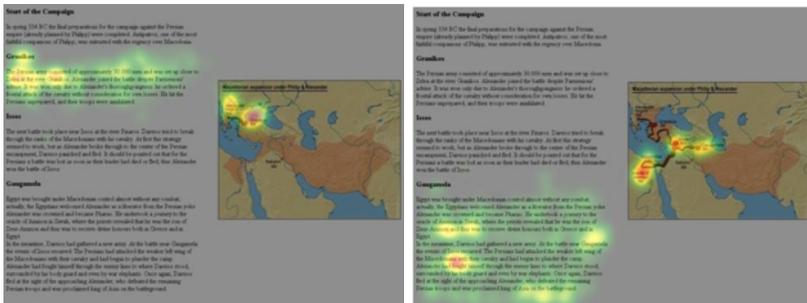
Eine andere Variante ist EyeWindows, das die gewünschte Anwendung durch Anschauen in den Fokus rückt [15]. Hierbei wurden zwei Prototypen untersucht, bei dem Applikations-Fenster ohne Überlappung mit oder ohne freiem Platz nebeneinander angeordnet sind. Mehrere Fenster waren parallel sichtbar, wobei das Aktive deutlich größer war als der Rest. Neu selektierte Fenster wurden gezoomt. Das Resultat von Fono und Vertegaal [15] war, dass die automatische augenbasierte Aktivierung in etwa doppelt so schnell war wie die Manuelle. Die Kombination von Eye-Tracking mit tastaturbasierter Aktivierung wurde hierbei bevorzugt [15].

Zusätzlich konnte gezeigt werden, dass eine Kombination aus Eye-Tracking und EEG recht zuverlässig zur impliziten Selektion auf graphischen Benutzeroberflächen nutzbar ist, auch wenn hier weitere Forschung nötig ist [41].

5.2 Einblenden relevanter Zusatzdaten

Ein weiteres Beispiel ist iTourist. Die Anwendung ermöglicht die Planung einer Städtereise, indem man durch die Stadtkarte navigiert. Dabei werden Extraintformationen zu Orten und Gebäuden wie Bilder, die Dauer von Punkt A zu B sowie wissenswerte Hintergrundinformationen ausschließlich aufgrund der gemessenen Augendaten automatisch angezeigt bzw. vorgelesen. [42]

Bei der Online-Lernplattform AdeLe können zusätzliche Erklärungen bei mittels Eye-Tracking erkannten Verständnisproblemen eingeblen-det und Bilder oder Karten je nach aktuellem Lesefortschritt im Seitenbereich aktualisiert werden (siehe Abb. 3) [18].

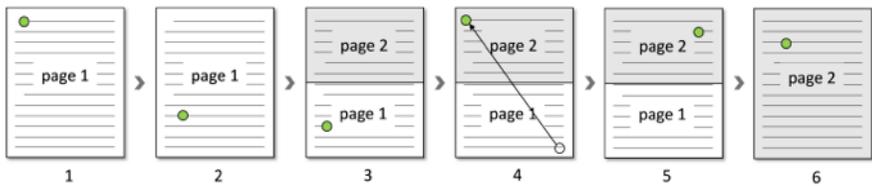


■ **Abbildung 3** Der Leser bekommt bei der Online-Lernplattform AdeLE automatisch je nach Lesefortschritt (links) die dazugehörige Karte auf der rechten Seite eingeblen-det. So kann man z. B. den Verlauf von Feldzügen, wie hier von Alexander dem Großen, gut nachvollziehen. Bildnachweis: Gütl et al. [18].

5.3 Unterstützung des Leseflusses

Darüber hinaus kann der Lesefluss durch implizite Interaktion mittels Eye-Tracking verbessert werden. Zum einen kann beim automatischen Scrollen auf Digitalgeräten die Scrollgeschwindigkeit dynamisch ans Lesetempo angepasst werden [29]. Zum anderen kann der kommende Inhalt (Halbseite bzw. Zeile) einer statischen Seite unauffällig im gerade nicht gelesenen Teil aktualisiert werden. Dies existiert beim Lesen von Text [11] und Musiknoten [3].

So wird bei AutoPager eine halbe Seite von einspaltigem Text nach Überschreiten eines Schwellenwerts graduell unbemerkt eingublendet durch das Nutzen des Change Blindness Effekts. Dadurch ist die neue Seite immer entweder im oberen oder unteren Bildschirmteil vorhanden, sodass man ohne Zutun ein ganzes Buch lesen kann. Dies wurde Scrolltechniken vorgezogen. [11]



■ **Abbildung 4** Übergang von einer gelesenen Seite zur nächsten bei AutoPager. Es wird schrittweise eine Halbseite unbemerkt erneuert. Bildnachweis: Wilson et al. [11].

Bonnici et al. [3] hingegen bereiten zuerst Klaviernoten auf, indem sie wiederholte Passagen mehrfach anhängen, damit die Noten am Stück durchgespielt werden können. Während des Spielens ersetzen sie dann jeweils gelesene Musik-Zeilen, wobei dem Nutzer die Kontrolle über das System bleibt. [3]

5.4 Einsatz im Gaming- und Grafik-Bereich

Im Gaming-Bereich kann implizite Interaktion mittels Eye-Tracking genutzt werden, um das Spielerlebnis zu verbessern und die Aufgabenkomplexität zu reduzieren. So könnte bei einem Rollenspiel automatisch vom Defensiv- in den Angriffsmodus geschaltet werden, wenn der Spieler statt den eigenen Bereich, wo Feinde angreifen können, den Bildschirmteil des Endgegners anschaut. [35]

In der 3D-Graphik wird es ermöglicht Rechenzeit zu sparen und zugleich die Realitätsnähe von Bildmaterial zu erhöhen, indem nur Elemente im Bereich der Fovea, einer kleinen Region im Auge mit größter Sehschärfe [28], hochauflösend gerendert werden und alles außerhalb mit reduzierter Auflösung [17].

5.5 Anpassungen an Sprachkenntnisse

Die Anpassung an Sprachkenntnisse kann sowohl im Interaktiv- als auch Diagnosesystem (siehe 7.3) Anwendung finden. In diesem Abschnitt gehe ich auf Echtzeit-Anpassungen ein.

Zum einen kann auf Verwirrtheit bzw. Unfähigkeit des Nutzers, den Text zu lesen reagiert werden, indem z. B. andere Sprachen angeboten werden [23]. Zum anderen können schwer lesbare Wörter erkannt werden [26] und könnten in Kombination mit dem Übersetzungsservice von Toyama et al. [50], das ursprünglich mit Gesten angestoßen wird, verwendet werden [50]. Die Online-Lernplattform AdeLe kann zusätzliche Erklärungen bei erkannten Verständnisproblemen einblenden [18].

Vor kurzem hat Karolus et al. [23] bei einer Studie mit 21 Teilnehmern herausgefunden, dass die Sprachfähigkeit von Nutzern recht gut durch Fixationen und Blinzeldauer ermittelt werden kann. Da für maximal 4,5 Sekunden gemessen wurde, kann es als Echtzeit-Interaktivsystem angesehen werden. So korrelierte das Sprachlevel C2 mit hoher Gesamtblinzelrate sowie langer durchschnittlicher Blinzeldauer und geringer Refixierungsrate sowie kurzer durchschnittlicher Fixationsdauer; umgekehrt bei Sprachstufe A1. [23] Dieses Wissen könnte z. B. genutzt werden, um graphische Oberflächen von Ticketautomaten auf Sprachfähigkeit reagieren zu lassen und entweder alternative Sprachen anzubieten oder die verwendete Sprachkomplexität zu reduzieren.

5.6 Erleichtertes Fortfahren nach Abwesenheit

Als letzter Unterpunkt sollen drei AUIs genannt werden, die mit Abwesenheitserkennung bzw. erleichtertem Wiederaufnehmen danach zu tun haben.

Das Attentive User Interface EyeBookmark ermöglicht das einfache Weiterlesen eines Buches nach einer Unterbrechung durch das Hervorheben des letzten Lesepunktes oder - je nach Einstellung - bis hin zum letzten Absatz. Die Granularität von Satz bis Absatz wurde von Studienteilnehmern bevorzugt. [22] SmoothGaze von Cheng et al. [8] funktioniert auf ähnliche Weise. Es ermöglicht das Lesen über mehrere Geräte hinweg, indem es den letzten Fixationspunkt anzeigt. Es blendete zusätzlich Anmerkungen über unbekannte Worte, die auf einem anderen Gerät nachgeschlagen wurden, im Text ein [8].

Ähnlich markiert das System Gazemarks den auf einer Karte zuletzt betrachteten Ort mit einem kreisförmigen Licht-Kegel, wobei das Außenrum ausgegraut ist. Bei ortsspezifischen Aufgaben, wo zwischen zwei Displays gewechselt werden musste, wurde das System als sehr hilfreich empfunden [24].

Eine weiteres Beispiel ist der AttentiveTV (siehe Bild 5), der bei visueller Unaufmerksamkeit von Menschen ein Video automatisch pausiert [48].

7:10 Implizite Interaktion mittels Eye-Tracking



■ **Abbildung 5** Der AttentiveTV pausiert automatisch einen Film, wenn er keine Aufmerksamkeit des Zuschauers mehr bekommt. Bildnachweis: Shell et al. [48].

Dieses Feature kann auch beim Lesen von einer Smartwatch im Gehen praktisch sein, wenn man sich nicht permanent auf das Lesen konzentrieren kann. Dingler et al. [12] entwarf ein System, das, sobald der Lesefluss des Nutzers auf einer Smartwatch gestört ist, pausiert oder auch zum Satzanfang zurückspult. Normalerweise musste nach damaligem Stand per Hand das automatische Scrollen unterbunden bzw. der Text zurückgescrollt werden. [12]

6 Verzögerte Interaktivsysteme zur Impliziten Interaktion

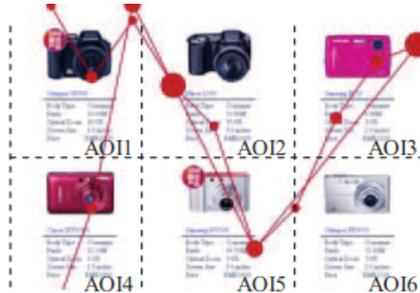
Neben Echtzeit-Interaktivsystemen, die unmittelbar auf den Nutzer reagieren [1], ordne ich Systeme, die weder sofort reagieren noch den Diagnosesystemen zuordbar sind, als verzögerte Interaktivsysteme ein. Darunter fallen alle Systeme, die zuerst einige Sekunden bis Minuten Daten sammeln und dann noch während der Nutzung den Inhalt entsprechend für den Nutzer anpassen.

6.1 Angepasste Web-Werbeanzeigen

Alt et al. [1] fand mit 12 Studienteilnehmern eine signifikante Aufmerksamkeitserhöhung durch die Adaption visueller Werbung basierend auf der Augenverweildauer im Vergleich zu randomisierter Werbung heraus. Dazu wurden in einer ersten Phase die Verweildauer auf Objekten einer Amazon-Webseite aufgezeichnet. Sobald auf eine andere Unterseite navigiert wurde, wurden Werbeanzeigen aktualisiert. Ein Vorteil ist, dass keine Datenspeicherung notwendig ist und flexibles Ein- und Ausschalten ermöglicht wird. [1]

Bei Cheng et al. [9] war das Ziel Nutzerpräferenzen zu Digi-Cams implizit herauszufinden, um passende Produkte zu empfehlen. Dadurch soll der Aufwand für explizite Angaben des Nutzers gemindert werden.

Bei der Studie sollten neun Teilnehmer mehrmals solange verschiedene digitale Kameraangebote mehrerer Kameragenerationen am PC durchschauen (Abb. 6), bis sie eine Kaufpräferenz hatten. Mit einer korrekten Prognoserate von 87,5 % wurde das Modell zum angegebenen Interesse mit Hilfe der Metriken Fixationsmenge, -Dauer und erster Fixation und Kleinheit des Pupillendurchmessers recht gut im Rahmen eines Optimierungsproblems berechnet. Es wurde sowohl das Layout (Bild oder Textpräferenz) als auch der Inhalt beim Wechsel zur nächsten Kamerageneration angepasst. [9]



■ **Abbildung 6** Prototyp von Cheng et al. [9] zur Erkennung von Nutzerpräferenzen für Digitalkameras. Dieser misst Fixationen (rote Punkte) auf Produkten. Bildnachweis: Cheng et al. [9].

6.2 Adaptive Online-Lernplattform

Die Web-Lernplattform AdeLe von Gütl et al. [18] umfasst alle drei Systeme, Echtzeit-Interaktiv und verzögerte Interaktiv- sowie Diagnosesysteme. Zu live-Features wurde bereits im vorigen Kapitel in 5.2 berichtet. Hier sei ergänzend zu sagen, dass die Plattform feingradige Nutzerprofile während der Nutzung anhand von Eye-Tracking-Daten und Browserverhalten erstellt. So kann es die Inhalte je nach erkannter Repräsentationspräferenz, sprich ob der Anwender Text oder Bild bevorzugt, unterschiedlich darstellen [18].

6.3 Suchmaschinenoptimierung

Mehrere Autoren haben gezeigt, dass implizite Interaktion mittels Eye-Tracking für Suchmaschinenoptimierung geeignet sein könnte, wenn auch aktuell die Verfahren noch nicht ganz perfekt sind [5, 16, 40]. Dabei war es für das Bild- und Videoempfehlensystem von Xu et al. [55] hilfreich, die Blickdaten mit weiterem impliziten Wissen zu kombinieren, wie z. B. Suchhistorie, Klickdaten, Scrolling, Anzeigezeit, erstellte Lesezeichen und ähnliche Metriken [55].

7:12 Implizite Interaktion mittels Eye-Tracking

Giordano et al. [16] extrahiert erst für den Nutzer interessante Begriffe, fasst sie dann zu Clustern zusammen, und blendet darauf basierend Vorschläge ein. 73,3 % der Nutzer (22 von 30) waren mit dem System zufrieden. [16]

Puolamäki et al. [40] kombiniert vielversprechend implizites Relevanzfeedback der Titel von Wissenschaftsartikeln mit einer kollaborativen Präferenzdatenbank. Dies verbesserte die Suche von solchen Artikeln signifikant im Vergleich zur zufälligen Anzeige, besaß aber noch Verbesserungspotenzial. [40]

7 Diagnosesysteme zur Impliziten Interaktion

Diagnosesysteme im Bereich Eye-Tracking sind laut Alt et al. [1] solche Systeme, die aufgezeichnete Daten nach der Nutzerinteraktion auswerten und damit arbeiten beziehungsweise Wissen generieren [1]. Dazu zählen sowohl die automatische Bildbearbeitung nach dem Betrachten, die Diagnose von Sprachkenntnissen und anpassbare Online-Lernplattformen.

7.1 Implizites Relevanz-Feedback

Diagnosesysteme können unter anderem dazu dienen, implizites Relevanzfeedback zu erhalten. So kann die Wichtigkeit eines gelesenen Textes für den Leser mit Hilfe von Eye-Tracking evaluiert werden [5]. Dafür soll das Verhältnis von gelesenen zu übersprungenen Abschnitten ein guter Indikator sein [6]. Ebenso kann hiermit der Erfolg von Webinhalten ermittelt werden [1].

Es ist relativ klar ersichtlich, welches Bild bei einer Aufgabe für den Nutzer am relevantesten ist, weil es z. B. länger betrachtet wurde als der Rest [19, 39]. Pallez et al. [39] entwickelte eine Annäherungsformel, die mit ihrer Relevanzvoraussage in 85 % der Fälle richtig lag [39]. Auf ähnliche Weise ist die Kaufabsicht eines Kunden für ein Produkt tendenziell ermittelbar [31]. Das Interesse an einem Spiel aus einem Play-Store wurde in einer Studie zum Beispiel mit 90,32 % korrekt vorhergesagt [30].

Diese Relevanzresultate können wiederum genutzt werden, um Prognosen zur Wichtigkeit von Inhalten für Anwender in Interaktivsysteme zu integrieren. Zwei weitere Diagnosesysteme werden im Folgenden näher beleuchtet.

7.2 Automatische Bildbearbeitung

Das erste Anwendungsgebiet innerhalb der Bildbearbeitung ist das automatisierte Beschriften und Kategorisieren von Bildern. Walber et al. [52] zeigten vor wenigen Jahren, dass im Rahmen einer Bildersuche bestimmten Regionen eines Bildes, die für den Nutzer zum Suchbegriff passten, Begriffe zugeordnet werden können. Jedoch besteht eine starke Varianz je nach Bild. [52]

Der zweite Anwendungsbereich ist das automatisierte Zuschneiden von betrachteten Bildern. Da geschossene Photos normalerweise im Nachhinein angeschaut werden, stehen implizite Relevanzdaten vom Eye-Tracker zur Verfügung [45]. Der Vorteil der impliziten Variante im Vgl. zur expliziten ist eine enorme Zeitersparnis der zum Teil lästigen Fleißarbeit [1]. Zudem war dieser Ansatz deutlich besser als bisherige automatische Zuschnitte (Abb. 7), die teilweise eher abgelenkt haben, jedoch weiterhin schlechter als manuell zugeschnittene Bilder waren [45]. Konkret wurde bei Santella et al. [45] so vorgegangen: Das Bild wurde in Segmente unterteilt und diese nach im Vorder- und Hintergrund eingestuft. Fixationen wurden dann diesen Bereichen zugeordnet, was ein gewisses Interesse an manchen Regionen indizierte. Mittels Kompositionsregeln wurde abschließend ein Zuschnitt erstellt [45].



■ **Abbildung 7** Beispiel von Eye-Tracking basiertem Zuschneiden von Bildern mit dem Algorithmus von Santella et al. [45]. Links zeigt das Originalbild, in der Mitte ist das recht gut zugeschnittene Bild per Blickbeobachtung. Rechts ist ein nicht so erfolgreich zugeschnittenes Bild zu sehen, das per visueller Prägnanzmethoden berechnet wurde, das hier Wichtiges mit Unwichtigem verwechselte.

7.3 Diagnose von Sprachkenntnissen

Ein Vorteil von der Diagnose von Sprachkenntnissen (siehe auch 5.5) mittels Eye-Tracking ist, dass individuelles Lernen verstärkt werden könnte. Auch wenn die Vorstudie durch zwei Teilnehmer bedingt allgemeingültig ist, konnten Englischlehrer anhand von automatisch erstellten Eye-Tracking Daten beim Lesen von zwei Fremdsprachentexten (Fixationsdauer, Lesegeschwindigkeit, schwierige Wörter und Sätze, mehrfach lesen, Fixationsbereiche), die individuelle Lese- und Sprachfähigkeit der Schüler recht gut ablesen. [36]

Bei einer initialen Studie mit fünf Teilnehmern waren schwere englische Wörter durch Eye-Tracking-Aufnahmen erkennbar. Jedoch korrelierte in dem Fall die Sprachfähigkeit nicht perfekt mit den Daten. Der Ausreißer liege vermutlich an einem während der Studie sehr müde gewesenem Studenten [26].

8 Limitations

Neben vielen vielversprechenden Aussichten der impliziten Interaktion mittels Eye-Tracking gibt es auch ethische und technische Bedenken.

Der Haupteinwand dürfte in der Privatsphärenverletzung liegen [1, 2] und im Kontrollverlust [12, 35]. Das Privatsphärenproblem kann zum einen gelöst werden, wenn die Daten nur lokal zur Nutzzeit gespeichert und nicht übertragen und ausgewertet werden. Zum anderen durch Privatsphäreneinstellungen und transparente Datenverarbeitung. Dem gefühlten Kontrollverlust könnte mit einer Ein- und Ausschaltvorrichtung, der Kombination mit expliziter Interaktion sowie noch unbemerkbarer bequemer Technik begegnet werden. So hat man sich inzwischen auch an einige Helfer im Auto wie das ABS oder den Spurhalteassistenten gewöhnt ohne diese stark als Kontrollverlust anzusehen, sondern eher als Hilfe. Auf ähnliche Weise könnte man sich allmählich an die neuen impliziten technischen Möglichkeiten gewöhnen.

Ein weiterer Punkt ist aktuell noch die Gefahr technischer Probleme wie Trackingverlust oder Rauschen, aber auch die Kalibrierung kann umständlich sein [35]. Eine hohe Zuverlässigkeit ist aber für die Akzeptanz notwendig [24]. Auch wenn technisch noch nicht alles perfekt ausgereift ist, wird passable Hardware immer günstiger und könnte in Laptops etc. standardmäßig verbaut werden [1]. Schon vor etwa 15 Jahren wurde anhand der damaligen Forschung prognostiziert, dass Eye-Tracking in Zukunft auch mit günstigen Geräten und mit impliziter Kalibrierung erfolgen kann [13, 14, 18, 53]. Aktuell sind bereits in wenigen Laptops integrierte Eye-Tracker verbaut², jedoch noch längst nicht serienmäßig. Interessanterweise gibt es einige Versuche Eye-Tracking allein mit integrierter Gerätekamera zu ermöglichen [31, 34, 54].

9 Zusammenfassung und Ausblick

Zusammenfassend lässt sich sagen, dass einige faszinierende und vielversprechende Anwendungsgebiete zur implizite Interaktion mittels Eye-Tracking, aber auch Privatsphäre-, Kontrollverlust- und Zuverlässigkeitsbedenken der Technik existieren. In der Seminararbeit wurde die Forschung nach Echtzeit- und verzögerte Interaktivsysteme sowie Diagnosesysteme separiert betrachtet.

Bei Echtzeit-Interaktivsystemen, die unmittelbar auf den Nutzerinput reagieren und ihr Interface anpassen (Adaptive User Interface), haben wir gesehen, dass bei der Städteplanung [42] und bei einer Lernplattform Zusatzinformationen vorteilhaft eingeblendet werden können [18].

² <https://windowsreport.com/eye-tracking-laptops/>, zuletzt geprüft am 12.08.2019

Attentive User Interfaces als Unterkategorie davon ermöglichen es, auf die Aufmerksamkeit des Nutzers in vielfacher Weise zu reagieren. So kann nach Unterbrechungen problemlos weitergelesen bzw. -gearbeitet werden, indem die zuletzt genutzte (Text-) Stelle graphisch hervorgehoben wird [8, 22, 24]. Das Reagieren auf erkannte Schwierigkeiten mit einem Text oder einer Sprache [18, 23, 26] kann interessante Anwendungen ermöglichen. Ich könnte mir einen Ticketautomaten vorstellen, der schnell erkennt, ob die angezeigt Sprache verstanden wird oder nicht und ggf. schnell zur Sprachauswahl wechselt. Dafür muss die Technik allerdings noch ausgereifter werden und mit schwierigen Lichtverhältnissen umgehen können. Darüber hinaus könnte ein Sofortübersetzung von unbekanntem Wörtern beim Lesen einer Fremdsprache hilfreich sein.

Interessant war bei verzögerten Interaktivsystemen, die erst ein paar Sekunden bis Minuten Daten sammeln, um dann das System anzupassen, dass Werbe- [1, 9] und Suchmaschinenadaptionen und -optimierungen gut funktionierten, wenn auch noch nicht perfekt [16, 40, 55]. Ich würde einen starken Anstieg der Nutzung sehen, sobald die Technik ausgereift genug und standardmäßig in Geräten vorhanden ist und zudem die Skepsis der Nutzer gesunken ist. Speziell wenn keine Nutzerprofile angelegt werden müssten [1], könnte es manche Datenschutzliebhaber bewegen umzusteigen.

Diagnosesysteme haben ein großes Potenzial, weil sie das Sprachverständnis [23, 26, 36], die Relevanz von Bildern [45], Produkten [30, 31], Texten [5] und Webseiten [1], die Präferenz von Text oder Bildern [18] mit recht hoher Wahrscheinlichkeit erkennen können und damit Adaptive Interfaces gestaltet werden können. Mit Hilfe dieses Wissens werden bereits Webseiten nach optimalem Aufbau evaluiert. In die Zukunft gedacht, wären diese Features bei einer Online-Lernplattform zum Lernen von Fremdsprachen sehr gewinnbringend. Denn zum einen könnte sich das Interface auf Lernvorlieben und -fortschritt anpassen, zum anderen könnten beim Leseverständnis sehr genau schwierige Passagen und Fortschritte in Lesegeschwindigkeit und Verständnis für jeden Nutzer aufzeigt werden - stärker als es mit bisherigen Mitteln möglich wäre.

Möglicherweise wird Eye-Tracking bald über integrierte Kameras funktionieren [31, 34, 54] und somit einen riesigen Markt für die implizite Interaktion eröffnen und manche beschriebene Features alltagstauglich machen. Insgesamt können wir uns also auf interessante Neuerungen in der Zukunft freuen.

Literatur

- 1 Florian Alt, Alireza Sahami Shirazi, Albrecht Schmidt, and Julian Menenöh. Increasing the user's attention on the web. In Lone Malmberg, editor, *Proceedings of the 7th Nordic Conference on Human-Computer In-*

7:16 Implizite Interaktion mittels Eye-Tracking

- teraction Making Sense Through Design*, page 544, New York, NY, 2012. ACM. doi:10.1145/2399016.2399099.
- 2 Richard Atterer, Monika Wnuk, and Albrecht Schmidt. Knowing the user's every move. In Leslie Carr, David de Roure, Arun Iyengar, Carole Goble, and Mike Dahlin, editors, *Proceedings of the 15th international conference on World Wide Web*, ACM Digital Library, page 203, New York, NY, 2006. ACM. doi:10.1145/1135777.1135811.
 - 3 Alexandra Bonnici, Stefania Cristina, and Kenneth P. Camilleri. Preparation of music scores to enable hands-free page turning based on eye-gaze tracking. In Kenneth Camilleri and Alexandra Bonnici, editors, *Proceedings of the 2017 ACM Symposium on Document Engineering*, pages 201–210, New York, NY, 2017. ACM. doi:10.1145/3103010.3103012.
 - 4 Frederick J. Brigham, Evangelia Zaimi, Juanita Jo Matkins, Jennifer Shields, Jackie McDonnough, and Jennifer J. Jakubecy. The eyes may have it:reconsidering eye-movement research in human cognition. In Thomas E. Scruggs and Margo A. Mastropieri, editors, *Technological applications*, volume 15 of *Advances in Learning and Behavioral Disabilities*, pages 39–59. Emerald Group Publishing Limited, Bingley, 2009. doi:10.1016/S0735-004X(01)80005-7.
 - 5 Georg Buscher, Andreas Dengel, Ralf Biedert, and Ludger v. Elst. Attentive documents. *ACM Transactions on Interactive Intelligent Systems*, 1(2):1–30, 2012. doi:10.1145/2070719.2070722.
 - 6 Georg Buscher, Andreas Dengel, and Ludger van Elst. Eye movements as implicit relevance feedback. In Mary Czerwinski, editor, *CHI '08 Extended Abstracts on Human Factors in Computing Systems*, ACM Digital Library, page 2991, New York, NY, 2008. ACM. doi:10.1145/1358628.1358796.
 - 7 Mon Chu Chen, John R. Anderson, and Myeong Ho Sohn. What can a mouse cursor tell us more?: Correlation of eye/mouse movements on web browsing. In *CHI '01 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '01, pages 281–282, New York, NY, USA, 2001. ACM. URL: <http://doi.acm.org/10.1145/634067.634234>, doi:10.1145/634067.634234.
 - 8 Shiwei Cheng, Jing Fan, and Anind K. Dey. Smooth gaze: a framework for recovering tasks across devices using eye tracking. *Personal and Ubiquitous Computing*, 22(3):489–501, 2018. doi:10.1007/s00779-018-1115-8.
 - 9 Shiwei Cheng, Xiaojian Liu, Pengyi Yan, Jianbo Zhou, and Shouqian Sun. Adaptive user interface of product recommendation based on eye-tracking. In Elisabeth André and Joyce Y. Chai, editors, *IUI 2010 workshop*, pages 94–101, New York, NY, 2010. ACM. doi:10.1145/2002333.2002348.
 - 10 Mark Claypool, Phong Le, Makoto Wased, and David Brown. Implicit interest indicators. In James C. Lester, editor, *2000 International conference on intelligent user interfaces*, pages 33–40, New York, New York, USA, 2001. ACM Press. doi:10.1145/359784.359836.

- 11 Andrew d. Wilson and Shane Williams. Autopager. In Stephen N. Spencer, Bonita Sharif, and Krzysztof Krejtz, editors, *Proceedings, ETRA 2018*, pages 1–5, New York, New York, 2018. The Association for Computing Machinery, Inc. doi:10.1145/3204493.3204556.
- 12 Tilman Dingler, Rufat Rzayev, Valentin Schwind, and Niels Henze. Rsvp on the go. In Michael Beigl, editor, *Proceedings of the 2016 ACM International Symposium on Wearable Computers*, pages 116–119, New York, NY, 2016. ACM. doi:10.1145/2971763.2971794.
- 13 Heiko Drewes, Alexander de Luca, and Albrecht Schmidt. Eye-gaze interaction for mobile phones. In Peter H. J. Chong, editor, *Mobility conference 2007*, page 364, Singapore, 2007. Research Publishing. doi:10.1145/1378063.1378122.
- 14 Andrew T. Duchowski, Nathan Cournia, and Hunter Murphy. Gaze-contingent displays: a review. *Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society*, 7(6):621–634, 2004. doi:10.1089/cpb.2004.7.621.
- 15 David Fono and Roel Vertegaal. Eyewindows: Evaluation of eye-controlled zooming windows for focus selection. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, pages 151–160, New York, NY, USA, 2005. ACM. URL: <http://doi.acm.org/emedien.ub.uni-muenchen.de/10.1145/1054972.1054994>, doi:10.1145/1054972.1054994.
- 16 Daniela Giordano, Isaak Kavasidis, Carmelo Pino, and Concetto Spampinato. Content based recommender system by using eye gaze data. In Carlos H. Morimoto, editor, *Proceedings of the Symposium on Eye Tracking Research and Applications*, ACM Digital Library, page 369, New York, NY, 2012. ACM. doi:10.1145/2168556.2168639.
- 17 Brian Guenter, Mark Finch, Steven Drucker, Desney Tan, and John Snyder. Foveated 3d graphics. *ACM Trans. Graph.*, 31(6):164:1–164:10, November 2012. URL: <http://doi.acm.org/10.1145/2366145.2366183>, doi:10.1145/2366145.2366183.
- 18 Christian Gütl, Maja Pivec, Christian Trummer, Victor Manuel García-Barrios, Felix Mödritscher, and Umgeher, Juergen Pripfl and Martin. Adele (adaptive e-learning with eye-tracking): Theoretical background, system architecture and application scenarios. *European Journal of Open, Distance and E-learning*, 8(2), 2005. URL: http://www.eurodl.org/materials/contrib/2005/Christian_Gutl.pdf.
- 19 Seyed Navid Haji Mirza and Ebroul Izquierdo. Finding the user’s interest level from their eyes. In David Vallet, editor, *SAPMIA '10*, page 25, New York, N.Y., 2010. Association for Computing Machinery. doi:10.1145/1878061.1878070.
- 20 David R. Hardoon and Kitsuchart Pasupa. Image ranking with implicit feedback from eye movements. In Stephen N. Spencer, editor, *ETRA 2010*,

7:18 Implizite Interaktion mittels Eye-Tracking

- page 291, New York, NY, 2010. ACM SIGGRAPH. doi:10.1145/1743666.1743734.
- 21 Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost van de Weijer. *Eye Tracking - a comprehensive guide to methods and measures*. Oxford University Press, Oxford, UK, September 2011. URL: https://www.researchgate.net/publication/254913339_Eye_Tracking_A_Comprehensive_Guide_To_Methods_And_Measures.
 - 22 Jaemin Jo, Bohyoung Kim, and Jinwook Seo. Eyebookmark. In Bo Begole, Jinwoo Kim, Kori Inkpen, and Woontack Woo, editors, *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 2963–2966, New York, NY, 2015. ACM. doi:10.1145/2702123.2702340.
 - 23 Jakob Karolus, Paweł W. Wozniak, Lewis L. Chuang, and Albrecht Schmidt. Robust gaze features for enabling language proficiency awareness. In Gloria Mark, Susan Fussell, Cliff Lampe, m.c schraefel, Juan Pablo Hourcade, Caroline Appert, and Daniel Wigdor, editors, *Explore, innovate, inspire*, pages 2998–3010, New York, NY, 2017. Association for Computing Machinery Inc. (ACM). doi:10.1145/3025453.3025601.
 - 24 Dagmar Kern, Paul Marshall, and Albrecht Schmidt. Gazemarks. In Elizabeth D. Mynatt, Scott E. Hudson, and Geraldine Fitzpatrick, editors, *CHI Conference*, page 2093, New York, N.Y., 2010. Association for Computing Machinery. doi:10.1145/1753326.1753646.
 - 25 Arto Klami, Craig Saunders, Teófilo E. de Campos, and Samuel Kaski. Can relevance of images be inferred from eye movements? In Michael S. Lew, editor, *Proceedings of the 1st ACM International Conference on Multimedia Information Retrieval*, page 134, New York, 2008. Association of Computing Machinery. doi:10.1145/1460096.1460120.
 - 26 Kai Kunze, Hitoshi Kawaichi, Kazuyo Yoshimura, and Koichi Kise. Towards inferring language expertise using eye tracking. In Wendy Mackay, Stephen Brewster, and Susanne Bødker, editors, *CHI 2013*, page 217, New York, 2013. ACM. doi:10.1145/2468356.2468396.
 - 27 Meng-Lung Lai, Meng-Jung Tsai, Fang-Ying Yang, Chung-Yuan Hsu, Tzu-Chien Liu, Silvia Wen-Yu Lee, Min-Hsien Lee, Guo-Li Chiou, Jyh-Chong Liang, and Chin-Chung Tsai. A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational Research Review*, 10:90–115, 2013. doi:10.1016/j.edurev.2013.10.001.
 - 28 Michael F. Land and Benjamin W. Tatler. *Looking and Acting: Vision and eye movements in natural behaviour*. Oxford University Press, Oxford, UK, July 2009. URL: <http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780198570943.001.0001/acprof-9780198570943>, doi:10.1093/acprof:oso/9780198570943.001.0001.

- 29 Christian Lander, Marco Speicher, Denise Paradowski, Norine Coenen, Sebastian Biewer, and Antonio Krüger. Collaborative newspaper. In Sven Gehring, Antonio Krüger, Florian Alt, Nick Taylor, and Stefan Schneegass, editors, *Pervasive Displays*, pages 163–169, New York, 2015. ACM. doi:10.1145/2757710.2757734.
- 30 Dong-Gun Lee, Kyeong-Ho Lee, and Soo-Young Lee. Implicit shopping intention recognition with eye tracking data and response time. In *Proceedings of the 3rd International Conference on Human-Agent Interaction*, HAI '15, pages 295–298, New York, NY, USA, 2015. ACM. URL: <http://doi.acm.org.emedien.ub.uni-muenchen.de/10.1145/2814940.2815001>, [TitelanhanddieserDOIinCitavi-Projektübernehmen], doi:10.1145/2814940.2815001.
- 31 Yixuan Li, Pingmei Xu, Dmitry Lagun, and Vidhya Navalpakkam. Towards measuring and inferring user interest from gaze. In Rick Barrett, editor, *Proceedings of the 26th International Conference on World Wide Web Companion*, pages 525–533, Republic and Canton of Geneva, 2017. International World Wide Web Conferences Steering Committee. doi:10.1145/3041021.3054182.
- 32 Jaime Madell and Sylvie Hébert. Eye movements and music reading: Where do we look next. *An Interdisciplinary Journal*, 26(2):157–170, 2008. URL: <https://doi.org/10.1525/mp.2008.26.2.157>, doi:10.1525/mp.2008.26.2.157.
- 33 Christian Gregor Mall. Explizite Mensch-Computer-Interaktion mittels Eye-Tracking. In Fiona Draxler, Matthias Hoppe, Jakob Karolus, Thomas Kosch, Pascal Knierim, and Albrecht Schmidt, editors, *2nd Seminar on Ubiquitous Interaction (UBIACTION 2018-2)*, page 3:1–3:20, Munich, Germany, June 2018.
- 34 Emiliano Miluzzo, Tianyu Wang, and Andrew T. Campbell. Eyephone. In Landon Cox, editor, *Proceedings of the second ACM SIGCOMM workshop on Networking, systems, and applications on mobile handhelds*, ACM Digital Library, page 15, New York, NY, 2010. ACM. doi:10.1145/1851322.1851328.
- 35 Diego Navarro and Veronica Sundstedt. Simplifying game mechanics. In Diego Gutierrez and Hui Huang, editors, *SIGGRAPH Asia 2017 Technical Briefs*, pages 1–4, New York, NY, USA, 2017. ACM. doi:10.1145/3145749.3149446.
- 36 Ayano Okoso, Kai Kunze, and Koichi Kise. Implicit gaze based annotations to support second language learning. In A. J. Brush, Adrian Friday, Julie Kientz, James Scott, and Junehwa Song, editors, *UbiComp'14 adjunct*, pages 143–146, New York, New York, 2014. Association for Computing Machinery. doi:10.1145/2638728.2638783.

7:20 Implizite Interaktion mittels Eye-Tracking

- 37 Reinhard Oppermann. From user-adaptive to context-adaptive information systems (von benutzeradaptiven zu kontextadaptiven informationssystemen). *i-com*, 4(3/2005):45, 2005. doi:10.1524/icom.2005.4.3.4.
- 38 Oskar Palinko, Francesco Rea, Giulio Sandini, and Alessandra Sciutti. Eye tracking for human robot interaction. In Pernilla Qvarfordt and Dan Witzner Hansen, editors, *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*, pages 327–328, New York, NY, 2016. ACM. doi:10.1145/2857491.2888590.
- 39 Denis Pallez, Marcel Cremene, Thierry Baccino, and Ovidiu Sabou. Analyzing human gaze path during an interactive optimization task. In Elisabeth André and Joyce Y. Chai, editors, *IUI 2010 workshop*, pages 12–19, New York, NY, 2010. ACM. doi:10.1145/2002333.2002336.
- 40 Kai Puolamäki, Jarkko Salojärvi, Eerika Savia, Jaana Simola, and Samuel Kaski. Combining eye movements and collaborative filtering for proactive information retrieval. In Gary Marchionini, editor, *SIGIR 2005*, page 146, New York, 2005. Association for Computing Machinery. doi:10.1145/1076034.1076062.
- 41 Felix Putze, Johannes Popp, Jutta Hild, Jürgen Beyerer, and Tanja Schultz. Intervention-free selection using eeg and eye tracking. In Yukiko Nakano, editor, *ICMI'16*, pages 153–160, New York, NY, 2016. The Association for Computing Machinery, Inc. doi:10.1145/2993148.2993199.
- 42 Pernilla Qvarfordt and Shumin Zhai. Conversing with the user based on eye-gaze patterns. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, pages 221–230, New York, NY, USA, 2005. ACM. URL: <http://doi.acm.org/10.1145/1054972.1055004>, doi:10.1145/1054972.1055004.
- 43 Keith Rayner and Alexander Pollatsek. Eye Movements, the Eye-Hand Span, and the Perceptual Span During Sight-Reading of Music. *Current Directions in Psychological Science*, 6(2):49–53, 1997. URL: <https://doi.org/10.1111/1467-8721.ep11512647>, doi:10.1111/1467-8721.ep11512647.
- 44 Ozge Samanci, Blacki Li Rudi Migliozi, and Daniel Sabio. Plink blink: Collaborative music production via blinking eyes. In *Proceedings of the 11th Conference on Advances in Computer Entertainment Technology*, ACE '14, pages 61:1–61:4, New York, NY, USA, 2014. ACM. URL: <http://doi.acm.org.emedien.uni-muenchen.de/10.1145/2663806.2663809>, doi:10.1145/2663806.2663809.
- 45 Anthony Santella, Maneesh Agrawala, Doug DeCarlo, David Salesin, and Michael Cohen. Gaze-based interaction for semi-automatic photo cropping. In Rebecca Grinter, editor, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, page 771, New York, NY, 2006. ACM. doi:10.1145/1124772.1124886.

- 46 Albrecht Schmidt. Implicit human computer interaction through context. *Personal and Ubiquitous Computing*, 4(2):191–199, 2000. doi:10.1007/BF01324126.
- 47 Barış Serim and Giulio Jacucci. Explicating "implicit interaction". In Stephen Brewster, Geraldine Fitzpatrick, Anna Cox, and Vassilis Kostakos, editors, *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, pages 1–16, New York, New York, USA, 2019. ACM Press. doi:10.1145/3290605.3300647.
- 48 Jeffrey S. Shell, Roel Vertegaal, and Alexander W. Skaburskis. Eyepliances. In Gilbert Cockton and Panu Korhonen, editors, *CHI '03 extended abstracts on Human factors in computing systems - CHI '03*, page 770, New York, New York, USA, 2003. ACM Press. doi:10.1145/765891.765981.
- 49 Greg J. Siegle, Naho Ichikawa, and Stuart Steinhauer. Blink before and after you think: blinks occur prior to and following cognitive load indexed by pupillary responses. *Psychophysiology*, 45(5):679–687, 2008. doi:10.1111/j.1469-8986.2008.00681.x.
- 50 Takumi Toyama, Daniel Sonntag, Andreas Dengel, Takahiro Matsuda, Masakazu Iwamura, and Koichi Kise. A mixed reality head-mounted text translation system using eye gaze input. In Tsvi Kuflik, Oliviero Stock, Joyce Chai, and Antonio Krüger, editors, *IUI'14*, pages 329–334, [New York, NY], 2014. ACM. doi:10.1145/2557500.2557528.
- 51 Roel Vertegaal. Attentive user interfaces. *Communications of the ACM*, 46, 2003. doi:10.1145/636772.636794.
- 52 Tina Walber, Chantal Neuhaus, and Ansgar Scherp. Tagging-by-search. In Tsvi Kuflik, Oliviero Stock, Joyce Chai, and Antonio Krüger, editors, *IUI'14*, pages 257–266, [New York, NY], 2014. ACM. doi:10.1145/2557500.2557517.
- 53 Kang Wang, Shen Wang, and Qiang Ji. Deep eye fixation map learning for calibration-free eye gaze tracking. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*, ETRA '16, pages 47–55, New York, NY, USA, 2016. ACM. URL: <http://doi.acm.org/10.1145/2857491.2857515>, doi:10.1145/2857491.2857515.
- 54 Pingmei Xu, Krista A. Ehinger, Yinda Zhang, Adam Finkelstein, Sanjeev R. Kulkarni, and Jianxiong Xiao. TurkerGaze: Crowdsourcing saliency with webcam based eye tracking. *CoRR*, abs/1504.06755, 2015. URL: <http://arxiv.org/abs/1504.06755>, arXiv:1504.06755.
- 55 Songhua Xu, Hao Jiang, and Francis C.M. Lau. Personalized online document, image and video recommendation via commodity eye-tracking. In Pearl Pu, editor, *Proceedings of the 2008 ACM conference on Recommender systems*, ACM Digital Library, page 83, New York, NY, 2008. ACM. doi:10.1145/1454008.1454023.
- 56 Shumin Zhai, Carlos Morimoto, and Steven Ihde. Manual and gaze input cascaded (magic) pointing. In *Proceedings of the SIGCHI Conference on Hu-*

7:22 Implizite Interaktion mittels Eye-Tracking

man Factors in Computing Systems, CHI '99, pages 246–253, New York, NY, USA, 1999. ACM. URL: <http://doi.acm.org/10.1145/302979.303053>, doi:10.1145/302979.303053.

Augenbasierte Activity Recognition

Margaretha Lucha

Ludwig-Maximilians-Universität München, München, Deutschland
M.Lucha@campus.lmu.de

Zusammenfassung

Diese Arbeit fasst die bisherige Forschung zu augenbasierter Activity Recognition zusammen, die die Schaffung alltagstauglicher kontextbewusster Systeme zum Ziel hat. Es wird dargestellt, warum Forscher in der Analyse der Augenbewegungen einer Person Potential für die Activity Recognition sehen, und welche Arten von Augenbewegungen hierfür verwendet werden können. Zudem wird den Fragen auf den Grund gegangen, welche Eyetracking-Methoden für den Alltag geeignet sind und welche Tätigkeiten bzw. kognitiven Prozesse bisher erkannt werden konnten. Abschließend wird in Aussicht gestellt, was noch geschehen muss, bis solche Systeme in unseren Alltag einziehen können.

2012 ACM Computing Classification Human computer interaction (HCI) → Ubiquitous and mobile computing

Keywords and phrases Activity Recognition; Augenbasierte Activity Recognition; Eyetracking; Elektrookulografie (EOG); Context Awareness.

1 Motivation

Wäre es nicht praktisch, einen Computerassistenten zu haben, der erkennt, dass man gerade mit einer Person spricht, deren Namen man vergessen hat und der einen daraufhin an den Namen erinnert? Ein solches imaginäres Zukunftsszenario schildern Bulling und Roggen in einer ihrer Arbeiten zur augenbasierten Activity Recognition [3].

Systeme zu schaffen, die dem Nutzer Informationen und Hilfestellungen automatisch genau dann liefern, wenn er sie benötigt, ihn also abhängig von seinen Bedürfnissen bei seiner aktuellen Tätigkeit unterstützen können, das ist die Vision, um die es bei der Forschung zu Activity Recognition geht [6, 1, 2].

Eine Anwendung, die dies schon teilweise umsetzt, ist beispielsweise Google Now¹. Die Anwendung wird mit dem Slogan „Die richtigen Informationen zur richtigen Zeit“¹ beworben und fasst Informationen, die der Nutzer aktuell

¹ <https://www.google.com/intl/de/landing/now/>



© Margaretha Lucha;

licensed under Creative Commons License CC-BY

Cite as: Margaretha Lucha. Augenbasierte Activity Recognition. In *3rd Seminar on Ubiquitous Interaction (UBIACTION 2019)*. Editors: Pascal Knierim, Jakob Karolus, Fiona Draxler, Ville Mäkelä, Tonja Machulla, Florian Lang, Thomas Kosch, Albrecht Schmidt. July 16, 2019. Munich, Germany. pp. 8:1–8:22.

8:2 Augenbasierte Activity Recognition

benötigt (z. B. das aktuelle Wetter oder die Verkehrslage), unter Berücksichtigung von Informationen wie der Tageszeit, seinem Aufenthaltsort oder seiner Termine an einem Ort zusammen [25].

Activity Recognition soll jedoch auch dabei helfen, Tätigkeiten und Verhaltensweisen des Nutzers zu dokumentieren (Lifelogging). Dies kann einerseits als Gedächtnisstütze fungieren und Erinnerungen an Erlebnisse erleichtern. Andererseits können durch eine nachträgliche Analyse des Verhaltens Fehlverhalten, schlechte Angewohnheiten etc. ermittelt werden. So kann der Nutzer beispielsweise dabei unterstützt werden einen gesünderen Lebensstil zu etablieren. [9, 1, 21, 15]

Das im iPhone integrierte Feature Bildschirmzeit² ermöglicht den Nutzern einen Überblick, wie viel Zeit sie am Smartphone verbringen und welche verschiedenen Anwendungen sie nutzen. Zudem lassen sich Zeitlimits für die Appnutzung setzen. Doch ebenso wie Google Now kann dieses Feature nur bedingt die aktuelle Tätigkeit des Nutzers bestimmen, denn es wird nicht ermittelt, was genau mit der App gemacht wurde, oder ob der Nutzer zu einem Zeitpunkt Hilfestellung benötigt hätte [29]. Auch Google Now ermittelt nur ungefähre Tätigkeiten und bietet eher allgemeine Hilfestellungen.¹

Um die oben genannten Ziele zu verwirklichen, benötigt man daher Systeme, die in Echtzeit präzise zahlreiche Tätigkeiten des Nutzers erkennen können. Die vorliegende Arbeit erläutert das Potential von Augenbewegungen für das maschinelle Erkennen physischer und psychischer Vorgänge und gibt einen Überblick über die bisherige Forschung im Bereich der augenbasierten Activity Recognition auch unter Berücksichtigung hierfür geeigneter Eyetracking-Methoden. Zudem wird diskutiert, was noch passieren muss, damit diese Systeme in unser alltägliches Leben integriert werden können.

2 Activity Recognition

Unter Activity Recognition versteht man den Versuch, maschinell - u. a. anhand von Sensoren und Machine-Learning-Algorithmen - zu erkennen, was eine Person zu einem bestimmten Zeitpunkt macht. Dies umfasst einerseits das Erkennen physischer Gesten oder ganzer Handlungen (z. B. das Trinken aus einer Tasse), andererseits aber auch das Nachvollziehen kognitiver Prozesse (z. B. ob eine Person konzentriert ist oder sich an etwas erinnert [3]). Ende der 90er Jahre begann man für Messungen Sensoren zu verwenden, die am Körper getragen werden konnten, um die Forschung hingehend alltäglicher Situationen zu erweitern. [2]

² <https://support.apple.com/de-de/HT208982>

Für das Nachvollziehen von Tätigkeiten gibt es verschiedene Ansätze. Beispielsweise wurden Beschleunigungssensoren verwendet, um zu bestimmen, wie viel sich eine Person an einem Tag bewegte und wie diese Bewegungen aussahen (z. B. Radfahren, Treppensteigen, Sitzen etc.) [14]. Diese Art der Activity Recognition hat in Form von Fitnessapps, wie z. B. der vorinstallierten Health App³ auf dem iPhone oder am Handgelenk getragenen Fitnesstrackern⁴, bereits Einzug in unseren Alltag gefunden. Ein anderer Ansatz ist, mittels Elektromyografie (EMG) die Muskelaktivität zu bestimmen. So lässt sich beispielsweise feststellen, ob eine Person kaut, also isst [13]. Kombiniert man einen Beschleunigungssensor und ein Gyroskop, so lassen sich Handgesten erkennen, durch die Tätigkeiten wie das Umblättern einer Buchseite oder das Schneiden mit einem Messer identifiziert werden können [2]. Mit Hilfe eines Mikrofons und den Geräuschen, die für eine Tätigkeit charakteristisch sind, lässt sich darüberhinaus bestimmen, ob jemand beispielsweise gerade Auto fährt oder sich mit einer anderen Person unterhält [24]. In der vorliegenden Arbeit liegt der Fokus hingegen darauf, wie die Augenbewegungen zur Activity Recognition genutzt werden können. [2]

Dieses Forschungsfeld entstand, da man herausfand, dass Menschen ihren Blick, also ihre visuelle Aufmerksamkeit, auf aktuell (z. B. für eine Aufgabe) relevante Stellen richten und sich die Augenbewegungen bei der Bearbeitung verschiedener Aufgaben unterscheiden [37, S. 174, S. 192ff.][12, S. 12f., S. 219f.].

Augenbasierte Activity Recognition Systeme könnten dem Nutzer in vielen Anwendungsbereichen helfen. Durch das Erfassen des Leseverhaltens und der Konzentrationsstärke könnte ein solches System beispielsweise bestimmen, ob ein Leser gerade unterbrochen werden kann oder nicht, oder gegebenenfalls Hilfestellungen bieten, wie das automatische Vergrößern der Schrift bei Sehproblemen oder das Erklären eines Wortes, das das System als dem Leser unbekannt identifiziert. [6, 5, 29]

Auch sehen Forscher so beispielsweise die Möglichkeit, zu erfassen, wie sozial aktiv Personen sind, eine Information, die unter anderem besonders für Angehörige oder medizinische Betreuer älterer Menschen oder Autisten von Interesse sein könnte [9, 36].

Bezüglich des maschinellen Erkennens physischer und psychischer Prozesse des Menschen bestehen jedoch gewisse Herausforderungen.

Wie eine Tätigkeit ausgeführt wird, hängt von einer Vielzahl an Faktoren ab (z. B. Müdigkeit) und kann stark zwischen verschiedenen Personen variieren. So ist beispielsweise die Gangart in vielen Fällen individuell unterschiedlich. Oft ist es daher nötig, die Machine-Learning-Algorithmen mit personenspezifischen

³ <https://www.apple.com/de/ios/health/>

⁴ <https://consumer.huawei.com/de/wearables/band3-pro/>

Daten zu trainieren. Beim Herausarbeiten der Tätigkeiten umfasst die NULL-Klasse alle Aktivitäten, die für eine bestimmte Anwendung nicht von Interesse sind und nicht genauer klassifiziert werden müssen. Schwierigkeiten bei der Zuordnung ergeben sich jedoch, wenn uninteressante Aktivitäten ähnliche Muster aufweisen, wie die Aktivitäten von Interesse. Auch das Sammeln von Referenzdaten (Ground Truth) für die Klassifizierung ist vor allem bei Messungen im Alltag (vs. Labor) ein aufwändiger Vorgang, da die Aktivitäten entweder direkt, wenn sie passieren, dokumentiert werden müssen, manchmal auch von der Testperson selbst, oder nachträglich alles durchgegangen und beschriftet werden muss. Zudem kann die Korrektheit der Klassifizierung verringert werden, wenn Tätigkeiten in Echtzeit erkannt werden müssen. [2]

3 Augenbewegungen und ihr Potential für die Activity Recognition

Für diese Arbeit ist besonders das Verständnis von Sakkaden, Fixationen und Blinzeln wichtig - den drei Augenbewegungsarten, die bisher am häufigsten für die augenbasierte Activity Recognition verwendet wurden [4, 30].

Da nur mit einem kleinen Bereich des Auges (der Fovea) scharf gesehen werden kann, bewegen sich unsere Augen ständig. Diese schnellen Bewegungen der Augen von einem Punkt zum anderen nennt man Sakkaden. Je nach Abstand zum Ziel benötigen sie ca. 10 bis 100 ms. Sie treten sowohl gewollt, als auch reflexartig auf. Ein Punkt von Interesse wird fixiert, wenn das Auge mit Hilfe von Stabilisierungsmechanismen auf diesen Punkt gerichtet bleibt. Dies geschieht zwischen zwei Sakkaden. Wenn die Augen für den Bruchteil einer Sekunde geschlossen werden, um einen dünnen Film Tränenflüssigkeit auf der Vorderseite des Auges (Cornea) zu verteilen, spricht man von Blinzeln. [37, S. 123][12, S. 139][36, 4]

Auf Grund bisheriger Erkenntnisse und weil unsere Augen in fast all unsere Handlungen involviert sind, sehen Forscher in den Augenbewegungen ein großes Potential für die Activity Recognition [7]. So werden unsere Augenbewegungen beispielsweise von unseren momentanen Zielen bzw. Aufgaben gelenkt [37, S. 174, S. 192ff.], und Eigenschaften wie die Fixationsdauer unterscheiden sich bei unterschiedlichen Aufgaben [26, 29]. Auch momentane kognitive Prozesse, z. B. während des Lesens oder der (visuellen) Suche, haben Einfluss auf unsere Augenbewegungen. Es hat sich gezeigt, dass insbesondere Sakkaden viele Informationen über derartige interne Prozesse liefern. So unterscheidet sich beim Lesen und bei der visuellen Suche beispielsweise die Länge der Sakkaden. [26] Längere Fixationen lassen auf höhere kognitive Aktivität schließen [12, S. 168], ebenso wie bestimmte Blinzelmuster [28]. Betrachtet man das Blinzelnverhalten einer Person, kann man außerdem Müdigkeit erkennen [10].

Zu den kognitiven Prozessen gehört auch die Aufmerksamkeit. In der Regel, besonders wenn komplexere Informationsverarbeitung nötig ist, liegt die Aufmerksamkeit einer Person dort, worauf die Augen gerade fokussieren [26, 5]. In einigen Studien wurde auch die Blinzelrate als Maß für Aufmerksamkeit genutzt [7].

Dass sich die Augenbewegungsmuster von Personen mit geistigen Behinderungen, wie Autismus oder Demenz, von den Augenbewegungsmustern anderer Personen unterscheiden, macht ebenfalls den Zusammenhang zwischen Augenbewegungen und kognitiven Prozessen deutlich [36].

Es konnte auch gezeigt werden, dass verschiedene Tätigkeiten, z. B. Lesen, Gehen oder Sprechen, unterschiedliche mittlere Fixationsdauern und Sakkadenamplituden bedeuten [11].

4 Eyetracking im Alltag

Um die Augenbewegungen zur Activity Recognition nutzen zu können, benötigt man robuste Geräte, die auch im Alltag die Augenbewegungen aufzeichnen können [5]. Durch den starken Fokus der Forschung auf die Chancen, die Eyetracking für viele Anwendungen bietet, in Bereichen, wie der Mensch-Computer Interaktion oder der Psychologie [12, S. 207, S. 241, S. 275], wurden Eyetracker in den letzten Jahren kontinuierlich weiterentwickelt und verbessert [36, 4, 19, 29]. Bezüglich der Alltagstauglichkeit von Eyetrackern konnten zwar schon große Fortschritte erzielt werden, jedoch ist ihre Entwicklung noch nicht so weit, dass sie so nahtlos in den Alltag integriert werden könnten wie beispielsweise Smartphones [27, 34]. Eyetracker, die für den Alltag geeignet sein sollen, müssen eine Vielzahl an Anforderungen erfüllen und stellen Entwickler daher vor gewisse Herausforderungen.

4.1 Herausforderungen

Das ultimative Ziel der Entwickler ist es, einen Eyetracker zu bauen, der so unauffällig ist, dass der Nutzer diesen nicht mehr bewusst wahrnimmt. Ein solcher Eyetracker muss im Alltag bequem zu tragen sein, auch über einen längeren Zeitraum hinweg, darf nicht verrutschen und muss uneingeschränkte Bewegungsfreiheit erlauben. [19, 29]

Des Weiteren muss das Gerät sehr leistungsfähig sein [15] und idealerweise ganztägige Aufnahmen [4] und Echtzeitverarbeitung und -analyse der Daten ermöglichen [35, 36]. Auch Unempfindlichkeit gegenüber wechselnden Lichtverhältnissen ist ein bedeutender Faktor [1, 19, 34]. Ebenso muss die Nutzerfreundlichkeit in Bezug auf eine einfache und schnelle Bedienung [34] und auf die Funktionsfähigkeit für Brillen- bzw. Kontaktlinsenträger gewährleistet werden [12, S. 88][19].

4.2 Methoden

Im Folgenden werden die beiden mobilen Eyetracking-Methoden vorgestellt, die am häufigsten zur Bestimmung der Augenbewegungen für die Activity Recognition verwendet wurden, auch unter Berücksichtigung wie diese den oben genannten Herausforderungen begegnen.

4.2.1 Videobasiertes Eyetracking

Dies ist die bekannteste Variante des Eyetrackings [36][12, S. 51]. Videobasierte Eyetracker bestehen aus einem Infrarotlicht und einer Kamera, die auf das Auge gerichtet sind. Anhand der aufgenommenen Bilder werden die Position des Zentrums der Pupille und die Position der Reflexion des Infrarotlichts auf der Cornea (Hornhaut auf der Vorderseite des Auges [37, S. 5f.]) ermittelt. Bewegt sich das Auge, verändert sich der Abstand zwischen Pupille und Reflexion, da sich die Position der Pupille verändert, die Position der Reflexion aber nicht. So lassen sich Augenbewegungen bestimmen. In manchen Fällen ist zusätzlich auch noch eine Kamera, die das Sichtfeld der Person filmt, am Gerät befestigt. [12, S. 54-58][23, 36, 19]

Neben Blickpunkten können auch u. a. Blinzelbewegungen, Fixationen und Sakkaden sehr präzise erkannt werden [12, S. 54-58][36, 1].

Bezüglich Tragbarkeit und Unaufdringlichkeit besteht jedoch das Problem, dass Kamera und Lichtquelle das Sichtfeld des Nutzers einschränken und die Geräte nach einer gewissen Zeit unbequem werden [35, 36, 34]. Da sie in der Regel viel Rechenleistung benötigen, sind sie nicht geeignet für lange Aufnahmen [36, 27]. Manchmal wird das Gerät deshalb beispielsweise per Kabel an einen Laptop angeschlossen. Dies schränkt jedoch die Beweglichkeit des Nutzers ein und erhöht das Tragegewicht [34]. Wenn die Sonne scheint, ist viel Infrarotlicht vorhanden, was beim videobasierten Eyetracking zu Fehlern führen kann [19]. Ebenso erschweren Kontaktlinsen und Brillen, die Reflexion zu erkennen [12, S. 88].

Aktuelle Bemühungen der Entwickler haben oft zum Ziel, den Eyetracker wie eine gewöhnliche Brille aussehen zu lassen [34] ohne das Sichtfeld einzuschränken.

In aktuellen Eyetracking-Studien wird häufig der Eyetracker Pupil⁵ von Pupil Labs verwendet, s. z. B. [30, 27]. Dessen Entwickler bemühten sich besonders um ein leichtes Headset mit kleinen Kameras, um die Tragbarkeit zu maximieren und die Einschränkung des Blickfeldes zu minimieren. Bei Sonnenlicht gibt es jedoch weiterhin Probleme [19] und die Nutzer sind sich des Gerätes weiterhin bewusst [27]. Anfang 2019 kündigte Pupil Labs eine

⁵ <https://pupil-labs.com/pupil/>

neue Version des Eyetrackers an, die optisch kaum von einer gewöhnlichen Brille zu unterscheiden ist [22].

Auf Grund der oben genannten Probleme, die zur Zeit der ersten Studien im Bereich der augenbasierten Activity Recognition noch präsenter waren als heute, wurde videobasiertes Eyetracking zunächst kaum verwendet [6, 4]. Erst in späteren Studien, mit der Weiterentwicklung videobasierter Eyetracker, wurde auch diese Messmethode für die Activity Recognition verwendet [30, 21]. Zunächst, aber auch später noch, nutzten die Forscher in vielen Studien die Elektrookulografie [16, 9, 4].

4.2.2 Elektrookulografie

Bei der Elektrookulografie (EOG) werden die Dipoleigenschaften des Auges genutzt. Der positive Pol befindet sich hierbei an der Cornea und der negative Pol an der Retina. Werden um das Auge herum Elektroden befestigt (auf jeder Seite des Auges und eine Referenzelektrode auf der Stirn), können Spannungsänderungen gemessen werden. Diese entstehen, wenn sich die Retina einer Elektrode und die Cornea der entgegengesetzten Elektrode annähert, also durch Augenbewegungen. Durch die Verteilung der Elektroden um das Auge herum können sowohl horizontale als auch vertikale Spannungsänderungen aufgezeichnet werden. [36, 5, 34][12, S. 52]

Die Elektrookulografie wurde in einer Vielzahl an Studien zur augenbasierten Activity Recognition als Messmethode verwendet auf Grund ihrer besseren Alltagstauglichkeit (besonders in frühen Studien) durch das im Vergleich zu videobasierten Eyetrackern geringere Gewicht und die größere Bewegungsfreiheit des Nutzers [6, 4, 5, 36, 34]. Da sich die Sensoren direkt auf der Haut befinden [32], wird außerdem das Sichtfeld nicht eingeschränkt [36]. Wurden sie gefragt, gaben die meisten Teilnehmer an, keine Einschränkungen empfunden zu haben. [27, 5, 4]

Zudem ist der Rechenaufwand bei dieser Methode geringer und sie benötigt nicht so viel Strom wie die videobasierte Variante [34, 32, 36], daher können auch wesentlich längere Aufnahmen gemacht werden [36, 4, 1]. Zwar kann mittels EOG nicht bestimmt werden, wohin eine Person blickt, jedoch reicht es für die Activity Recognition, die Augenbewegungen und -bewegungsmuster relativ zu erfassen. Dies geschieht mit hoher Genauigkeit, auch in Bewegung. [12, S. 52][36, 5]. Des weiteren können die Signale auch bei Sonnenlicht, Dunkelheit und geschlossenen Augen erkannt werden [6, 15].

Bei der Messung kann es manchmal zu langsamen Änderungen im Signal kommen, die nichts mit Augenbewegungen zu tun haben. Diese Änderung wird als Baseline Drift bezeichnet und muss entfernt werden, da dieser das Signal beschädigt [6]. Bei EOG ist außerdem zu beachten, dass der Kontakt der Elektroden mit der Haut gut sein muss, da es sonst zu Problemen kommen

kann. Da es für den Nutzer möglicherweise unangenehm ist, die Elektroden auf der Gesichtshaut zu befestigen, gibt es auch für diese Messmethode den Ansatz, sie in Form einer Brille noch alltagstauglicher zu machen [4, 5]. Beispielsweise gibt es mittlerweile die EOG-Brille J!NS MEME⁶, die auch in einigen Studien zur Activity Recognition schon zum Einsatz kam [17, 16].

5 Augenbasierte Activity Recognition - Anwendungen

Im Folgenden sollen bisherige Studien aus der Forschung zur augenbasierten Activity Recognition vorgestellt und auch ihre Vorgehensweisen kurz erläutert werden. Insbesondere soll der Verlauf der Forschung im Hinblick auf die erkannten Tätigkeiten aufgezeigt werden.

5.1 Erkennen ausgewählter Tätigkeiten

Zu Beginn versuchte man vor allem ausgewählte, konkrete Tätigkeiten anhand der Augenbewegungen zu erkennen. Zunächst, aber auch in einigen späteren Studien [16], fokussierten sich die Forscher vor allem darauf, Lesen zu identifizieren. Augenbewegungen im Kontext Lesen (stilles Lesen) wurden seit 1879 umfangreich erforscht. Hierbei erkannte man, dass Augenbewegungen beim Lesen ein spezielles, regelmäßiges Muster aufweisen, das sie von Augenbewegungen bei anderen Tätigkeiten unterscheidet. Dies bezieht sich auf Eigenschaften, wie die durchschnittliche Sakkadenlänge und Fixationsdauer innerhalb einer Zeile, Regressionen (kurze Sakkaden gegen die Leserichtung auf Grund einer zu langen Sakkade o. ä.), aber auch auf den Sprung des Blicks vom Ende einer Zeile zum Anfang der nächsten Zeile. [26, 6]

Zudem ist Lesen eine Tätigkeit, die im Alltag sehr häufig und in vielen Ausführungen und Kontexten praktiziert wird und daher sehr relevant für die Activity Recognition ist [5]. Andere Aktivitäten, wie Surfen im Internet, sind hingegen schwieriger zu identifizieren, weil mit ihnen mehr verschiedene und unvorhersehbare Augenbewegungen verbunden sind [6, 7].

5.1.1 Methodik

Viele der im Folgenden aufgeführten Studien verwendeten eine featurebasierte Machine-Learning-Methode, um aus den Augenbewegungen die zugehörigen Tätigkeiten (bzw. kognitiven Prozesse, s. Abschnitt 5.2) abzuleiten [7, 5, 3, 4].

Generell wurden zunächst die in den Studien überwiegend mit EOG gemessenen Augenbewegungen [6, 3, 21] mit Hilfe von Algorithmen zur Herausarbeitung einzelner Bestandteile verarbeitet. Die meisten Studien analysierten

⁶ <https://jins-meme.com/en/products/es/>

sowohl Sakkaden, Fixationen als auch das Blinzeln, z. B. [3, 16]. Manchmal berücksichtigten die Forscher jedoch zusätzlich auch Kopfbewegungen oder konzentrierten sich nur auf einen Typ Augenbewegungen (in diesem Fall wurde eine andere Klassifizierungsmethode verwendet) [5, 17].

Da es schwer ist, die Augenbewegungen direkt zu analysieren, nutzt die featurebasierte Klassifizierungsmethode vier Featuregruppen - hergeleitet aus den drei grundlegenden Augenbewegungsarten (Sakkaden, Fixationen, Blinzeln) plus Augenbewegungsabfolgen [4, 2] - aus denen die Features ausgewählt werden, die die jeweilige Tätigkeit am besten beschreiben. Lesen lässt sich beispielsweise gut durch Features identifizieren, die das damit verbundene gleichmäßige Bewegungsmuster ausdrücken. Diese Methode funktioniert relativ gut, ist aber vom Rechenaufwand her aufwändig. [5]

Für die Überprüfung der Korrektheit der Zuordnung, aber auch für das Training des Machine-Learning-Algorithmus werden Referenzdaten (Ground Truth) benötigt. Daher muss dokumentiert werden, was die Person tatsächlich tut [4, 3]. Dies wurde in den ersten Studien überwiegend manuell gelöst. Teilweise wurde die Annotation in Echtzeit vorgenommen, entweder vom Teilnehmer selbst [3] oder von einer zweiten Person, die den Teilnehmer genau beobachtete [5], teilweise auch nachträglich anhand von Videoaufnahmen (wieder entweder vom Teilnehmer selbst [16] oder von einem Assistenten [17]).

5.1.2 Zugehörige Arbeiten

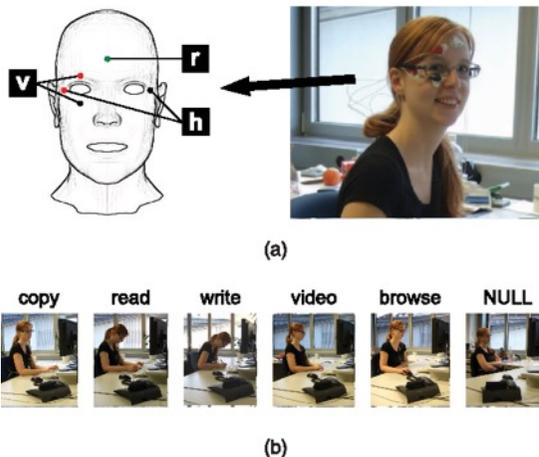
Eine der ersten Studien (2008) zur augenbasierten Activity Recognition befasste sich damit, Lesen in fünf verschiedenen Alltagssituationen zu erkennen (am Tisch sitzend, einen Flur und eine Straße entlanglaufend, stehend auf die Trambahn wartend und Trambahn fahrend). Hierbei durften die Teilnehmer (N = 8) ihre (bilderlose) Lektüre selbst wählen. Mit EOG wurden Sakkaden und Fixationen ermittelt und die hier genutzte Machine-Learning-Methode zur Klassifizierung unterschied sich etwas von der erst später entwickelten, oben beschriebenen Methode [7]. Die Erkennungsrate (True Positives + True Negatives) lag bei 80.2% mit einem Recall (True Positive Rate; True Positives/(True Positives + False Negatives)) von 71%. Teilweise wurde die Dauer einer Aktivität kürzer (underfill) oder länger (overfill) im Vergleich zu den Referenzdaten eingestuft, was möglicherweise an ungenauen Annotationen lag, die in Echtzeit durch eine zweite Person durchgeführt wurden. Beim Stehen konnte Lesen am besten erkannt werden (Erkennungsrate: 84.5%), beim Gehen am schlechtesten (Erkennungsrate: 77.1%). [6]

Eine spätere Studie von 2012 erweiterte diese Untersuchungen durch eine zusätzliche Analyse der Blinzeln- und Kopfbewegungen, denn die Autoren vermuteten, dass durch die Analyse weiterer Modalitäten die Genauigkeit der Bestimmung erhöht werden kann, da sich Augenbewegungen manchmal

8:10 Augenbasierte Activity Recognition

auch an Körperbewegungen orientieren [26]. Das untersuchte Szenario, die Teilnehmerzahl und die Annotationsweise, glich der Vorgängerstudie. Zusätzlich wurde jedoch noch gemessen, ob der Kopf gesenkt war (= Lesen) oder nicht. Dies sollte auch der Abgrenzung des Lesens von anderen Handlungen helfen. Mit der featurebasierten Methode konnte eine Genauigkeit (True Positives/(True Positives + False Positives)[2]) von 87.7% (Recall: 87.9%) erreicht werden. Die Genauigkeit konnte um 14,7% verbessert werden im Vergleich zu einer Analyse ohne Kopfbewegungen. Die Autoren merkten jedoch an, dass Lesen nicht immer zwingend mit einem gesenkten Kopf einhergehen muss, weshalb eventuell noch andere Modalitäten in Erwägung gezogen werden sollten. [5]

Bulling et al. versuchten 2009 eine Reihe typischer Büroaktivitäten (in einer Büroumgebung) zu erkennen und so [6] um vier Tätigkeiten zu erweitern. Insgesamt untersuchten sie: das Kopieren eines Textes am Computer (egal auf welche Art), das Lesen in einem Buch, das Schreiben (per Hand), das Ansehen eines Videos und das Surfen im Internet (s. Abbildung 1). Sie untersuchten nun auch Tätigkeiten, die - im Unterschied zu Lesen [26] - bezogen auf ihre zugehörigen Augenbewegungen teilweise stärker variierten und in gewisser Weise unvorhersehbarer waren. Beim Ansehen eines Videos beispielsweise sind die Augenbewegungen auch abhängig von den Inhalten des Videos. Das Nachvollziehen dieser Tätigkeiten ist zudem relevant, da sie auch in anderen Situationen auftreten können. So ähnelt das Kopieren eines Textes beispielsweise den regelmäßigen Aufmerksamkeitsverlagerungen zwischen Straße und Spiegel beim Autofahren.



■ **Abbildung 1** (a) Platzierung der Elektroden für die Messung der Augenbewegungen und (b) Versuchsablauf der Bürostudie von Bulling et al. [8, S. 748]

Während der Studie saßen die Teilnehmer ($N = 8$) an einem Arbeitsplatz mit Computer und sollten jede Aktivität fünf Minuten ausführen. Außerdem gab es auch eine Phase, in der sie Pause bzw. alles außer den untersuchten Tätigkeiten machen konnten. Diese Phase wurde zusammen mit kurzen Blicken nach oben, auf Grund von Ablenkungen o. ä., zur NULL-Klasse gezählt. Sakkaden, Fixationen und Blinzeln wurden mittels EOG gemessen und mit der featurebasierten Methode analysiert. Die Dokumentation der Referenzdaten wurde auch hier wieder manuell von einem Assistenten erledigt. Die Activity Recognition erreichte eine Erkennungsrate von durchschnittlich 76.1% (Recall: 70.5%). Das Ansehen eines Videos konnte sogar mit 88% Genauigkeit (Recall: 80%) erkannt werden. Fehler entstanden manchmal bei Aktivitäten, die einen oder mehrere der anderen Tasks beinhalteten, z. B. gehört zu Surfen im Internet u. a. auch das Lesen der Webinhalte, so wurde manchmal Lesen für Browsen gehalten oder andersherum. [7]

Auf Basis dieser Studie wurde untersucht, ob sich die Analyse der Augenbewegungen auch für kontextbewusste Hörgeräte eignet. Hierbei soll sich das Hörgerät automatisch an eine Unterhaltungssituation anpassen können oder störende Stimmen ausblenden. Denn wie die vorherige Studie zeigte, lässt sich anhand der Augenbewegungen ermitteln, ob jemand gerade einer Bürotätigkeit nachgeht oder nicht [7]. Die reine Analyse der drei grundlegenden Augenbewegungsarten erreichte beim Erkennen der Bedürfnisse eine Erkennung von 86%. Die Kombination aus Körper- und Augenbewegungen erzielte mit 91% eine noch höhere Rate. [32]

Eine Studie von 2014 nutzte eine völlig andere Herangehensweise und testete, inwieweit sich die Google Glass⁷ zur Activity Recognition eignet. Die Autoren verwendeten die kommerziell erwerbliche Brille zur Messung der Blinzelhäufigkeit und zusätzlich der Kopfbewegungen zur besseren Unterscheidung, um fünf Tasks zu identifizieren (sich mit jemandem unterhalten, lesen, ein Video ansehen, ein mathematisches Problem lösen und sägen). Zur besseren Erkennung wurde der verwendete Machine-Learning-Algorithmus mit nutzerspezifischen Daten trainiert. Zur Ermittlung der Referenzdaten wurden die tatsächlichen Binzelbewegungen nachträglich anhand eines Videos dokumentiert. Auch hier erhöhten die zusätzliche Analyse der Kopfbewegungen die Erkennungsrate von 67% auf 82%, was jedoch für praktische Anwendungen trotzdem noch zu wenig wäre, mutmaßten die Autoren. [17]

Mit Hilfe einer kommerziellen EOG-Brille wurde außerdem versucht, die Lesestudien hinsichtlich der Erfassung von Lesen in natürlicheren, unkontrollierten Situationen zu erweitern. Die Aufnahmen wurden ohne Beschränkungen bezüglich der Textquelle über die Dauer von zwei Wochen gemacht und erziel-

⁷ <https://www.google.com/glass/start/>

ten eine Erkennungsrate von 73.8%. [16]

5.2 Erkennen von kognitiven Prozessen

Zum Verständnis des gesamten Kontextes, in dem sich eine Person befindet, müssen neben den physischen auch die psychischen Vorgänge berücksichtigt werden [3, 4].

Auch hierfür scheinen die Augenbewegungen geeignet zu sein, denn wie in Abschnitt 3 erläutert, können Augenbewegungen auch Einblicke in kognitive Prozesse, z. B. bei Tätigkeiten wie Lesen oder Suchen, unter anderem in Bezug auf die Aufmerksamkeit, Motivationen oder die kognitive Belastung ermöglichen [12, S. 12][26][37, S. 192ff.].

Bulling und Roggen untersuchten deshalb, ob auf Basis der Augenbewegungen festgestellt werden kann, dass ein Bild wiedererkannt wurde. Das Abrufen von Informationen aus dem Gedächtnis ist ebenfalls ein kognitiver Prozess. Den Teilnehmern ($N = 7$) wurden Bilder aus vier Kategorien gezeigt (abstrakte Bilder, Landschaften, Gebäude und Gesichter), wovon einige Bilder mehr als einmal gezeigt wurden. Die Bilder wurden einzeln für jeweils zehn Sekunden dargeboten. Die Autoren interessierten sich hierbei vor allem für das Erkennen der Erinnerung an Gesichter auf Grund dessen Relevanz für mögliche zukünftige Anwendungen (s. Abschnitt 1). Mittels EOG wurden auch hier wieder Sakkaden, Fixationen und Blinzeln aufgenommen und mit der featurebasierten Methode analysiert. Um zu überprüfen, ob während des Experiments tatsächlich Erinnerungen stattfanden, sollten sieben andere Teilnehmer in einer zusätzlichen Studie angeben, ob sie sich an das jeweilige Bild erinnerten oder nicht. Da in ca. 97% der Fälle wiederkehrende Bilder erkannt wurden, schlussfolgerten die Autoren, dass es zwar sein könne, dass ein Bild nicht erkannt wird, dies jedoch selten der Fall sei. Die beste Erkennungsrate, ob sich ein Teilnehmer an ein Bild erinnerte oder nicht, wurde für abstrakte Bilder erzielt (84.3%, Recall: 89.3%). Bei Bildern von Gesichtern hingegen betrug der Recall 85.7%. Interessant für zukünftige Assistenzsysteme, die auch innerhalb kürzester Zeit Unterstützung leisten sollen, ist auch, dass der Recall im Falle der Gesichtsbilder bei nur dreisekündiger Darbietung des Bildes fast den gleichen Wert hatte wie bei einer zehneckündigen. Interessant ist außerdem, dass die zur Analyse genutzten Features vor allem aus Sakkaden und Bewegungsmustern abgeleitet wurden. Bezüglich der Fixationen wurde nur die Varianz der Fixationsdauer betrachtet, jedoch nicht die mittlere Anzahl der Fixationen, ein Feature, das laut [4] sogar möglicherweise als alleiniger Indikator für das Erkennen bzw. Nichterkennen eines Bildes genutzt werden könnte. [3]

Eine andere Studie beschäftigte sich damit, zu erkennen, inwiefern eine Person Gelesenes sprachlich versteht. Fünf Teilnehmer sollten einen Text auf einem Blatt Papier lesen und im Anschluss die Wörter markieren, die sie schwierig fanden (Ground Truth). Durch eine Analyse der Fixationen, die hier mit einem videobasiertem Eyetracker erfasst wurden, konnten ebendiese Wörter identifiziert werden. [21]

Diese Studien zeigen beispielhaft, dass auch kognitive Prozesse anhand der Augenbewegungen nachvollzogen werden können [4, 21].

5.3 Erkennen von Tagesabläufen

Basierend auf den oben aufgeführten Erkenntnissen wurde in weiterführenden Studien versucht, dem eigentlichen Ziel näherzukommen, nämlich alle möglichen Kontexte und Tätigkeiten, auch unter Berücksichtigung der kognitiven Komponente, maschinell zu erkennen. Nun wurden Messungen auch über längere Zeiträume bzw. den ganzen Tag ausgeführt und nicht nur für einige Minuten. Dies war nun auch auf Grund der technischen Weiterentwicklung der Eyetracker leichter möglich [19, 18].

5.3.1 Methodik

Die Methoden zur Klassifizierung von Tätigkeiten unterschieden sich von denen vorhergehender Studien, da nun der Fokus auf dem Erkennen von Klassen lag. Diese sollten (kombiniert) viele verschiedene Tätigkeiten beschreibbar machen.

Wörterbasiert

Bei dieser Methode werden Sakkaden als Characters codiert, z. B. L für Sakkade nach links usw., und zu einem String zusammengesetzt. Mittels Machine Learning werden daraufhin die Wörter (= Bewegungsmuster) ausgewählt, die einen Kontext am besten beschreiben. Grundlage davon ist die Annahme, dass unterschiedliche Augenbewegungsmuster auftreten, wenn ein Kontext (z. B. soziale Interaktion) aktiv ist oder nicht. Basierend darauf soll der gesamte Kontext bestimmt werden, wobei immer ein Teil der Daten einer Person zum Training des Algorithmus verwendet wird. Diese Methode ist in der Regel weniger komplex als die featurebasierte Methode. [9]

Themenbasiert

Bei diesem Machine-Learning-Ansatz werden die Augenbewegungen als String codiert und Bewegungsmuster als Menge an Wörtern, die zu unterschiedlichen Themen gehören, dargestellt. Ein Thema ist sozusagen eine Aktivitätsklasse (z. B. soziale Interaktion). Mittels Machine Learning wird bestimmt, wann ein

Thema aktiv ist, und so werden Aktivitäten nachvollzogen. Aktivitäten können sich hierbei auch aus mehreren Themen zusammensetzen. Bei dieser Methode ist kein Training nötig, die Tätigkeiten müssen nur lange genug ausgeführt werden. [30]

In Studien zum Erkennen von Tagesabläufen wurde sowohl EOG [9] als auch videobasiertes Eyetracking [30] verwendet. Ground-Truth-Daten wurden wieder manuell, entweder nachträglich per Video [30] (nur für die Evaluation) oder vom Teilnehmer selbst während des Experiments dokumentiert [9].

5.3.2 Zugehörige Arbeiten

In einer Studie von Bulling et al. (2013) wurde untersucht, ob sich die folgenden vier high-level Kontexteigenschaften aus den Augenbewegungen erschließen lassen: sozial (soziale Interaktion vs. keine), kognitiv (konzentriert arbeiten (z. B. lesen) vs. Müßiggang (passiver Konsum, z. B. einen Film ansehen, bzw. Tätigkeiten, die nicht zielgerichtet sind, z. B. ausgehen), physisch (physisch aktiv vs. nicht), räumlich (drinnen vs. draußen aufhalten). Diese binären Eigenschaften sollten weniger eingrenzend sein und so ein breiteres Spektrum an Situationen ermittelbar machen. Die Augenbewegungen (hier nur Sakkaden) der Teilnehmer (N = 4) wurden über einen ganzen Tag hinweg (normaler Arbeitstag) mittels EOG aufgezeichnet. Annotation wurde von den Teilnehmern, unterstützt durch eine App, selbst übernommen. Zudem wurde der Machine-Learning-Algorithmus für die Klassifizierung mit personenbezogenen Daten trainiert um gute Activity Recognition zu gewährleisten. Obwohl die Art der Interaktionen mit anderen Personen zwischen den Teilnehmern mit großer Sicherheit variierte, konnte soziale Interaktion am besten erkannt werden (Genauigkeit: 85.3%, Recall: 98%). Die Komponente der körperlichen Betätigung wurde am schlechtesten erkannt (Genauigkeit: 75.3%, Recall: 74.4%), da der Fall *keine körperliche Betätigung* manchmal schlecht erkannt wurde. Ungenaue Ergebnisse lagen teilweise auch an schlechten Signalen. Die durchschnittliche Genauigkeit lag bei 76.8% (Recall: 85.5%). [9]

Auch in einer Studie von Steil und Bulling (2015) wurden die Augenbewegungen über einen ganzen Tag hinweg aufgezeichnet (10h). Den Teilnehmern (N = 10) wurden keinerlei Einschränkungen bezüglich ihres Tagesablaufs auferlegt, da beliebige Tätigkeiten anhand charakteristischer Augenbewegungen erkannt werden sollten. Da diese Studie 2015 durchgeführt wurde, konnte man nun auch einen alltagstauglicheren videobasierten Eyetracker verwenden. Der genutzte themenbasierte Machine-Learning-Ansatz ermöglichte eine Durchführung, die relativ „unsupervised“ war. Die Annotation auf der Grundlage von Videoaufnahmen wurde für neun beispielhafte Kontexteigenschaften lediglich zu Evaluationszwecken durchgeführt. Dazu gehörten: draußen sein, soziale In-

teraktion, konzentriertes Arbeiten, Fortbewegung (zu Fuß, mit dem Auto etc.), Lesen, Arbeiten am Computer, Medienkonsum (Filme etc.), Essen, aber auch spezielle Tätigkeiten, zu denen beispielsweise Schuhe binden oder Rucksack packen, zählten. Am besten konnte Lesen erkannt werden (Performance: 74.75%, $\text{Performance} = 2 * ((\text{Genauigkeit} * \text{Recall}) / (\text{Genauigkeit} + \text{Recall}))$). Auch konzentriertes Arbeiten (Performance: 70.01%) und Arbeiten am Computer (Performance: 64.18%) konnten wie in vorhergehenden Studien leichter erkannt werden als andere Tätigkeiten. Die Performances für andere Tätigkeiten lagen bei ca. 50%. Generell galt, je länger eine Tätigkeit ausgeführt wurde, desto besser konnte sie erkannt werden. Die reine Analyse von Sakkaden funktionierte beispielsweise bei sozialer Interaktion gut, eine gemeinsame Analyse der Sakkaden und Fixationsdauer tat dies für fokussiertes Arbeiten, und eine Analyse von Sakkaden, Fixationsdauer und Blinzeln tat dies für Lesen (für diese Erkenntnisse wurden jedoch nur die Daten eines Teilnehmers beispielhaft betrachtet). [30]

6 Diskussion

Die in den Studien verwendeten tragbaren Sensoren zeichneten die Augenbewegungen der Versuchsteilnehmer auch in Alltagssituationen in der Regel zuverlässig auf und verdeutlichten das Potential tragbarer Eyetracker. Zudem waren schon grundlegende Augenbewegungsarten ausreichend, um Tätigkeiten und Kontexte nachzuvollziehen, und die Blickrichtung wurde für eine Analyse gar nicht zwingend benötigt. Daher war neben videobasiertem Eyetracking auch die Elektrookulografie eine geeignete Messmethode. Teilweise genügte sogar schon die Analyse einer Bewegungsart für die Klassifizierung. Unter Berücksichtigung weiterer Eigenschaften konnte jedoch oft eine größere Genauigkeit gewährleistet werden. Die Annahme, dass Activity Recognition auch mit Hilfe der Augenbewegungen möglich sei, konnte also bestätigt werden. Deutlich wurde außerdem, dass anhand der Augen im Vergleich zu anderen Modalitäten (z. B. Körperbewegungen) ein breiteres Spektrum an physischen und sogar auch an psychischen Prozessen identifiziert werden kann. Darüber hinaus wurden in keiner Studie geschlechtsbezogene Unterschiede erkannt. [4, 6, 9, 30, 3]

Das Problem, dass Tätigkeiten, die keine gleichmäßigen Augenbewegungsmuster mitsichbringen oder auch die Muster anderer Tätigkeiten beinhalten, zu Schwierigkeiten bei der Klassifizierung führten, konnte oft nicht umgangen werden. Auch Tätigkeiten trotz kurzer Unterbrechungen (z. B. durch Heben des Blickes) zu erkennen, bereitete manchmal gewisse Schwierigkeiten. Als mögliche Lösungen wurden eine genauere Annotation oder das Betrachten längerer Zeiträume vorgeschlagen. [7, 5]

Dass die Dauer einer Aktivität kürzer oder länger eingestuft wurde als sie eigentlich war, lag teilweise an ungenauen Annotationen, die oft manuell gemacht wurden [6, 5]. Möglichwerweise wäre auch ein anderer Machine-Learning-Ansatz sinnvoll. Werden wie in vielen hier genannten Studien low-level Eigenschaften, wie z. B. die mittlere Fixationsdauer, zur featurebasierten Klassifizierung verwendet, kann es leicht zu Überschneidungen kommen und die Tätigkeiten können nicht klar voneinander abgegrenzt werden [29]. Diese Verbesserungsansätze wurden in späteren Studien auch teilweise schon umgesetzt [30].

Bezüglich kognitiver Prozesse ist ebenfalls noch weitere Forschung nötig. So gibt es bisher noch nicht viele kognitive Prozesse, die in Studien der Activity Recognition betrachtet wurden bzw. auf Grund des aktuellen Forschungsstandes und einem Mangel an alltagstauglichen Messmethoden betrachtet werden konnten. Zudem sollte auch untersucht werden, wie der Einfluss kognitiver Prozesse von anderen Einflüssen auf die Augenbewegungen separiert werden kann [4, 3].

Weitere Einschränkungen sind, dass die Aktivitäten in den ersten Studien teilweise nur relativ kurz (nur ein paar Minuten) ausgeführt wurden [7, 17]. Zudem fanden manche der Experimente in sehr kontrollierten Umgebungen statt [21, 16] bzw. die Teilnehmer mussten sich an gewisse Handlungsvorgaben halten (z. B. [7, 6]), und Annotation war zwingend nötig. Erst 2015 testeten Steil und Bulling einen weniger einschränkenden Ansatz [30, 29]. Generell war die Teilnehmerzahl außerdem in jeder Studie eher gering, z. B. [6, 9].

Der Erfolg der Klassifizierung war zu einem Teil auch von der jeweiligen Person abhängig, bezogen auf ihr visuelles Verhalten und Art und Dauer der Ausführung [30].

In allen Studien wurde mindestens eine der drei grundlegenden Augenbewegungsarten (Sakkaden, Fixationen, Blinzeln) zur Klassifizierung verwendet. Die genauen Eigenschaften dieser Augenbewegungsarten, die ausgewählt wurden, variierten jedoch oft zwischen den Studien, z. B. [29, 4, 3, 6]. Oft führte eine Kombination aus verschiedenen Features verschiedener Bewegungsarten zu gutem Erkennen, und Autoren vermuteten, dass zusätzliche Features anderer Modalitäten, wie beispielsweise der Pupillendurchmesser, die Bestimmungsgenauigkeit sogar noch verbessern könnten [7, 35, 17].

7 Ausblick

Um augenbasierte Activity Recognition für kontextbewusste Systeme im Alltag nutzen zu können, müssen noch bessere Eyetracker und Machine-Learning-Ansätze gefunden werden. Ein idealer Eyetracker sollte präzise und robust sein, jedoch auch unaufdringlich und für alle Personen tragbar [1, 15]. Darüberhinaus sollten eingebaute Batterien Eyetracking über den ganzen Tag erlauben [6].

Bezüglich alltagstauglicher Eyetracker wurden in den letzten Jahren schon große Fortschritte erzielt, doch Alltagstauglichkeit konnte bisher noch nicht zu 100% erreicht werden [16]. So gibt es nun schon eine EOG-Brille, die sowohl vom Tragecomfort wie auch vom Aussehen her einer normalen Brille ähnelt und Signale über einen langen Zeitraum aufzeichnen kann (s. Abschnitt 4.2)[16]. Auch tragbarere und weniger aufdringliche videobasierte Eyetracker wurden entwickelt, z. B. nähern sich die Tobii Pro Glasses 2.0⁸ der Vision einen Eyetracker zu schaffen, der wie eine normale Brille aussieht, ebenfalls an. Jedoch hat dieser Eyetracker nur eine begrenzte Batterielaufzeit von 120 Minuten [33]. Auch Pupil Labs kündigte 2019 einen Eyetracker an, der aussieht wie eine gewöhnliche Brille [22].

Außerdem sollte untersucht werden, ob Activity Recognition durch das Hinzuziehen zusätzlicher Augenbewegungsarten, wie z. B. Smooth Pursuits, verbessert werden kann. Diese zeigen besonders im Hinblick auf kognitive Prozesse Potential, da eine Analyse der Smooth Pursuits beispielsweise Auskunft über geistige Krankheiten oder den kognitiven Workload geben kann [35, 20]. Auch könnte noch mehr bezüglich der Robustheit von Activity Recognition Systemen gegenüber Ablenkungen und verschiedenen Erscheinungsbildern von Situation erforscht werden [4]. Zudem sollten Studien mit einer größeren Teilnehmerzahl durchgeführt werden für aussagekräftigere Ergebnisse.

Bulling et al. regten außerdem an, noch mehr binäre Kontexteigenschaften zu sammeln und zu untersuchen, um den Kontext einer Person noch präziser bestimmen zu können [9].

Sollte Annotation für die Klassifizierung benötigt werden, sollte dieser Prozess noch verbessert und möglicherweise automatisiert werden, um präzisere Activity Recognition zu erzielen. Möglich wären auch die Nutzung zusätzlicher Sensoren, die z. B. Körperbewegungen aufzeichnen [6]. Auch die Nutzung von Videoaufnahmen wurde häufig empfohlen [5]. [9]

Für die Umsetzung von Systemen, die auf aktuelle physische und psychische Zustände eines Menschen reagieren können, haben die Rücksichtnahme auf Ethik und die Gewährleistung der Privatsphäre allergrößte Priorität [4]. Auch in diesem Bereich wird schon intensiv geforscht, z. B. [31].

Augenbasierte, kontextbewusste Systeme haben ein großes Potential, den Alltag des Nutzers in vielerlei Hinsicht zu erleichtern [9, 1, 6, 32]. Obwohl bis zur Verwirklichung augenbasierter Activity Recognition im Alltag noch etwas Forschung nötig sein wird, scheint die zu Anfang geschilderte Vision, dass es bald Systeme gibt, die sich automatisch an die Tätigkeit und die damit einhergehenden Bedürfnisse des Nutzers anpassen, doch schon in immer greifbarere Nähe zu rücken.

⁸ <https://www.tobiipro.com/de/produkte/tobii-pro-glasses-2/>

Literatur

- 1 Andreas Bulling. Human visual behaviour for collaborative human-machine interaction. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, UbiComp/ISWC'15 Adjunct, pages 901–905, New York, NY, USA, 2015. ACM. doi:10.1145/2800835.2815378.
- 2 Andreas Bulling, Ulf Blanke, and Bernt Schiele. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys*, 46(3):33:1–33:33, Jan 2014. doi:10.1145/2499621.
- 3 Andreas Bulling and Daniel Roggen. Recognition of visual memory recall processes using eye movement analysis. In *Proceedings of the 13th international conference on Ubiquitous computing*, UbiComp '11, pages 455–464, New York, NY, USA, 2011. ACM. doi:10.1145/2030112.2030172.
- 4 Andreas Bulling, Daniel Roggen, and Gerhard Tröster. What's in the eyes for context-awareness? *IEEE Pervasive Computing*, 10(2):48–57, Apr 2011. doi:10.1109/MPRV.2010.49.
- 5 Andreas Bulling, Jamie A. Ward, and Hans Gellersen. Multimodal recognition of reading activity in transit using body-worn sensors. *ACM Transactions on Applied Perception*, 9(1):2:1–2:21, Mar 2012. doi:10.1145/2134203.2134205.
- 6 Andreas Bulling, Jamie A. Ward, Hans Gellersen, and Gerhard Tröster. Robust recognition of reading activity in transit using wearable electrooculography. In Jadwiga Indulska, Donald J. Patterson, Tom Rodden, and Max Ott, editors, *Pervasive Computing*, Pervasive 2008, pages 19–37, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg. doi:10.1007/978-3-540-79576-6_2.
- 7 Andreas Bulling, Jamie A. Ward, Hans Gellersen, and Gerhard Tröster. Eye movement analysis for activity recognition. In *Proceedings of the 11th international conference on Ubiquitous computing*, UbiComp '09, pages 41–50, New York, NY, USA, 2009. ACM. doi:10.1145/1620545.1620552.
- 8 Andreas Bulling, Jamie A. Ward, Hans Gellersen, and Gerhard Tröster. Eye movement analysis for activity recognition using electrooculography. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(4):741–753, Apr 2011. doi:10.1109/TPAMI.2010.86.
- 9 Andreas Bulling, Christian Weichel, and Hans Gellersen. Eyecontext: recognition of high-level contextual cues from human visual behaviour. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 305–308, New York, NY, USA, 2013. ACM. doi:10.1145/2470654.2470697.
- 10 Philipp P. Caffier, Udo Erdmann, and Peter Ullsperger. Experimental evaluation of eye-blink parameters as a drowsiness measure. *European*

- journal of applied physiology*, 89(3–4):319–325, Jun 2003. doi:10.1007/s00421-003-0807-5.
- 11 Roxanne L. Canosa. Real-world vision: Selective perception and task. *ACM Transactions on Applied Perception*, 6(2):11:1–11:34, Mar 2009. doi:10.1145/1498700.1498705.
 - 12 Andrew T. Duchowski. *Eye tracking methodology: theory and practice*. Springer, London, 2nd edition, 2007. doi:10.1007/978-1-84628-609-4.
 - 13 L.M. Gallo and S. Palla. Activity recognition in long-term electromyograms. *Journal of Oral Rehabilitation*, 22(6):455–462, Jun 1995. doi:10.1111/j.1365-2842.1995.tb00801.x.
 - 14 A. Godfrey, R. Conway, D. Meagher, and G. ÓLaighin. Direct measurement of human movement by accelerometry. *Medical Engineering & Physics*, 30(10):1364–1386, Jan 2009. URL: https://www.researchgate.net/publication/23461269_Direct_measurement_of_human_movement_by_accelerometry, doi:10.1016/j.medengphy.2008.09.005.
 - 15 Yoshio Ishiguro, Adiyana Mujibiya, Takashi Miyaki, and Jun Rekimoto. Aided eyes: eye activity sensing for daily life. In *Proceedings of the 1st Augmented Human International Conference on*, AH '10, pages 25:1–25:7, New York, NY, USA, 2010. ACM. doi:10.1145/1785455.1785480.
 - 16 Shoya Ishimaru, Kensuke Hoshika, Kai Kunze, Koichi Kise, and Andreas Dengel. Towards reading trackers in the wild: detecting reading activities by eog glasses and deep neural networks. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers on*, UbiComp '17, pages 704–711, New York, NY, USA, 2017. ACM. doi:10.1145/3123024.3129271.
 - 17 Shoya Ishimaru, Kai Kunze, Koichi Kise, Jens Weppner, Andreas Dengel, Paul Lukowicz, and Andreas Bulling. In the blink of an eye: combining head motion and eye blink frequency for activity recognition with google glass. In *Proceedings of the 5th Augmented Human International Conference on*, AH '14, pages 15:1–15:4, New York, NY, USA, 2014. ACM. doi:10.1145/2582051.2582066.
 - 18 Shoya Ishimaru, Yuji Uema, Kai Kunze, Koichi Kise, Katsuma Tanaka, and Masahiko Inami. Smarter eyewear: using commercial eog glasses for activity recognition. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct Publication*, UbiComp '14 Adjunct, pages 239–242, New York, NY, USA, 2014. ACM. doi:10.1145/2638728.2638795.
 - 19 Moritz Kassner, William Patera, and Andreas Bulling. Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, UbiComp

- '14 Adjunct, pages 1151–1160, New York, NY, USA, 2014. ACM. doi:10.1145/2638728.2641695.
- 20 Thomas Kosch, Mariam Hassib, Paweł W. Woźniak, Daniel Buschek, and Florian Alt. Your eyes tell: Leveraging smooth pursuit for assessing cognitive workload. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 436:1–436:13, New York, NY, USA, 2018. ACM. doi:10.1145/3173574.3174010.
 - 21 Kai Kunze, Hitoshi Kawaichi, Kazuyo Yoshimura, and Koichi Kise. Towards inferring language expertise using eye tracking. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems on*, CHI EA '13, pages 217–222, New York, NY, USA, 2013. ACM. doi:10.1145/2468356.2468396.
 - 22 Pupil Labs. Pupil invisible - beta launch. Blog entry, Jan 2019. Retrieved August 1, 2019 from <https://pupil-labs.com/news/pupil-invisible-beta-launch>.
 - 23 Feng Li, Susan Kolakowski, Jeff Pelz, and Chester F Carlson. A model-based approach to video-based eye tracking. *Journal of Modern Optics*, 55:503–531, Feb 2008. doi:10.1080/09500340701467827.
 - 24 Hong Lu, Wei Pan, Nicholas D. Lane, Tanzeem Choudhury, and Andrew T. Campbell. Soundsense: scalable sound sensing for people-centric applications on mobile phones. In *Proceedings of the 7th international conference on Mobile systems, applications, and services*, MobiSys '09, pages 165–178, New York, NY, USA, 2009. ACM. doi:10.1145/1555816.1555834.
 - 25 Cornelia Möhring. Was ist google now? Alles, was sie wissen müssen. Heise Website, Jun 2018. Retrieved August 1, 2019 from <https://www.heise.de/tipps-tricks/Was-ist-Google-Now-Alles-was-Sie-wissen-muessen-4078179.html>.
 - 26 Keith Rayner. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3):372–422, Nov 1998. doi:10.1037/0033-2909.124.3.372.
 - 27 Thiago Santini, Hanna Brinkmann, Luise Reitstätter, Helmut Leder, Raphael Rosenberg, Wolfgang Rosenstiel, and Enkelejda Kasneci. The art of pervasive eye tracking: unconstrained eye tracking in the austrian gallery belvedere. In *Proceedings of the 7th Workshop on Pervasive Eye Tracking and Mobile Eye-Based Interaction*, PETMEI '18, pages 5:1–5:8, New York, NY, USA, 2018. ACM. doi:10.1145/3208031.3208032.
 - 28 Greg J. Siegle, Naho Ichikawa, and Stuart Steinhauer. Blink before and after you think: Blinks occur prior to and following cognitive load indexed by pupillary responses. *Psychophysiology*, 45(5):679–687, Sep 2008. doi:10.1111/j.1469-8986.2008.00681.x.
 - 29 Namrata Srivastava, Joshua Newn, and Eduardo Velloso. Combining low and mid-level gaze features for desktop activity recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(4):189:1–189:27, Dec 2018. doi:10.1145/3287067.

- 30 Julian Steil and Andreas Bulling. Discovery of everyday human activities from long-term visual behaviour using topic models. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '15, pages 75–85, New York, NY, USA, 2015. ACM. doi:10.1145/2750858.2807520.
- 31 Julian Steil, Marion Koelle, Wilko Heuten, Susanne Boll, and Andreas Bulling. Privaceye: privacy-preserving head-mounted eye tracking using ego-centric scene image and eye movement features. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, ETRA '19, pages 26:1–26:10, New York, NY, USA, 2019. ACM. doi:10.1145/3314111.3319913.
- 32 Bernd Tessendorf, Andreas Bulling, Daniel Roggen, Thomas Stiefmeier, Manuela Feilner, Peter Derleth, and Gerhard Tröster. Recognition of hearing needs from body and eye movements to improve hearing instruments. In Kent Lyons, Jeffrey Hightower, and Elaine M.Editors Huang, editors, *Pervasive Computing*, pages 314–331, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg. doi:10.1007/978-3-642-21726-5_20.
- 33 Tobii AB. *Tobii Pro Glasses 2 Product Description*, Aug 2018. Retrieved August 1, 2019 from <https://www.tobii.com/siteassets/tobii-pro/product-descriptions/tobii-pro-glasses-2-product-description.pdf>.
- 34 Marc Tonsen, Julian Steil, Yusuke Sugano, and Andreas Bulling. Invisibleeye: Mobile eye tracking using multiple low-resolution cameras and learning-based gaze estimation. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(3):106:1–106:21, Sep 2017. doi:10.1145/3130971.
- 35 Mélodie Vidal, Andreas Bulling, and Hans Gellersen. Analysing eog signal features for the discrimination of eye movements with wearable devices. In *Proceedings of the 1st international workshop on Pervasive eye tracking & mobile eye-based interaction*, PETMEI '11, pages 15–20, New York, NY, USA, 2011. ACM. doi:10.1145/2029956.2029962.
- 36 Mélodie Vidal, Jayson Turner, Andreas Bulling, and Hans Gellersen. Wearable eye tracking for mental health monitoring. *Computer Communications*, 35(11):1306–1311, Jun 2012. doi:10.1016/j.comcom.2011.11.002.
- 37 A. L. Yarbus. *Eye Movements and Vision*. Plenum Press, New York, NY, 1967.

Mobile Wiederaufnahme unterbrochener Aufgaben mit Attention-Aware Interfaces

Lars Reisig

Ludwig-Maximilians-Universität München, München, Deutschland
lars.reisig@campus.lmu.de

Zusammenfassung

Moderne Smartphones stellen in den unterschiedlichsten Umgebungen kontinuierlich neue Informationen bereit. Da ihre Nutzer jedoch nur ein begrenztes Maß an Informationen gleichzeitig verarbeiten können, führen besonders digitale Push-Benachrichtigungen zu regelmäßigen Unterbrechungen im Alltag. Häufig sinken hierdurch die Produktivität und das Wohlbefinden, stellenweise wird zudem auch die Fortsetzung von Aufgaben vergessen. In dieser Arbeit wird wissenschaftlich analysiert, anhand welcher Faktoren sich diese Auswirkungen vermeiden lassen könnten. Darauf aufbauend werden Gestaltungsrichtlinien für zukünftige Systeme hergeleitet und vor dem Hintergrund aufmerksamkeitssensitiver Schnittstellen (AUIs) auf ihre Umsetzbarkeit untersucht. Die Ergebnisse deuten darauf hin, dass AUIs schon heute in der Lage wären, die Auswirkungen von Unterbrechungen strategisch zu kompensieren oder diese bereits im Vorhinein zu vermeiden. Einen noch größeren Effekt könnte in der Zukunft jedoch die Kombination beider Ansätze erzielen.

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

Keywords and phrases Multitasking; task resumption; mobile computing; divided attention; interruption management, attentive-awareness, mobile computing

1 Einleitung

“Was wollte ich eigentlich hier?” - “Wie bin ich hierher gekommen?” - “Was habe ich mir denn hierbei gedacht?” - “Wo war ich denn überhaupt?”

All dies sind Fragen, die sich jeder Mensch in der heutigen Zeit schon mindestens einmal gestellt hat und sie entspringen alle demselben Problem: Wir wurden an irgendeiner Stelle in unserem Ablauf unterbrochen. Neu ist diese Problematik sicherlich nicht [15, 21, 56], doch die stetige Weiterentwicklung ubiquitärer Konnektivität hat bislang zu einem rapiden Anstieg dieses



© Lars Reisig;

licensed under Creative Commons License CC-BY

Cite as: Lars Reisig. Mobile Wiederaufnahme unterbrochener Aufgaben mit Attention-Aware Interfaces. In *3rd Seminar on Ubiquitous Interaction (UBIACTION 2019)*. Editors: Pascal Knierim, Jakob Karolus, Fiona Draxler, Ville Mäkelä, Tonja Machulla, Florian Lang, Thomas Kosch, Albrecht Schmidt. July 16, 2019. Munich, Germany. pp. 9:1–9:28.

9:2 "Hey Siri, wo war ich stehen geblieben?"

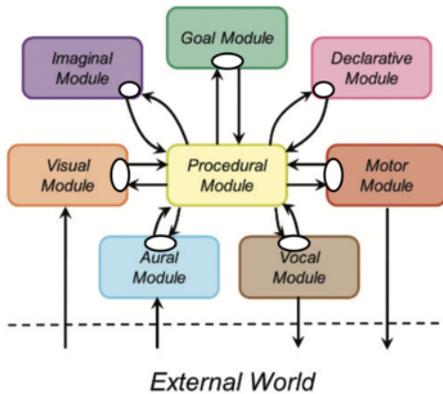
Phänomens geführt [55].

Wir befinden uns in einer Zeit der Digitalisierung und das Internet of Things trägt zu einer immer globaleren Vernetzung unserer Alltagsgegenstände bei. Dennoch ist das Smartphone auch weiterhin das meistgenutzte Kommunikationsmedium [37]. Im Laufe der Zeit hat es sich von einem anfänglichen Mobiltelefon mit Internetzugang zu einem wahrhaften Multitool entwickelt, welches uns mit Funktionalitäten rund um unsere Mails, Termine, Unterhaltung, aber auch Gesundheit und Hobbys in beinahe jeder alltäglichen Situation unterstützt [11, 73]. Menschen scheinen heutzutage gar abhängig von mobiler Erreichbarkeit zu sein [37], setzen sich aber somit ununterbrochen dem damit einhergehenden Einfluss aus. Während die Smartphones der letzten Jahre stetig an Leistung und daraus resultierender Funktionalität zugenommen haben, ist die kognitive Leistungsfähigkeit ihrer Nutzer stetig auf einem vergleichsweise geringen Level geblieben [52, 63]. Als Resultat müssen die Nutzer mobiler Geräte häufig zwischen verschiedenen Aufgaben wechseln, was mitunter nicht freiwillig geschieht, sondern von mittlerweile mehr als 60 mobilen Benachrichtigten am Tag verschuldet wird [73]. Dies birgt gewisse Risiken, da sich die Nutzer oftmals im Nachhinein wieder an die Aufgaben erinnern müssen, die sie an früheren Zeitpunkten einmal unterbrochen haben. Schlägt dies fehl, werden Aufgaben gänzlich vergessen und nicht mehr fortgesetzt [6]. Selbst wenn Aufgaben selbstständig wieder aufgenommen werden, kommt es häufig zu dem Problem, den ursprünglichen Kontext wieder in den Vordergrund zu rufen, sodass auch für die erneute Einarbeitung noch einmal Zeit aufgewendet werden muss [36]. Im Fazit setzen sich heutige Nutzer somit nicht nur ubiquitärer Konnektivität sondern vor allem ubiquitären Unterbrechungen aus [9].

Obwohl mittlerweile schon vielfach die strategische Kontrolle der Auslöser, meist Notifications, erprobt wurde [3], können oder wollen einige dieser Unterbrechungen gar nicht gänzlich von den Nutzern vermieden werden [59]. Für all diese unvermeidbaren Situation benötigen Nutzer eine Unterstützung bei der späteren Wiederaufnahme ihrer Aufgaben, welche jedoch bislang in einem äußerst unzufrieden stellenden Maße von mobilen Systemen bereitgestellt wird. Ich möchte mich in dieser Arbeit daher zunächst mit den kognitiven Gründen für diese Problematik befassen und entsprechende Design-Richtlinien sowohl für die Unterstützung vor, aber insbesondere auch nach einer Unterbrechung zusammentragen, um Entwicklern mobiler Systeme einen konzentrierten Überblick zu verschaffen. Darauf aufbauend werde ich dann bestehende und neue Ansätze diskutieren, wie diese allgemeingültigen Richtlinien unter Verwendung von Attention-Aware-Interfaces tatsächlich in unserem heutigen Alltag realisiert werden könnten.

2 Multitasking im mobilen Kontext

Unterbrechungen sind keine alleinstehenden Ereignisse, sondern Teil eines mehrschrittigen menschlichen Multitasking-Prozesses. Multitasking bezeichnet hierbei die Fähigkeit, mehrere Aufgaben in einer bestimmten Zeit auszuführen [17], und wurde in Vergangenheit bereits vielfach wissenschaftlich untersucht [2, 4, 5, 49, 60, 64, 62].



■ **Abbildung 1** Aufbau der ACT-R 6.0 Architektur adaptiert von Wirzberger und Russwinkel [74]. Jede kognitive Ressource wird hierbei über ein eigenes Modul abgebildet. Der Austausch von Informations-Einheiten wird durch das prozedurale Modul gesteuert.

einer konkreten Intention heraus, weswegen ACT-R auch diese in einem *Goal-Modul* abbildet. Gemäß der *Threaded Cognition*-Theorie von Salvucci et al. [62] wird jede Aufgabe hierbei als simultan agierender Arbeitsprozess (Thread) verwaltet, sodass auch mehrere Aufgaben gleichzeitig ausgeführt werden könnten. Eingeleitet von diesen Threads greift das prozedurale Modul dann auf die benötigten Module zu und leitet den Informationsaustausch zwischen den Modulen und somit auch der Außenwelt ein.

Sofern diese Threads nicht oder nur in kurzen Zyklen (von mehreren Sekunden) auf dieselben Module (beispielsweise die Augen oder die Hände) zurückgreifen, wird der Vorgang als **gleichzeitiges Multitasking** bezeichnet [64]. Auf diese Weise wird unter anderem die gleichzeitige Unterhaltung mit einem Beifahrer während einer Autofahrt ermöglicht. Auch im Umgang mit Smartphones können derartige Momente auftreten (beispielsweise beim Blick

Häufig wurde hierzu die *ACT-R* Architektur von Anderson et al. [4] als digitales Modell der menschlichen Kognition herangezogen. Wie in Abbildung 1 ersichtlich, basiert dieses Modell auf der Theorie, dass alle kognitiv relevanten Ressourcen als eigenständige Module unabhängig voneinander operieren. Jedes Modul (mit Ausnahme des prozeduralen Moduls) verwaltet seine Informationen in definierten Einheiten (Chunks). Diese stellt es über Schnittstellen (Buffer) mit einer begrenzten Kapazität für das prozedurale Modul bereit.

Da in diesem Modell auch die sensorischen und motorischen Module getrennt voneinander agieren, wären simultane Handlungen per Definition möglich. Menschliche Handlungen entstehen jedoch zumeist erst aus einer

9:4 "Hey Siri, wo war ich stehen geblieben?"

auf den Boden zur Orientierung). Diese gelten jedoch im Rahmen dieser Arbeit als harmlos, da die primäre Aufgabe relativ schnell wieder fortgesetzt wird [20, 28, 64] und demnach nur eine sehr geringe Gefahr besteht, die Aufgaben zu vergessen.

Im umgekehrten Fall spricht man von **sequentiellm Multitasking**, da sich die Threads hierbei für einen längeren Zeitraum (von mehrere Minuten bis Stunden) gegenseitig unterbrechen. Obwohl dies im Umgang mit Smartphones bereits durch die Tatsache gegeben ist, dass mobile Apps nach einer Unterbrechung tendenziell für einen längeren Zeitraum beendet werden [59], so gibt es in diesem Fall noch eine weitere limitierende Ressource. Borst et al. [10] haben auf Basis der *Memory-for-Goals*-Theorie von Altmann et al. [2] herausgefunden, dass die Informationen eines Aufgabenkontextes (beispielsweise der Inhalt eines Buches), getrennt vom zuständigen Thread, in einer dedizierten kognitiven Ressource verwaltet werden. ACT-R bildet diese im *Problem State* bzw. Imaginal Module ab, welches jedoch nur einen Aufgabenkontext zur selben Zeit beinhalten kann. Demzufolge könnte selbst bei gleichzeitiger Darstellung zweier mobiler Apps nur eine von beiden zielgerichtet fokussiert werden, wenn beide jeweils eigene Hintergrundinformationen besitzen. Diese Art der Limitierung wird auch als „Problem State Bottleneck“ bezeichnet [9]. Da es sich bei den meisten Situationen, in denen mobile Geräte eingesetzt werden, um diese Art des Multitaskings handelt, wird der sequentielle Multitasking-Prozess im Folgenden noch einmal konkreter analysiert.

2.1 Sequentielles Multitasking



■ **Abbildung 2** Der zeitliche Ablauf eines sequentiellen Multitasking-Prozesses. Adaptiert von Trafton et al. [69]

Ein sequentieller Multitasking-Prozess verläuft auf Basis der aktuellen Literatur [26, 63, 69] in vier zeitlich aufeinanderfolgenden Abschnitten (siehe Abbildung 2). Charakteristisch sind hierbei insbesondere die beiden Zeitabschnitte *interruption lag* und *resumption lag*. Sie unterteilen den gesamten Prozess in die Phasen der **Unterbrechung** sowie der **Wiederaufnahme**.

2.1.1 Unterbrechung einer Aufgabe

Ein sequentieller Multitasking-Prozess wird stets durch einen Hinweis auf eine sekundäre Aufgabe ausgelöst. Häufig wird dieser Hinweis jedoch selbst als Interruption (Unterbrechung) bezeichnet, da er die weitere Ausführung einer primären Aufgabe unterbindet. Im Laufe der Zeit haben sich in der Wissenschaft einige verschiedene Definitionen für dieses Auftreten ergeben [20]. Neben der oft verwendeten Unterteilung in die sogenannten internen und externen Unterbrechungen [43], wird stellenweise auch die folgende Einteilung in die sogenannten vermeidbaren und unvermeidbaren Unterbrechungen vorgenommen [63].

- **Vermeidbare Unterbrechungen** werden durch externe Einflüsse ausgelöst, welche bereits im Vorfeld hätten kontrolliert und vermieden werden können. Dies trifft im Kontext mobiler Unterbrechungen vor allem auf die sogenannten Push-Notifications, also Benachrichtigungen simultan ausgeführter Apps, zu [38, 56]. Studien haben ermittelt, dass insbesondere Kommunikationsanwendungen, wie soziale Netzwerke, Messenger oder Mail-Apps den größten Anteil hiervon ausmachen. Darüber hinaus werden aber auch komplett unterschiedliche Anwendungsszenarien, wie beispielsweise Lifestyle-, Gesundheits- und Produktivitäts-Anwendungen sowie Spiele voneinander in der Ausführung unterbrochen [48, 59]. Es sei an dieser Stelle erwähnt, dass falsch platzierte Notifications schon im Desktop-Umfeld als Problem identifiziert und untersucht wurden [15, 28]. Im mobilen Umfeld wird dieses Problem jedoch noch einmal verstärkt [55], weswegen auf diesem Gebiet eine gesonderte Betrachtung notwendig ist.
- **Unvermeidbare Unterbrechungen** können entweder durch die interne Motivation oder externe Einflüsse ausgelöst werden. Im ersteren Fall führen oft Überanstrengungen, Wartezeiten oder Angewohnheiten dazu, dass wir uns selbst (spontan) unterbrechen [16]. Externe unvermeidbare Einflüsse sind häufig durch die diversen Einsatzgebiete heutiger Smartphones begründet. Sie treten beispielsweise in Szenarien aus dem Verkehr oder der Arbeitsumgebung auf und erzwingen eine sofortige Unterbrechung der aktuellen Aufgabe [31].

Da die unterbrechende sekundäre Aufgabe oft einen eigenen Kontext besitzt, muss der ursprüngliche Aufgabenkontext (*Problem State*) vor Beginn der sekundären Aufgabe aus dem Problem State-Modul ausgelagert werden [2]. Hierzu leitet die sekundäre Aufgabe den Transport des *Problem States* ins deklarative Modul ein. Dort zerfallen die Chunks von nun an in exponentiellem Maße zur vergangen Zeit [2, 4]. Dies führt wiederum zu einem stetigen Abfall ihrer *Basisaktivierung*, wodurch die Wahrscheinlichkeit für eine spätere erfolgreiche Wiederaufnahme ebenfalls sinkt [2, 10].

Interruption Lag

Die Zeitspanne, bis der Nutzer nach einem entsprechenden Hinweis mit der sekundären Aufgabe beginnt, wird als *Interruption Lag* bezeichnet. In dieser Zeit wird sowohl die bewusste als auch die unterbewusste Vorbereitung auf eine spätere Wiederaufnahme der primären Aufgabe vorgenommen [64].

- Bei der **bewussten Vorbereitung** wird die primäre Aufgabe aufgeräumt oder noch bis zu einem erreichbaren Meilenstein fortgesetzt. Dies kann sich im zu Ende Lesen einer Buchseite oder dem Ausschreiben eines Satzes äußern [64].
- Die **unterbewusste Vorbereitung** erfolgt durch den *Rehearsal*-Prozess der primären Aufgabe [2]. Der primäre Thread, welcher weiterhin im *Goal* Modul besteht, greift hierzu regelmäßig auf den ausgelagerten *Problem State* im deklarativen Gedächtnis zu. Mit jedem Zugriff wird dabei die entsprechende *Basisaktivierung* wieder graduell verstärkt, sodass ebenfalls die Wahrscheinlichkeit für eine erfolgreiche spätere Wiederaufnahme ansteigt (siehe Abschnitt 2.1.2). Dieser Effekt hält jedoch nur für mehrere Sekunden an [63], sodass längere Unterbrechungen von dieser Vorbereitung mitunter nicht profitieren können.

Da unvermeidbare Unterbrechungen meist ohne entsprechende Ankündigung stattfinden, kann der Nutzer in diesen Fällen meist keine bewusste Vorbereitung mehr vornehmen. Der unterbewusste Rehearsal-Prozess findet in diesen Fällen zwar dennoch, aber zeitgleich mit der sekundären Aufgabe statt, was zu Interferenzen und einer geringeren Effektivität des Prozesses führen kann [63].

2.1.2 Wiederaufnahme einer Aufgabe

Nachdem die sekundäre Aufgabe im sequentiellen Multitasking-Prozess beendet wurde, setzt die Phase der Wiederaufnahme (*resumption lag*) ein, welche bis zum erneuten Beginn der ursprünglichen Aufgabe anhält. Hierbei ist es entweder möglich, den ursprünglichen *Problem State* aus dem deklarativen Gedächtnis zurückzurufen (*Recall*) [2] oder wiederum notwendig, diesen neu zu rekonstruieren [8, 60].

Recall

In welchem Maße das Zurückrufen bzw. Zurückkehren zur ursprünglichen Aufgabe tatsächlich gelingt, hängt der Theorie nach streng von der momentanen *Gesamtaktivierung* des *Problem States* im deklarativen Gedächtnis ab [2]. Neben der bereits zuvor genannten *Basisaktivierung* wird in diese auch die soge-

nannte *assoziierte Aktivierung* eingerechnet. Diese gründet auf der Erkenntnis, dass sich kontextuell zusammenhänge Chunks im deklarativen Gedächtnis gegenseitig aktivieren. Sieht oder hört ein Nutzer also Informationen, die in einem Bezug zur primären Aufgabe stehen, so wird hierdurch auch der dazugehörige *Problem State* im deklarativen Gedächtnis aktiviert. Ein Beispiel könnte sein, dass Menschen dazu neigen, an einen Ursprungsort zurückzukehren, um sich an eine Idee zu erinnern, die dort entstanden ist. Die assoziierte Aktivierung ist demnach das Maß an Einfluss äußerer Hinweisen auf die ursprüngliche Aufgabe. Bei Erfolg kann diese Art der Wiederaufnahme gemäß Salvucci [63] auf zwei Weisen eingeleitet werden:

- Die *Basisaktivierung* ist noch stark genug, sodass sich der Nutzer nach Beendigung der sekundären Aufgabe wieder selbstständig an die primäre Aufgabe erinnert.
- Die Umgebung bzw. konkrete Elemente darin erinnern den Nutzer an die primäre Aufgabe, wodurch das Zurückrufen des ursprünglichen *Problem States* einleitet wird. In diesen Fällen ist die *assoziierte Aktivierung* ausschlaggebend.

Rekonstruktion

Obwohl der Recall-Prozess bei kurzen Unterbrechungen zum Erfolg führen kann, schlägt dieser besonders im Falle längerer sowie unvermeidbarer Unterbrechungen von komplexen Aufgaben fehl [60]. Der Grund hierfür liegt darin, dass komplexe *Problem States* aus zahlreichen Chunks bestehen und stellenweise nicht vollständig fehlerfrei zurücktransportiert werden können [63]. Desweiteren führt eine längere Unterbrechung der primären Aufgabe zu einem rapiden Abfall der *Basisaktivierung* ihres *Problem States*. Da immer nur der am stärksten aktivierte Kontext zurückgerufen werden kann, erreicht in diesen Fällen auch eine assoziierte Aktivierung nicht die notwendige *Gesamtaktivierung*, um das automatische Zurückrufen einzuleiten [2, 9]. In diesen Fällen muss der *Problem State* der primären Aufgabe neu rekonstruiert werden. Hierzu werden häufig Elemente aus dem Aufgabenumfeld neu analysiert und eingeordnet. Ein prominentes Beispiel ist das erneute Lesen eines Artikels, wenn der bisherige Inhalt und/oder die frühere Stopp-Position nicht mehr bekannt ist [60].

Da diese Art der Wiederaufnahme komplett auf entsprechende äußere Einflüsse angewiesen ist, Smartphones jedoch oftmals nur in kurzen geeigneten Situationen wie beispielsweise der U-Bahn eingesetzt werden [31], kann die Zeitspanne bis zur nächsten Verwendung besonders lang ausfallen [60]. Mitunter wird die Wiederaufnahme sogar komplett vergessen [15]. Es ist im mobilen Umfeld daher zwingend notwendig, entsprechende unterstützende Maßnahmen für die Phase der Wiederaufnahme zu berücksichtigen.

2.2 Folgen von Unterbrechungen (Kosten)

Obwohl Unterbrechungen häufig negativ diskutiert werden, schätzen mobile Nutzer die informative Natur gewisser Unterbrechungen bzw. Notifications [59]. Dennoch existiert mittlerweile die allgemeine Erkenntnis, dass fehlplatzierte oder uninteressante Unterbrechungen negative Auswirkungen auf die jeweilig primäre Aufgabe haben [3, 6, 21, 25, 42, 44, 47]. Um diesen unmittelbaren Einfluss genauer einordnen und vergleichen zu können, hat sich das Konzept der Kosten (oftmals auch switch cost oder disruption cost) durchgesetzt. Dabei handelt es sich um quantifizierbare Effekte, die unmittelbar mit einer Unterbrechung in Verbindung gebracht werden können. Obwohl auch hierfür kein gemeinsamer Standard verfügbar ist, können bisher verwendete Indikatoren und somit auch identifizierte negative Auswirkungen anhand der folgenden Kategorien von Mehrota und Musolesi [47] eingeteilt werden:

- **Auswirkungen auf das Gedächtnis**
 - u. a. Erinnerungsvermögen an primäre Aufgabe
- **Zusammenhang mit der primären Aufgabe**
 - u. a. Fehlerrate, Zeit und Aufwand für die Wiederaufnahme, Zeit für den Abschluss
- **Zusammenhang mit dem emotionalen Zustand des Nutzers**
 - u. a. Stress, Angst und Belästigung bei Eintreffen einer Unterbrechung
- **Auswirkungen auf das Nutzererlebnis**

2.2.1 Längerfristige Auswirkungen

Neben den zuvor genannten kurzfristigen Auswirkungen kann der kontinuierliche falsche Einsatz von Unterbrechungen auch längerfristige negative Konsequenzen für den Nutzer nach sich ziehen. Kushlev et al. führen in diesem Zuge Hyperaktivität und Unaufmerksamkeit als begünstigte Negativeffekte mobiler Unterbrechungen an [34]. Die Auswertung der sogenannten "Do not disturb challenge" von Pielot et al. hat außerdem ergeben, dass Nutzer von Smartphones relativ schnell das stressige Gefühl entwickeln, etwas zu verpassen, wenn sie keinen Zugriff mehr auf ihre mobilen Geräte erhalten [56]. Mark et. al konnten überdies für den stationären Arbeitsplatz die Tendenz entdecken, dass Menschen aus bloßer Vorahnung einer zünftigen Unterbrechung ihr Arbeitstempo erhöhen [44]. Jedoch geht auch dies mit einem dauerhaft erhöhten Stress-Level einher. Pielot et al. haben außerdem herausgefunden, dass auch diejenigen Nutzer, die ihre Notifications komplett unterdrückt hielten, vergleichsweise schnell auf diese reagierten [55]. Dies belegt einerseits die Annahme, dass Smartphonennutzer heutzutage völlig selbstständig auf ihre Geräte schauen. Andererseits unterstreicht es die eingangs in dieser Arbeit genannte These, dass heutige Maßnahmen zum Verwalten von Notifications nicht angemessen sind

und stattdessen dem jeweiligen Kontext adaptiv angepasst werden müssten. Entsprechende Ansätze hierzu werden in Kapitel 4 vorgestellt.

3 Richtlinien zur Kostenreduzierung

Im vorigen Kapitel wurden Kosten als zentrale Maßeinheiten für die Disruptivität von Unterbrechungen eingeführt. Nun gilt es, diese weitestgehend zu minimieren, um Smartphonennutzer bei der mobilen Wiederaufnahme ihrer Aufgaben zu unterstützen. Ein kompletter Verzicht auf Unterbrechungen würde sich jedoch ähnlich negativ wie der aktuelle Zustand auswirken [19], sodass spezifische kontrollierende Maßnahmen ergriffen werden müssen [29]. Um Entwickler bei der zukünftigen Implementierung derartiger Maßnahmen zu unterstützen, werden im Folgenden allgemeingültige Richtlinien definiert, welche im Zuge dessen beachtet werden sollten. Da auf vermeidbare Unterbrechungen bereits im Vorfeld präventiv reagiert werden kann, unvermeidbare Unterbrechungen hingegen eine nachträgliche kurative Unterstützung bei der Wiederaufnahme verlangen, wird gemäß Leiva et al. [38] eine Unterteilung in vorsorgende (präventive) sowie nachsorgende (kurative) Einsatzmöglichkeiten vorgenommen.

3.1 Präventiv

Viele der heutigen mobilen Notifications verlangen keine direkte Reaktion und sind demnach vermeidbar, verleiten uns aber, ihr nachzugehen, woraus eine Unterbrechung entstehen kann [16, 59]. Im Folgenden werden auf Basis von Borst et al. sowie Salvucci [9, 63] theoretische Faktoren aufgeführt, welche schon im Vorfeld die Kosten der Unterbrechung reduzieren und somit die Notwendigkeit nachsorgender Strategien eindämmen könnten. Aus jedem Faktor wird dann eine konkrete Gestaltungsrichtlinie hergeleitet.

■ Komplexität beider Aufgaben

Komplexe Aufgaben besitzen einen größeren *Problem State*. Daraus resultierend müssen mehr Informationen ins deklarative Gedächtnis und aus diesem zurück übertragen werden, sollte es zu einer Unterbrechung kommen. Dieser Vorgang ist fehleranfällig und wirkt sich negativ auf die Wiederaufnahme des primären Tasks aus [14, 50]. Bei Unterbrechungen ohne eigenen *Problem State* (beispielsweise bloßes Abwarten) tritt dieses Problem nicht ein, da der ursprüngliche Inhalt im Modul erhalten bleiben kann [9, 10]. *Richtlinie: Unterbrechungen ohne Problem State können jederzeit ausgeführt werden. Unterbrechungen mit Problem State sollten jedoch nicht in komplexen Situationen angezeigt werden.*

9:10 "Hey Siri, wo war ich stehen geblieben?"

■ Zeitpunkt der Unterbrechung

Nach Miyata und Norman kann eine Aufgabe mehrere Ebenen an Unteraufgaben besitzen [49]. Im Laufe der Ausführung durchlaufen diese Aufgaben wiederum sieben aufeinanderfolgende Phasen [51]. Studien, wie [1, 6, 24] haben gezeigt, dass die Unterbrechung von Aufgaben zwischen diesen Phasen eine geringere negative Auswirkung auf die Fortsetzung hat als die Unterbrechung innerhalb einer dieser Phasen. Dies ist mit der größeren Komplexität des jeweiligen *Problem States* während einer Phase zu begründen. Wird einem Menschen die Wahl des Zeitpunktes gelassen, so wartet er in der Regel selbstständig ab, bis ein Wendepunkt zwischen zwei Aufgaben erreicht ist [26, 61]. *Richtlinie: Benachrichtigungen sollten zwischen zwei (Unter-)Aufgaben präsentiert werden. Kann dies nicht garantiert werden, sollte der Nutzer die Möglichkeit erhalten, selbstständig die Entscheidung zu treffen, wann darauf reagiert wird.*

■ Verhältnis beider Aufgaben

Bezieht sich der *Problem State* einer unterbrechenden Aufgabe auf die primäre Aufgabe, so muss der State nicht ausgelagert werden und sowohl die Zeiten für Unterbrechung als auch Wiederaufnahme sind reduziert [13]. Es wurde des Weiteren festgestellt, dass auch eine Unterbrechung mit einem eigenen, jedoch ähnlichen, *Problem State* die spätere Wiederaufnahme der ursprünglichen Aufgabe fördert [3]. Dies ist unter anderem mit der regionalen Nähe beider Konzepte im deklarativen Gedächtnis zu begründen. Aufgrund der verteilten Aktivierung (*spreaded Activation*) im deklarativen Gedächtnis aktiviert der neue Kontext somit auch den ursprünglichen *Problem State*. *Richtlinie: Benachrichtigungen/Warnungen einer unterbrechenden Aufgabe sollten eine Relevanz zum Problem State des aktuell ausgeführten primären Tasks besitzen.*

■ Länge der unterbrechenden Aufgabe

Wie bereits zuvor erwähnt, zerfallen *Problem States* im deklarativen Gedächtnis über die Zeit hinweg. Die Wahrscheinlichkeit, dass dieser im Anschluss erfolgreich wieder aufgenommen werden kann, nimmt hierbei ebenfalls ab [50], sodass in vielen Fällen eine Rekonstruktion des ursprünglichen Kontexts notwendig ist [60]. *Richtlinie: Unterbrechende Aufgaben sollten in kurzer Zeit zu erfüllen sein und keine längerfristige Auseinandersetzung erfordern. Ist dies unvermeidbar, sollten entsprechende nachsorgende Vorkehrungen getroffen werden, um die Wiederaufnahme zu erleichtern (siehe Abschnitt 3.2)*

■ Länge der Benachrichtigung

Nach Trafton et al. kann die längere Anzeige eines Hinweis dazu beitragen, die Länge der Wiederaufnahme zu reduzieren [69]. Gemäß Kapitel 2.1.1 liegt dies an den Vorbereitungsprozessen, welche in dieser Zeitspanne

stattfinden. Diese können erwirken, dass der ursprüngliche *Problem State* bei der Wiederaufnahme nicht neu rekonstruiert werden muss, sondern zurückgerufen werden, und somit die Wiederaufnahme verkürzen kann. *Richtlinie: Ankündigungen bzw. Warnungen einer Unterbrechung sollten so lange neben dem primären Task angezeigt werden, bis der ursprüngliche Kontext im deklarativen Gedächtnis ausreichend aktiviert wurde. Kann dieser Zeitpunkt nicht bestimmt werden, sollte eine Spanne von (mindestens) acht Sekunden [63] gewählt oder dem Nutzer die Möglichkeit überlassen werden, der Warnung zu folgen.*

3.2 Kurativ

Sowohl das Zurückrufen, aber insbesondere auch das aktive Rekonstruieren einer bisherigen Aufgabe benötigen dem Abschnitt 2.1.2 nach eine Unterstützung durch externe Hinweise. Dies betrifft besonders die unvermeidbaren Unterbrechungen, da diese meist schon aufgrund ihrer Disruptivität eine Rekonstruktion benötigen, und nicht im Vorherein vermieden werden können [31, 68, 69]. Karlson et al. definierten ausgehend von dieser Notwendigkeit die folgenden zwei Ansätze [31]:

■ Erinnerung an eine Aufgabe

Durch das Platzieren von verwandten Informationen zu einer Aufgabe im Umfeld des Benutzers kann das Zurückrufen des *Problem States* initiiert werden [63]. Erfolgt dies schon während der sekundären Aufgabe, wird hierdurch zusätzlich noch der unterbewusste Rehearsal-Prozess unterstützt. Im Falle eines dennoch missglückten Recall-Prozesses kann hierdurch die Rekonstruktion eingeleitet werden, sodass die Aufgabe zumindest in keinem Fall vergessen wird. *Richtlinie: Schon im Verlauf der sekundären Aufgabe sollte nach Möglichkeit ein auffälliger Hinweis zur ursprünglichen Aufgabe mit ausreichender Diskrepanz [49] im Umfeld des Nutzers angezeigt werden.*

■ Assistierte Rehydration des ursprünglichen Task-Kontexts

Hierbei handelt es sich ebenfalls um eine Erinnerung, welche jedoch im Gegensatz zum vorherigen signalisierenden einen beschreibenden Ansatz verfolgt [49]. Eine derartige Unterstützung bei der Rekonstruktion des Ausgangszustandes kann unter Umständen zu einer schnelleren Wiederaufnahme der ursprünglichen Aufgabe führen [8]. Um jedoch einen erneuten Information Overload des Nutzers zu vermeiden, sollte diese Information nur dann eingesetzt werden, wenn sie einen realen Wunsch oder Nutzen erfüllt [2, 8]. Hierbei sollte ebenfalls darauf geachtet werden, lediglich den mentalen Kontext des Nutzers abzubilden, was jedoch von außen betrachtet nicht ohne weiteres möglich ist [9]. *Richtlinie: Im Falle eines zu scheiternden oder mental zu aufwendigen Recall-Prozesses sollten kon-*

9:12 "Hey Siri, wo war ich stehen geblieben?"

solidierte Informationen über den ursprünglichen Kontext der primären Aufgabe am Zeitpunkt der Unterbrechung bereitgestellt werden.

Da beide Strategien ursprünglich von Iqbal und Horvitz [26] für den stationären Desktop-Arbeitsplatz entwickelt wurden, stellt deren Umsetzung im mobilen Umfeld eine besondere Herausforderung dar. Klassische Desktopumgebungen bieten schon von Haus aus eine unterstützende Erinnerung an. Durch die überlagernde Anzeige von Fenstern werden stets Hinweise zu anderen Aufgaben gegeben [49] und in der Taskleiste dient das Icon einer Anwendung als persistenter visueller Hinweis [31], welcher wiederum zu einer dauerhaften verteilten Reaktivierung führt. Mobilien Umgebungen fehlen diese Hinweise, weshalb die Unterstützung hier umso bedeutsamer, aber auch kritischer ist. Es muss besonders darauf geachtet werden, dass die Erinnerungen auch tatsächlich als Unterstützung und nicht als weiterer disruptiver Faktor wahrgenommen werden. [35, 59].

4 Kostenreduzierung mittels Attention-Aware Interfaces

Mobile Benachrichtigungen werden in der Regel unmittelbar auf dem jeweiligen Endgerät angezeigt, obwohl häufig keine direkte Reaktion notwendig wäre [55]. Dies geht oftmals mit erhöhten Kosten, wie einer verzögerten Wiederaufnahme, einem erhöhten Stresslevel oder verminderter Performanz einher, da eine Unterbrechung verursacht wird [44]. Nachdem im vorherigen Kapitel bereits entsprechende Gestaltungsrichtlinien zur Kostenreduzierung präsentiert wurden, soll nun untersucht werden, wie diese tatsächlich im Alltag realisiert werden könnten. Hierzu wird das Forschungsfeld der "Attention-aware Interfaces" herangezogen, welches sich in der Vergangenheit speziell mit dieser Thematik beschäftigt hat. Nach einer initialen Definition derartiger Schnittstellen, werden im Anschluss konkrete präventive sowie kurative Einsatzmöglichkeiten dieser Technologie auf Basis der jeweiligen Gestaltungsrichtlinien diskutiert.

4.1 Definition & Funktionsweise

Attention-Aware Interfaces (oftmals auch als Attentive User Interfaces oder Attention Management System bezeichnet) zeichnen sich dadurch aus, dass sie auf Grundlage des aktuellen attentiven Zustandes des Nutzers den Informationsfluss eines Systems dahingehend zu steuern, dass Nutzer optimal bei der Ausführung ihrer aktuellen Aufgaben unterstützt werden und eine kognitive Überladung vermieden wird [57, 71, 72]. Hierfür werden gemäß Roda et al. Daten über den *aktuellen Fokus des Nutzers*, seine *aktuellen Ziele* sowie *Faktoren aus der Umgebung des Nutzers* benötigt [57]. Anderson et al. ordnen diese Daten noch einmal folgendermaßen ein [3]:

- **Physikalische Geräteparameter**
z. B. Beschleunigung, Orientierung
- **Interaktion mit dem Gerät**
z. B. Kamera- bzw. Mikrofonaufzeichnungen, Nähe zum Gerät, Nutzungsverhalten
- **Ortsbezogene Informationen**
z. B. Bewegungsverhalten
- **Zeitliche Informationen**
z. B. Zeitpunkt, Aktuelle Termine/Veranstaltungen
- **Physiologische Parameter**
z. B. EEG, EKG, Augenbewegung, Körpertemperatur, Hautleitfähigkeit

Auf Basis dieser Daten kann dann ein maschinelles Modell trainiert werden, welches wiederum prognostiziert, worauf die Aufmerksamkeit in der Vergangenheit, der Realität, sowie der Zukunft gerichtet sein wird [66, 72] und wie die begrenzte Ressource der Aufmerksamkeit optimal für den aktuellen Nutzungskontext eingesetzt werden kann [71]. Diese Information kann ebenfalls für die (implizite) Kommunikation mit anderen Menschen oder einem digitalen System verwendet werden [66].

4.2 Präventive Einsatzmöglichkeiten

Systeme, die sich mit der Steuerung vermeidbarer Unterbrechungen befassen, werden weitestgehend als Interruption bzw. Notification Management-Systeme bezeichnet [17, 73]. Da zu diesen Systemen bereits sehr viel Wissenschaft betrieben wurde [3, 47, 73], fokussieren die nun folgenden Abschnitte keine konkreten Projekte, sondern nur die Einhaltung der Gestaltungsrichtlinien aus Kapitel 3. Für den Einsatz von Interruption Management-Systemen existieren bereits diverse Strategien [17], welche basierend auf den Arbeiten von Anderson et al. [3] und Visuri et al. [73] in drei Kategorien unterteilt werden können. Obwohl diese zumeist unabhängig voneinander erforscht wurden, könnte auch eine Kaskadieren dieser Strategien wie folgend angedacht werden:

4.2.1 Verwerfen

Diese Strategie folgt keiner Gestaltungsrichtlinie, beruht aber auf dem natürlichen Umgang der Nutzer mit Notifications. Mithilfe individueller Präferenzmodelle versucht man hierbei zu prognostizieren, ob eine Benachrichtigung überhaupt relevant für den jeweiligen Nutzer ist. Insbesondere Mehrota et al. haben hierzu untersucht, wie das Akzeptanzverhalten von Nutzern auf bestimmte Benachrichtigungen gemessen und in Management-Strategien übersetzt werden könnte [45, 48]. Gewisse uninteressante oder redundante Fälle

können nach diesem Verfahren schon im Vornherein aussortiert werden, was wiederum die gesamte Last aller Benachrichtigungen reduziert.

4.2.2 Verschieben

Diese am häufigsten untersuchte [6, 53, 54] Strategie basiert auf der Erkenntnis, dass die disruptiven Kosten zwischen zwei Aufgaben am geringsten sind und wurde bereits mit der Richtlinie *Zeitpunkt der Unterbrechung* bedacht. Wurde eine Benachrichtigung mittels der vorherigen Strategie als relevant eingestuft, können Systeme mit dieser Strategie die Benachrichtigung bereits versetzen, bevor es der Nutzer bemerkt und somit die Kosten senken. Denkbar wäre an dieser Stelle jedoch auch die Integration weiterer präventiver Richtlinien.

- *Zeitpunkt & Verhältnis* Aufgrund des ständigen Bedürfnisses heutiger Nutzer nach neuen Informationen sollten Benachrichtigungen möglichst kurz verzögert werden [57]. Bis zu einem geeigneten Zeitpunkt können sich jedoch mehrere Benachrichtigungen angesammelt haben. Kombiniert man ein individuelles attentives Nutzermodell mit Informationen zur aktuell ausgeführten Aufgabe, so könnten verwandte Benachrichtigungen schon am nächst möglichen Zeitpunkt ausgelöst werden, selbst wenn sie chronologisch erst zu einem späteren Zeitpunkt eingetroffen sind. Die weiteren Benachrichtigungen würden solange reserviert werden, bis der Nutzer sie in seinen aktuellen Handlungskontext integrieren kann.
- *Zeitpunkt & Länge (Unterbrechung)* Analog zur vorherigen Überlegung könnten Unterbrechungen auch nach ihrer prognostizierten Länge geordnet und entsprechend angezeigt werden. Sehr simple Unterbrechungen, wie zum Beispiel eine kurze Mail ohne Handlungsauftrag könnten so auch schon während einer laufenden Aufgabe angezeigt werden.

4.2.3 Verlagern

Interruptibility Modelle werden mittlerweile immer häufiger an individuelle Nutzerpräferenzen angepasst [45]. Es könnten demnach Auslöser zu ungünstigen Momenten entstehen, welche jedoch aufgrund ihrer aktuellen Relevanz eigentlich zugelassen werden müssten. Um hierbei zumindest die Richtlinie *Länge Benachrichtigung* einzuhalten, können diese Benachrichtigungen alternativ ins ubiquitäre Umfeld und auf andere Ausgaberegionen verlagert werden, sodass die primäre Aufgabe in dieser Zeit noch weiter betrachtet werden kann. Anderson et al. unterscheiden hierbei zwischen Ansätzen, die entweder eine andere Ausgabemodalität verwenden, um beispielsweise den visuellen Sinn zu entlasten oder gar auf andere Geräte ausweichen [3]. Kubitza et al. sowie Mehrota et al. haben dieses Prinzip adaptiert und in ihrem Projekt eintreffende Benachrichtigungen an umliegende Displays umgeleitet [33, 46].

4.3 Kurative Einsatzmöglichkeiten

Während viele vermeidbare Unterbrechungen bereits durch die zuvor genannten Maßnahmen verhindert werden können, benötigen besonders die unvermeidbaren Unterbrechungen im heutigen Alltag entsprechende nachsorgende Unterstützung. Auch hierfür wurden in Kapitel 3.2 bereits entsprechende Richtlinien definiert, welche jedoch ähnlich wie die präventiven Strategien bisher nicht realisiert werden konnten, sodass häufig lediglich auf die entsprechende Notwendigkeit verwiesen werden konnte [26, 43, 70].

4.3.1 Erinnerung an eine Aufgabe

Im mobilen Umfeld ist eine simultane Anzeige aller Anwendungen wie im stationären Desktop-Umfeld aufgrund der begrenzten Anzeigefläche nicht möglich. Die Herausforderung besteht somit darin, einen geeigneten äquivalenten Kanal für solche Hinweise im mobilen Kontext zu identifizieren. Im Folgenden werden einige Möglichkeiten hierzu diskutiert.

Smartphone Notifications

Einige mobile Apps, beispielsweise aus dem Reisesegment, nutzen übliche Notifications, um Nutzer an das Fortsetzung eines Buchungsprozesses zu erinnern. Diese Art der Hinweise kann jedoch schnell zu noch disruptiveren Kosten führen, weswegen Attention Awareness in diesem Fall besonders die zuvor genannten präventiven Maßnahmen ermöglichen könnte, um einen wiederkehrenden Negativeffekt zu verhindern. Besonders die Kombination mit der präventiven Richtlinie *Ähnlichkeit* sollte hierbei beachtet werden, um den Hinweis vielmehr als Unterstützung anstatt einer Ablenkung wirken zu lassen.

Plattformübergreifende Hinweise

Apples mobiles Betriebssystem iOS zeigt mittlerweile bei einem App-Wechsel einen Link zur jeweils ursprünglichen App an. Obwohl diese Möglichkeit bislang von vielen Nutzern positiv angenommen wurde, stören sich einige an der Persistenz und Position dieses Hinweises. Da Attention Aware Interfaces oftmals Informationen über den umgebenden Kontext besitzen, könnte der Hinweis auch automatisch auf demjenigen Display in der Umgebung angezeigt werden, auf das der Nutzer nach Abschluss der Unterbrechung als nächstes seine Aufmerksamkeit richtet. Viele Aufgaben werden ohnehin auf mehrere Geräte aufgeteilt [12], sodass sich dieser Ansatz auch in Kombination mit einem stationären Arbeitsplatz anbietet. Mehrota et al. verfolgten diesen Ansatz bereits, jedoch ohne den Einsatz von Attention Awareness [46]. Da Smartphones heutzutage mit vielerlei Sensorik ausgestattet sind, könnte ein solcher Hinweis im öffentlichen Umfeld dann beispielsweise auch über das Scannen eines QR-Code oder einen Sprachbefehl aufgegriffen werden.

4.3.2 Assistierte Rehydratation des Task-Kontexts

Besonders im mobilen Umfeld sind Hinweise notwendig, die eine vergangene Aufgabe nicht nur nennen, sondern auch aktiv zur Rekonstruktion des dazugehörigen Kontexts beitragen. Auch diesbezüglich wurden bereits einige Konzepte für stationäre [18, 32, 39, 40, 58, 65, 67] und mobile Umgebungen [12, 22, 30, 41, 75] veröffentlicht, wobei häufig gewisse Parallelen erkennbar sind.

Visuelle Konsolidierung von Aufgaben

Im bisherigen Verlauf der Arbeit wurde bereits die vielschichtige hierarchische Unterteilung von Aufgaben in Unteraufgaben thematisiert. Dieses Prinzip hat sich unter anderem der Tasktracer von Dragunov et al. zu Nutze gemacht, indem er einzelne Fenster nicht der entsprechenden Anwendung, sondern nach dem Ziel, an welchem in diesen Fenstern gearbeitet wird, zusammen konsolidiert [18]. Der Bento-Browser von Hanh et al. adaptiert diese Idee auf mobile Browser-Tabs [22]. Dieser Ansatz ermöglicht, dass durch die virtuell getrennten Arbeitsbereiche jede konsolidierte Aufgabe als Trigger für die jeweils anderen dient. Somit wird bei jeder Betrachtung bzw. Durchführung einer Aufgabe stets das gemeinsame übergeordnete Ziel zurückgerufen bzw. rekonstruiert. Während beide Projekte bisher einer vorherigen händischen Konsolidierung bedürfen, könnten Attentive User Interfaces dies automatisch übernehmen, da das zugrundeliegende Modell ohnehin das individuelle Nutzerverhalten abbildet.

Anzeige des letzten Arbeitsstandes

Viele mobile oder stationäre Anwendungen behalten ihren Zustand (beispielsweise die vertikale Scrollposition oder Eingaben in einem Formular) bei, wenn diese geschlossen und wieder erneut geöffnet wird. Oftmals reicht der Anblick jedoch nicht aus, um sich wieder an den ursprünglichen *Problem State* zu erinnern. Dennoch hat eine Studie von Karlson et al. gezeigt, dass Smartphone Nutzer mittlerweile Bildschirmaufnahmen eines Arbeitsstandes als Selbsthilfe anfertigen, um sich anhand dieser wieder an die ursprünglichen Gedankengänge erinnern zu können [31]. Attentive User Interfaces könnten diese automatisch bereitstellen. GazeMarks von Kern et al. nutzt beispielsweise Blickdaten, um den Nutzer beim Fortsetzen einer Aufgabe auf den letzten Fokuspunkt hinzuweisen [32]. Auch dieser Ansatz wurde anschließend für das mobile Umfeld adaptiert [30, 41], jedoch mit speziellem Fokus auf das Lesen von Texten. Obwohl dieser Ansatz bereits Attention Awareness integriert, könnte er mit einer Wiederholung mehrerer Frames wie es Leiva et al. vorgeschlagen hatten, ergänzt werden [38]. Unabhängig von der Art der Darstellung sollte diese, gemäß der Richtlinie, nur dann eingesetzt werden, wenn der Nutzer sie tatsächlich benötigt. Attentive User Interfaces könnten hierbei erkennen, ob der Nutzer Inhalte bewusst sucht und dabei bisheriges Verhalten repliziert, um

erst dann den letzten Arbeitsstand zu visualisieren.

4.4 Kombinatorischer Ansatz

Die vorherigen zwei Abschnitte haben gezeigt, dass “Attentive User Interfaces” schon heute sowohl präventiv als auch kurativ für den Umgang mit Unterbrechungen eingesetzt werden können. Einen sicherlich noch größeren Mehrwert könnte diese Technologie jedoch durch Kombination dieser beiden Ansätze bringen. Schon heute nutzen Menschen alltägliche Hilfestellungen wie die berühmte Schleife um den Finger, um sich an eine spätere (Wiederaufnahme einer) Aufgabe zu erinnern. Dass dies oftmals zu abstrakt ist, erkannten Miyata und Norman schon 1986 und schlugen stattdessen sinnhaftigere Hinweise vor [49]. Ein vielversprechender Ansatz, welcher in der Zwischenzeit mehrfach wissenschaftlich belegt wurde [2, 7, 69], besteht in der Verwendung eines gleichen Signals bei Unterbrechung (als Priming [2]) und Wiederaufnahme. In beiden Fällen muss hierbei jedoch auf eine ausreichende Erkennbarkeit geachtet werden, welche wiederum nicht den jeweiligen sensorischen Kanal überladen darf [2]. Ein AUI könnte diese Voraussetzungen automatisch erfüllen, da es per Definition [66] schon beim Eintreffen einer Unterbrechung die folgenden Eigenschaften des Hinweises bestimmen könnte:

- **Form des Hinweises:** Welcher sensorische Kanal (visuell, auditiv, taktil etc.) wird sowohl während der Unterbrechung als auch nach der Wiederaufnahme verfügbar sein?
- **Inhalt des Hinweises:** Welche Information wird zum Zeitpunkt der Wiederaufnahme eindeutig sein und keine anderen Kontexte aktivieren?
- **Intensität des Hinweises:** Wie stark muss der Hinweis präsentiert werden, um ihn nicht verpassen zu können?

Ein Hinweis auf Basis dieser Daten könnte den Nutzer bestmöglich durch den sequentiellen Multitasking Prozess hinweg begleiten.

5 Fazit und Ausblick

Multitasking macht einen erheblichen Anteil unseres heutigen Alltags aus. Vor allem im mobilen Umfeld führen Unterbrechungen in ungünstigen Momenten auf kurze Sicht zu einer verzögerten Wiederaufnahme und auf lange Sicht zu erwiesenermaßen mehr Stress. In dieser Arbeit wurden theoretisch hergeleitete Gestaltungsrichtlinien für den Einsatz von Notifications sowie für die Unterstützung bei der Wiederaufnahme von mobilen Aufgaben dargelegt. Darauf aufbauend konnten Möglichkeiten aufgezeigt werden, wie das Forschungsfeld der Attentive User Interfaces diese Richtlinien adaptieren und zu weniger disruptiven Unterbrechungen führen könnte.

Für eine weitergehende Betrachtung dieser Thematik wäre es interessant, das Zusammenspiel von vor- und nachsorgenden Maßnahmen noch tiefergehender zu beleuchten sowie zu evaluieren. Insbesondere unter dem Gesichtspunkt neuer mobiler Trends wie Wearable oder Physiological Computing sollten neue Möglichkeiten für den Einsatz dieser Technologie betrachtet werden. Auch gänzlich neue Arbeitsmodelle wie das Microtasking [23, 27] könnten in diesem Zuge in einen neuen ganzheitlichen Ansatz integriert werden.

Da der Einsatz von Notification Management-Systemen für den breiten Markt jedoch ohnehin noch ein Novum darstellt, wäre es überhaupt einmal interessant, die längerfristigen Auswirkungen auf das Verhalten, besonders die hartnäckigen Angewohnheiten, und das Vertrauen der Nutzer zu beobachten. Schauen Nutzer auch mit intelligenten Notification Systemen noch regelmäßig nach neuen Nachrichten oder vertrauen sie darauf, dass ihr Smartphone sich zu gegebener Zeit schon bei Ihnen meldet? Neben der Beantwortung all dieser Fragen würde ich mir für die nahe Zukunft auch einen engeren Kontakt zwischen Industrie und Forschung wünschen, um derartige Lösung auch in der Allgemeinheit zugänglich zu machen. Letzten Endes hätte nämlich niemand etwas davon, wenn wir in Zukunft mehr Zeit für Unterbrechungen als unsere eigentliche Arbeit aufwenden.

Literatur

- 1 Piotr D. Adamczyk and Brian P. Bailey. If not now, when?: the effects of interruption at different moments within task execution. In *Proceedings of the 2004 conference on Human factors in computing systems - CHI '04*, pages 271–278, Vienna, Austria, 2004. ACM Press. URL: <http://portal.acm.org/citation.cfm?doid=985692.985727>, doi:10.1145/985692.985727.
- 2 E Altmann. Memory for goals: an activation-based model. *Cognitive Science*, 26(1):39–83, February 2002. URL: [http://doi.wiley.com/10.1016/S0364-0213\(01\)00058-1](http://doi.wiley.com/10.1016/S0364-0213(01)00058-1), doi:10.1016/S0364-0213(01)00058-1.
- 3 Christoph Anderson, Isabel Hübener, Ann-Kathrin Seipp, Sandra Ohly, Klaus David, and Veljko Pejovic. A Survey of Attention Management Systems in Ubiquitous Computing Environments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(2):1–27, July 2018. URL: <http://dl.acm.org/citation.cfm?doid=3236498.3214261>, doi:10.1145/3214261.
- 4 John R. Anderson. *How Can the Human Mind Occur in the Physical Universe?* Oxford University Press, October 2007. URL: <http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780195324259.001.0001/acprof-9780195324259>, doi:10.1093/acprof:oso/9780195324259.001.0001.
- 5 John R. Anderson, Daniel Bothell, Michael D. Byrne, Scott Douglass, Christian Lebiere, and Yulin Qin. An Integrated Theory of the Mind. *Psycholo-*

- gical Review*, 111(4):1036–1060, 2004. URL: <http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-295X.111.4.1036>, doi:10.1037/0033-295X.111.4.1036.
- 6 Brian P. Bailey and Joseph A. Konstan. On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Computers in Human Behavior*, 22(4):685–708, July 2006. URL: <https://linkinghub.elsevier.com/retrieve/pii/S074756320500107X>, doi:10.1016/j.chb.2005.12.009.
 - 7 Shadan Sadeghian Borojeni, Abdallah El Ali, Wilko Heuten, and Susanne Boll. Peripheral Light Cues for In-Vehicle Task Resumption. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction - NordiCHI '16*, pages 1–4, Gothenburg, Sweden, 2016. ACM Press. URL: <http://dl.acm.org/citation.cfm?doid=2971485.2971498>, doi:10.1145/2971485.2971498.
 - 8 Jelmer P. Borst, Trudy A. Buwalda, Hedderik van Rijn, and Niels A. Taatgen. Avoiding the problem state bottleneck by strategic use of the environment. *Acta Psychologica*, 144(2):373–379, October 2013. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0001691813001789>, doi:10.1016/j.actpsy.2013.07.016.
 - 9 Jelmer P. Borst, Niels A. Taatgen, and Hedderik van Rijn. What Makes Interruptions Disruptive?: A Process-Model Account of the Effects of the Problem State Bottleneck on Task Interruption and Resumption. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, pages 2971–2980, Seoul, Republic of Korea, 2015. ACM Press. URL: <http://dl.acm.org/citation.cfm?doid=2702123.2702156>, doi:10.1145/2702123.2702156.
 - 10 JP Borst, NA Taatgen, and H Van Rijn. Problem representations in multitasking: An additional cognitive bottleneck. In *Proceedings of the 9th international conference on cognitive modeling. Manchester: University of Manchester*, 2009.
 - 11 Matthias Böhmer, Christian Lander, Sven Gehring, Duncan P. Brumby, and Antonio Krüger. Interrupted by a phone call: exploring designs for lowering the impact of call notifications for smartphone users. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, pages 3045–3054, Toronto, Ontario, Canada, 2014. ACM Press. URL: <http://dl.acm.org/citation.cfm?doid=2556288.2557066>, doi:10.1145/2556288.2557066.
 - 12 Shiwei Cheng, Jing Fan, and Anind K. Dey. Smooth Gaze: a framework for recovering tasks across devices using eye tracking. *Personal and Ubiquitous Computing*, 22(3):489–501, June 2018. URL: <http://link.springer.com/10.1007/s00779-018-1115-8>, doi:10.1007/s00779-018-1115-8.
 - 13 Edward Cutrell, Mary Czerwinski, and Eric Horvitz. Notification, Disruption, and Memory: Effects of Messaging Interruptions on Memory and Per-

- formance. *Disruption, and Memory: Effects of Messaging Interruptions on Memory and Performance*. (12 2000), 2000.
- 14 Edward B Cutrell, Mary Czerwinski, and Eric Horvitz. Effects of instant messaging interruptions on computing tasks. In *CHI'00 extended abstracts on Human factors in computing systems*, pages 99–100. ACM, 2000.
 - 15 Mary Czerwinski, Eric Horvitz, and Susan Wilhite. A diary study of task switching and interruptions. In *Proceedings of the 2004 conference on Human factors in computing systems - CHI '04*, pages 175–182, Vienna, Austria, 2004. ACM Press. URL: <http://portal.acm.org/citation.cfm?doid=985692.985715>, doi:10.1145/985692.985715.
 - 16 Laura Dabbish, Gloria Mark, and Víctor M. González. Why do i keep interrupting myself?: environment, habit and self-interruption. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*, page 3127, Vancouver, BC, Canada, 2011. ACM Press. URL: <http://dl.acm.org/citation.cfm?doid=1978942.1979405>, doi:10.1145/1978942.1979405.
 - 17 Saber Darmoul, Ali Ahmad, Mageed Ghaleb, and Mohammed Alkahtani. Interruption Management in Human Multitasking Environments. *IFAC-PapersOnLine*, 48(3):1179–1185, 2015. URL: <https://linkinghub.elsevier.com/retrieve/pii/S2405896315004838>, doi:10.1016/j.ifacol.2015.06.244.
 - 18 Anton N. Dragunov, Thomas G. Dietterich, Kevin Johnsrude, Matthew McLaughlin, Lida Li, and Jonathan L. Herlocker. TaskTracer: a desktop environment to support multi-tasking knowledge workers. In *Proceedings of the 10th international conference on Intelligent user interfaces - IUI '05*, page 75, San Diego, California, USA, 2005. ACM Press. URL: <http://portal.acm.org/citation.cfm?doid=1040830.1040855>, doi:10.1145/1040830.1040855.
 - 19 Nicholas Fitz, Kostadin Kushlev, Ranjan Jagannathan, Terrel Lewis, Devang Paliwal, and Dan Ariely. Batching smartphone notifications can improve well-being. *Computers in Human Behavior*, 101:84–94, December 2019. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0747563219302596>, doi:10.1016/j.chb.2019.07.016.
 - 20 Keaton A. Fletcher, Sean M. Potter, and Britany N. Telford. Stress Outcomes of Four Types of Perceived Interruptions. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 60(2):222–235, March 2018. URL: <http://journals.sagepub.com/doi/10.1177/0018720817738845>, doi:10.1177/0018720817738845.
 - 21 Tony Gillie and Donald Broadbent. What makes interruptions disruptive? A study of length, similarity, and complexity. *Psychological Research*, 50(4):243–250, April 1989. URL: <http://link.springer.com/10.1007/BF00309260>, doi:10.1007/BF00309260.

- 22 Nathan Hahn, Joseph Chee Chang, and Aniket Kittur. Bento Browser: Complex Mobile Search Without Tabs. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, pages 1–12, Montreal QC, Canada, 2018. ACM Press. URL: <http://dl.acm.org/citation.cfm?doid=3173574.3173825>, doi:10.1145/3173574.3173825.
- 23 Nathan Hahn, Shamsi T. Iqbal, and Jaime Teevan. Casual Microtasking: Embedding Microtasks in Facebook. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, pages 1–9, Glasgow, Scotland Uk, 2019. ACM Press. URL: <http://dl.acm.org/citation.cfm?doid=3290605.3300249>, doi:10.1145/3290605.3300249.
- 24 Shamsi T. Iqbal and Brian P. Bailey. Investigating the effectiveness of mental workload as a predictor of opportune moments for interruption. In *CHI '05 extended abstracts on Human factors in computing systems - CHI '05*, page 1489, Portland, OR, USA, 2005. ACM Press. URL: <http://portal.acm.org/citation.cfm?doid=1056808.1056948>, doi:10.1145/1056808.1056948.
- 25 Shamsi T. Iqbal and Brian P. Bailey. Leveraging characteristics of task structure to predict the cost of interruption. In *Proceedings of the SIGCHI conference on Human Factors in computing systems - CHI '06*, page 741, Montrééal, Québec, Canada, 2006. ACM Press. URL: <http://portal.acm.org/citation.cfm?doid=1124772.1124882>, doi:10.1145/1124772.1124882.
- 26 Shamsi T. Iqbal and Eric Horvitz. Disruption and recovery of computing tasks: field study, analysis, and directions. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07*, page 677, San Jose, California, USA, 2007. ACM Press. URL: <http://portal.acm.org/citation.cfm?doid=1240624.1240730>, doi:10.1145/1240624.1240730.
- 27 Shamsi T. Iqbal, Jaime Teevan, Dan Liebling, and Anne Loomis Thompson. Multitasking with Play Write, a Mobile Microproductivity Writing Tool. In *The 31st Annual ACM Symposium on User Interface Software and Technology - UIST '18*, pages 411–422, Berlin, Germany, 2018. ACM Press. URL: <http://dl.acm.org/citation.cfm?doid=3242587.3242611>, doi:10.1145/3242587.3242611.
- 28 Quintus R Jett. Work Interrupted: A Closer Look at the Role of Interruptions in Organizational Life. *Academy of Management Review*, page 14, 2003.
- 29 Jing Jin and Laura A. Dabbish. Self-interruption on the computer: a typology of discretionary task interleaving. In *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*, page 1799, Boston, MA, USA, 2009. ACM Press. URL: <http://dl.acm.org/citation.cfm?doid=1518701.1518979>, doi:10.1145/1518701.1518979.
- 30 Jaemin Jo, Bohyoung Kim, and Jinwook Seo. EyeBookmark: Assisting Recovery from Interruption during Reading. In *Proceedings of the*

- 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, pages 2963–2966, Seoul, Republic of Korea, 2015. ACM Press. URL: <http://dl.acm.org/citation.cfm?doid=2702123.2702340>, doi:10.1145/2702123.2702340.
- 31 Amy K. Karlson, Shamsi T. Iqbal, Brian Meyers, Gonzalo Ramos, Kathy Lee, and John C. Tang. Mobile taskflow in context: a screenshot study of smartphone usage. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*, page 2009, Atlanta, Georgia, USA, 2010. ACM Press. URL: <http://portal.acm.org/citation.cfm?doid=1753326.1753631>, doi:10.1145/1753326.1753631.
 - 32 Dagmar Kern, Paul Marshall, and Albrecht Schmidt. Gazemarks: gaze-based visual placeholders to ease attention switching. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*, page 2093, Atlanta, Georgia, USA, 2010. ACM Press. URL: <http://portal.acm.org/citation.cfm?doid=1753326.1753646>, doi:10.1145/1753326.1753646.
 - 33 Thomas Kubitzka, Alexandra Voit, Dominik Weber, and Albrecht Schmidt. An IoT infrastructure for ubiquitous notifications in intelligent living environments. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct - UbiComp '16*, pages 1536–1541, Heidelberg, Germany, 2016. ACM Press. URL: <http://dl.acm.org/citation.cfm?doid=2968219.2968545>, doi:10.1145/2968219.2968545.
 - 34 Kostadin Kushlev, Jason Proulx, and Elizabeth W. Dunn. "Silence Your Phones": Smartphone Notifications Increase Inattention and Hyperactivity Symptoms. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, pages 1011–1020, Santa Clara, California, USA, 2016. ACM Press. URL: <http://dl.acm.org/citation.cfm?doid=2858036.2858359>, doi:10.1145/2858036.2858359.
 - 35 Katherine Labonté, Sébastien Tremblay, and François Vachon. Effects of a Warning on Interruption Recovery in Dynamic Settings. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 60(1):1304–1308, September 2016. URL: <http://journals.sagepub.com/doi/10.1177/1541931213601302>, doi:10.1177/1541931213601302.
 - 36 Byung Lee, Kwanghun Chung, and Sung-Hee Kim. Interruption Cost Evaluation by Cognitive Workload and Task Performance in Interruption Coordination Modes for Human–Computer Interaction Tasks. *Applied Sciences*, 8(10):1780, September 2018. URL: <http://www.mdpi.com/2076-3417/8/10/1780>, doi:10.3390/app8101780.
 - 37 Yu-Kang Lee, Chun-Tuan Chang, You Lin, and Zhao-Hong Cheng. The dark side of smartphone usage: Psychological traits, compulsive behavior and technostress. *Computers in Human Behavior*, 31:373–383, February 2014. URL: <https://linkinghub.elsevier.com/retrieve/pii/S074756321300397X>, doi:10.1016/j.chb.2013.10.047.

- 38 Luis Leiva, Matthias Böhmer, Sven Gehring, and Antonio Krüger. Back to the app: the costs of mobile application interruptions. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services - MobileHCI '12*, page 291, San Francisco, California, USA, 2012. ACM Press. URL: <http://dl.acm.org/citation.cfm?doid=2371574.2371617>, doi:10.1145/2371574.2371617.
- 39 Luis A. Leiva. MouseHints: easing task switching in parallel browsing. In *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems - CHI EA '11*, page 1957, Vancouver, BC, Canada, 2011. ACM Press. URL: <http://portal.acm.org/citation.cfm?doid=1979742.1979861>, doi:10.1145/1979742.1979861.
- 40 Yikun Liu and Mark S. Pfaff. Wiredin: using visual feedback to support task resumption. In *Proceedings of the extended abstracts of the 32nd annual ACM conference on Human factors in computing systems - CHI EA '14*, pages 2539–2544, Toronto, Ontario, Canada, 2014. ACM Press. URL: <http://dl.acm.org/citation.cfm?doid=2559206.2581264>, doi:10.1145/2559206.2581264.
- 41 Alexander Mariakakis, Mayank Goel, Md Tanvir Islam Aumi, Shwetak N. Patel, and Jacob O. Wobbrock. SwitchBack: Using Focus and Saccade Tracking to Guide Users' Attention for Mobile Task Resumption. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, pages 2953–2962, Seoul, Republic of Korea, 2015. ACM Press. URL: <http://dl.acm.org/citation.cfm?doid=2702123.2702539>, doi:10.1145/2702123.2702539.
- 42 Gloria Mark, Mary Czerwinski, and Shamsi T. Iqbal. Effects of Individual Differences in Blocking Workplace Distractions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, pages 1–12, Montreal QC, Canada, 2018. ACM Press. URL: <http://dl.acm.org/citation.cfm?doid=3173574.3173666>, doi:10.1145/3173574.3173666.
- 43 Gloria Mark, Victor M. Gonzalez, and Justin Harris. No task left behind?: examining the nature of fragmented work. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '05*, page 321, Portland, Oregon, USA, 2005. ACM Press. URL: <http://portal.acm.org/citation.cfm?doid=1054972.1055017>, doi:10.1145/1054972.1055017.
- 44 Gloria Mark, Daniela Gudith, and Ulrich Klocke. The cost of interrupted work: more speed and stress. In *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*, page 107, Florence, Italy, 2008. ACM Press. URL: <http://portal.acm.org/citation.cfm?doid=1357054.1357072>, doi:10.1145/1357054.1357072.
- 45 Abhinav Mehrotra, Robert Hendley, and Mirco Musolesi. PrefMiner: mining user's preferences for intelligent mobile notification management. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '16*, pages 1223–1234, Heidelberg, Ger-

- many, 2016. ACM Press. URL: <http://dl.acm.org/citation.cfm?doid=2971648.2971747>, doi:10.1145/2971648.2971747.
- 46 Abhinav Mehrotra, Robert Hendley, and Mirco Musolesi. NotifyMeHere: Intelligent Notification Delivery in Multi-Device Environments. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval - CHIIR '19*, pages 103–111, Glasgow, Scotland UK, 2019. ACM Press. URL: <http://dl.acm.org/citation.cfm?doid=3295750.3298932>, doi:10.1145/3295750.3298932.
 - 47 Abhinav Mehrotra and Mirco Musolesi. Intelligent Notification Systems: A Survey of the State of the Art and Research Challenges. *CoRR*, abs/1711.10171, 2017. URL: <http://arxiv.org/abs/1711.10171>.
 - 48 Abhinav Mehrotra, Mirco Musolesi, Robert Hendley, and Veljko Pejovic. Designing content-driven intelligent notification mechanisms for mobile applications. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15*, pages 813–824, Osaka, Japan, 2015. ACM Press. URL: <http://dl.acm.org/citation.cfm?doid=2750858.2807544>, doi:10.1145/2750858.2807544.
 - 49 Yoshiro Miyata and Donald A Norman. Psychological issues in support of multiple activities. *User centered system design: New perspectives on human-computer interaction*, pages 265–284, 1986.
 - 50 Christopher A. Monk, J. Gregory Trafton, and Deborah A. Boehm-Davis. The effect of interruption duration and demand on resuming suspended goals. *Journal of Experimental Psychology: Applied*, 14(4):299–313, 2008. URL: <http://doi.apa.org/getdoi.cfm?doi=10.1037/a0014402>, doi:10.1037/a0014402.
 - 51 Donald A. Norman. The seven Stages of Action. In *The design of everyday things*, pages 40–44. Basic Books, New York, New York, revised and expanded edition edition, 2013.
 - 52 Tadashi Okoshi, Hiroki Nozaki, Jin Nakazawa, Hideyuki Tokuda, Julian Ramos, and Anind K. Dey. Towards attention-aware adaptive notification on smart phones. *Pervasive and Mobile Computing*, 26:17–34, February 2016. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1574119215001881>, doi:10.1016/j.pmcj.2015.10.004.
 - 53 Nia Peters. Interruption timing prediction via prosodic task boundary model for human-machine teaming. In *Future of Information and Communication Conference*, pages 501–522. Springer, 2019.
 - 54 Martin Pielot, Bruno Cardoso, Kleomenis Katevas, Joan Serrà, Aleksandar Matic, and Nuria Oliver. Beyond Interruptibility: Predicting Opportune Moments to Engage Mobile Phone Users. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(3):1–25, September 2017. URL: <http://dl.acm.org/citation.cfm?doid=3139486.3130956>, doi:10.1145/3130956.

- 55 Martin Pielot, Karen Church, and Rodrigo de Oliveira. An in-situ study of mobile phone notifications. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services - MobileHCI '14*, pages 233–242, Toronto, ON, Canada, 2014. ACM Press. URL: <http://dl.acm.org/citation.cfm?doid=2628363.2628364>, doi:10.1145/2628363.2628364.
- 56 Martin Pielot and Luz Rello. The Do Not Disturb Challenge: A Day Without Notifications. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '15*, pages 1761–1766, Seoul, Republic of Korea, 2015. ACM Press. URL: <http://dl.acm.org/citation.cfm?doid=2702613.2732704>, doi:10.1145/2702613.2732704.
- 57 Claudia Roda and Julie Thomas. Attention aware systems: Theories, applications, and research agenda. *Computers in Human Behavior*, 22(4):557–587, July 2006. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0747563205001135>, doi:10.1016/j.chb.2005.12.005.
- 58 Adam Rule, Aurélien Tabard, Karen Boyd, and Jim Hollan. Restoring the Context of Interrupted Work with Desktop Thumbnails. In *37th Annual Meeting of the Cognitive Science Society*, pages 1–6, Pasadena, United States, July 2015. Cognitive Science Society. URL: <https://hal.archives-ouvertes.fr/hal-01213708>.
- 59 Alireza Sahami Shirazi, Niels Henze, Tilman Dingler, Martin Pielot, Dominik Weber, and Albrecht Schmidt. Large-scale assessment of mobile notifications. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, pages 3055–3064, Toronto, Ontario, Canada, 2014. ACM Press. URL: <http://dl.acm.org/citation.cfm?doid=2556288.2557189>, doi:10.1145/2556288.2557189.
- 60 Dario D. Salvucci. On reconstruction of task context after interruption. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*, page 89, Atlanta, Georgia, USA, 2010. ACM Press. URL: <http://portal.acm.org/citation.cfm?doid=1753326.1753341>, doi:10.1145/1753326.1753341.
- 61 Dario D. Salvucci and Peter Bogunovich. Multitasking and monotasking: the effects of mental workload on deferred task interruptions. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*, page 85, Atlanta, Georgia, USA, 2010. ACM Press. URL: <http://portal.acm.org/citation.cfm?doid=1753326.1753340>, doi:10.1145/1753326.1753340.
- 62 Dario D. Salvucci and Niels A. Taatgen. Threaded cognition: An integrated theory of concurrent multitasking. *Psychological Review*, 115(1):101–130, 2008. URL: <http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-295X.115.1.101>, doi:10.1037/0033-295X.115.1.101.

- 63 Dario D. Salvucci and Niels A. Taatgen. Sequential Multitasking, Problem State, and Task Suspension and Resumption. In *The multitasking mind*, Oxford series on cognitive models and architectures, pages 111–140. Oxford University Press, New York, 2011. OCLC: ocn607986732.
- 64 Dario D. Salvucci, Niels A. Taatgen, and Jelmer P. Borst. Toward a unified theory of the multitasking continuum: from concurrent performance to task switching, interruption, and resumption. In *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*, page 1819, Boston, MA, USA, 2009. ACM Press. URL: <http://dl.acm.org/citation.cfm?doid=1518701.1518981>, doi:10.1145/1518701.1518981.
- 65 Stacey D. Scott, Stephane Mercier, M. L. Cummings, and Enlie Wang. Assisting Interruption Recovery in Supervisory Control of Multiple Uavs. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(5):699–703, October 2006. URL: <http://journals.sagepub.com/doi/10.1177/154193120605000518>, doi:10.1177/154193120605000518.
- 66 Jeffrey S. Shell, Ted Selker, and Roel Vertegaal. Interacting with groups of computers. *Communications of the ACM*, 46(3):40, March 2003. URL: <http://portal.acm.org/citation.cfm?doid=636772.636796>, doi:10.1145/636772.636796.
- 67 Greg Smith, Patrick Baudisch, George Robertson, Mary Czerwinski, Brian Meyers, Dan Robbins, Eric Horvitz, and Donna Andrews. GroupBar: The TaskBar Evolved. In *Proceedings of OZCHI 2003*, pages 34–43, 2003.
- 68 J. Gregory Trafton, Erik M. Altmann, and Derek P. Brock. Huh, what was I Doing? How People Use Environmental Cues after an Interruption. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 49(3):468–472, September 2005. URL: <http://journals.sagepub.com/doi/10.1177/154193120504900354>, doi:10.1177/154193120504900354.
- 69 J. Gregory Trafton, Erik M Altmann, Derek P Brock, and Farilee E Mintz. Preparing to resume an interrupted task: effects of prospective goal encoding and retrospective rehearsal. *International Journal of Human-Computer Studies*, 58(5):583–603, May 2003. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1071581903000235>, doi:10.1016/S1071-5819(03)00023-5.
- 70 Johanne R. Trippas, Flora D. Salim, Mark Sanderson, Damiano Spina, Falk Scholer, Ahmed Hassan Awadallah, Peter Bailey, Paul N. Bennett, Ryen W. White, Jonathan Liono, and Yongli Ren. Learning About Work Tasks to Inform Intelligent Assistant Design. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval - CHIIR '19*, pages 5–14, Glasgow, Scotland UK, 2019. ACM Press. URL: <http://dl.acm.org/citation.cfm?doid=3295750.3298934>, doi:10.1145/3295750.3298934.
- 71 Vertegaal and Ivan Poupyrev. Attentive User Interfaces. *Editorial, Special Issue on Attentive User Interfaces, Communications of ACM*, 2003.

- 72 Roel Vertegaal, Jeffrey S. Shell, Daniel Chen, and Aadil Mamuji. Designing for augmented attention: Towards a framework for attentive user interfaces. *Computers in Human Behavior*, 22(4):771–789, July 2006. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0747563205001160>, doi:10.1016/j.chb.2005.12.012.
- 73 Aku Visuri, Niels van Berkel, Tadashi Okoshi, Jorge Goncalves, and Vasilis Kostakos. Understanding smartphone notifications’ user interactions and content importance. *International Journal of Human-Computer Studies*, 128:72–85, August 2019. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1071581919300205>, doi:10.1016/j.ijhcs.2019.03.001.
- 74 Maria Wirzberger and Nele Russwinkel. Modeling Interruption and Resumption in a Smartphone Task: An ACT-R Approach. *i-com*, 14(2), January 2015. URL: <https://www.degruyter.com/view/j/icom.2015.14.issue-2/icom-2015-0033/icom-2015-0033.xml>, doi:10.1515/icom-2015-0033.
- 75 W.L. Yeung and Simon Y.W. Li. Prototyping the Machine-Human Dialogues in a Smartphone Voice Call Application With Task Resumption Support. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '16*, pages 1788–1793, Santa Clara, California, USA, 2016. ACM Press. URL: <http://dl.acm.org/citation.cfm?doid=2851581.2892464>, doi:10.1145/2851581.2892464.

Public AR Displays: Herausforderungen und Möglichkeiten

Helena Bayerl

Ludwig-Maximilians-Universität München, München, Deutschland
h.bayerl@campus.lmu.de

Zusammenfassung

Digitale öffentliche Displays haben aufgrund ihrer Lage das Potenzial, Informationen und Botschaften an große Personengruppen weiterzugeben. Jedoch verfehlen sie aus unterschiedlichen Gründen ihr Ziel. Diese wissenschaftliche Arbeit befasst sich mit den verschiedenen Problemen, die diese öffentlichen Bildschirme mit sich bringen und wie sie sich auf User hinsichtlich ihres Verhaltens auswirken. Es wird anhand von verschiedenen Theorien erklärt, wie Aufmerksamkeit gesteigert werden kann und wie Passanten zur Interaktion motiviert werden können. Außerdem werden verschiedene Prinzipien zur Gestaltung von öffentlichen Bildschirmen vorgestellt. Anschließend wird diskutiert, wie mithilfe von Augmented Reality Technologien den Problemen von öffentlichen Displays entgegengetreten werden kann.

2012 ACM Computing Classification Human-centered computing → Displays and imagers

Keywords and phrases Öffentliche Displays; Augmented Reality; Interaktion; Benutzerschnittstelle.

1 Einleitung

Auf dem Weg zum Zug, im Einkaufszentrum oder an Bushaltestellen begegnen sie uns überall, fallen uns jedoch nicht immer auf - öffentliche Displays. Früher gab es nur Litfaßsäulen, welche mit Papierplakaten mühselig beklebt wurden. Heutzutage besitzen sie Bildschirme mit einem echten Computersystem dahinter. Sie zeigen ununterbrochen die neuesten Nachrichten, Werbung oder Entertainmentinhalte an und haben sich vollständig in das alltägliche Leben integriert. Doch die meisten Menschen gehen an diesen Displays vorbei ohne sie überhaupt zu beachten, obwohl mit manchen sogar interagiert werden kann. Ihr mögliches Potenzial wird von den Betreibern somit nicht komplett ausgenutzt und es kann nicht auf die unterschiedlichen Bedürfnisse der Passanten eingegangen werden. Mithilfe von Augmented Reality und einem persönlichen



© Helena Bayerl;

licensed under Creative Commons License CC-BY

Cite as: Helena Bayerl. Public AR Displays: Herausforderungen und Möglichkeiten. In *3rd Seminar on Ubiquitous Interaction (UBIACTION 2019)*. Editors: Pascal Knierim, Jakob Karolus, Fiona Draxler, Ville Mäkelä, Tonja Machulla, Florian Lang, Thomas Kosch, Albrecht Schmidt. July 16, 2019. Munich, Germany. pp. 10:1–10:18.

10:2 Public AR Displays: Herausforderungen und Möglichkeiten

Gerät kann hier die User Experience verbessert und für Nutzer attraktiv gestaltet werden.

2 Motivation

Die Digitalisierung findet überall und besonders im öffentlichen Raum statt. Einkaufszentren, Flughäfen und Bahnhöfe erleben einen Wandel von traditionellen Werbeanzeigen auf Papier zur digitalen Anzeige, die eine ganz neue Form der Multimediapräsentationen und auch neue Benutzererfahrungen bietet. Öffentliche Displays werden heutzutage mit vielen verschiedenen Informationen angereichert. Seien es die täglichen Neuigkeiten aus der Umgebung, Entertainment oder Werbeanzeigen. Letztendlich entspricht der Inhalt nicht dem, was die Passanten sehen möchten.

Doch was möchten die Menschen sehen? Der eine interessiert sich für die Sportnachrichten, ein anderer möchte den neuesten Trailer seines Lieblingsfilms sehen. Somit haben alle Menschen unterschiedliche Interessen, weswegen ein öffentliches Display im besten Fall all diesen Interessensgebieten entspricht. Mithilfe von Augmented Reality können öffentliche Displays mit Informationen erweitert werden und personalisierten Inhalt anzeigen.

3 Grundlegende Probleme mit öffentlichen Bildschirmen

Öffentliche Displays werden von den meisten Passanten ignoriert oder gar nicht erst beachtet, da sie meistens uninteressanten Inhalt anzeigen, die Personen auf dem Weg zum nächsten Termin sind oder weil es gerade anfängt zu regnen. In diesem Abschnitt wird erläutert was die Gründe für die Ignoranz von öffentlichen Displays sind.

3.1 Interaction Blindness

Eines der größten Probleme ist die Interaction Blindness. Interaction Blindness bedeutet, dass Menschen mit Displays oder Hotspots nicht interagieren, weil sie nicht wissen, welche Interaktionsmöglichkeiten diese Systeme bieten. Für Passanten wirkt ein öffentlicher Bildschirm wie eine normale Anzeige, da sie es nicht gewohnt sind in einer öffentlichen Situation zu interagieren. Und somit verfehlt das installierte Projekt sein Ziel. [23]

3.2 Display Blindness

Doch ebenso können Displays auch ganz übersehen werden. Das Ignorieren oder das Übersehen von Systemen, öffentlichen Bildschirmen oder anderen

Installationen nennt man Display Blindness. Dieser Effekt kann ebenso auftreten, wenn Passanten uninteressanten Kontent auf einem Display erwarten und dieser von Anfang an ignoriert wird. [20]

3.3 Informationsüberflutung

Ebenso kann eine Display Blindness durch Informationsüberflutung stattfinden. Informationsüberflutung tritt oft an öffentlichen Orten auf. Es beschreibt die Situation, in der zu viele Informationen auf eine Person einwirken und diese nicht in der Lage ist, die Informationsmenge zu erkennen, zu verstehen oder damit umzugehen. Die Person hat Probleme wichtige Informationen zu identifizieren und selektiert oder ignoriert große Informationsmengen. [18]

3.4 Wetter

Das Wetter ist unbeständig, unvorhersehbar und kann die Innen- sowie Außeninstallation beeinflussen.

Sonnenlicht kann die Sicht auf den Display enorm einschränken. Aber auch die Sensorperformance wird dadurch beeinträchtigt. In dem Test wurde eine Kinect in einem Zelt verwendet. Falsch gerenderte Bilder entstanden durch die starke Sonneneinstrahlung. Hierbei musste ein schwarzer Vorhang aufgehungen werden, um die Einstrahlungen zu minimieren. Regen und Feuchtigkeit können die Hardware zerstören. Es sollte stets drauf geachtet werden, dass die Systeme wasserdicht installiert sind.

Wetterbedingungen beeinflussen ebenso Menschen. Bei schönem, sonnigem Wetter können die Interaktionen ansteigen. Kaltes, windiges Wetter lässt die Menschenanzahl draußen sinken und somit auch den Willen zur Interaktion. [19]

3.5 Öffentliche Umgebungen

Events, wie Festivals oder Vorträge, können ebenso die Benutzung von öffentlichen Bildschirmen beeinflussen. Es kann vorkommen, dass tagsüber die Passanten hauptsächlich Studenten sind und abends Besucher eines nahegelegenen Theaters. Bei der Installation von Displays sollte darauf geachtet werden den richtigen Platz zu finden, da Veranstaltungen die soziale Umgebung maßgeblich beeinflussen. Vor und nach solchen Veranstaltungen sind nahe gelegene Plätze stark besucht. Wenn ein interaktives System in der Nähe von solchen Veranstaltungen installiert ist, versuchen manche Menschen möglicherweise mit einem Bildschirm zu interagieren, andere jedoch fühlen sich bei solch großen Menschenmengen unwohl, da sie das Gefühl haben im Weg zu stehen.

Es muss immer das Ausmaß und die Art der bevorstehenden Ereignisse in einer Region beachtet werden. Events können das soziale Umfeld, die

10:4 Public AR Displays: Herausforderungen und Möglichkeiten

Demografie und die Anzahl der Personen verändern. Aber auch der Verkehr oder eine naturbelassene Gegend, wie Wald oder Wiesen, kann die Elektronik oder die Netzwerkverbindung negativ beeinflussen oder stören.

Ebenso müssen verschiedene Aspekte bei Innenrauminstallationen beachtet werden. Das Beachten von den physikalischen Eigenschaften des Raums, z. B. die Anordnung, die Türen, die Säulen, die Aufzüge und die Maschinen kann beim Identifizieren helfen, wie sich diese Eigenschaften auf den Personenfluss und die Raumnutzung auswirken.

Die Einsatzorte sowie die umliegenden Gebiete sind in der Regel mit einer Vielzahl von Personen mit jeweils unterschiedlichen Rollen besetzt. Mäkela et al. installierten die Information Wall am Campus einer Universität [19]. Abends hat der Nachtwächter die Projektoren ausgeschaltet, ohne das Wissen der Installateure.

In Universitäten bauen Studentenorganisationen gelegentlich Stände auf, um Eintrittskarten für Partys zu verkaufen oder um Mitglieder anzuwerben. Dabei kann es vorkommen, dass die Stände die Displayinteraktion behindern. Möbel oder anderes Equipment werden manchmal ebenso vor öffentlichen Displays abgestellt.

Beobachten und Identifizieren von verschiedenen Bevölkerungsgruppen und deren Routinen und wie diese die Installation beeinträchtigen können, kann helfen diese zu verbessern.

Vandalismus ist ein ernstzunehmendes Thema und ist der vorsätzliche Eingriff in die Installation und die Zerstörung von Equipment. Berichten zufolge wurde schon Sicherheitsglas von Installationen zerstört, sodass das System für mehrere Wochen ausgeschaltet war und teure Reparaturen veranlasst werden mussten. Müller et al. berichteten von Usern, die expressive Interaktionen durchführten, sodass die Forscher sogar eingreifen mussten, damit keine Beschädigungen am Display entstanden [17].

Es kann helfen, wenn man sich in das Gedankenset eines Vandalen hinversetzt. Es sollten einfache und offensichtliche Ziele von Vandalismus identifiziert und angepasst werden. Einfache Ziele, wie Kabel, sollten versteckt werden. [19]

4 Wie erhält man die Aufmerksamkeit von Passanten?

Müller et al.[20] beschreibt, wie sich Passanten im Bezug auf öffentliche Displays verhalten.

Im Gegensatz zu anderen interaktiven Computersystemen fängt die Interaktion bei den öffentlichen Displays bereits beim Passieren an. Potenzielle Nutzer gehen am Display vorbei ohne dieses wirklich wahrzunehmen. In Abbildung 1 von Müller et al. werden die verschiedenen Phasen des Nutzerverhaltens dargestellt [20].

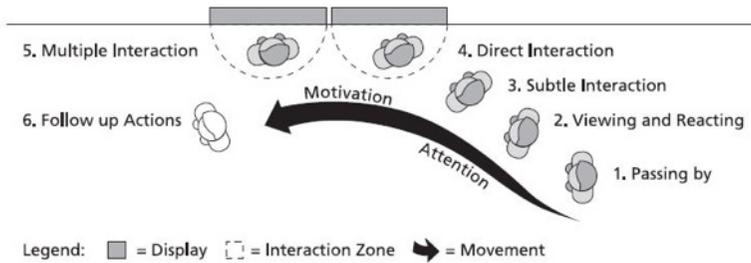


Abbildung 1 Audience Funnel - Diese Abbildung beschreibt die Verhaltensphasen des Publikums mit einem Display [20].

In der ersten Phase wird die Anwendung lediglich von potenziellen Usern passiert. In der zweiten Phase sehen sie das Display an und / oder reagieren darauf mit beispielsweise einem Grinsen oder Kopfdrehen. Subtile Interaktion definiert die Interaktion des Nutzers mit dem Bildschirm durch Gesten oder Bewegungen und die Nutzer anschließend einen Effekt wahrnehmen können. Sobald sich die Person vor dem Bildschirm platziert und die Interaktion aktiver wird, spricht man von der direkten Interaktion. Der User muss jedoch bei jeder dieser Phasen eine gewisse Schwelle überwinden. Beispielsweise werden einige Passanten den Bildschirm gar nicht ansehen, nicht alle werden in subtiler oder direkter Interaktion darauf reagieren. Hier sollte von Anfang an eine gezielte Steigerung der Aufmerksamkeit stattfinden, um die ersten Hürden zu überwinden. [20]

4.1 Aufmerksamkeitssteigerung

In vielen Fällen ist es für ein Computersystem nicht nötig die Aufmerksamkeit des Users zu erlangen. Im Kontrast dazu stehen jedoch die öffentlichen Displays, da der Besitzer nicht gleichzeitig der primäre Nutzer ist. Sie konkurrieren mit der Umwelt, wie Verkehr, Menschen oder Plakatwände, ständig um die Aufmerksamkeit. [20]

Es gibt verschiedene Arten, wie mit ubiquitären Computersystemen die Aufmerksamkeit erhalten werden kann. Stimuli, die einen potenziellen Bedarf an sofortiger Reaktion erfordern, erregen Aufmerksamkeit. Dazu gehören abruptes Auftreten von neuen Objekten [12] oder die Änderung von Farbe, Helligkeit oder Kontrast [9].

Der Bottom-Up Prozess beschreibt das Auftreten externer Stimuli, wie beispielsweise Errornachrichten, um die Aufmerksamkeit auf sich zu ziehen. Bei dem Top-Down Prozess werden interne Bedürfnisse, wie beispielsweise das Ziel eine bestimmte Form oder Buchstaben zu finden, angesteuert.

10:6 Public AR Displays: Herausforderungen und Möglichkeiten

Einerseits argumentiert Weiser et al., ubiquitäre Computersysteme sollten zwischen dem Zentrum der Aufmerksamkeit und der Peripherie der Aufmerksamkeit wandern [28]. User können sich nur auf eine Tätigkeit fokussieren, während mehrere Dinge im Peripheriebereich gleichzeitig wahrgenommen werden. Sobald eine Aufgabe aus der Peripherie der Aufmerksamkeit wichtig erscheint, sollte es für den Nutzer einfach sein, den Fokus darauf zu legen und stetig hin und her wechseln können, sobald es nötig ist.

Andererseits wird argumentiert, dass solche Systeme die Menschen aktiver einbeziehen sollten [25]. In jedem Fall muss ein öffentlicher Display Publikum anziehen, da sonst der Sinn dieser Installation verloren geht.

4.2 Keine Aufmerksamkeit wecken

Change Blindness ist die Tendenz von Menschen, Änderungen in Bildern oder Anwendungen zu übersehen, insbesondere wenn diese unmittelbar nach einer visuellen Unterbrechung angezeigt werden oder die Informationen in dem Moment der Änderung verdeckt sind. Dieser Effekt könnte ebenso verwendet werden um die Aufmerksamkeit zu minimieren, um beispielsweise Updates oder Installationen im Hintergrund durchzuführen. [20]

5 Motivieren

Traditionelle Leinwände bestanden aus Papier und passierende Menschen konnten die Informationen darauf lediglich durch Lesen erfassen. Doch seitdem die Leinwände interaktiv wurden, muss für den Passanten ein Anreiz geschaffen werden, diese zu benutzen. Typischerweise gehen User nicht auf die Straße und suchen direkt nach einem öffentlichen Display, das sie benutzen können. Sie finden diese eher zufällig beim Warten oder Vorbeigehen vor und müssen durch externe Faktoren dazu motiviert werden sie zu benutzen. Und genau dieser Eintritt von interaktiven Displays in den öffentlichen Raum ist Teil einer größeren Tendenz: Der Gebrauch von Computern ist nicht mehr limitiert auf die bloße Abarbeitung von Aufgaben. Während aufgabenorientierte Theorien lediglich das „Wie“ einer Aktivität, nicht aber das „Warum“ betrachten, muss die Frage zur Motivation angegangen werden.

Im Folgenden werden die motivierenden Prinzipien von Michelis vorgestellt, die als Gestaltungsrichtlinien für öffentliche Bildschirme dienen sollen, sodass User Spaß an der Interaktion mit Computersystemen finden [16].

5.1 Herausforderung und Kontrolle

Diese Faktoren beschreiben die Motivation durch Herausforderung und Kontrolle. Der Nutzer ist der Meinung eine herausfordernde Interaktion bewältigen

zu können. Dieser Zustand führt zu einer erhöhten Motivation diese Interaktion auszuführen.

„Im Flow sein“ ist sicherlich vielen ein Begriff. „Flow“ bedeutet ebenso sich in einem mentalen Zustand zu befinden, indem man vollkommen eingetaucht ist, sich voller Energie fühlt und komplett fokussiert ist [7].

Bezogen auf die Mensch-Computer-Interaktion lässt sich somit feststellen, dass der „Flow“ zwischen „zu wenig Herausforderung“, was Langeweile hervorzubringen würde, und „zu viel Herausforderung“, also Angst vor dem Versagen, liegt. Es muss ein optimales Level der Kompetenzen zur Bewältigung der Aufgabe erreicht werden.

Neben der Herausforderung spielt auch die Kontrolle eine wichtige Rolle im Bezug auf die Motivation zur Interaktion. Taucht plötzliche eine neue Aufgabe auf, so besteht der Bedarf, diese gleich zu lösen. Interaktive Umgebungen sollten zudem dem User die Möglichkeit bieten auch eigene dringende Ziele lösen zu können [6]. Zudem sollte beim Entwickeln von Interaktionen immer darauf geachtet werden, ein direktes und klares Feedback nach der Handlung zu geben, da dies die intrinsische Motivation steigert.

Die motivierende Wirkung von Kontrolle beruht in erster Linie auf dem Erkennen einer Ursache-Wirkungs-Beziehung, sowie auf starken Effekten und der Wahlfreiheit bei der Ausführung der Interaktion. Für die Motivation ist die Wahrnehmung von Kontrolle wichtiger als die tatsächliche Kontrolle. Das subjektive Gefühl der Kontrolle kann sogar motivierend wirken, wenn die Person keine tatsächliche Kontrolle besitzt. [15]

5.2 Neugier und Erforschung

Eine der wichtigsten Grundlagen für intrinsisch motivierendes Verhalten ist die Neugier durch neuartige Reize, die etwas Unklares, Unvollständiges oder Unsicheres darstellen. Der Wunsch Unsicherheiten zu vermeiden motiviert Menschen zum Handeln. Neugier wird als Vorreiter von explorativen Verhalten angesehen, durch das Menschen sich zuvor nicht verfügbare Informationen über ihre Umgebung zugänglich machen und Unsicherheiten vermeiden. [8]

Die Interaktion sollte in diesem Fall ebenfalls nicht zu komplex oder zu leicht gestaltet werden. Interaktive Elemente sollten neu und überraschend sein, aber dennoch immer leicht verständlich. Das gewünschte Verhalten für die Interaktion kann durch überraschende Elemente aktiviert und durch konstruktive Elemente aufrechterhalten werden.

Um die Motivation durch Neugier zu steigern, erscheint es zunächst ausreichend, dem Einzelnen ein Gefühl der Unvollständigkeit, Diskrepanz oder Zerstreung zu vermitteln und durch die Interaktion die Chance zu bieten, diese Empfindungen abzubauen. Während der Interaktion sollte jedoch besonders deutlich gemacht werden, wie Vollständigkeit erreicht werden kann. [15]

5.3 Wahl

Die Wahl als motivierender Faktor basiert auf Beobachtungen, dass die Motivation für ein Verhalten zunimmt, wenn dabei zwischen Verhaltensalternativen gewählt werden kann. Zwischen Alternativen zu wählen gibt dem User das Gefühl sein Verhalten zu kontrollieren und aktiv Entscheidungen zu treffen [11]. Die Präsenz von Interaktionsalternativen kann stark motivierend wirken und ermutigt den Nutzer dazu, bestimmte Interaktionen auszuführen und aufrechtzuerhalten [27].

5.4 Fantasie und Metaphern

Im Allgemeinen scheinen imaginäre Settings einen motivierenden Einfluss auf das Verhalten zu haben. In der Fantasie können neue Fähigkeiten erreicht werden und die Beschränkungen der Realität werden aufgelöst. Inwieweit interaktive Umgebungen Fantasie anregen, bestimmt deren Attraktivität und weckt Interesse an der Aufnahme der Interaktion. Durch das Implementieren von Metaphern werden Fantasieelemente direkt in die Mensch-Computer-Interaktion integriert. Da Metaphern sich auf physikalische oder andere Systeme beziehen, können sie dazu benutzt werden, dem User die Interaktion vor der tatsächlichen Verwendung verständlich zu machen. [14]

Wenn neue Formen von Interaktion mit alten traditionellen verknüpft werden, wird es für die Benutzer einfacher bereits etablierte Verhaltensweisen zu übertragen.

5.5 Zusammenarbeit

Im Gegensatz zu den vorher genannten Motivationsfaktoren ist die Kollaboration auf die Interaktion mit anderen Menschen bezogen. Die Möglichkeit andere Menschen durch die Interaktion zu beeinflussen ist hierbei ein wichtiger Effekt [6]. Dies muss nicht zwingend eine Kooperation vor Ort sein, sondern kann sich ebenso auf eine Zusammenarbeit über das Internet und verschiedenen Standorten beziehen. Die Motivation Zusammenzuarbeiten wird beispielsweise durch eine Visualisierung von den eigenen Handlungen verstärkt.

Neben der kooperativen Orientierung, bei dem die Präferenzen von anderen ebenso wichtig sind, wie die eigenen, gibt es noch die kompetitive Orientierung, bei der der User versucht seine eigenen Präferenzen, im Verhältnis zu den anderen, zu maximieren. Kollaboration ist besonders motivierend, wenn das eigene Verhalten von anderen erkannt und ebenso anerkannt wird. Inwieweit die Zusammenarbeit motivierend wirkt, hängt von der persönlichen Erfahrung des Einzelnen ab und kann von Situation zu Situation stark variieren. [13]

6 Interaktion in der Öffentlichkeit

Ein weiteres großes Problem an der Interaktion mit öffentlichen Displays ist, dass sie nun mal in der Öffentlichkeit stattfindet. Menschen möchten möglicherweise einen bestimmten Eindruck hinterlassen, sich nicht von Displays oder anderen Personen stören lassen, keine privaten Informationen preisgeben und höflich gegenüber anderen sein.

6.1 Selbstpräsentierung

Goffman beschreibt in seiner Arbeit das soziale Verhalten von Menschen als Szenenspiel. Jeder Mensch spielt eine bestimmte Rolle und das oberste Ziel ist es, in dieser Rolle zu bleiben. Beispielsweise kann sich ein Verkäufer im Verkaufsraum und im Lagerbereich sehr unterschiedlich verhalten. Ein Polizist wird wahrscheinlich die Interaktion mit einem verspielten öffentlichen Display vermeiden, um seine Rolle beizubehalten. Ebenso vermeiden Menschen Gesten, von denen sie glauben, dass sie ihrer Rolle widersprechen, wie in unserer Kultur das Verbeugen oder Knien. [10]

Ein öffentliches Display kann als Bühne angesehen werden. Wie Menschen damit interagieren hängt von ihren Charaktereigenschaften ab. Eine introvertierte Person würde wohl eine Interaktion mit so einem Bildschirm vermeiden, um keine Aufmerksamkeit auf sich zu ziehen.

6.2 Selektive Kontrolle des Zugriffs zum Selbst

Der Datenschutz kann in die selektive Kontrolle des Zugriffs auf sich selbst und die Kontrolle über die persönlichen Daten unterteilt werden. Privatsphäre als selektive Kontrolle des Zugangs zum Selbst wurde von Altman als dialektischer und dynamischer Grenzregulationsprozess definiert. [1]

Beispielsweise möchten Personen möglicherweise nicht offensiv von einer öffentlichen Anzeige angesprochen werden, da dies als Spam empfunden wird. In ähnlicher Weise haben sie eventuell Angst davor, in der Öffentlichkeit zu stehen und von anderen angesprochen zu werden, wenn sie mit einem öffentlichen Display interagieren.

6.3 Kontrolle über persönliche Daten

Das Gesetz in vielen Ländern garantiert den Datenschutz und die Kontrolle über die persönlichen Daten. Anonymität als „Unbestimmtheit in allen Situationen“ sollte stets garantiert werden.

6.4 Soziales Verhalten

Schließlich möchten Menschen möglicherweise höflich zu anderen sein. Beispielsweise kann es passieren, dass sie auf einer Durchgangsstraße stehen, wenn sie mit einem bestimmten öffentlichen Bildschirm interagieren. Ist die Straße sehr überlaufen mit Menschen, kann es sein, dass die Interaktion abgebrochen wird, um anderen nicht im Wege zu stehen.

6.5 Die Natur vom öffentlichen Raum

Der öffentliche Raum ist dadurch gekennzeichnet, dass er nicht von Einzelpersonen oder kleinen Gruppen kontrolliert wird. Dies bedeutet insbesondere, dass die User eines öffentlichen Bildschirms den Raum um ihn herum nicht kontrollieren können. Wenn sich beispielsweise eine Gruppe von Personen vor dem Bildschirm aufhält und dessen Verwendung verhindert, kann ein User nicht viel dagegen tun.

Aufgrund der Vielzahl an Möglichkeiten im öffentlichen Raum, haben Passanten meistens andere Ziele, wenn sie an einem öffentlichen Display vorbei laufen. Sie sind können auf dem Weg zur Arbeit sein, gehen einkaufen oder besuchen jemanden. Wenn das Ziel freizeitbezogen ist, z.B. wenn Passanten nur spazieren, kann die Wahrscheinlichkeit einer Interaktion viel höher sein. [20]

7 Aktuelle Interaktionsmöglichkeiten mit Displays

Durch die Vielzahl an Sensoren, die es am Markt gibt, ist es möglich viele verschiedene Interaktionsmöglichkeiten anzubieten. Seien es durch Touch- oder Infrarotsensoren, Bluetooth oder RFID Scanner, sie alle ermöglichen eine Variation an Interaktionen. Interaktion mit Computersystemen kann in explizit und implizit unterteilt werden. In der expliziten Interaktion teilt der Computer dem User in verschiedenen abstrakten Kommunikationsformen mit, wie gehandelt werden kann. Mit einer Mouse oder Keyboard können explizite Interaktionen durchgeführt werden. Implizite Interaktion nutzt den Kontext und nehmen die Umgebung über Sensoren wahr, beispielsweise durch Augenbewegungen. [20]

In diesem Abschnitt werden nun die verschiedenen Interaktionsmöglichkeiten beschrieben.

- Präsenz - Kameras, Mikrofone, Bluetooth oder RFIPD Scanner erfassen durch implizite Interaktion die Anwesenheit von Personen, um sie zur Interaktion zu bewegen.
- Körperposition - Ähnlich wie bei der Präsenz werden durch Kameras oder Drucksensoren nicht nur die Präsenz, aber auch die Position einer Person

bestimmt. Das kann zu einer hochwertigeren Anzeige oder aktuelleren Inhalten im Verhältnis zur Userposition führen.

- Körperhaltung - Orientierung, Position und Nähe zum Display kann genutzt werden, um zu beurteilen, ob der User sich dem Display nähert oder vorbei geht. Das kann mit Motion Tracking, 3D Kameras oder Niederfrequenzwellen gemessen werden.
- Gesichtsausdruck - Es gibt eine Fülle an Soft- und Hardware, die Gesichtsausdrücke erkennt. Sie kann bestimmen, ob jemand glücklich, traurig, überrascht oder wütend ist.
- Blick - Mithilfe von Eye Tracking kann der Pfad der Blickrichtungen gemessen werden. Aus der Perspektive der Interaktion können somit Präferenzen im Inhalt bestimmt werden und so das Interface verbessert werden.
- Sprache - Ein Mikrofon in der Nähe eines Displays kann nicht nur zum Erkennen von Keywörtern in Konversationen genutzt werden (wie beispielsweise bei Werbung), sondern auch zur Bestimmung der Anzahl an Personen in der Umgebung. Daraufhin kann der Inhalt an Einzelpersonen oder Gruppen angepasst werden.
- Gesten - Obwohl viele Techniken, wie Akzelerometer, Touch Sensoren, Mouse und Blick Tracking die Gestenerkennung ermöglichen, ist die beliebteste immernoch das Tracking mithilfe von Kameras. Handgesten werden für indirekte und direkte Interaktion, wie beispielsweise für das Manipulieren von Objekten, verwendet.
- Fernsteuerung - Die Steuerung von Displays ist nicht immer durch direkte Interaktion möglich, insbesondere wenn das Display zu weit entfernt oder zu groß ist. Die Fernsteuerung ermöglicht Benutzern das Durchsuchen, Hinzufügen oder Ändern von Inhalten. Derzeitige Ansätze basieren hauptsächlich auf Mobiltelefone, die eine Verbindung zum Display herstellen, z.B. über Bluetooth oder HTTP.
- Keys - Explizite Interaktion kann ebenso klassisch durch Keyboard und Mouse erreicht werden.
- Touch - Touch Interfaces gibt es schon seit vielen Jahren. Ihre Beliebtheit stieg mit dem iPhone und anderen mobilen Smartphones. Touch ermöglicht direkte Manipulation von Objekten mithilfe von expliziter Interaktion. Multitouch Displays erlauben mehreren Nutzern parallel am Display zu interagieren.

8 Erweiterung mit Augmented Reality

Augmented Reality (AR) bezeichnet die Erweiterung und Anreicherung der echten Realität mit künstlichen Informationen, Objekten oder Multimediainhalten. Hierzu können Smartphones, Tablets oder Augmented Reality Brillen

verwendet werden. Das Display eines Head-worn Display sollte unempfindlich gegenüber Tageslicht und klein sein, sowie so wenig Akku wie möglich verbrauchen. Augmented Reality Anwendungen werden hauptsächlich zur Schulung, in Ausstellungen oder für mobile Spiele verwendet. Eines der bekanntesten Beispiele hierfür ist PokemonGo. Die vorher genannten Probleme, die es bereits seit Jahren mit interaktiven Displays gibt, können so möglicherweise mithilfe von Augmented Reality reduziert werden.

8.1 Interaktionsmöglichkeiten

Die Interaktion mit AR Anwendungen für Smartphones und Tablets unterscheidet sich kaum von der herkömmlichen Anwendungen. Die Eingabe via Touchscreen ist die bekannteste und beliebteste, da viele Nutzer täglich so ihr Smartphone bedienen. Die Tonerkennung ist ebenso eine beliebte Interaktionsart mit dem Smartphone in AR Anwendungen. Durch Sprachbefehle können die User ihrem Smartphone mitteilen, was sie als nächstes tun möchten. Gestenerkennung wird durch die Vorder- und Rückkameras am Smartphone ermöglicht. [21]

Vertrautheit mit dem System ist wichtig. Die Interaktion mit einem Mobilgerät sollte stets gleichartig zu anderen Apps sein. Bei einem Experiment von Parker et al. [24] wurde festgestellt, dass User eher mittels Smartphone hingezoomt haben, als sich physikalisch zum öffentlichen Display hin oder weg zu bewegen. Die Gründe dafür sind die Präferenz für einen verdeckten Gerätezoom, der das Individuum nicht so offensichtlich identifiziert wie das physische Zoomen.

Bei Augmented Reality Systemen, die am Kopf getragen werden, ist es sehr wichtig auf eine akkurate Ausführung der Befehle zu achten. Augen-Tracking, das Erkennen von Blinzeln, Sprachbefehle und Gesten mit Händen oder Fingern werden von AR Brillen erkannt. Beide Varianten - sowohl Brillen als auch Smartphones - können ebenso Marker erkennen, die beispielsweise für Positionierung oder Starten einer Anwendung benutzt werden kann [5]. Aufgrund der hohen Preislage, des Tragekomforts und den wenigen Anwendungsfällen für den Alltag, haben sich solche Head-worn Displays noch nicht im Alltag eines normalen Users etabliert. Außerdem fallen Benutzer mit einer AR-Brille auf, weswegen sie sozial noch nicht anerkannt sind. Hauptsächlich werden solche Systeme auf Ausstellungen oder im Schulungsbereich verwendet, weswegen sie in dieser Arbeit nicht berücksichtigt werden.

Somit wäre eine potentielle Lösung zum Schutz der Privatsphäre, eine Kombination aus öffentlichen Displays mit Mobile Augmented Reality. Dieser Ansatz hat Potenzial als wirksames Mittel für die Interaktion von Menschen mit öffentlichen Bildschirmen, da er die sozialen Bedenken hinsichtlich der Interaktion in der Öffentlichkeit verringern könnte.

8.2 Lösungsansätze

Eine intelligente Vorrichtung bietet in den meisten Fällen eine private Ansicht in einem viel kleineren Formfaktor zusätzlich zu der aktuellen Ansicht, die auf dem öffentlichen Bildschirm angezeigt wird. Beispielsweise sendet die Augmented Video Wall von Baldauf et al. ein vom Benutzer ausgewähltes Video auf das Smartphone des Benutzers aus einer Liste von Videos, die auf dem großen Bildschirm angezeigt werden [2]. Die meisten Museumsanwendungen haben auch den Einsatz eines Smartphones untersucht, um den Nutzern den Zugang zu zusätzlichen Informationen über die auf dem öffentlichen Bildschirm angezeigten Artefakte zu ermöglichen.

Durch solche privaten Ansichten können in erster Linie interessante und angepasste Inhalte für den einzelnen Nutzer dargestellt werden und stets durch personalisierte Eingaben weiterentwickelt werden. Andererseits wird ebenso die Privatsphäre des einzelnen geschützt, indem diese nicht an dem öffentlichen Bildschirm präsentiert werden.

Werbungssysteme können Smartphones nutzen, um mit den auf dem öffentlichen Bildschirm angezeigten Werbeinhalten zu interagieren. Benutzer sind somit in der Lage, die Inhalte, an denen sie interessiert sind (Kommentare, Produktempfehlungen usw.), zusätzlich zu den wichtigsten Inhalten auf dem Bildschirm auf ihrem Smartphone zum privaten Surfen abzurufen. Die derzeitigen Interaktionstechniken beschränken sich jedoch auf systemdefinierte Inhalte und können ihre Inhalte nicht an die Präferenzen der Benutzer anpassen. Angesichts der großen Datenmengen aus Social Media Netzwerken und Sensoren kann die Datenanalyse in die aktuellen Systeme für die Werbung integriert werden, so dass diese in der Lage sind, die Präferenzen der Nutzer vorherzusagen und ihre Inhalte entsprechend anzupassen. [22]

Eine andere Möglichkeit zur Inhaltsbestimmung ist die Verlagerung der Inhaltsgenerierung auf Benutzer. Es kann möglicherweise zu einer Erhöhung des Nutzerengagements führen, da sich die Nutzer bereitwillig in eine Interaktion einbringen könnten. In diesem Fall könnte der Bildschirm eine Mischung von vordefiniertem Inhalt anzeigen, die den Benutzer auffordert, benutzerdefinierte Inhalte in Form von Texten und Bildern zu übermitteln. Beispielsweise kann ein User seinen Kommentar zu einem spezifisches Artefakt durch ein Tablet eingeben, das in einem Museum gezeigt wird [4]. Der eingegebene Kommentar wird öffentlich gemacht und kann so andere User dazu ermutigen an weiterführende Diskussionen teilzunehmen [22].

Eine ähnliche Arbeit, erlaubte die Fernbedienung einer Anzeige mit 3D-Interaktionen durch mobiles Augmented Reality und wurde als alternative Methode zur Interaktion mit einem großen Display vorgestellt. Dieses Verfahren ermöglichte es den Personen, mit der öffentlichen Anzeige aus der Ferne mit ihrem eigenen Gerät zu interagieren, indem sie den Touchscreen

10:14 Public AR Displays: Herausforderungen und Möglichkeiten

ihrer Vorrichtung berührten, um Objekte auf dem öffentlichen Bildschirm zu manipulieren. [3]

Die Interaktion mit Mobilgeräten kann auch dazu verwendet werden, die Hindernisse für die Teilnahme und die Interaktion mit öffentlichen Bildschirmen abzubauen. Schroeter et al. entwickelten ein System, das es der Öffentlichkeit ermöglichte, ihre Meinung zu bestimmten Fragen, per SMS oder Twitter zu Versenden, wodurch die Barriere für die Teilnahme gesenkt wurde. Die Nachrichten erschienen auf einem großen öffentlichen Bildschirm, den alle lesen konnten. Diese Interaktion war jedoch nicht privat, da der Name oder die ID des Benutzers zusammen mit seiner Nachricht veröffentlicht wurde. [26]

Eine solche Art der Interaktion hat außerdem den Vorteil, dass mehrere Personen vor Ort interagieren können und somit auch eine räumliche Entlastung vor dem Bildschirm entsteht (und gegebenenfalls bei schlechtem Wetter trotzdem interagiert werden kann). Zusätzlich kann sich so auch ein Honeypot-Effekt bilden, bei dem Passanten aufgrund einer Menschenmenge auf die Interaktionsmöglichkeiten aufmerksam gemacht werden und ebenfalls an den Bildschirm herantreten.

Ein Problem bei der Personalisierung von Inhalten für öffentliche Anzeigen besteht jedoch darin, Inhaltsanforderungen von Benutzern zu identifizieren, ohne ihre Privatsphäre zu beeinträchtigen.

9 Fazit und Ausblick

Öffentliche interaktive Displays gibt es schon seit vielen Jahren, doch ihr Potenzial wird nicht ausgeschöpft. Viele gute Ansätze sind bereits vorhanden und haben sich mittlerweile etabliert. Es wurde festgestellt, dass die Interaktion über ein mobiles Gerät ein wirksames Mittel zum Datenschutz darstellt. Sie ist privat und ermöglicht Usern von weitem zu interagieren. Außerdem könnte es die Interaktion mit mehreren Benutzern unterstützen, was es zu einem äußerst flexiblen Ansatz macht.

Durch die ubiquitären Computersysteme um uns herum, ist es unerlässlich, dass auch öffentliche Displays eine Verbesserung hinsichtlich der Technologie erleben müssen, was jedoch kostspielig und aufwendig sein kann. Es müssten verschiedene Sensoren, wie beispielsweise Bluetooth Low Energy Beacons, NFC Chips oder Kinects zur Gestenerkennung eingebaut werden, um eine bessere User Experience zu schaffen und um maßgeschneiderte, intuitive und schnelle Lösungen, sowie Plattformen anbieten zu können. Jedoch muss auch beachtet werden: Inhalte für jeden relevant zu machen, bedeutet ebenso die Privatsphäre noch mehr zu schützen. Bei der Entwicklung von ubiquitären Systeme sollte stets darauf geachtet werden, dass der User und seine Bedürfnisse im Mittelpunkt stehen.

Literatur

- 1 Irwin Altman. *The environment and social behavior: privacy, personal space, territory, crowding*. Brooks/Cole Pub. Co, Monterey, Calif, 1975.
- 2 Matthias Baldauf and Peter Fröhlich. The augmented video wall: multi-user ar interaction with public displays. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pages 3015–3018. ACM, 2013.
- 3 Matthias Baldauf, Katrin Lasinger, and Peter Fröhlich. Private public screens: detached multi-user interaction with large displays through mobile augmented reality. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia - MUM '12*, page 1, Ulm, Germany, 2012. ACM Press. URL: <http://dl.acm.org/citation.cfm?doid=2406367.2406401>, doi:10.1145/2406367.2406401.
- 4 Andrea Bellucci, Paloma Diaz, and Ignacio Aedo. A see-through display for interactive museum showcases. In *Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces*, pages 301–306. ACM, 2015.
- 5 Marica Bertarini. ETH Zurich Department of Computer Science Zurich, Switzerland maricab@student.ethz.ch. page 8.
- 6 Mark A. Blythe, editor. *Funology: from usability to enjoyment*. Number 3 in Human-computer interaction series. Kluwer Acad. Publ, Dordrecht, 2003. OCLC: 249080675.
- 7 Mihaly Csikszentmihalyi. *Flow: The psychology of optimal experience*. Harper Perennial Modern Classics. Harper [and] Row, New York, nachdr. edition, 2009. OCLC: 553803226.
- 8 EL Deci. *Intrinsic Motivation*. Springer US, Boston, MA, 1975. OCLC: 863864486. URL: <http://nbn-resolving.de/urn:nbn:de:1111-201201172997>.
- 9 James T Enns, Erin L Austen, Vincent Di Lollo, Robert Rauschenberger, and Steven Yantis. New Objects Dominate Luminance Transients in Setting Attentional Priority. page 16.
- 10 Erving Goffman et al. *The presentation of self in everyday life*. Harmondsworth London, 1978.
- 11 Sheena S. Iyengar and Mark R. Lepper. Rethinking the value of choice: A cultural perspective on intrinsic motivation. *Journal of Personality and Social Psychology*, 76(3):349–366, 1999. URL: <http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.76.3.349>, doi:10.1037/0022-3514.76.3.349.
- 12 John Jonides and Steven Yantis. Uniqueness of abrupt visual onset in capturing attention. *Perception & Psychophysics*, 43(4):346–354, July 1988. URL: <http://www.springerlink.com/index/10.3758/BF03208805>, doi:10.3758/BF03208805.
- 13 Lucy A Joyner and Jim TerKeurst. Accounting for User Needs and Motivations in Game Design. page 13.
- 14 Julie E. Kendall and Kenneth E. Kendall. Metaphors and Methodologies: Living beyond the Systems Machine. *MIS Quarterly*, 17(2):149, June

10:16 Public AR Displays: Herausforderungen und Möglichkeiten

1993. URL: <https://www.jstor.org/stable/249799?origin=crossref>, doi:10.2307/249799.
- 15 Thomas W. Malone. Toward a Theory of Intrinsically Motivating Instruction*. *Cognitive Science*, 5(4):333–369, October 1981. URL: http://doi.wiley.com/10.1207/s15516709cog0504_2, doi:10.1207/s15516709cog0504_2.
 - 16 Daniel Michelis. *Interaktive Großbildschirme im öffentlichen Raum*. Springer Fachmedien, Wiesbaden, 2009. OCLC: 828797330. URL: <http://public.eblib.com/choice/publicfullrecord.aspx?p=750340>.
 - 17 Jörg Müller, Gilles Bailly, Thor Bossuyt, and Niklas Hillgren. Mirrortouch: Combining touch and mid-air gestures for public displays. In *Proceedings of the 16th International Conference on Human-computer Interaction with Mobile Devices & Services, MobileHCI '14*, pages 319–328, New York, NY, USA, 2014. ACM. URL: <http://doi.acm.org/10.1145/2628363.2628379>, doi:10.1145/2628363.2628379.
 - 18 Jörg Müller, Dennis Wilmsmann, Juliane Exeler, Markus Buzeck, Albrecht Schmidt, Tim Jay, and Antonio Krüger. Display blindness: The effect of expectations on attention towards digital signage.
 - 19 Ville Mäkelä, Sumita Sharma, Jaakko Hakulinen, Tomi Heimonen, and Markku Turunen. Challenges in Public Display Deployments: A Taxonomy of External Factors. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, pages 3426–3475, Denver, Colorado, USA, 2017. ACM Press. URL: <http://dl.acm.org/citation.cfm?doid=3025453.3025798>, doi:10.1145/3025453.3025798.
 - 20 Jörg Müller, Florian Alt, Daniel Michelis, and Albrecht Schmidt. Requirements and design space for interactive public displays. In *Proceedings of the international conference on Multimedia - MM '10*, page 1285, Firenze, Italy, 2010. ACM Press. URL: <http://dl.acm.org/citation.cfm?doid=1873951.1874203>, doi:10.1145/1873951.1874203.
 - 21 Nazri, Nur Intan Adhani M and Rambli, Dayang Rohaya Awang and Doolan, Daniel C. INTERACTION EXPERIENCES IN MOBILE AUGMENTED REALITY. *Jurnal Teknologi*, 2015.
 - 22 Pai Chet Ng, James She, Kang Eun Jeon, and Matthias Baldauf. When Smart Devices Interact With Pervasive Screens: A Survey. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 13(4):1–23, August 2017. URL: <http://dl.acm.org/citation.cfm?doid=3129737.3115933>, doi:10.1145/3115933.
 - 23 Timo Ojala, Vassilis Kostakos, Hannu Kukka, Tommi Heikkinen, Tomas Lindén, Marko Jurmu, Simo Hosio, Fabio Kruger, and Daniele Zanni. Multipurpose Interactive Public Displays in the Wild: Three Years Later. *Computer*, 45(5):42–49, May 2012. URL: <http://ieeexplore.ieee.org/document/6174993/>, doi:10.1109/MC.2012.115.

- 24 Callum Parker, Judy Kay, Matthias Baldauf, and Martin Tomitsch. Design implications for interacting with personalised public displays through mobile augmented reality. In *Proceedings of the 5th ACM International Symposium on Pervasive Displays - PerDis '16*, pages 52–58, Oulu, Finland, 2016. ACM Press. URL: <http://dl.acm.org/citation.cfm?doid=2914920.2915016>, doi:10.1145/2914920.2915016.
- 25 Yvonne Rogers. Moving on from Weiser's Vision of Calm Computing: Engaging UbiComp Experiences. In David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Dough Tygar, Moshe Y. Vardi, Gerhard Weikum, Paul Dourish, and Adrian Friday, editors, *UbiComp 2006: Ubiquitous Computing*, volume 4206, pages 404–421. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006. URL: http://link.springer.com/10.1007/11853565_24, doi:10.1007/11853565_24.
- 26 Ronald Schroeter and Marcus Foth. Discussions in space. In *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group on Design: Open 24/7 - OZCHI '09*, page 381, Melbourne, Australia, 2009. ACM Press. URL: <http://portal.acm.org/citation.cfm?doid=1738826.1738903>, doi:10.1145/1738826.1738903.
- 27 Linda Klebe Trevino and Jane Webster. Flow in Computer-Mediated Communication: Electronic Mail and Voice Mail Evaluation and Impacts. *Communication Research*, 19(5):539–573, October 1992. URL: <http://journals.sagepub.com/doi/10.1177/009365092019005001>, doi:10.1177/009365092019005001.
- 28 Mark Weiser and John Seely Brown. The Coming Age of Calm Technology. In *Beyond Calculation*, pages 75–85. Springer New York, New York, NY, 1997. URL: http://link.springer.com/10.1007/978-1-4612-0685-9_6, doi:10.1007/978-1-4612-0685-9_6.

UI-Design für ältere Menschen: Physische und kognitive Probleme

Oliver Hein

Ludwig-Maximilians-Universität München, München, Deutschland
oliver.hein@campus.lmu.de

Zusammenfassung

Die folgende Arbeit beschäftigt sich mit User-Interface-Design für ältere Menschen und die damit einhergehenden Probleme. Die hier angesprochenen Probleme betreffen früher oder später jeden Nutzer. Daher ist Forschung und das Finden von Lösung zu Problemen auf diesem Gebiet nötig. Die Probleme sind unterteilt in physische und kognitive Probleme. Zu jeder Gruppe werden jeweils drei konkrete altersbedingt Probleme erläutert, und es folgt ein möglicher Lösungsansatz. Alle Lösungsansätze wurden bereits in anderen Studien und Arbeiten erforscht und beschrieben. Anschließend werden die Lösungen miteinander verglichen, und es wird versucht, mögliche Kombinationen zu erschließen, oder auch gegenseitige Beeinträchtigungen zu identifizieren. Diese Arbeit endet mit dem Fazit, dass einige Probleme noch genauerer Untersuchung benötigen, um adäquate und effektive Lösungsansätze bereitzustellen. Die bisher betriebene Forschung lieferte einige interessante Ergebnisse, doch sind viele dieser an und durch alte Technologien getestet, weshalb auch eine Neuauflage mancher Arbeiten mit modernen Geräten und Methoden bestimmt neue interessante Ergebnisse liefern würde. Außerdem untersuchten viele Arbeiten nur einzelne Problemstellungen, daher wären neue Studien in Bezug auf umfassendere Systeme wünschenswert, welche auch die soziale Komponente mit einbeziehen.

2012 ACM Computing Classification Human-centered computing → Interaction design → Interaction design process and methods → Scenario-based design

Keywords and phrases UI; UX; Benutzeroberflächen; Senioren; ältere Menschen; physische und kognitive Probleme.

1 Motivation

Beim Lesen des Titels dieser Arbeit, könnte man denken, es handle sich hierbei hauptsächlich um die Probleme der derzeitigen Seniorengeneration, welche manchmal vielleicht etwas technisch avers wirkt. Jedoch werden die hier angesprochenen Problematiken jeden von uns früher oder später einmal

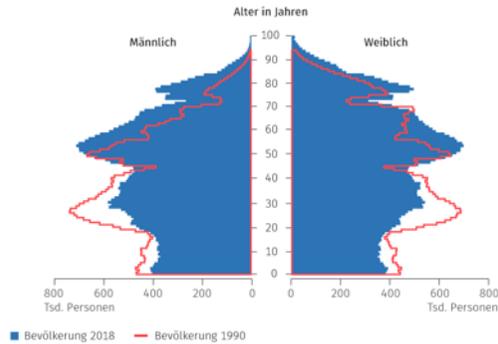


© Oliver Hein;

licensed under Creative Commons License CC-BY

Cite as: Oliver Hein. UI-Design für ältere Menschen: Physische und kognitive Probleme . In *3rd Seminar on Ubiquitous Interaction (UBIACTION 2019)*. Editors: Pascal Knierim, Jakob Karolus, Fiona Draxler, Ville Mäkelä, Tonja Machulla, Florian Lang, Thomas Kosch, Albrecht Schmidt. July 16, 2019. Munich, Germany. pp. 11:1–11:12.

11:2 UI-Design für ältere Menschen



■ **Abbildung 1** Altersaufbau der deutschen Bevölkerung 2018 im Vergleich zu 1990 [3].

betreffen. Wie in Abbildung 1 zu sehen ist, wird unsere Gesellschaft immer älter. Daher wird eine Auseinandersetzung mit diesen Erschwernissen früher oder später unausweichlich. Wie technische Herausforderungen älteren Menschen im Alltag oft schwer fallen, kann man zum Beispiel beim Kauf einer Fahrkarte an einem Ticketautomaten oder der Bedienung eines neuen Mobiltelefons gut beobachten. Ist es bei dem Ticketautomaten das zu unübersichtliche Menü, so ist es bei dem Mobiltelefon die zu kleine Benutzeroberfläche, welche den älteren Menschen Probleme bereiten kann. Schuld daran ist jedoch nicht der Nutzer, sondern die Gestaltung des Systems, welches meist nicht an die Bedürfnisse eines älteren Klientels angepasst ist. Daher ist es von äußerster Wichtigkeit, passende Lösungen zu finden, welche diese ubiquitären Probleme zu verbessern, um das Leben für alle nach wie vor und auch in Zukunft so komfortabel wie möglich zu gestalten [13]. Diese Arbeit orientiert sich an aus der Literatur bekannten Problemstellungen, welche sich hauptsächlich auf die Ursache des Alterns zurückführen lassen [4]. Diese sind unterteilt in physische und kognitive Probleme.

2 Physische Herausforderungen und Lösungsansätze

Dieses Kapitel widmet sich den physischen Herausforderungen in Bezug auf das Alter. Zuerst wird immer ein Problem dargelegt, für welches die darauf folgende Studie versucht eine Lösungsmöglichkeit zu bieten.

2.1 Sehschwächen

Sehschwächen können bereits ab einem Alter von 40 Jahren spürbar eintreten. Altersweitsichtigkeit ist ein ganz natürlicher Prozess, der bereits ab der Geburt beginnt, doch erst ab einem Alter von etwa 40 Jahren deutlich spürbar



■ **Abbildung 2** Prototyp Interface des Home Screens [19].

wird. Während diesem Prozess kommt es zu einem altersbedingten Verlust der Nahanpassungsfähigkeit der Augen. Dadurch verschiebt sich der Nahpunkt des Auges von etwa 25 cm immer weiter nach hinten [11]. Aber auch andere Probleme, wie zum Beispiel Makuladegeneration, Grüner und Grauer Star, können auftreten [15, 1]. Diese können sogar zur Erblindung führen.

Angepasste Benutzeroberflächen

Hutchison et al. haben bereits 1997 die Erkenntnis gewonnen, dass für viele ältere Nutzer der Text der Oberflächen zu klein dargestellt wird, was sich auf die verschlechterte Nahanpassungsfähigkeit der Augen der Nutzer zurückführen lässt [9]. Daher haben sie eine Benutzeroberfläche entworfen, bei der die Text- und Schaltflächengröße manuell eingestellt werden kann. Das ursprüngliche Layout-Design blieb jedoch erhalten. Somit konnten die Nutzer die Oberfläche ihren individuellen Problemen anpassen. Diese Methode ist jedoch auch sehr zeitintensiv.

2016 wurde eine Arbeit veröffentlicht, welche sich mit dem Design von Mobiltelefonoberflächen beschäftigte [19]. Hier wurde versucht, unter Berücksichtigung gegebener Designmethoden, eine Oberfläche zu gestalten, die möglichst nutzerfreundlich gegenüber einer älteren Zielgruppe ist. Abbildung 2 zeigt das Prototyp Interface des Home Screens. Wie man sehen kann werden hier die am wichtigsten und am meisten genutzten Funktionen direkt auf dem Home Screen positioniert. Die Symbole werden möglichst groß angezeigt, jedoch ist durch die ausgewählte Anzahl an Funktionen die Anzeigefläche auch nicht überladen. Die Funktionen sind "Ambulance", womit man direkt einen Notruf absetzen kann, "Spoken Assistance", welches einen Audioclip mit kontextbezogenem Inhalt entwirft, "Make a Call", womit man einen Anruf tätigen kann und "Send SMS", womit man eine SMS versenden kann. Um mehr Apps zu

11:4 UI-Design für ältere Menschen

sehen, reicht es nach links oder rechts zu wischen. In der Fußzeile befinden sich drei extra Funktionen. Eine, um zum Home Screen zurückzukehren, eine weitere, zum Sperren des Geräts und eine Taste, mit welcher man einen Schritt zurück gehen kann. Dabei sind alle Funktionen mit besonders einfachen und ausdrucksstarken Symbolen ausgestattet, sodass die älteren Nutzer sofort den Zweck dahinter verstehen.

2.2 Haptische Verschlechterung

Im Alter verschlechtern sich auch die motorischen und haptischen Fähigkeiten. Die Reaktionszeiten, wie auch die Genauigkeit der ausgeführten Bewegungen, werden signifikant langsamer und ungenauer [16]. Dies erschwert älteren Menschen vor allem die Nutzung von touch-basierten Benutzeroberflächen, wie man sie zum Beispiel bei einem Smartphone findet.

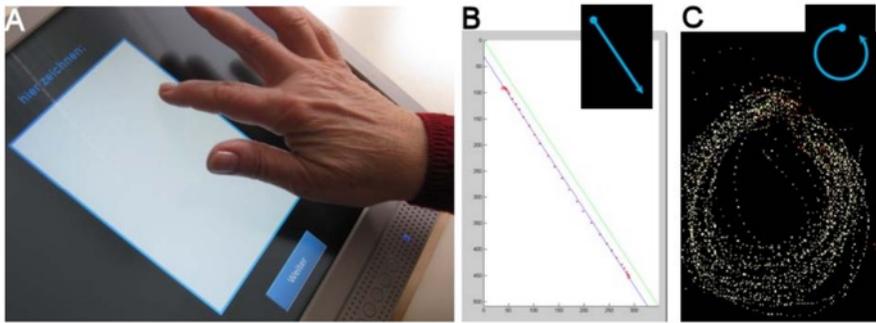
Gestensteuerung

Anstatt genaue Eingaben per Fingerdruck vom Nutzer zu verlangen, können bestimmte Funktionen auch anhand der Erkennung bestimmter Fingergesten ausgeführt werden. In einer Arbeit von Stoessel et al. [18] aus dem Jahr 2010 wurde untersucht, ob Fingergesten, insbesondere für ältere Nutzer (Alter 60+), eine geeignete Eingabemethode darstellen. Dabei wurde primär auf altersbedingte Veränderungen der sensorischen, kognitiven und motorischen Fähigkeiten geachtet. In einer dazu durchgeführten Studie wurde eine Gruppe älterer Nutzer mit einer jüngeren Benutzergruppe verglichen. Bei Ausübung der 42 verschiedenen getesteten Fingergesten, wovon zwei in Abbildung 3 zu sehen sind, wurde die Geschwindigkeit und Eingabegenauigkeit der einzelnen Nutzer gemessen.

Die Ergebnisse zeigten, dass ältere Nutzer etwas langsamer reagierten, aber nicht signifikant schlechter als die jüngere Benutzergruppe. Dabei war die Größe des Bildschirms und die Gestenkomplexität im Vergleich der beiden Gruppen irrelevant. Diese gestenbasierte Interaktion stellt also kein Hindernis für ältere Nutzer da, sondern könnte als geeignete Eingabemethode sogar eine Hilfe sein. Allerdings kommt hierbei auch die Frage auf, ob diese zusätzlichen körperlichen Anstrengungen die Zielgruppe nicht zu sehr belasten würden.

2.3 Hörverlust

So wie alle Sinne, verschlechtert sich mit zunehmendem Alter auch das Gehör, vor allem im Hochfrequenzbereich. Dies hat nicht nur Auswirkungen auf den Betroffenen, es beeinträchtigt auch verbale Kommunikation mit anderen [8]. Auch wenn dies meistens die klassischen Mensch-Maschine-Interaktionen nicht verhindert, so stellen sie bei manchen Handlungen doch eine Beeinträchtigung dar.



■ **Abbildung 3** A) Versuchsaufbau. B) Beispiel für einen Gestenpfad mit Originaldaten (rote Punkte), lineare Anpassung (blaue Linie) und Referenzlinie (hellgrüne Linie). C) Beispielmessung eines zirkulären Gestenmusters [18].

Hörgeräte

Hörgeräte und Cochlea-Implantate sind die am häufigsten verwendeten Geräte zur Behandlung von Altersschwerhörigkeit. Studien zu Hörgeräten zeigen, dass sie eine wirksame Methode zur Behandlung von leicht bis mittelschwerem Hörverlust sind, solange der Patient motiviert und in der Lage ist, das Gerät zu verwenden. Je nach Art und Schwere können auch Mittelohrimplantate geeignete Lösungen für die Behandlung von Altersschwerhörigkeit sein, da diese auch bei schwerer Altersschwerhörigkeit nachgewiesenermaßen sehr positive Ergebnisse liefern [17]. Moderne Hörgeräte verfügen über rechenstarke und energieeffiziente Audioprozessoren, welche über ein Mikrofon alle umgebenden Klänge aufnehmen und den Hörverlust durch eine pegel- und frequenzabhängige Verstärkung ausgleichen. Heutige Geräte besitzen auch leistungsfähige Algorithmen zur Reduktion von Rückkopplungen und Umgebungsgeräuschen. Auch können neuere Hörgeräte wichtige Hörsituationen erkennen und automatisch das beste Programm dafür wählen. So bieten sie den Nutzern ein optimales Hören [12]. Eine neuere Entwicklung sind integrierte Miniaturfunksysteme, über die die Hörgeräte auf beiden Seiten permanent Daten austauschen. So können sie gemeinsam eine optimale Einstellung des Systems finden [6].

3 Kognitive Herausforderungen und Lösungsansätze

Dieses Kapitel widmet sich den Problemen durch abnehmende kognitive Leistungsfähigkeit im Alter. Die Struktur dieses Kapitels folgt dem selben Schema wie Kapitel 2.

3.1 Aufmerksamkeitsschwund

Die Aufmerksamkeitsspanne nimmt im Alter erwiesenermaßen signifikant ab [2]. So fällt es älteren Menschen oft schwerer, sich an jüngst geschehene Vorkommnisse zu erinnern, oder sich länger darauf zu konzentrieren.

Eldergames

Basierend auf diesen Herausforderungen wurde Eldergames entwickelt, ein von der EU finanziertes Projekt, welches dem Erhalt von durch Alterung beeinträchtigten kognitiven Funktionen dient und durch die Multiplayer-Modi auch die Geselligkeit der Nutzer erhöht. Der hierfür entwickelte Prototyp konnte aufgrund seiner Einfachheit, Benutzerfreundlichkeit der Benutzeroberfläche und der Multiplayer-Architektur ein angenehmes soziales kognitives Training bieten [5]. Das Spiel ist in zwei Gruppen unterteilt: Memospiele und Minispiele. Minispiele sind einzelne Spiele, die vom System zufällig ausgewählt werden, um eine bestimmte kognitive Fähigkeit (Gedächtnis, Aufmerksamkeit, Argumentation und Kategorisierung) zu trainieren. Das Memospiel ist ein soziales Spiel, bei dem ein Spieler seine/ihre Karten - die Bilder, Töne oder Berechnungen darstellen - mit den entsprechenden Karten im mittleren Feld des Bildschirms koppeln muss. Richtiges oder falsches Verknüpfen führt zu einem Minispiel, dessen erfolgreicher Abschluss auf eine Strafe verzichtet, oder zum Erhalt eines Bonus, welcher für ein abschließendes Multiplayer-Spiel verwendet werden kann. Abbildung 4 zeigt eine Gruppe von Senioren bei einer Multiplayerrunde.

Die Punktzahlen aller Spieler werden während des Spiels angezeigt und zusammen mit mehreren Indizes der Leistung der Spieler dauerhaft in einem Archiv gespeichert. Experten können auf dieses Archiv zugreifen, um die Entwicklung oder den Rückgang der Benutzerfähigkeiten zu überwachen. Tatsächlich wurde ein Test zum Vergleich zwischen den Leistungen von 59 älteren Teilnehmern in den Minispielen hinsichtlich Gedächtnis, Argumentation, selektiver Aufmerksamkeit, geteilter Aufmerksamkeit und Kategorisierung durchgeführt. Es wurden keine signifikanten Unterschiede zwischen den Tests festgestellt, was zeigt, dass Eldergames als Überwachungsinstrument für kognitive Funktionen verwendet werden können.

3.2 Verschlechterung des Arbeitsgedächtnisses

Die Fähigkeit des Arbeitsgedächtnisses, welches ermöglicht, Informationen vorübergehend zu speichern und gleichzeitig zu manipulieren, nimmt ab einem Alter von bereits 20 bis 25 Jahren ab [14]. Dieses löst in etwa die selben Effekte aus wie der Aufmerksamkeitsschwund, nur dass hier auch die Entscheidungsfindung für den Betroffenen erschwert wird.



■ **Abbildung 4** Gruppe von Senioren bei einem Eldergames Multiplayerspiel [10].

Optionsreduzierung und Bestätigung

In einer Studie von 2009 wurden zwei Strategien zur Entlastung des Arbeitsgedächtnisses für Benutzer von Sprachschnittstellen evaluiert. Hierfür wurde ein Versuch aufgebaut, bei dem die Nutzer über ein Sprachdialogsystem einen Termin vereinbaren sollten. Die erste Strategie bestand darin, pro Runde weniger Optionen zur Verfügung zu stellen, die zweite Strategie bestand aus Bereitstellung von Bestätigungen. 48 Benutzer buchten Termine mit neun verschiedenen Dialogsystemen, die sich in der Anzahl der angebotenen Optionen und der verwendeten Bestätigungsstrategie unterschieden. Die Teilnehmer führten außerdem vier kognitive Tests durch und bewerteten die Benutzerfreundlichkeit jedes Dialogsystems anhand eines standardisierten Fragebogens.

Als das System mehr Optionen pro Runde bereitstellte und explizite Bestätigungsunterdialoge vermieden wurden, buchten ältere und jüngere Benutzer Termine schneller, ohne den Erfolg ihrer Aufgabe zu beeinträchtigen. Benutzer mit geringerer Informationsverarbeitungsgeschwindigkeit konnten sich seltener an alle relevanten Aspekte des Termins erinnern. Die Arbeitsspeicherspanne hatte keinen Einfluss auf den Terminrückruf. Ältere Benutzer waren jedoch mit den Dialogsystemen etwas weniger zufrieden als jüngere Benutzer. [21]

3.3 Verschlechterung des Langzeitgedächtnisses

Das Langzeitgedächtnis ist das dauerhafte Speichersystem für verschiedenste Arten von Information im Gehirn. Doch auch dieses lässt im Laufe der Zeit nach, ähnlich dem Arbeitsgedächtnis [7]. So fällt es älteren Menschen oft schwer, sich an bestimmte Ereignisse oder Informationen zu erinnern.

Personal Digital Assistant

Folgende Studie von Vemuri et al. [20] präsentiert eine tragbare, rechnergestützte Gedächtnishilfe in Form eines Personal Digital Assistant (PDA). Dieser ist in der Lage ubiquitäre Tonaufnahmen vorzunehmen und stellt auch passende Abrufwerkzeuge zur Verfügung. In einer Selbststudie hat einer der Autoren seine Gespräche mit Abteilungskollegen über einen Zeitraum von zwei Jahren aufgenommen. Nach der Aufnahmephase hörte der Autor alle Aufzeichnungen an, konstruierte dazu Fragen und stellten sie drei zufällig ausgewählten Gesprächspartnern. Wenn die befragte Person sich nicht richtig erinnern konnte, durfte sie das iRemember PDA zum Durchsuchen der Daten nutzen, um die richtige Antwort zu finden, während der Autor ihr Suchverhalten beobachtete.

Es konnte gezeigt werden, dass mit Hilfe dieses Geräts der Informationsabruf im Gedächtnis der Nutzer erheblich verbessert werden kann, und sie sich auch noch nach mehrerer Monaten genauer an vergangene Gespräche erinnern konnten im Vergleich zu einer Kontrollgruppe, bei welcher der PDA nicht genutzt wurde. Ein Problem hierbei ist jedoch, dass diese Art der Analyse sehr zeitaufwendig ist, und auch bestimmtes Fachwissen erfordert, welches die durchschnittlichen Nutzer meist nicht aufweisen.

4 Diskussion

In diesem Kapitel werden Vor- und Nachteile einzelner Lösungsansätze verschiedener Arbeiten verglichen, inwieweit diese auch kombinierbar sein könnten, oder aber wo es zu neuen Komplikationen und Beeinträchtigungen führen kann.

Vergleich

Auch wenn die meisten Arbeiten unterschiedlichste Ansätze verfolgen, so sind viele Designkonzepte von Oberflächen, welche genutzt werden, doch sehr ähnlich. Da die meisten Probleme des Alters bereits lange bekannt sind, ist auch bekannt, auf welche Problemstellungen zu achten ist. So fällt zum Beispiel auf, dass viele Arbeiten besonderen Wert auf simple und übersichtliche Gestaltung legen. Daher folgen die meisten entworfenen Interfaces den Designmotiven Schlichtheit, Übersichtlichkeit und Benutzerfreundlichkeit. Dies zeigt sich durch die Reduzierung auf die wichtigsten Informationen und Funktionen, sowie Vereinfachung der Aufgabenstellung durch einheitliche Designmuster.

Was die Gestaltung von Hardware angeht, welche speziell zur Unterstützung von Problemen des Alters genutzt wird, so finden sich auch hier einige Gemeinsamkeiten. Die größten Ziele sind auch hier eine einfache Handhabung, sowie eine zuverlässige unproblematische Unterstützung. Eine große Unterscheidung ist jedoch, dass einige entwickelte Systeme nur zur Überwachung und Kontrolle älterer Menschen genutzt werden können, wie zum Beispiel Eldergames, wohingegen andere Systeme den Nutzer aktiv in seinem normalen Leben unterstützen, aber keine Daten zur Analyse sammeln können, wie etwa Hörgeräte.

Kombination

Aufgrund der Unterschiede der einzelnen Probleme, könnten sich auch bei manchen Lösungsansätzen gute Kombinationsmöglichkeiten mit anderen Arbeiten finden. So wäre zum Beispiel die Arbeit zu speziell angepassten Benutzeroberflächen für ältere Menschen [19] gut mit der Arbeiten zu Fingergestensteuerung [18] kombinierbar. So könnten nicht nur die bereitgestellten Informationen verbessert werden, sondern auch die möglichen Interaktionen. Ein weiteres Konzept, welches sich auch gut mit der Nutzung von Smartphones realisieren lassen würde, wäre die Gedächtnishilfe in Form eines Personal Digital Assistant [20]. Würde dieses System auf ein Smartphone portiert werden, würden wesentlich mehr Möglichkeiten zur Überwachung/Aufzeichnung des Nutzerverhaltens entstehen, wie auch zur (Echtzeit-)Analyse dieser. Diese wäre in einem solchen Konzept wesentlich nutzerfreundlicher umsetzbar und so auch für ältere Menschen einfacher zu handhaben.

Gegenseitige Beeinträchtigung

Doch die Unterschiedlichkeit der einzelnen Probleme führt nicht nur zu Kombinationsmöglichkeiten, sondern kann leider auch zu gegenseitiger Beeinträchtigung führen. So sind die beiden Arbeiten zu speziell angepassten Benutzeroberflächen für ältere Menschen [19] und der Gedächtnishilfe in Form eines Personal Digital Assistant [20] beide eher schlecht kompatibel mit hörgeschädigten Nutzern. Außerdem lässt sich keine adäquate Lösung finden, die dieses Problem ohne zu großen Aufwand beheben kann.

Ein weiteres Problem ist durch das verschlechterte Sehvermögen gegeben. Da die älteren Nutzer vergrößerte Oberflächeninhalte bevorzugen, die (mobilen) Endgeräte jedoch immer kleiner werden, müssen hier entweder Abstriche in Funktionalität oder Mobilität gemacht werden. Doch nicht nur der technische Aspekt spielt hier eine Rolle, sondern auch der soziale. Obwohl zum Beispiel spezielle Seniorentelefone den meisten Anforderungen gerecht werden, finden sie wenig Anklang in der Gesellschaft. Es sollte also mehr versucht werden, nicht ein eigenständiges System speziell für ältere Menschen zu entwickeln, sondern von Anfang an ein Konzept, welches auf die Bedürfnisse aller Nutzer abgestimmt ist.

5 Fazit

Abschließend wird ein Überblick präsentiert, der die sinnvollsten und problematischsten Projekte Dritter und die darauf erarbeiteten Lösungsansätze oder auch identifizierten Komplikationen aufzeigt. Auf dem Gebiet der altersbedingten Probleme wurden bereits viele solche identifiziert und analysiert. Auch gibt es zu vielen Problemen bereits entwickelte Lösungsansätze. Die Schwierigkeit mit diesen Ansätzen ist allerdings, dass viele bereits Jahre zurück liegen, und mit heutigen Technologien weitaus effektiver und einfacher umgesetzt werden könnten, was auch zu einer Verbesserung der Studienergebnisse führen kann.

Auch wenn es eine große Gruppe an Schwierigkeiten gibt, welche im Alter auftreten, so haben viele ältere Nutzer auch kognitiv sehr individuelle Probleme, welche nur anhand von individueller und spezialisierter Analyse und Anpassung überwunden werden können. Dazu benötigt es jedoch technisch sehr ausgereifte Systeme, die zum Beispiel mit Hilfe von künstlicher Intelligenz und neuronalen Netzen die benötigten Analysen erstellen und ein angepasstes Design dazu entwickeln können, um die voranschreitenden Verschlechterungen des Gedächtnisses zu verlangsamen oder im besten Fall zu stoppen.

Ein System, das sich als sehr effektiv erwiesen hat, ist Eldergames. Hier werden die Nutzer spielerisch getestet, um ihre kognitiven Leistungen zu überwachen. Im Vordergrund für den Nutzer steht jedoch der Spaß am Spielen, wie auch die reale Verbundtheit zu anderen Mitspielern, welche alle gemeinsam um das Gerät sitzen können. Doch hier liegt auch ein großer Nachteil dieses Systems. Es ist ein physischer Gegenstand, es gibt keine Chance dieses System online von einem anderen Computer aus zu nutzen, oder einen Multiplayermodus zu realisieren, welcher auch rein digital funktioniert, und so nicht auf die physische Anwesenheit der Mitspieler angewiesen ist. Doch gerade dass hier absichtlich Rücksicht auf die soziale Komponente gelegt wird, steigert das Nutzererlebnis erheblich. Für viele ältere Menschen wird Einsamkeit zu einem Problem, daher ist dies ein guter Weg dem entgegen zu wirken.

Der hier betrachtete Bereich ist, zumindest was die Problemidentifikation betrifft, gut erforscht. Jedoch beziehen sich die meisten Arbeiten nur auf kleine Teilaspekte aus diesem Bereich. So werden zwar spezielle Schwierigkeiten effektiv angegangen, helfen jedoch nicht, was das große Bild angeht, eine dauerhafte und ubiquitäre Lösung zu finden. So wird zum Beispiel bei Mobiltelefonen hauptsächlich auf die Gestaltung der Benutzeroberfläche geachtet, anstatt das ganze Betriebssystem den Anforderungen anzupassen. Daher wäre es vermutlich sinnvoll, bei zukünftigen Arbeiten zu versuchen, mehrere Herausforderungen gleichzeitig anzugehen, um so eine Lösung zu finden, welche auf mehrere Anwendungsbeispiele übertragen werden kann.

Literatur

- 1 Penny A Asbell, Ivo Dualan, Joel Mindel, Dan Brocks, Mehdi Ahmad, and Seth Epstein. Age-related cataract. *The Lancet*, 365(9459):599–609, 2005.
- 2 Robin A Barr and Leonard M Giambra. Age-related decrement in auditory selective attention. *Psychology and Aging*, 5(4):597, 1990.
- 3 Statistisches Bundesamt. Demografischer Wandel, 2018. URL: https://www.destatis.de/DE/Themen/Querschnitt/Demografischer-Wandel/_inhalt.html.
- 4 Connor Dodd, Rukshan Athauda, and MT Adam. Designing user interfaces for the elderly: a systematic literature review. In *Proceedings of the Australasian Conference on Information Systems*, pages 1–11, 2017.
- 5 Luciano Gamberini, Francesco Martino, Bruno Seraglia, Anna Spagnoli, Malena Fabregat, Francisco Ibanez, Mariano Alcaniz, and Javier Montesa Andrés. Eldergames project: An innovative mixed reality table-top solution to preserve cognitive functions in elderly people. In *2009 2nd Conference on Human System Interactions*, pages 164–169. IEEE, 2009.
- 6 Volkmar Hamacher and Uwe Raß. Hightech im ohr: physikalische und technische grundlagen moderner hörgeräte. *Physik in unserer Zeit*, 37(2):90–96, 2006.
- 7 Lynn Hasher. Aging and long-term memory: Deficits are not inevitable. *Lifespan cognition: Mechanisms of change*, page 162, 2006.
- 8 Karen S Helfer and Laura A Wilber. Hearing loss, aging, and speech perception in reverberation and noise. *Journal of Speech, Language, and Hearing Research*, 33(1):149–155, 1990.
- 9 Douglas Hutchison, Caroline Eastman, and Terry Tirrito. Designing user interfaces for older adults. *Educational Gerontology*, 23(6):497–513, 1997. URL: <https://doi.org/10.1080/0360127970230601>, arXiv:<https://doi.org/10.1080/0360127970230601>, doi:10.1080/0360127970230601.
- 10 Fares Kayali, Naemi Luckner, Oliver Hödl, Geraldine Fitzpatrick, Peter Purgathofer, Tanja Stamm, Daniela Schlager-Jaschky, and Erika Mosor. *Elements of Play for Cognitive, Physical and Social Health in Older Adults*, volume 7946, pages 296–313. 01 2013. doi:10.1007/978-3-642-39062-3_19.
- 11 Ronald Klein. Age-related eye disease, visual impairment, and driving in the elderly. *Human Factors*, 33(5):521–525, 1991. PMID: 1769672. URL: <https://doi.org/10.1177/001872089103300504>, arXiv:<https://doi.org/10.1177/001872089103300504>, doi:10.1177/001872089103300504.
- 12 Birger Kollmeier. Cocktail-partys und hörgeräte: Biophysik des gehörs. *Physik Journal*, 1(4):40, 2002.
- 13 Aaron Marcus. Universal, ubiquitous, user-interface design for the disabled and elderly. *Interactions*, 10(2):23–27, March 2003. URL: <http://doi.acm.org/10.1145/637848.637858>, doi:10.1145/637848.637858.
- 14 Timothy A Salthouse. The aging of working memory. *Neuropsychology*, 8(4):535, 1994.

11:12 UI-Design für ältere Menschen

- 15 Wolfgang F Schrader. Altersbedingte makuladegeneration. *Der Ophthalmologe*, 103(9):742–748, 2006.
- 16 Manuel Sebastián and Soledad Ballesteros. Effects of normal aging on event-related potentials and oscillatory brain activity during a haptic repetition priming task. *NeuroImage*, 60:7–20, 12 2011. doi:10.1016/j.neuroimage.2011.11.060.
- 17 GM Sprinzl and H Riechelmann. Current trends in treating hearing loss in elderly people: a review of the technology and treatment options—a mini-review. *Gerontology*, 56(3):351–358, 2010.
- 18 Christian Stöbel, Hartmut Wandke, and Lucienne Blessing. Gestural interfaces for elderly users: Help or hindrance? In Stefan Kopp and Ipke Wachsmuth, editors, *Gesture in Embodied Communication and Human-Computer Interaction*, pages 269–280, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.
- 19 Judy Van Biljon and Karen Renaud. Validating mobile phone design guidelines: Focusing on the elderly in a developing country. In *Proceedings of the Annual Conference of the South African Institute of Computer Scientists and Information Technologists*, page 44. ACM, 2016.
- 20 Sunil Vemuri, Chris Schmandt, and Walter Bender. iremember: a personal, long-term memory prosthesis. In *Proceedings of the 3rd ACM workshop on Continuous archival and retrieval of personal experiences*, pages 65–74. ACM, 2006.
- 21 Maria Wolters, Kallirroi Georgila, Johanna D Moore, Robert H Logie, Sarah E MacPherson, and Matthew Watson. Reducing working memory load in spoken dialogue systems. *Interacting with Computers*, 21(4):276–287, 2009.