

# Ubiaction 2018/2019

**UBIACTION 2018/2019**

Edited by Pascal Knierim, Thomas Kosch, Jakob Karolus, Matthias Hoppe,  
Fiona Draxler, Tobias Benz, Lewis Chuang, and Albrecht Schmidt

Haptics.EEG.Presence.EyeTracking.MR.VR.AR.  
TUI.SmartEnvironments.Self-Perception.VUI.  
Security.IX.UltrasonicHaptics.  
AutomotiveUI.

# UBIACTION 2018/2019

2st Seminar on Ubiquitous Interaction  
June 28, 2018, Munich, Germany

3rd Seminar on Ubiquitous Interaction  
January 31, 2019, Munich, Germany

Edited by

Pascal Knierim

Albrecht Schmidt

## ■ Vorwort

### *Editors*

Pascal Knierim, Thomas Kosch, Jakob Karolus, Matthias Hoppe, Fiona Draxler,  
Tobias Benz, Lewis Chuang, and Albrecht Schmidt

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

*ACM Classification 1998 - H.5 INFORMATION INTERFACES AND PRESENTATION*

ISBN-13: 978-1722881986

*Publication date*  
31. January 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.

Die kontinuierliche Entwicklung von neuen Technologien eröffnet vor Allem in der Forschung die Pforten zu neuen Möglichkeiten der Interaktion zwischen Mensch und Computer. In der heutigen Gesellschaft ist die Allgegenwärtigkeit von Computern nahezu unabdingbar und werden als Kommunikations-, Mobilitäts- und auch als Bezahlmittel verwendet. Das Forschungsgebiet wird auch als „Ubiquitous Computing“ bezeichnet und untersucht unter anderem die Zugänglichkeit und Einfachheit der Interaktion mit mobilen und tragbaren Geräten und deren Einfluss auf die Erfahrungen des Menschen. Im Rahmen des Seminars „Menschzentrierte Interaktion mit ubiquitären Computersystemen“ am Lehrstuhl für „Human-Centered Ubiquitous Media“ der Ludwig-Maximilians-Universität München haben Studierende Vorträge zu verschiedenen Themengebieten vorbereitet. Diese wurden während der zweiten „Ubiaction“-Veranstaltung am 28.06.2018 gehalten. Zusätzlich wurden Ergebnisse aus dem Praktikum „Web Programmierung (PWP)“ vorgestellt und Prototypen aus dem Praktikum „Developing interactive capture devices“ ausgestellt. Herr Prof. Dr. Albrecht Schmidt eröffnete die Veranstaltung mit dem Vortrag „Menschliche Fähigkeiten durch digitale Technologien verstärken“. Die schriftlichen Ausarbeitungen der Studierenden aus den Sommer und Winter Semestern wurden in diesem Band als zweiter Band der Serie „Ubiaction“ zusammengeführt.

## ■ Inhaltsverzeichnis

Tangible User-Interfaces im Kontext ubiquitärer Computersysteme <i>Gregor Anzer</i> .....	1:1–1:14
Eingabemedien zur bedeutungsvollen Interaktion mit TUIs <i>Melanie Hauser</i> .....	2:1–2:14
Explizite Mensch-Computer-Interaktion mittels Eye-Tracking <i>Christian Gregor Mall</i> .....	3:1–3:19
Das Potential der Potentiale - EEG und Interaktion <i>Anna Hubert</i> .....	4:1–4:17
Adaptives Lernen mit Gehirn-Computer-Schnittstellen <i>Raimund Krämer</i> .....	5:1–5:14
Interaktive intelligente Umgebungen <i>Jennifer Busta</i> .....	6:1–6:18
Entwicklung eines Prozesses für Voice User Interface Design <i>Sabine Christina Kaupp</i> .....	7:1–7:28
Nutzbare Sicherheit: Erkenntnisse aus dem Bereich der Authentifizierung <i>Sebastian Halder</i> .....	8:1–8:13
Reduzierung der kognitiven Belastung durch natürliche Interaktionsprinzipien in VR-Anwendungen <i>Thomas B. M. Puls</i> .....	9:1–9:19
„Ich sehe mich, also bin ich., - Selbstwahrnehmung und Avatar- design in Virtueller Realität <i>Nadine Sarah Schüler</i> .....	10:1–10:17
Präsenz und Messung von Präsenz in Virtual Reality <i>Chantal Marie Huttenloher</i> .....	11:1–11:16
Freihand basierte Interaktion in Augmented Reality <i>Duc Anh Le</i> .....	12:1–12:17
Bildung Augmentieren: Anwendung und Herausforderungen beim Lernen mit Augmented Reality <i>Dimitri Hein</i> .....	13:1–13:18

# Tangible User-Interfaces im Kontext ubiquitärer Computersysteme

Gregor Anzer

Ludwig-Maximilians-Universität München, München, Deutschland  
gregor.anzer@campus.lmu.de

Das Unsichtbare anfassen: Ultrasonic Haptic Feedback. <i>Verena Schlott</i> .....	14:1–14:17
Haptische Rückmeldung in gemischter Realität <i>Anita C. Baier</i> .....	15:1–15:20
Cockpit Design für autonome Fahrzeuge <i>Daniel F. Steidl</i> .....	16:1–16:19
Evaluation der Fähigkeit von haptischen Interfaces für dynamische Materialsimulation in Virtueller Realität <i>Yannick Weiß</i> .....	17:1–16:21

---

## Zusammenfassung

Diese Arbeit beschäftigt sich mit dem Konzept von Tangible User-Interfaces (TUI) im Kontext ubiquitärer Computersysteme. Wie hängen diese Begriffe von einander ab? Gibt es Unterschiede? Es wird auf die Begrifflichkeiten eingegangen und aktuelle, sowie auch alte Beispiele verwendet, um zu verdeutlichen, welche Rolle TUIs im Kontext ubiquitärer Computersysteme spielen können. Der Trend zu ubiquitären Systemen, genauso wie zu Calm Computing, verlangen nach einer anderen Herangehensweise an Mensch-Maschine-Interaktion, als bei den zur Zeit verbreiteten Konzepten der Interaktion. Dabei muss man auch klar Unterschiede klarstellen zwischen haptics, touch, und tangible.

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

**Keywords and phrases** Ubiquitous Computing; Tangible User Interface; TUI.

## 1 Entwicklung und Begriffe

Ein Verständnis der folgenden Begriffe ist notwendig: TUI, Calm Technology, Ubiquitous Computing. Die letzten beiden Begriffe sind eng miteinander verbunden. Der erste Begriff beschreibt ein User-Interface Bedienkonzept. Welche Rolle kann dieses Konzept im Kontext der beiden anderen Begriffe spielen? Vertraut sind wir im Alltag mit WYSIWYG (What you see is what you get), der direkten Manipulation und Anzeige von Information, und GUIs (Grafische Benutzeroberflächen) als Bedienkonzepte. Der Rest wird im folgenden erläutert.

### 1.1 Ubiquitous Computing

Einer der wegweisenden Forscher im Bereich Interaktion war Mark Weiser. Er arbeitete Zeit seines Lebens an der Antwort auf der Frage, wie der Mensch in Zukunft mit Computern interagieren wird. Seine Visionen der zukünftigen



© Gregor Anzer;

licensed under Creative Commons License CC-BY

Cite as: Gregor Anzer. Tangible User-Interfaces im Kontext ubiquitärer Computersysteme. 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. 1:1–1:14.

Mensch-Maschine-Interaktion haben sich schon vielfach bewährt. Die Forschung um den Begriff Ubiquitous Computing postulierte zum Beispiel die heutzutage gängigen Computing-Formate Smartphone, und Tablet. Damals als Tabs und Pads bezeichnet, postulierte Weiser eine Unmenge an solchen Geräten in jedem Raum. Der Preis unserer Mobilgeräte mit Bildschirmen ist jedoch noch nicht so weit gesunken, dass man diese kleinen Geräte als “scrap computer” verwenden kann. Meine These zu derlei Entwicklung ist, dass unsere Tonnen von Elektroschrott nicht mehr so früh zu Müll werden sollten. Es gibt zahlreiche Projekte, wo ältere Geräte mit bestimmten Verwendungszwecken versehen, zu neuem Wert verholfen wird. Zum Beispiel kann ein iPad, das für Webbrowsing und neuere Anwendungen zu langsam geworden ist, immernoch für die Steuerung von vernetzter Heimelektronik, als digitaler Notizblock, oder ähnliches verwendet werden. Vielleicht wird in Zukunft weniger weggeworfen und mehr für spezielle Zwecke wiederverwendet werden, besonders wenn die Ressourcenknappheit seltener Metalle weiter zunimmt. Doch die Vision von Ubiquitous Computing geht über die Definition von Computerformaten hinaus. Es beschreibt ein ganzheitliches Konzept. So argumentiert Weiser, dass die Maschinen selbst nicht mehr sichtbarer Teil unserer Interaktion mit Computern sein werden, sondern zum Teil der Umgebung werden, mit der wir auf selbstverständliche Art und Weise interagieren. Als Metapher beschreibt Weiser die Lesebrille, welche zwar präsent ist, aber vom Träger nicht mehr bewusst wahrgenommen wird. Diese Metapher erweckt auch den Gedanken an Computing als “Augmentation” der menschlichen Fähigkeiten; also eine Vermehrung, Verbesserung, oder Erweiterung. Eine weitere Metapher, die meines Erachtens nach nicht komplett zutrifft, ist der Vergleich mit der Wahrnehmung von Straßenschildern. Hier erklärt Weiser, dass das Erlernen des Informationsgehaltes des Schildes irgendwann dazu befähigt das Schild nicht mehr aktiv lesen zu müssen, um die Information wahrzunehmen. Im Gegensatz zu der Metapher mit der Lesebrille, ist das Erlernen von Straßenschildern ein aktiver Prozess und braucht eine gewisse Zeit. Man kann sich jedoch eingestehen, dass seit Otl Aichers Meilenstein des Icon Designs zu den Olympischen Spielen 1972, das Design von öffentlicher Beschilderung eine universell verständliche Sprache bekommen hat, was den Lernprozess auf ein Minimum reduziert. Wir sind zweidimensionale Vereinfachungen des Menschen und der Umgebung gewohnt, und lesen sie blitzschnell. Eine Brille zu verwenden benötigt jedoch keinerlei Training, nur eine Anschaffungs- und Eingewöhnungsphase. In beiden Fällen verweist Weiser allerdings auf die unbewusste Verwendung von Information in der Umgebung. Sei es durch erlerntes Wissen, oder durch natürliche, visuelle Verarbeitung. Computing an sich soll somit allgemein omnipresent werden, was im Gegensatz zu der anfänglich der 90er Jahre verbreiteten Idee des “personal computers” steht.. Auch dem Prinzip der Virtual Reality steht Weiser

kritisch gegenüber. Er empfindet Virtual Reality als eine ungemein aufwendige Simulation der realen Welt, bei gleichzeitigem kompletten Missachten der realen Umgebung um den Anwender. Er sieht darin keinen verbesserten Nutzen der existierenden Umgebung. Hier wird wieder deutlich, dass Weiser die menschlichen Fähigkeiten verbessern und erweitern will. Ein Abtauchen in eine Simulation empfindet er als nicht wertvoll. Eine simple Art, wie Computer nützlicher werden können, ist das Hinzufügen einer simplen Information: Die Lage. Wenn ein Computergerät weiß, in welchem Raum es sich befindet, kann es schon sehr viel nützlicher sein, ohne künstliche Intelligenz verwenden zu müssen. Dies ist besonders wichtig für die weiteren Kapitel über TUI. Man kann sich vorstellen, dass ein Smartphone, das weiß, ob es sich zuhause befindet, andere Funktionen primär zu erfüllen hat, wie ein Smartphone im Auto. Ein große Hürde in der Umsetzung von “Ubiquitärem Computing” sieht Weiser in der Vernetzung des allzeit verfügbaren Computings. Heutige Technologien sind bereits sehr fortgeschritten, und schaffen Sicherheit und ständige Vernetzung. Eine nahtlose und einfache Erfahrung ist das allzeit vernetzte Arbeiten jedoch noch nicht. Netzwerktechnik ist immer noch ein Fall für Techniker und Spezialisten. Und die Sicherheit unserer vernetzten Welt ist immer wieder kritisiert und diskutiert. Ubiquitous Computing wird und hat unsere Arbeit schon längst positiv beeinflusst, aber wie wird eine Interaktion mit der ubiquitären Welt von morgen aussehen? Wie interagiert man mit einer Maschine, die sich nahtlos in die Umgebung integriert hat? Ist es Touch, Sprache oder Gesten? Vielleicht finden sich die Antworten in der Konzeption zu TUIs. [8]

## 1.2 Calm Technology

Mark Weiser arbeitete ebenfalls an einem Konzept, dass er Calm Technology nannte. Es ist ein Verständnis von Technologie, dass ein Gefühl von Vertrautheit und Ruhe in die Interaktion mit Maschinen bringt. Es wird der Begriff der Peripherie eingeführt, um die nicht sofort aktiv wahrgenommene Umwelt einzubeziehen. Weiser konzipierte eine Umgebung mit hunderten Displays in unserer direkten Umgebung. Calm Computing spezialisiert diesen Gedanken weiter, indem es auf subtilere Art die Umgebung mit Information und Möglichkeiten zur Interaktion anreichert. [9] Eine solche subtilere Art der Informationsvermittlung ist bei einer solchen Umgebung auch notwendig. Wenn hunderte Displays ständig um die Aufmerksamkeit der Personen in ihrer Umgebung konkurrieren, wäre ein jeder Mensch überfordert und gestresst.

esigns that encalm and inform meet two human needs not usually met together. Information technology is more often the enemy of calm: pagers, cellphones, news services, the World Wide Web, email, TV, and radio bombard us frenetically. Can we really look to technology itself

for a solution? [10]

Die periphere Wahrnehmung kann sowohl als räumliche, visuelle, auditive, oder anderweitig periphere Wahrnehmung verstanden werden. Ein simples Beispiel wäre, wenn man das MacOS Betriebssystem verwendet und eine App möchte auf subtile Art auf sich aufmerksam machen, dann springt das Applikationssymbol im Dock für einen Augenblick auf und ab. Unsere periphere Wahrnehmung ist sehr schlecht, was Farben oder langsame Änderungen betrifft, aber eine schnelle kurzzeitige Bewegung ist für den Menschen instinktiv mit dem Wechsel der Aufmerksamkeit verbunden. Hier rückt eine mögliche Interaktion in den Vordergrund, um sodann gleich wieder in den Hintergrund zu verschwinden. Mark Weiser verwendet als Beispiel für den Begriff der Peripherie das Motorengeräusch beim Autofahren. Jeder kennt das normale Motorengeräusch seines Autos. Es im Normalbetrieb zu hören, gibt Zuversicht und Sicherheit. Eine kleine Änderung im Geräusch ist zwar keine große Ablenkung, aber dennoch wahrnehmbar und bietet subtil die Möglichkeit darauf zu reagieren. Vielleicht fehlt Öl im Motor, oder ist eine Inspektion fällig? Ein anderes Beispiel des Forschers aus der analogen Welt ist die Anbringung von Bürofenstern in einer Arbeitsumgebung. Selbst wenn man nicht ständig aus dem Bürofenster zu seinen Kollegen sieht, gibt das geschäftige Treiben und die Sicht zu den Kollegen ein Gefühl von Sicherheit und Ruhe, dadurch, dass man weiß, dass alles wie gewohnt geschäftig ist. Herannahende Kollegen sind auch aus der Peripherie erkennbar und weisen subtil auf bevorstehende Interaktion mit der Umgebung hin. Der Begriff definiert auch, dass Technik jederzeit aus der Peripherie in das Zentrum der Wahrnehmung rücken kann und damit interagiert werden kann. Genauso wie Calm Technology das Zentrum der Aufmerksamkeit wird, so verschwindet es auch wieder in die definierte Peripherie. Ein wichtiger Begriff auf den später noch mehr eingegangen wird, ist der Begriff "Affordance". Weiser zitiert Norman und Gibson, die sich ausgiebig mit Affordance beschäftigt haben. Es besteht eine Relation zwischen einem Objekt und einer Person. Eine Person hat gewisse Fähigkeiten und Vorwissen. Ein Objekt hat physische Eigenschaften. Eine Person nimmt daraufhin eine gewisse Funktion des Gegenstands wahr, je nach Vorwissen und den Eigenschaften des Objektes. Weiser sieht ihn als nicht genug definiert für die Peripherie. Dem ist zuzustimmen, da peripher wahrgenommene Objekte oft nicht in ihrer ganzen Dimensionalität wahrgenommen werden. Die Wahrnehmung ist auch oft nicht gezielt und somit ungenau. Wie ist daraus eine Beziehung zwischen den beiden Entitäten abzuleiten? Zusammenfassend kann man sagen, dass Calm Technology somit helfen kann, Technik ruhiger und vertraut zu machen. [9]

### 1.3 Tangible User-Interfaces

Hiroshi Ishii und Brygg Ullmer prägten den Begriff des Tangible User-Interfaces, kurz TUI. [3] Es führt haptische Bedienelemente unabhängig von der physischer Welt in die Welt der Mensch-Maschine-Interaktion ein. Präziser, es beschreibt ein Bedienkonzept und die damit verbundene Implementierung von greifbaren Bedienelementen. Es entstand aus der Beobachtung, dass wir immer mehr mit Computern verbunden sind, aber die Schnittstelle zusehends auf grafischen Bedienelementen besteht, die die erlernten Mechanismen der Interaktion des Menschen nicht beachtet. Computer sind auch heute noch hauptsächlich Rechteckige Bildschirme mit Elementen auf die man mit einem Zeiger (Maus oder Finger) deutet. TUIs sollen die digitale Welt wieder mit physischen Objekten und der realen Umgebung verbinden. Eine Art Befreiung von den Limitierungen der rechteckigen Bildschirme vor unseren Augen. Ein TUI-Objekt kann direkt ein digitales Objekt repräsentieren und ermöglicht somit eine direkte Manipulation und Ansicht von Daten, genauso, wie man es aus der realen Umgebung kennt. In Tangible Bits [3] werden zunächst drei Kategorien von Tangible Objects definiert:

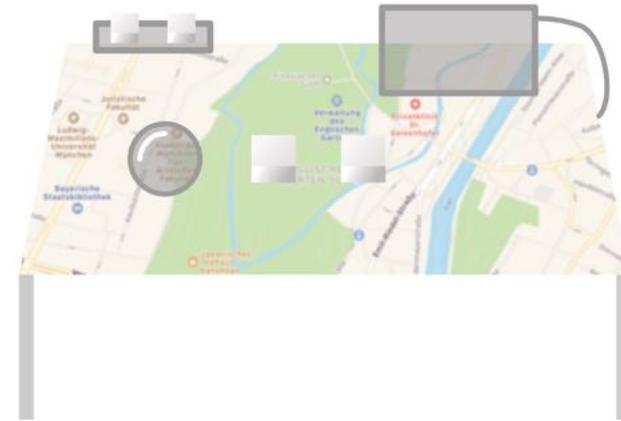
- Interaktive Flächen - interactive surfaces
- TUI Objekte - atoms and bits
- Periphere Medien - Ambient Media

Die erste Kategorie beschreibt interaktive Flächen. Diese können zum Beispiel ein Tisch oder eine Wand sein, welche die erste Schnittstelle zwischen Realem und Digitalem darstellen. Die Atom und Bits genannten Objekte sind dann die genannten Repräsentationen von digitaler Information als reales Objekt. So könnte ein reales Buch seinen Inhalt auf einer TUI-Fläche in digitaler Form preisgeben. Die Peripheren Ambient Medium genannten Teile des TUI-Konzeptes sind fast analog zu Mark Weisers Calm Technology zu sehen. Hier wird die Peripherie mit Informationsgehalt angereichert. Die Idee von TUIs ist insgesamt stark inspiriert von Ubiquitous Computing und somit von Mark Weisers Vision der Zukunft der Mensch-Maschine-Interaktion [3], doch arbeiten Ullmer und Ishii nicht daran, die bestehenden Konzepte zu ersetzen. Es geht ihnen vielmehr darum, das verlorene Haptische aus der Zeit vor GUIs wieder in den die Interaktion mit Computern zu überführen, wie in folgendem Zitat verdeutlicht.

Our work has been stimulated by Weiser's vision, but it is also marked by important differences. The Tab/Pad/Board vision is largely characterized by exporting a GUI-style interaction metaphor to large and small computer terminals situated in the physical environment. While this approach clearly has a place, our interest lies in looking towards the bounty of richly-afforded physical devices of the last few

millennia and inventing ways to re-apply these elements of “tangible media” augmented by digital technology. Thus, our vision is not about making “computers” ubiquitous per se, but rather about awakening richly-afforded physical objects, instruments, surfaces, and spaces to computational mediation, borrowing perhaps more from the physical forms of the pre-computer age than the present. [3]

Die Forscher sehen in AR eine gute Möglichkeit für Tangibles, jedoch nicht in der Form, wie man sich AR vorstellt. Eine Brille die visuell Informationen über die reale Welt legt, sondern eine Welt mit greifbaren Elementen als Input für Computer. [3] Durch die Verknüpfung von realen Objekten mit digitaler Information, sehen sie die Realität als zusätzlich angereichert an. Ullmer und Ishii nennen einige Implementierungen von Mensch-Maschine-Interaktionskonzepten, um TUI zu verdeutlichen. Das Clearboard ist eine geteilte Zeichenfläche, die die Präsenz der beiden Teilnehmer in den Vordergrund stellt, indem sie die zeichnenden Personen gleichzeitig darstellt. Örtlich voneinander getrennt, können beide Nutzer eine digitale Zeichenfläche bearbeiten. Ausschlaggebend für den Fokus auf greifbare Objekte war die Arbeit von Fitzmaurice und Ishii an Bricks [1]. Bricks hat Ähnlichkeiten mit einer Maus als Input. Wohingegen die Maus jedoch zeitlich gesehen mithilfe der GUI jederzeit andere Funktionen erfüllt, werden die Bricks, kleine quaderförmige Elemente, bestimmten digitalen Informationen angehängt und dienen fortan der Manipulation dieser Informationen, sie sind also räumlich bestimmten Funktionen angehängt. Dieses Konzept hat vielerlei Vorteile im Vergleich zu Multitouch und GUI mit Maus. Multitouch ist aufgrund der physisch vorhandenen bricks evident. Die Funktion ist offensichtlicher. Ein paralleler Input, auch von mehreren Personen ist kein Problem. Bei Multitouch Anwendungen, wie sie heutzutage üblich sind, besteht oft die Annahme, dass eine einzelne Person den Input liefert. Allgemein ist die Bedienung über derlei Interaktion somit besser mit dem menschlichen Verständnis von Räumlichkeit und menschlicher Interaktion vereinbar. [1] Die Arbeit von Fitzmaurice ist die klarste Referenz von Ishii und Ullmer, um das Konzept von TUI zu verstehen. Die Experimente in der Arbeit zeigen, wie schneller und effizienter das Manipulieren von physischen Objekten ist, um Digitales zu verändern. Die darauffolgende Arbeit der Forscher an metaDESK erinnert sehr an die Arbeit an Bricks. MetaDESK ist eine schräge interaktive Tischfläche mit rückwärtiger Projektion. [7] Es werden GUI- und TUI-Objekte verwendet. Die TUI-Objekte werden Linse, Phicon, Ablagefläche, Griff und Instrument genannt. Rein intuitiv betrachtet erfüllen sie die Funktionen, die ihre Namen bereits beschreiben. Im Prinzip stellen die meisten Elemente eine verhaptisierung von GUI-Elementen dar. Ein greifbares Icon wird hier Phicon genannt. Siehe Abbildung 1. Die Funktion der TUI-Elemente ist jedoch drastisch anders. So gibt es eine echte passive Linse,



■ **Abbildung 1** Der metaDESK ist ein Multitouch-Tisch mit rückseitiger Projektion und Infrarot Objekterkennung.

die Teile des Displays vergrößert. Eine aktive Linse, bestehend aus einem zweiten Display, welches mit einem frei bewegbaren Arm ausgestattet ist. Mit zwei Phicons, welche man auf der Fläche voneinander wegzieht, kann man zum Beispiel eine Darstellung vergrößern. Auch für die Funktion von Rotation haben sich Ishii und Ullmer etwas überlegt. Mithilfe eines zylindrischen Objektes, kann man die Ansicht drehen. Hier wird der Unterschied von haptischem Bedienelement, wie ein Drehknopf am Radio und einem TUI-Objekt besonders klar. Ein TUI Objekt auf dem metaDESK ist sich seiner Lage bewusst. Es repräsentiert nicht irgendwelche zeitlich-abhängige Funktionen, sondern ist sich aufgrund seiner Lage seiner Funktion bewusst. Der zylindrische Drehknopf auf der Darstellung am metaDESK rotiert die Karte genau um den Punkt, wo er platziert wurde. [7] Die reine haptische Eigenschaft ist nur dafür da, die Funktion ersichtlich zu machen. Man kann jedoch sekundäre Qualitäten des Haptischen ausmachen (Material, Reibung, Größe, Gewicht, Farbe, etc.), welche auch zur Benutzererfahrung beitragen.

## 2 Chancen für TUI in ubiquitären Systemen

Auch wenn die Mensch-Maschine-Interaktion sich immer mehr dem Ubiquitous Computing zuwendet, ist das davon inspirierte Tangible-UI-Konzept noch nicht

so stark verbreitet. Durchgesetzt haben sich bisher die bekannteren grafischen Benutzeroberflächen in Verbindung mit Multitouch und Eingabegeräten, wie Maus und Tastatur. Obwohl Ishii und Ullmer stark von Ubiquitous Computing inspiriert waren, bleibt der große Erfolg bislang aus. Die interaktive Umgebung die in Tangible-User-Interfaces angesprochen wurde ist ein klarer Verweis auf die Arbeiten von Mark Weiser und den anderen Forschern bei Xerox. Auch der bewusste Bezug auf Peripherie, wie zum Beispiel bei metaDESk und transBOARD zeigen auf, wie TUIs auf Ubiquitäre Systeme anzuwenden sind. Deshalb werden in den folgenden Unterkapiteln die besonderen Merkmale des Bedienkonzeptes TUI dargestellt. Konkret soll darauf geachtet werden, was die Vorteile und Nachteile sind, aber auch wie der Bezug zu Ubiquitous Computing (und Calm Technology) zu verstehen ist.

## 2.1 Affordance

Affordance ist ein Begriff innerhalb des TUI-Konzepts, der den physischen Attributen von Objekten gewisse Kenntnis über deren Funktionsweise zuspricht. Ullmer und Ishii verweisen auf die dazugehörige Arbeit von Donald Norman, aus der sie schließen, dass bestimmte physische Objekte eine Handlung von uns gerade zu erfordern. Norman entnahm den Begriff des Forschers J.J. Gibson, welcher den Begriff Affordance als erstes in diesem Kontext verwendete. [5] Für Norman und Gibson stellt Affordance eine Beziehung zwischen einem physischen Objekt und dem Menschen dar. Viele Faktoren auf beiden Seiten prägen diese Beziehung. Nachdem sich der Begriff der Affordance in der Designwelt durchgesetzt hatte, wollte Norman die Begrifflichkeiten weiter spezifizieren, da sie in seinen Augen oft falsch verwendet wurde. Auf der Seite des Menschen bildet sich, laut Norman, nun eine wahrgenommene Affordance. Sie muss nicht immer der wahren Funktion des Objektes entsprechen. So ist allgemein bekannt, wie man einen Türknauf dreht, einen Griff zieht, oder einen Knopf drückt. Jedoch passiert es schon öfter, dass eine Tür die gedrückt werden sollte, fest zu gezogen wird. Hier entsprach die Affordance des Türknaufes, nicht der eigentlichen Funktion der Tür. Diese markante Eigenschaft ist in vielen alltäglichen Interaktion mit Computern vorhanden. Wie in der Arbeit zu Bricks [1], sehen die Forscher die gesteigerte Effizienz in der Interaktion mit TUI-Objekten in engem Zusammenhang mit Affordance.

Auch in, bis jetzt, wenigen kommerziellen Produkten, hat sich das Prinzip anwenden lassen. Das Nest Thermostat dreht man, wie einen Drehknauf um die Temperatur im Haus zu regeln. Siehe Abbildung 2 links. Dass das Thermostat die Wärmepräferenzen, und andere Daten, nutzt um in Zukunft energiesparender die Temperatur zu regeln, merkt der Nutzer nicht. Das Computing wird somit unsichtbar, so wie es Weiser postuliert hatte. Das Surface Dial stellt einen Haptischen Drehregler dar, der auch gedrückt werden

kann. Man positioniert es auf dem Bildschirm. Der Regler wird vom Microsoft Surface Display (PixelSense) erkannt und zeigt unter dem Regler runde Menus an, welche der Drehregler sodann bedienen kann. Hier ist eine Erkennung des Reglers, genau so umgesetzt, wie Ishii und Ullmer, Jahrzehnte vor Microsoft, die Erkennung von Phicons angedacht hatten. Hier ist der Regler genau wie ein Phicon durch die räumliche Positionierung auf einer bestimmten Fläche einer Applikation (Vergleiche Tray) mit einer bestimmten Funktion oder Information gekoppelt. Ein sehr intuitives Beispiel, dass sich schwerer erklären lässt, als es sich (be)greifen lässt, ist das Schwedische Projekt Bounceslider (oder auch ForceFeedback Slider). [2, 2] Mit Hilfe von Reglern, welche einen motorisierten Federmechanismus implementieren, lassen sich Loops erzeugen. Der Regler ist dabei Bedienelement und Ausgabe zugleich (WYSIWYG). Das intuitive Verständnis des physischen Reglers ist direkt an die Soundinformation gekoppelt; eine direkte Umsetzung des TUI-Konzepts. Im Sinne von ubiquitous Computing stellt es eine einfach zu bedienende Interaktion dar, welche in der Umgebung aufgeht, anstatt einem bewusst zu machen, dass man mit einem Computer interagiert. Vorteilhaft ist hier die einfache Bedienung und das subtile Interface. Es passt in den Kontext des ubiquitären Computings, da es das Computing an sich in den Hintergrund stellt und das Erstellen von Musik ohne das Befassen mit dem Computer an sich ermöglicht. Es geht nur alleine um die Interaktion mit Musik; das direkte Manipulieren von Ton und Frequenz. Auch hier hilft Affordance, als eine Eigenschaft von physischen Objekten, die Interaktion zu vereinfachen. Man kann die Regler als TUI-Objekt definieren, da ein Regler an einen Loop eines Tons gebunden ist. Mehrere Regler ermöglichen mehrere gleichzeitige Töne.

## 2.2 Peripherie

Die Peripherie, wie sie Weiser definiert und Ullmer und Ishii verwenden, ist eine gute Möglichkeit zu zeigen, wie TUI und Ubiquitous Computing einander komplementieren können. So können diverse TUI-Objekte Teil der Peripherie werden. Bei Ubiquitous Computing, als auch bei TUI, wird der Begriff verwendet. Weiser zitiert seinen Kollegen John Seeley Brown aus seiner Forschung bei Xerox, der Peripherie als das versteht, was aufgrund von gelerntem nicht mehr aktiv wahrgenommen wird. [8] Weiser definiert den Begriff noch etwas genauer in seiner Arbeit zu Calm Technology. Siehe Abschnitt 1.2. Tangible User Interfaces enthält als Konzept ebenfalls die Verwendung des Begriffes. Hier ist die Peripherie ähnlich wie bei Calm Technology verwendet. Es geht um die peripher wahrgenommenen Informationen. Ishii und Ullmer nennen als peripher wahrgenommene Medien (ambient media) die Lichttemperatur, die Temperatur, Luftzug, Geräusche und vieles mehr. [3, 4] Somit können TUI-Objekte auf vielerlei Art Teil der Peripherie werden. Durch häufiges Benutzen, werden die

Objekte visuell weniger stark wahrgenommen. Durch das Nutzen peripherer Medien, kann TUI in die Peripherie überblenden. Markantes Beispiel hierfür ist die Tastatur des Computers. An sich ist diese kein TUI-Objekt, kann aber aufgrund der haptischen Interaktion als Beispiel herangezogen werden. Sie wird nach längerem Gebrauch ohne visuellen Fokus genutzt und bleibt in der Peripherie. Tangible-UI-Objekte könnten nach einer kurzen Eingewöhnung ebenfalls unbeachtet genutzt werden. Beispiele hierfür gibt es viele. Die Tangible Remote Controllers für wandgroße Displays von Yvonne Jansen sind haptische Bedienelemente mit denen dynamische Anfragen (dynamic queries) an Daten gestellt werden können. 1 Die Forscherin und ihr Team realisierten dies mittels



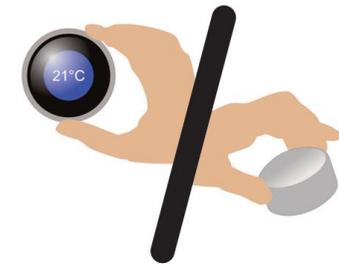
■ **Abbildung 3** Links ist das wandgroße Display mit einer Visualisierung eines Datensatzes. Rechts ist das Tablet mit eigens hergestellten Reglern. Sie steuern dynamische Anfragen die in der Visualisierung sichtbar werden. 1

eigens erbauter, haptischer Bedienelemente mit Markern. Siehe Abbildung 3. Diese wurden auf handelsüblichen Tablets befestigt. Der Fokus des Benutzers liegt in erster Linie natürlich auf den Daten des überdimensionalen Displays. Nur für kurze Momente verweilt das Auge auf den Reglern. Sie können sodann bedient werden, ohne die Augen von den Daten abwenden zu müssen. Dies stärkt den Erkenntnisgewinn, durch eine bessere Wahrnehmung der gezeigten Veränderung in den Daten. Das Erfühlen der vorhandenen Regler in der Peripherie ist ebenfalls ganz im Sinne von Weisers Calm Technology. Es gibt einem Vertrauen und Ruhe, denn eine pure grafische Darstellung ist viel volatiler als das beständige Gefühl eines haptischen Objektes. Die Implementierung auf einem Tablet entspricht auch ganz den Vorstellungen Weisers von Tablets in intelligenter Umgebung. In solch einer Umgebung ist ein geschicktes Vernetzen erforderlich, welches die Netzwerktechnik in den Hintergrund stellt und die Interaktion mit den räumlich nahe gelegenen Geräten ermöglicht. Nachteil in der hier aufgeführten Implementierung ist, dass die Interaktion auf eine Person mit einem Tablet beschränkt ist. Eine echte Kollaboration ist hier nicht gegeben. Es besteht nur eine Vernetzung von einem Tablet und dem Display, bzw. dem Computer. Aber auch in diesem Szenario tritt der Computer an sich eher in den Hintergrund, da die Information direkt durch das Tablet

manipuliert wird und die grafischen Elemente auf ein Minimum reduziert sind. Von der Arbeit des Computers oder der Präsenz eines Computers wird wenig vermittelt. Im Sinne von Ambient Media, also von peripheren Medien wie in TUI, ist dieses Projekt ein wenig entfernt. Als klassisches Beispiel für periphere Medien sei ambientROOM erwähnt. Diese Kollaboration von Ishii, Ullmer und der Firma Steelcase, bettet Information in die Peripherie ein, welche durch TUI-Objekte gesteuert wird. [3, 5] TUI-Objekte könnten hier die Quelle oder das Ziel von Informationen sein und durch periphere Medien dargestellt werden. Sie Abschnitt 1.2 für Informationsgehalt in der Peripherie. Beispielhaft symbolisiert ein Spielzeugauto die stetige Anzahl von Besuchen auf der Produktseite des besagten Autos, ein Produkt einer fiktiven Firma. Bewegt man das Auto in einen gewissen Bereich (eine dedizierte Ablage wie auf metaDESK z.B.), kann durch unterschiedlich starkes Regenrauschen, peripher dargestellt werden, ob die Produktseite gut, oder eher weniger stark besucht ist. Alternativ wurde die Frequenz der Besucher durch Wasserwellenspiegelungen an der Decke dargestellt. [3, 6]

### 2.3 Kollaboration

Das Projekt “From Big Data to Insights: Opportunities and Challenges for TEI in Genomics” ist ein gutes Beispiel für verbesserte Kollaboration durch TUIs. [6] Generell müssen Genforscher aus enorm großen Datensätzen Erkenntnisse gewinnen. Die Daten sind schwer darzustellen und Berechnungen dauern sehr lange. Eine Kollaboration ist unbedingt erforderlich. Normale Desktops sind nur begrenzt für mehrere gleichzeitige Nutzer ausgelegt. In verschiedenen Fällen wurde erforscht, wie man mit solchen Datensätzen umzugehen hat und wie TUI als Bedienkonzept die Arbeit erleichtern kann. Wie beim Beispiel mit den wandgroßen Displays und Remote Controls, fanden die Forscher TUI gut geeignet für dynamische Anfragen an die Datensätze. [4], jedoch war es nicht vorgesehen mehrere Nutzer gleichzeitig eingaben machen zu lassen. Bei Genomics waren die berührungsempfindlichen und mit Markern steuerbaren Projektionen besonders gut für kleine Gruppen geeignet. Schon das Arbeit zu



■ **Abbildung 2** Links das Nest Thermostat zur intelligenten Steuerung der Temperatur. Rechts das Surface Dial. Zu positionieren auf einem PixelSense Display von Microsoft.

Bricks hat festgestellt, dass die von Ishii und Ullmer TUI-Objekte genannten Inputmethoden gut für Kollaboration geeignet sind. Auch das Manipulieren der Visualisierung war, laut der Shaer, eine Stärke des TUI-Konzeptes, sogar mit extremen Zeitskalen in den Daten. Selbst für Gruppen mit unterschiedlichem Vorwissen (Studenten, Forscher, Kinder) sei die Präsentation mittels projizierter Tischflächen geeignet. Die Gruppen durften jedoch nicht zu groß werden und das Umschalten zwischen vielen Ansichten wurde noch nicht problemfrei gelöst. Auch eine Unzulänglichkeit des Systems sei die Anzahl von Objekten (Markern), sie versperren irgendwann die Sicht auf die Daten. [6] Als Lösung könnte man eine Multiuser-Implementierung der Remote Controls für wandgroße Displays konzipieren. Hier ist die Okklusion kein Problem mehr, da die Steuerung auf den Tablets stattfindet.

## 2.4 Interaktive intelligente Umgebung

In der Vision von Ubiquitous Computing wird eine interaktive, intelligente Umgebung beschrieben. Eine solche Umgebung hat gewisse Eigenschaften. Die perfekte Vernetzung von hunderten Geräten um Kollaboration und kontextuelle Arbeit zu ermöglichen. Die Pads, Tabs und Boards sind sich ihrer Lage bewusst. Auch in TUI-Konzepten sind interaktive, intelligente Umgebungen fester Bestandteil. Interaktive Flächen, TUI-Objekte und die Peripherie schaffen eine interaktive, und einigermaßen intelligente Umgebung. Die Ergänzungen Weisers über Calm Technology und das TUI Konzept führt außerdem gezielt Information und Interaktion in die Peripherie. Eine Interaktive Umgebung wie in Weisers Arbeit wäre ein guter Ort für TUI-Bedienelemente von Ishii und Ullmer. Bis jetzt gibt es noch keine vollständige Implementierung eines ganzheitlichen Konzeptes einer solchen interaktiven Umgebung. Wir befinden uns jedoch in der Endphase der PC-Ära und am Anfang der Ära des Ubiquitous Computing. [10] Kommerziell gibt es noch dennoch keine vollständig realisierte, intelligente Umgebung. Aus den bereits beschriebenen Projekten kann man sich aber ein Bild von einer möglichen ubiquitären Computing Umgebung machen. Im Jahr 2001 waren in einem durchschnittlichen US-amerikanischen Haushalt bereits über 40 Mikroprozessoren vorhanden. [10] Mittlerweile könnten es dank dem Trend zu Internet-of-things und intelligenter Objekte im Bereich Smart Home sicher schon viel mehr sein. Es gibt daher viele Möglichkeiten in einem durchschnittlichen Haushalt für die Verwendung von TUI.

## 3 Schlussfolgerungen, Kritik und Herausforderungen

Tangible UI Elemente haben enormes, unausgeschöpftes Potenzial. Die genannten Projekte zeigen welche Stärken und Schwächen das TUI als ergänzendes Bedienkonzept erreichen kann. Wenn die Vision des Ubiquitous Computings

mal erreicht werden soll, werden sicher viele mit Markern, Slidern, Phicons, und anderen TUI-Elementen vertraut sein, denn physische Objekte haben dank "Affordance" eine kleinere Lernkurve. Bei dem jetzigen Trend der immer größer werdenden Daten und der immer schwerer werdenden Gewinnung von Erkenntnissen aus eben diesen Daten, wird das Bedienkonzept der Tangible-User-Interfaces vieles erleichtern. Die intuitive Steuerung mittels TUIs und die erleichterte Kollaboration sind ein Vorteil. Ein weiterer ist dynamische Anfragen an Datensätze mittels TUI leichter und effizienter werden. Auch die Steuerung der Darstellung mittels haptischer Objekte ist mittels TUI verbessert. Die interaktive Umgebung und gezielter Einsatz der Peripherie schaffen Gelassenheit, bei gleichzeitig gesteigertem Informationsgehalt der Umgebung. Gegenüber der GUI und anderen Konzepten gibt es jedoch auch Nachteile, so skalieren derzeitige Implementierungen nicht mit einer großen Anzahl von Nutzern. Die Manipulation über verschiedene Ansichten hinweg bleibt teilweise problematisch und die physische Welt hat natürlich gegebene Limitierungen. So ist eine grafische Oberfläche viel anpassbarer als eine physische. Objekte erzielen auch oft eine Okklusion der Darstellung. Vieles lässt sich durch die Aussage Ullmers mitigieren, dass das TUI nicht gedacht ist, bestehende Konzepte zu verdrängen. [3] Die Lösung besteht darin die grafische Oberfläche nicht als Allheilmittel einzusetzen, sondern sich stattdessen an die Einfachheit und das physischen Objekten inherente Verständnis zu erinnern. So wird auch die Gefahr, vor der Mark Weiser gewarnt hat gebannt, dass man Obeflächen vielleicht bald nur noch mit Fokus auf Attraktivität entwerfe, anstatt die Nützlichkeit zu verbessern. In unserer Zeit, in der das Computing immer ubiquitärer wird, [10] ist ein vermehrter Einsatz von Tangible User Interfaces, wie beschrieben, überaus nützlich.

## 4 Zusammenfassung

Zusammenfassend ist das Konzept der TUIs stark vom Ubiquitären Computing inspiriert und ergänzt dieses in vielerlei Hinsicht. Auch Calm Technology, dass als Anhang zu Ubiquitärem Computing gesehen werden kann, profitiert von TUIs. Die genannten Projekte stellen Vorteile und Nachteile der TUIs heraus und verdeutlichen die direkte Referenz auf Ubiquitäres Computing von Mark Weiser. TUI hilft bei Datenverarbeitung und Erkenntnisgewinn, durch bessere Kollaboration, direkte Manipulation der Daten und deren Darstellung, vernetztes Arbeiten, stärkt die periphere Wahrnehmung und gibt ihr mehr Informationsgehalt. "Affordance" hilft TUIs auf Antrieb verständlicher und effizienter zu sein. TUI ist jedoch nur als Ergänzung sinnvoll und bringt auch einige Nachteile. Okklusion und fehlende Skalierung der Anwenderzahl sind bestehende Nachteile. In Zukunft können aber selbst diese Probleme durch

bessere Vernetzung und Standards (wie das TUIO Protokoll) gelöst werden.

---

### Literatur

---

- 1 George W Fitzmaurice, Hiroshi Ishii, and William AS Buxton. Bricks: laying the foundations for graspable user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 442–449. ACM Press/Addison-Wesley Publishing Co., 1995.
- 2 Romain Gabriel, Johan Sandsjö, Ali Shahrokni, and Morten Fjeld. Bouncelider: actuated sliders for music performance and composition. In *Proceedings of the 2nd international conference on Tangible and embedded interaction*, pages 127–130. ACM, 2008.
- 3 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*, pages 234–241. ACM, 1997.
- 4 Yvonne Jansen, Pierre Dragicevic, and Jean-Daniel Fekete. Tangible remote controllers for wall-size displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2865–2874. ACM, 2012.
- 5 Donald A Norman. Affordance, conventions, and design. *interactions*, 6(3):38–43, 1999.
- 6 Orit Shaer, Ali Mazalek, Brygg Ullmer, and Miriam Konkel. From big data to insights: opportunities and challenges for tei in genomics. In *Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction*, pages 109–116. ACM, 2013.
- 7 Brygg Ullmer and Hiroshi Ishii. The metadesk: models and prototypes for tangible user interfaces. In *Proceedings of the 10th annual ACM symposium on User interface software and technology*, pages 223–232. ACM, 1997.
- 8 Mark Weiser. The computer for the 21 st century. *Scientific american*, 265(3):94–105, 1991.
- 9 Mark Weiser and John Seely Brown. Designing calm technology. *PowerGrid Journal*, 1(1):75–85, 1996.
- 10 Mark Weiser and John Seely Brown. The coming age of calm technology. In *Beyond calculation*, pages 75–85. Springer, 1997.

# Eingabemedien zur bedeutungsvollen Interaktion mit TUIs

Melanie Hauser

Ludwig-Maximilians-Universität München, München, Deutschland  
melanie.hauser@campus.lmu.de

---

### Zusammenfassung

---

Heutzutage gewinnen eine Vielzahl neuer Technologien an Bedeutung, seien es ubiquitäre Computer, das Internet der Dinge oder am Körper tragbare Geräte. Diese neuartigen Computersysteme ermöglichen und erfordern teilweise sogar die Nutzung von Interaktionsformen, die über die traditionelle Vorstellung eines Eingabegeräts hinausgehen. Dabei spielen zunehmend auch greifbare Nutzerschnittstellen eine Rolle. Diese Tangible User Interfaces nutzen haptisch erfahrbare Objekte für die Interaktion mit einem Computersystem. Diese greifbaren Eingabemedien bieten eine enorme Bandbreite an Möglichkeiten, die kaum überschaubar ist. Um diese Vielfalt überschaubarer zu machen, wurde eine Taxonomie erstellt. Diese zeigt eine mögliche Kategorisierung der Eingabemedien in Tangible User Interfaces auf. In dieser Arbeit wird eine Vielzahl an Arbeiten kategorisieren, um einen Überblick zu geben.

**2012 ACM Computing Classification** Human-centered computing → Interaction devices

**Keywords and phrases** Eingabemedien, Nutzereingaben, bedeutungsvolle Interaktion mit Computersystemen, greifbare Nutzerschnittstellen / Tangible User Interfaces, Mensch-Maschine-Interaktion

## 1 Einleitung

Greifbare Nutzerschnittstellen oder Tangible UIs (von Engl.: Tangible User Interface, kurz TUI, vgl. [8]) bieten eine Vielzahl von Möglichkeiten. Unter anderem können Tangible UIs helfen, die Bedeutung der gezeigten Inhalte hervorzuheben, der Interaktion selbst Bedeutung hinzuzufügen oder eine emotionale Komponente zu erschaffen. Diese Punkte sind selten der Hauptzweck eines Tangible UIs, sondern entstehen meist als wünschenswertes Nebenprodukt durch die Wahl eines geeigneten Eingabemediums. Medium deshalb, weil in diesem Zusammenhang praktisch alles zur Eingabe verwendet werden kann, seien es Gegenstände oder auch Substanzen und Elemente wie Feuer,



© Melanie Hauser;

licensed under Creative Commons License CC-BY

Cite as: Melanie Hauser. Eingabemedien zur bedeutungsvollen Interaktion mit TUIs. 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. 2:1–2:14.

Wasser oder Erde. Die traditionelle Vorstellung eines Eingabegeräts ist daher für Tangible UIs zu eingeschränkt.

Ziel dieser Arbeit soll sein, diese Vielzahl an Möglichkeiten zu kategorisieren und einen Überblick zu geben, wie diese Interaktionsformen die Semantik des System, der Inhalte oder der Interaktion selbst unterstützen können.

Obwohl Tangible UIs eine so breite Palette an Möglichkeiten für Interaktionen ermöglichen, werden auch in diesen Systemen oftmals weniger bedeutungsvolle Interaktionsmedien verwendet. Aus rein pragmatischen Gesichtspunkten, sei es wegen besserer Kompatibilität oder Mangel an Zeit, sind handelsübliche Ein- und Ausgabegeräte häufig auch bei greifbaren Nutzerschnittstellen das Mittel der Wahl. Dennoch sollen hier nur solche Tangibles, das sind greifbare Interaktionsmedien für Tangible UIs, betrachtet werden, deren Verwendung auch einen semantischen Hintergrund hat.

Weiterhin bieten Tangible UIs vielerlei interessante Aspekte, in denen sie sich stark von anderen Interfaces unterscheiden. So kann beispielsweise das Display selbst eine taktile, 3-dimensionale Struktur besitzen oder die Ausgabe nicht nur rein visuelle Komponenten besitzen. Semantische Hintergründe spielen dabei an vielen Stellen eine Rolle, dennoch sollen an dieser Stelle nur die verschiedenen Medien betrachtet werden, die dem Nutzer für Eingaben zur Verfügung gestellt werden können.

Wie bereits erwähnt, ist die Menge denkbarer Eingabemedien sehr breit gefächert. Dadurch kann hier nur ein kleiner Einblick in die diversen Möglichkeiten gegeben werden. Gleichzeitig soll aber ein Rahmen geschaffen werden, der für viele dieser Möglichkeiten passend ist und somit zusätzliche Medien jederzeit leicht eingeordnet werden können.

## 2 Motivation

Gerade im Bereich der Mensch-Maschine-Interaktion gewinnt die User Experience, also die Nutzererfahrung, zunehmend an Bedeutung. Damit ist gemeint, wie angenehm und unterhaltsam die Bedienung eines Systems ist. Hierbei spielt auch die Erwartungshaltung eine Rolle. Im Kontext traditioneller Computersysteme, beispielsweise eines PCs oder eine Webseite, wird die User Experience besser, je eher das System so auf Eingaben reagiert, wie der Nutzer es erwartet. Da Tangible UIs aber oftmals auch eine künstlerische Komponente haben und weniger zur reinen Informationsbeschaffung oder zu anderen produktiven Tätigkeiten genutzt werden, trifft dies für Tangible UIs nicht immer zu. Besonders in greifbaren Systemen, die eine spielerische Komponente haben oder den Nutzer zum Nachdenken anregen sollen, können unvorhergesehene Reaktionen der User Experience ebenso zuträglich sein. Einige Beispiele für solche Systeme, die eher im Bereich der Kunst oder Musik als der traditionellen

Informationstechnik angesiedelt werden können, werden im späteren Verlauf dieser Arbeit aufgezeigt und erläutert.

Gleichzeitig spielt, wenn man von der User Experience spricht, auch Emotionalität eine Rolle. Heutzutage werden vor allem mobile Geräte nicht als reines Mittel zum Zweck gesehen. Sie sollen möglichst individualisierbar sein, ihren Besitzer „kennen“ und sind als ständige Begleiter für viele Menschen kaum noch aus ihrem Leben wegzudenken. Die Geräte haben oftmals einen ähnlichen persönlichen Stellenwert wie ein Tagebuch und werden durch die Verbreitung intelligenter Assistenten auch immer mehr vermenschlicht. Gerade hier findet sich großes Potenzial zur Verbesserung der User Experience, indem positive Gefühle und Erfahrungen bewusst betont werden. Der Einfluss von Gefühlen und vorangegangenen Erfahrungen auf das Konsumverhalten im Allgemeinen wurde bereits seit den späten 70er Jahren analysiert, unter anderem von Hirschman und Holbrook [10]. Aber auch aktuellere Studien, die speziell Erfahrungen von Nutzer im Bereich der Mensch-Maschine-Interaktion betrachten, betonen die Auswirkungen von Erfahrungen auf die Wahrnehmung von Produkten und Systemen [7, 15].

Bei realen Alltagsgegenständen oder Stoffen kann dieser Effekt besonders deutlich beobachtet werden, da hier die Meinung und Erwartungshaltung gegenüber dem Produkt regelmäßig erneuert und gegebenenfalls angepasst werden.

Diese beiden Aspekte, die Erwartungshaltung des Nutzers und die Bekanntheit des Mediums sowie die damit einhergehenden Konnotationen, bieten gleichzeitig eine Herausforderung, aber auch Potenzial für die Nutzung greifbarer Interaktionselemente in Tangible UIs. Zwar gibt es weitere Faktoren, die die User Experience beeinflussen können, aber diese beiden können in Tangible UIs besonders gut ausgenutzt werden und sollen deshalb im Folgenden vorrangig betrachtet werden.

## 3 Verwandte Arbeiten

Menschen neigen dazu, die Dinge zu ordnen und abstrakte Sammelbegriffe zu finden (vgl. [17]). Obwohl dabei die Gefahr der Oberflächlichkeit besteht, führen solche Kategorisierungen zu besserem Verständnis und schnelleren Übersicht. Es gibt daher viele weitere Arbeiten, die sich mit dem Entwurf von Taxonomien rund um Tangible UIs beschäftigen. Meist werden dabei aber nicht die greifbaren Nutzerschnittstellen im Ganzen analysiert, sondern lediglich ein klar abgegrenztes Teilgebiet.

So gibt etwa Döring einen Überblick über eine spezielle Art von Tangible UIs, nämlich denen, die eine flüchtige oder zeitlich begrenzte Komponente mit sich bringen [4]. Dabei wird beispielsweise Eis als Interaktions- und Dar-

stellungsmedium betrachtet. Da Eis bei entsprechenden Temperaturen zu schmelzen beginnt, sind aus Eis gehauene Gebilde vergänglich und können dadurch nur für begrenzte Zeit verwendet werden.

Auch Hurtienne und Israel haben eine Taxonomie zur Kategorisierung von Tangible UIs entworfen [12]. Der Fokus liegt dabei allerdings weniger auf den Gegenständen und Materialien, die für die Interaktion verwendet werden, sondern vielmehr auf den Strategien, die diesen Interaktionen zugrunde liegen.

Bei Fishkin werden Tangible UIs nach der Distanz zwischen Ein- und Ausgabe sowie nach Art der Metapher unterschieden [5]. Metapher bedeutet in dem Fall, in welchem Grad ein realer Gegenstand als Vorbild für die Tangible UI Elemente genutzt wird. Dieser Ansatz der Metapher wird in der vorliegenden Arbeit aufgegriffen. Als zweiter Einteilungsaspekt wurde aber ein anderer Gesichtspunkt gewählt als in der Arbeit von Fishkin.

Man findet weitere Arbeiten, in denen eine spezielle Art von Tangible UI oder gewisse Teile derer betrachtet und ein Versuch der Kategorisierung vorgenommen wird [3, 9, 11]. Die potenziellen Interaktionsgegenstände, speziell die für die Eingabe, stehen dabei als solche jedoch nicht im Fokus. Außerdem wird der Einfluss, den die Wahl des Mediums auf die Nutzererfahrung haben kann, in den bereits genannten Arbeiten bei der Einteilung nicht berücksichtigt. Daher soll hier, ausgehend von den Überlegungen zur User Experience, eine Übersicht über das Spektrum der Eingabemedien entstehen.

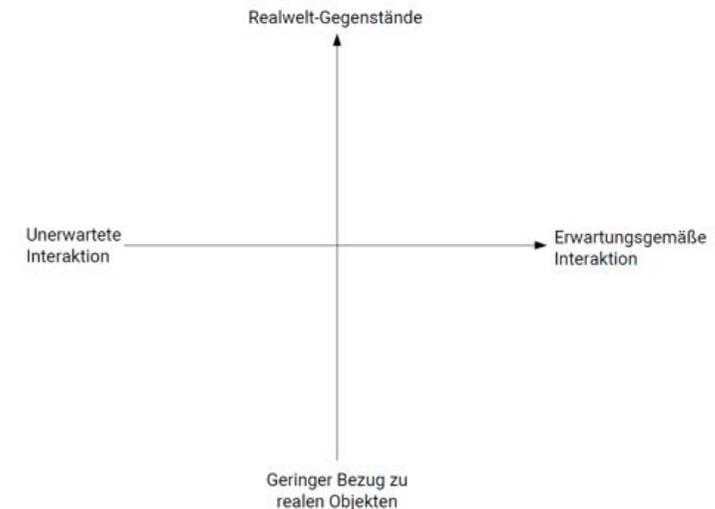
#### 4 Einteilung der Interaktionsmittel

In dieser Arbeit soll eine mögliche Einteilung von Eingabemedien betrachtet werden. Dabei sollte erwähnt werden, dass besonders im Fall von Tangible UIs Eingabe, Ausgabe und Gesamtsystem nicht immer von einander getrennt werden können. Oft wird dasselbe Medium für Ein- und Ausgabe genutzt. Ähnlich wie bei einem Smartphone, bei dem das Display Eingabe in Form von Touch Input, Ausgabe durch Vibration sowie die grafische Darstellung der Informationen vereint, übernimmt auch in Bereich der Tangible UIs oft eine Komponente mehrere dieser Aufgaben.

Realweltgegenstände bieten großes Potenzial für die Verwendung in Tangible UIs, sei es als Vorlage für ein Eingabemedium oder selbst als Teil des Systems. Da es sich hierbei zumeist um bekannte Objekte mit einem fest zugewiesenen Zweck handelt, hat der Nutzer eine gewisse Vorstellung davon, wofür der Gegenstand sinnbildlich steht, welche Aktionen an ihm durchgeführt werden können und eventuell auch, welche Reaktionen zu erwarten sind. Diese Erwartungshaltung des Nutzers kann man gut ausnutzen. Somit können sehr intuitiv nutzbare Eingabemedien erschaffen werden, die keinerlei oder sehr wenig weiterer Erklärung bedürfen. Andererseits bietet das Wissen um die

Funktion aber auch die Möglichkeit, mit der Erwartungshaltung des Nutzers zu spielen, diese zu durchkreuzen und überraschende neue Funktionen an die altbekannten Gegenstände zu binden.

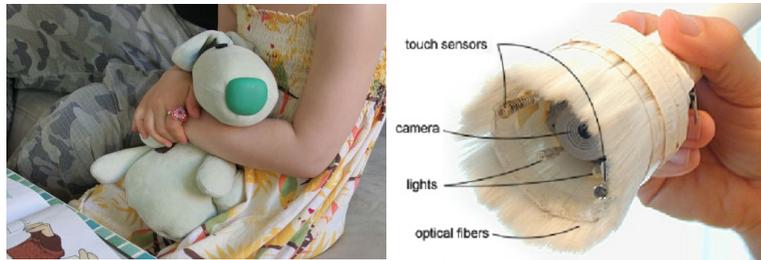
Darüber hinaus können Erwartungen, die der Nutzer an das Objekt hat, nicht nur Auswirkungen auf den ersten Eindruck und die Nutzbarkeit eines unbekanntes Systems haben. Auch bei langfristiger Verwendung spielen die anfängliche Erwartungshaltung und Annahmen, die der Nutzer vor der eigentlichen Interaktion getroffen hat, eine Rolle, selbst, wenn diese sich in der Benutzung nicht bewahrheitet haben [14].



■ **Abbildung 1** Die vorgeschlagene Kategorisierungsmatrix. Die X-Achse beschreibt, inwieweit die Erwartung, die der Nutzer an das Medium hat, erfüllt wird. Auf der Y-Achse findet sich der Realitätsgrad des Objekts wieder.

Zur Beschreibung der Eingabemedien wird deshalb eine 2-dimensionale Einordnung vorgeschlagen. Wie auch in Abbildung 1 zu sehen, wird auf der ersten Achse beschrieben, inwieweit der Bestimmungszweck der Vorlage bei der Eingabe abgebildet wird. An einem Ende der Skala findet man Medien, die bei der Benutzung Überraschung oder sogar (gewollte) Verwirrung verursachen, weil sie eine andere Art der Verwendung voraussetzen als erwartet. Am anderen Ende kann man all jene Medien ansiedeln, die die Funktion ihrer realen Vorbilder vollständig übernehmen.

Die zweite Achse zeigt an, wie sehr das Medium an sein Vorbild angelehnt



■ **Abbildung 2** Links: Plüschtier Sniff von Johansson [13]. Rechts: Aufbau der I/O Brush von Ryokai [20].

ist. Hier reicht das Spektrum von Gegenständen und Stoffen aus der realen Welt, die unverändert als TUI Element genutzt werden, bis hin zu komplett neu erschaffenen Gegenständen. Dies ist heutzutage industriell auf vielfältige Weisen möglich.

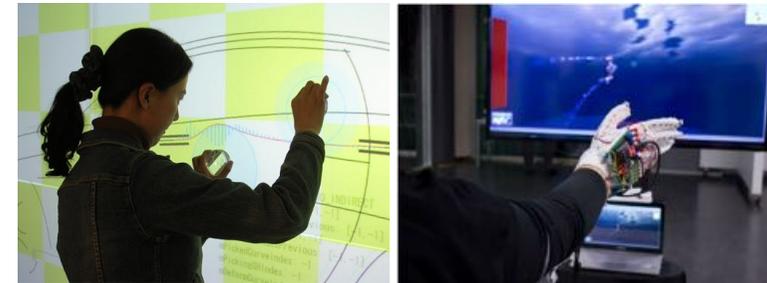
#### 4.1 Durchkreuzung der Erwartungshaltung

Am linken Ende der X-Achse findet man solche Eingabemedien, die die Erwartungshaltung komplett durchkreuzen oder der bekannten Funktion ein neues Element hinzufügen. Letzteres findet man z. B. bei Schoemann [21]. In dieser Arbeit wird mit einer Nadel, mit der gestickt wird, gleichzeitig Musik erzeugt. So wird ein bekanntes Konzept mit einer weiteren künstlerischen Tätigkeit verknüpft.

#### 4.2 Erfüllung der Erwartungshaltung

Am anderen Ende der horizontalen Achse findet man das genaue Gegenteil von Unterabschnitt 4.1. Bekannte Alltagsgegenstände erfüllen genau den Zweck, den man ihnen intuitiv zuordnen würde.

Kinderspielzeug als Eingabemedium, wie etwa bei Marco [16], kann meistens hier eingeordnet werden. Hier wird klassisches Spielen gefördert und dieses gleichzeitig durch digitale Repräsentationen unterstützt. Auch die emotionale Komponente spielt eine wichtige Rolle. Auch in „Sniff: Designing characterful interaction in a tangible toy“ [13] wird ein typisches Plüschtier per RFID um neue Interaktionen erweitert (Abbildung 2). Ein RFID-Lesegerät in der Schnauze des Plüschtiers kann RFID-Tokens erkennen, die an beliebigen Orten und Gegenständen angebracht werden können. Unter anderem kann der Plüschhund dann durch Vibration und akustische Ausgaben seine Freude über



■ **Abbildung 3** Links: Nutzerin bei der Interaktion mit einem Drehregler zur Veränderung der angezeigten Kurve. Rechts: Der haptische Controller Handschuh von Footit. Er wird genutzt, um verschiedene Elemente eines Computerspiels durch Gesten zu manipulieren.

das „erschnüffelte“ Objekt zum Ausdruck bringen.

Die I/O Brush [20] von Ryokai (Abbildung 2) hat Elemente von beiden Enden der Skala und kann deshalb eher mittig eingeordnet werden. Man malt auf einer - hier digitalen - Leinwand, kann aber neben Farbe auch Textur und Bewegung mit dem Pinsel aufnehmen und abbilden.

Knöpfe, Regler und Schalter werden häufig in diversen Kontexten zur Eingabe genutzt. Ihre Funktion ist von verschiedenen Geräten oder grafischen Oberflächen bekannt, wodurch diese im Bereich der greifbaren Interaktion häufig genutzt werden, um intuitive und effiziente Interaktionen zu ermöglichen. Aber auch im Kontext von Tangible UIs werden solche Objekte häufig verwendet. Dabei bieten Tangibles gegenüber ihrer Vorbilder eine größere Flexibilität. Beispielsweise können sie, je nach Bedarf, unterschiedlich platziert werden, wodurch ein Regler verschiedene Teile der Oberfläche ansteuern kann. Außerdem können so auch kontextabhängig unterschiedliche Funktionen in einem Eingabemedium vereint werden.

Solche Gegenstände bieten sich vor allem bei Inhalten technischer Natur an, zum Beispiel werden in dem Paper „Tangible nurbs-curve manipulation techniques using graspable handles on a large display“ Drehregler zur Manipulation von NURBS-Kurven genutzt [1], zu sehen in Abbildung 3. Bei de Siqueira et al. wird die Interaktion mit digitalen wissenschaftlichen Postern durch Tangibles ergänzt [2].

#### 4.3 Gegenstände aus der realen Welt

Oftmals können reale Gegenstände oder Materialien ohne jegliche Modifikation oder nur durch Hinzufügen unauffälliger Sensoren oder ähnlichem in Tangible

UIs genutzt werden. Auf die Spitze getrieben bedeutet das, dass die Welt, wie sie ist, und jeder Gegenstand auf ihr das Tangible ist. Jedes beliebige Objekt kann dann zur Eingabe genutzt werden. Peng setzt dies auf die Art um, dass Musik erzeugt werden kann, indem beliebige Gegenstände und Oberflächen berührt werden [18]. Abhängig von unterschiedlichen Attributen wie Material oder Farbe werden durch die Berührung verschiedene Töne erzeugt.

Bei Poupyrev bilden interaktive Pflanzen ein Gesamtsystem, das eine Rückbesinnung zur Natur unterstützen soll [19]. Diese sind dabei mit kaum sichtbarer Sensorik ausgestattet, wodurch diverse Interaktionen ermöglicht werden. Man kann sich den Pflanzen annähern oder über Blätter und Stängel streichen. Durch diese Interaktionen werden zum Beispiel verschiedene Lichteffekte hervorgerufen.

Auch Wasser erweist sich als vielseitiges Eingabemedium, das ohne Veränderung in technischen Systemen eingesetzt werden kann. Unter anderem setzt Yonezawa es als musikalisches Instrument ein [22].

Generell kann man festhalten, dass besonders in dieser Kategorie die Semantik der Eingabemedien eine große Rolle spielt. Durch den hohen Grad an Realität, den die Medien widerspiegeln, hat der Nutzer ein sehr konkretes Bild von der Interaktion. Dieses Bild wird geformt durch die Erfahrungen, die der Nutzer mit dem Gegenstand oder Stoff in der realen Welt gemacht hat, durch damit verbundene Gefühle oder Erinnerungen. Die Interaktion mit einem Tangible kann diese Emotionen und Erwartungen aufgreifen und die Sicht des Nutzers auf das gesamte System verändern. Diese Stärke von greifbaren gegenüber klassischer Eingabemedien kann den eigentlich Zweck der Interaktion unterstreichen, kann aber auch unabhängig davon die Nutzererfahrung verbessern.

#### 4.4 Gegenstände nach realem Vorbild

Trotz dieser Vorteile von besonders naturalistischen Tangibles, kann je nach Anwendungsszenario auch ein Gegenstand nach realem Vorbild von Nutzen sein. Der Gaming Handschuh von Footit [6] etwa wird als Spiele-Controller genutzt. Er ist klar als Handschuh erkennbar, wie auch in Abbildung 3 zu sehen ist, aber der semantische Fokus liegt eher auf den Händen. Der Nutzer bewegt die Spielelemente mit den Händen. Trotzdem erfüllt der Handschuh auch auf der Bedeutungsebene einen Zweck, da klar ersichtlich ist, wie dieser zu tragen ist und eine Brücke zwischen der Sensorik im Handschuh und der händischen Interaktion geschaffen wird.

### 5 Weitere Möglichkeiten der Kategorisierung

Die hier vorgestellte Kategorisierung ist nur eine von vielen möglichen Arten der Einordnung von Tangibles. Da sehr unterschiedliche Faktoren die Wahl des Eingabemediums beeinflussen können, gibt es diverse weitere denkbare Taxonomien. Zudem erschwert die enorme Bandbreite von Tangible UIs und auch von nutzbaren Eingabemedien im Speziellen den Entwurf einer solchen Taxonomie, da nicht nur umfassende Kategorien gefunden werden sollen, sondern auch die Anzahl dieser Kategorien überschaubar bleiben soll.

Ein Attribut, das zur Einordnung von Tangibles genutzt werden könnte, wäre der Zustand des Materials. Hierbei könnte zwischen den klassischen chemischen Zuständen fest, flüssig und gasförmig unterschieden werden, aber auch weitere Unterscheidungen wären möglich. Beispielsweise könnten, wie in einer der bereits erwähnten Arbeiten [4], flüchtige Stoffe und, im Gegensatz dazu, persistente eine Kategorie bilden. Auch harte oder weiche beziehungsweise nachgiebige Oberflächen könnten zur Einteilung herangezogen werden.

Weiterhin spielt bei jeglichen Systemen und so auch bei TUIs oftmals der Ort oder Kontext eine Rolle. So sind gewisse Interaktionen nur möglich oder sinnvoll, wenn sie an einem bestimmten Standort (draußen, in dunklen Räumen, an öffentlichen Plätzen) stattfinden. Eine Installation mit Lichteffekten beispielsweise wird bei grellem Tageslicht an Eindrucksstärke verlieren, ebenso wie ein öffentliches Display in einer privaten Wohnung seinen Bestimmungszweck nicht erfüllen kann. Aber auch die Zahl der anwesenden Nutzer, die Tageszeit oder der Systemstatus können die Interaktion beeinflussen und somit eine Einteilung für die so veränderlichen Interaktionsmedien bieten.

Zuletzt können auch die Nutzergruppen, auf die die Interaktionen abzielen, Kategorien bilden. Sollen vor allem junge oder alte Menschen die Tangibles benutzen? Personen, die erfahren im Umgang mit Tangible UIs sind oder solche, deren besondere Bedürfnisse durch das Interaktionsmedium erfüllt werden sollen?

Diese Aufzählung ließe sich noch beliebig ausweiten. Festzuhalten bleibt, dass es eine Vielzahl an Möglichkeiten gibt, greifbare Interaktionsmedien zu unterteilen und einzuordnen. Welche davon für ein bestimmtes System oder Anwendungsszenario den besten Überblick geben könnten oder überhaupt geeignet wären, hängt von vielen Faktoren ab, unter anderem letztlich auch von den persönlichen Präferenzen des Systemdesigners.

### 6 Diskussion

In dieser Arbeit ist eine Einteilung greifbarer Interaktionsmedien, wie sie für Nutzereingaben im Zusammenhang greifbarer Nutzerschnittstellen verwendet

werden, entstanden. Diese ist eine von diversen theoretisch möglichen Einteilungen und hat den Anspruch, einen groben Überblick über die Gegenstände und Stoffe zu geben, die für die Interaktion mit Tangible UIs verwendbar wären.

In Abbildung 4 ist zu sehen, wie die genannten Beispiele sich in die vorgeschlagene Übersicht einfügen würden. Im Allgemeinen kann man sagen, dass Medien aus dem linken oberen Quadranten die zuvor beschriebenen Effekte auf die User Experience am stärksten ausnutzen. Das heißt allerdings nicht, dass diese Medien zwangsläufig die beste User Experience zur Folge haben. Vielmehr muss man bedenken, dass vor allem unvorhersehbare Komponenten schnell das Gegenteil bewirken können. Künstlerische Systeme können von dieser Kategorie beispielsweise stark profitieren. Aber je weiter das genutzte System sich vom künstlerischen Selbstzweck hin zu produktiven Funktionen bewegt, umso eher kann eine andere Kategorie einen größeren Nutzen bieten. Systeme hingegen, die produktive Prozesse wie Informationsbeschaffung oder die Erstellung von Dokumenten unterstützen sollen, müssen vor allem intuitiv bedienbar sein. Überraschende Komponenten würden hier eher zu Verwirrung und Frust führen als zu einer gesteigerten User Experience.

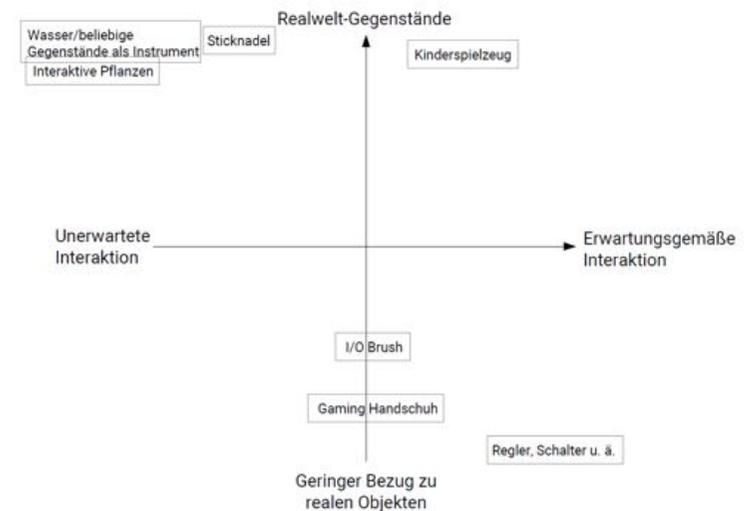
Um die Nützlichkeit dieser entwickelten Kategorisierung zu evaluieren, wären praktische Anwendungen derselben notwendig. Bei Einbeziehung in den Systementwurf würde sich zeigen, ob eine Betrachtung der Kategorien den Designprozess beschleunigen oder das resultierende System verbessern kann. Darüber kann auf dem aktuellen Stand noch keine Aussage getroffen werden.

## 7 Zusammenfassung und Ausblick

In dieser Arbeit wurde ein Überblick über mögliche Einteilungen von greifbaren Eingabemedien in Tangible UIs gegeben und eine konkrete Kategorisierung ausgearbeitet. Diese nutzt zwei Aspekte der Medien, um diese innerhalb eines 2-dimensionalen Koordinatensystems einzuordnen.

Diese Kategorisierung darf allerdings lediglich als Hilfestellung verstanden werden, um einen ersten Überblick über die vielfältigen Möglichkeiten zu gewinnen. Sie ist kein Ratgeber, der besagt, welches Medium oder welche Kategorie für ein konkretes Szenario am Besten geeignet ist. Dies muss individuell von Fall zu Fall entschieden werden, zumal es häufig nicht nur *das* eine perfekte Medium gibt. Eine solche generalisierte Übersicht kann separate Abwägungen für das jeweilige System nicht ersetzen.

Es hat sich gezeigt, dass eine genaue Einteilung entlang der Achsen nur bedingt möglich ist. Zum einen liegt die genaue Positionierung teilweise im Auge des Betrachters, da das subjektive Empfinden des Tangibles hierbei eine Rolle spielt. Zum anderen setzt ein einzelnes Medium manchmal mehrere



■ **Abbildung 4** Die vorgeschlagene Kategorisierungsmatrix. Die beispielhaften Interaktionsmedien aus dem vorangegangenen Kapitel wurden anhand der beiden gewählten Kriterien positioniert.

Funktionalitäten um, sodass es je nach Funktionalität verschiedenen Kategorien zugeordnet werden könnte.

Zuletzt sollte nicht unerwähnt bleiben, dass technische Aspekte, die hier keine Beachtung fanden, in der realen Umsetzung natürlich eine wesentliche Rolle spielen können. So ist die Umsetzbarkeit bei den meisten Projekten von größerer Wichtigkeit als die Semantik der Eingabemedien. In Sachen Bedeutsamkeit sehr relevante Gegenstände sind zwar erstrebenswert, rechtfertigen aber erhöhten Arbeitsaufwand und -kosten nicht in allen Fällen.

Die letztliche Entscheidung, welches Medium für die greifbare Eingabe genutzt werden sollte, kann nicht allein aufgrund einer Kategorisierung getroffen werden. Diese kann lediglich einen groben Überblick geben und somit einen Beitrag zur endgültigen Lösung leisten.

### Literatur

- 1 Seok-Hyung Bae, Takahiro Kobayash, Ryugo Kijima, and Won-Sup Kim. Tangible mubs-curve manipulation techniques using graspable handles on a large display. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology*, UIST '04, pages 81–90,

- New York, NY, USA, 2004. ACM. URL: <http://doi.acm.org.emedien.ub.uni-muenchen.de/10.1145/1029632.1029646>, doi:10.1145/1029632.1029646.
- 2 Alexandre G. de Siqueira, Brygg Ullmer, Mark Delarosa, Chris Branton, and Miriam K. Konkel. Hard and soft tangibles: Mixing multi-touch and tangible interaction in scientific poster scenarios. In *Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '18, pages 476–486, New York, NY, USA, 2018. ACM. URL: <http://doi.acm.org.emedien.ub.uni-muenchen.de/10.1145/3173225.3173252>, doi:10.1145/3173225.3173252.
  - 3 Tanja Döring. The interaction material profile: Understanding and inspiring how physical materials shape interaction. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '16, pages 2446–2453, New York, NY, USA, 2016. ACM. URL: <http://doi.acm.org.emedien.ub.uni-muenchen.de/10.1145/2851581.2892516>, doi:10.1145/2851581.2892516.
  - 4 Tanja Döring, Axel Sylvester, and Albrecht Schmidt. A design space for ephemeral user interfaces. In *Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction*, TEI '13, pages 75–82, New York, NY, USA, 2013. ACM. URL: <http://doi.acm.org.emedien.ub.uni-muenchen.de/10.1145/2460625.2460637>, doi:10.1145/2460625.2460637.
  - 5 Kenneth P. Fishkin. A taxonomy for and analysis of tangible interfaces. *Personal Ubiquitous Comput.*, 8(5):347–358, September 2004. URL: <http://dx.doi.org.emedien.ub.uni-muenchen.de/10.1007/s00779-004-0297-4>, doi:10.1007/s00779-004-0297-4.
  - 6 Jacques Foottit, Dave Brown, Stefan Marks, and Andy M. Connor. An intuitive tangible game controller. In *Proceedings of the 2014 Conference on Interactive Entertainment*, IE2014, pages 4:1–4:7, New York, NY, USA, 2014. ACM. URL: <http://doi.acm.org.emedien.ub.uni-muenchen.de/10.1145/2677758.2677774>, doi:10.1145/2677758.2677774.
  - 7 Jodi Forlizzi and Katja Battarbee. Understanding experience in interactive systems. In *Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, DIS '04, pages 261–268, New York, NY, USA, 2004. ACM. URL: <http://doi.acm.org.emedien.ub.uni-muenchen.de/10.1145/1013115.1013152>, doi:10.1145/1013115.1013152.
  - 8 Tangible Media Group. Tangibles at play. In *ACM SIGGRAPH 2006 Emerging Technologies*, SIGGRAPH '06, New York, NY, USA, 2006. ACM. URL: <http://doi.acm.org.emedien.ub.uni-muenchen.de/10.1145/1179133.1179166>, doi:10.1145/1179133.1179166.
  - 9 Jonna Häkkinä and Ashley Colley. Towards a design space for liquid user interfaces. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*, NordiCHI '16, pages 34:1–34:4, New York, NY, USA, 2016. ACM. URL: <http://doi.acm.org.emedien.ub.uni-muenchen.de/10.1145/2971485.2971537>, doi:10.1145/2971485.2971537.
  - 10 Elizabeth C Hirschman and Morris B Holbrook. Hedonic consumption: emerging concepts, methods and propositions. *The Journal of Marketing*, pages 92–101, 1982.
  - 11 Lars Erik Holmquist, Johan Redström, and Peter Ljungstrand. Token-based access to digital information. In *Proceedings of the 1st International Symposium on Handheld and Ubiquitous Computing*, HUC '99, pages 234–245, London, UK, UK, 1999. Springer-Verlag. URL: <http://dl.acm.org/citation.cfm?id=647985.743869>.
  - 12 Jörn Hurtienne and Johann Habakuk Israel. Image schemas and their metaphorical extensions: Intuitive patterns for tangible interaction. In *Proceedings of the 1st International Conference on Tangible and Embedded Interaction*, TEI '07, pages 127–134, New York, NY, USA, 2007. ACM. URL: <http://doi.acm.org.emedien.ub.uni-muenchen.de/10.1145/1226969.1226996>, doi:10.1145/1226969.1226996.
  - 13 Sara Johansson. Sniff: Designing characterful interaction in a tangible toy. In *Proceedings of the 8th International Conference on Interaction Design and Children*, IDC '09, pages 186–189, New York, NY, USA, 2009. ACM. URL: <http://doi.acm.org.emedien.ub.uni-muenchen.de/10.1145/1551788.1551824>, doi:10.1145/1551788.1551824.
  - 14 Sari Kujala and Talya Miron-Shatz. The evolving role of expectations in long-term user experience. In *Proceedings of the 19th International Academic MindTrek Conference*, AcademicMindTrek '15, pages 167–174, New York, NY, USA, 2015. ACM. URL: <http://doi.acm.org.emedien.ub.uni-muenchen.de/10.1145/2818187.2818271>, doi:10.1145/2818187.2818271.
  - 15 Effie Lai-Chong Law, Virpi Roto, Marc Hassenzahl, Arnold P.O.S. Vermeeren, and Joke Kort. Understanding, scoping and defining user experience: A survey approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 719–728, New York, NY, USA, 2009. ACM. URL: <http://doi.acm.org.emedien.ub.uni-muenchen.de/10.1145/1518701.1518813>, doi:10.1145/1518701.1518813.
  - 16 Javier Marco, Eva Cerezo, Sandra Baldasarri, Emanuela Mazzone, and Janet C. Read. User-oriented design and tangible interaction for kindergarten children. In *Proceedings of the 8th International Conference on Interaction Design and Children*, IDC '09, pages 190–193, New York, NY, USA, 2009. ACM. URL: <http://doi.acm.org.emedien.ub.uni-muenchen.de/10.1145/1551788.1551825>, doi:10.1145/1551788.1551825.
  - 17 Craig McGarty. *Categorization in social psychology*. Sage, 1999.
  - 18 Huaishu Peng. Touchsound: Making sounds with everyday objects. In *Proceedings of the Fifth International Conference on Tangible, Embedded,*

- and Embodied Interaction*, TEI '11, pages 439–440, New York, NY, USA, 2011. ACM. URL: <http://doi.acm.org.emedien.ub.uni-muenchen.de/10.1145/1935701.1935820>, doi:10.1145/1935701.1935820.
- 19 Ivan Poupyrev, Philipp Schoessler, Jonas Loh, and Munehiko Sato. Botanicus interacticus: Interactive plants technology. In *ACM SIGGRAPH 2012 Emerging Technologies*, SIGGRAPH '12, pages 4:1–4:1, New York, NY, USA, 2012. ACM. URL: <http://doi.acm.org.emedien.ub.uni-muenchen.de/10.1145/2343456.2343460>, doi:10.1145/2343456.2343460.
- 20 Kimiko Ryokai, Stefan Marti, and Hiroshi Ishii. I/o brush: Drawing with everyday objects as ink. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pages 303–310, New York, NY, USA, 2004. ACM. URL: <http://doi.acm.org.emedien.ub.uni-muenchen.de/10.1145/985692.985731>, doi:10.1145/985692.985731.
- 21 Sarah Schoemann and Michael Nitsche. Needle as input: Exploring practice and materiality when crafting becomes computing. In *Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '17, pages 299–308, New York, NY, USA, 2017. ACM. URL: <http://doi.acm.org.emedien.ub.uni-muenchen.de/10.1145/3024969.3024999>, doi:10.1145/3024969.3024999.
- 22 Tomoko Yonezawa and Kenji Mase. Tangible sound: musical instrument using fluid water. In *Proceedings of the 2000 International Computer Music Conference, Berlin*, ICMC 2000. Citeseer, 2000. URL: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.22.1147&rep=rep1&type=pdf>.

# Explizite Mensch-Computer-Interaktion mittels Eye-Tracking

Christian Gregor Mall

Ludwig-Maximilians-Universität München, München, Deutschland  
c.mall@campus.lmu.de

## Zusammenfassung

Diese Arbeit soll einen Überblick über explizite Interaktionstechniken mittels Eye-Tracking bieten. Diese Techniken stellen nicht nur für körperlich eingeschränkte Menschen, sondern auch für alle anderen eine neue und zum Teil bessere Form der Mensch-Computer-Interaktion dar, z. B. in Umgebungen, die einen hohen Hygienestandard voraussetzen, wie beispielsweise in OP-Sälen. Nach einer kurzen allgemeinen Einführung in das Thema Eye-Tracking werden drei verschiedene Methoden vorgestellt, wie man die Verfolgung der Augenbewegungen zur Interaktion mit Computern einsetzen kann. Danach wird darauf eingegangen, auf welche Weise man das Midas-Touch-Problem bereits beim Interface Design lösen kann sowie aufgezeigt, wie man das Auswählen von Interfaceelementen einfacher gestalten kann. Es folgt ein Abschnitt über das Thema Eye-Tracking in Kombination mit anderen Inputmodalitäten ehe in der Zusammenfassung resümiert wird, dass jede der vorgestellten Interaktionstechniken Stärken und Schwächen und somit ihre Daseinsberechtigung hat.

**2012 ACM Computing Classification** Human-centered computing → Interaction techniques

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

## 1 Einleitung

Eye-Tracking-Systeme werden schon seit dem späten 19. Jahrhundert eingesetzt. Um sicherzustellen, dass die Probanden ihre Köpfe nicht bewegten, verwendete Huey [15] (inspiriert von Ahrens [2]) Eisenstützen. Zuerst wurde das Auge mit Kokain betäubt (ebenso wie bei Delabarre [4]). Dann wurde an der Hornhaut ein mechanisch mit einem Schreiber verbundener Ring fixiert, um die Augenbewegungen aufzuzeichnen. Erst zu Beginn des 20. Jahrhunderts änderten sich diese eher barbarisch anmutenden Methoden und Eye-Tracking wurde für die Probanden komfortabler. Dodge und Klein [5] fotografierten



© Christian Gregor Mall;

licensed under Creative Commons License CC-BY

Cite as: Christian Gregor Mall. Explizite Mensch-Computer-Interaktion mittels Eye-Tracking. 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. 3:1–3:20.

die Reflexion einer externen Lichtquelle von der Fovea und revolutionierten auf diese Weise die Messung von Augenbewegungen. Heutige Eye-Tracking-Systeme verwenden zur Messung der Augenbewegungen meistens die Pupil-and-corneal-reflection-Methode. Für Blickfassungssysteme ist dabei vor allem die sogenannte 1. Purkinjereflexion relevant. Bei Purkinjereflexionen handelt es sich um Lichtreflexionen der Hornhaut bzw. Linse, welche durch eine oder mehrere Infrarotlichtquelle(n) hervorgerufen werden [12]. Des Weiteren unterscheidet man auch zwischen statischen Systemen (Tower-Mounted-Systeme, bei denen der Kopf des Probanden am Gerät fixiert wird oder Remote-Eye-Tracker, bei denen der Benutzer seinen Kopf in einem gewissen Radius frei bewegen kann) und kopfgetragenen Eye-Tracking-Systemen, welche ein höheres Maß an Mobilität erlauben. [12]

Für die Benutzung eines PCs hat sich seit dessen Einführung klar die Interaktion mittels Maus und Tastatur durchgesetzt. Nicht ohne Grund: Für Operationen in modernen Anwendungen bietet eine Maus viele wünschenswerte Eigenschaften. Sie ist bei der Auswahl von Buttons oder Icons sehr präzise, was es dem Benutzer möglich macht, auch sehr kleine Elemente sicher auszuwählen. Explizite Interaktionstechniken mittels Eye-Tracking ermöglichen auch Personen mit motorischen Einschränkungen, einen Computer zu bedienen. Auch in Bereichen, wo Hygiene von kritischer Wichtigkeit ist, z.B. bei Operationen, können Interaktionstechniken mittels Eye-Tracking ideal eingesetzt werden [19].

Da Menschen dazu neigen, ihren Blick den Objekten zuzuwenden, die sie interessieren, könnten aber auch alle anderen Benutzer von dieser intuitiveren Form der Interaktion profitieren. Von explizitem Input mittels Eye-Tracking spricht man, wenn Personen Augenbewegungen bewusst einsetzen, um einen Computer zu bedienen. Davon abzugrenzen sind beispielsweise Anwendungen, die auf die Blickbewegungen des Nutzers reagieren aber keine explizite Steuerung von ihm verlangen [26].

Eine Schwierigkeit dabei ist beispielsweise die Tatsache, dass die meisten Benutzeroberflächen nicht für diese Art der Interaktion ausgelegt sind, sodass die im Vergleich zu einer herkömmlichen Steuerung per Maus geringere Genauigkeit zum Problem werden kann. Eine weitere Herausforderung stellt die Tatsache dar, dass eine Blickzuwendung auf ein Element nicht zwangsläufig bedeuten muss, dass die Intention des Benutzers dessen Auswahl ist [17]. Wenn ein Benutzer mit seinem Blick längere Zeit ein Objekt fixiert, weil es einige Zeit dauert, bis die kognitive Verarbeitung abgeschlossen ist, diese Blickzuwendung vom System aber als Aktion gedeutet wird, kann das zum Problem werden.

Nachfolgend werden die wichtigsten Interaktionstechniken mittels Eye-Tracking vorgestellt. Danach wird auf Lösungsmöglichkeiten für die dabei am häufigsten auftretenden Probleme eingegangen, ehe im letzten Abschnitt

Ansätze behandelt werden, bei denen Eye-Tracking in Kombination mit anderen Inputmodalitäten zum Einsatz kommt.

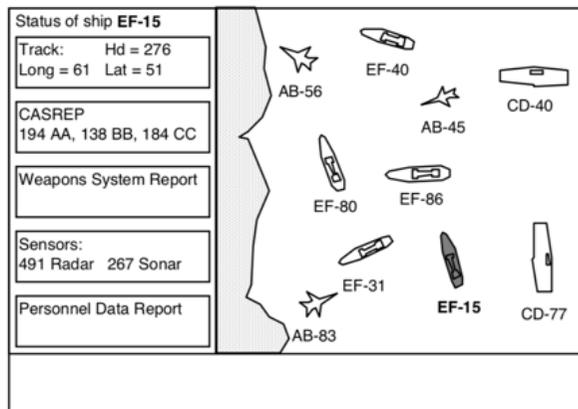
## 2 Fixationsbasierte Interaktion

Bei der fixationsbasierten Interaktion kommen Blickzuwendungen als direkter Mausersatz zum Einsatz. Es handelt sich dabei also um eine sehr natürliche und dem Nutzer im Grunde vertraute Art der Interaktion. Von einer Fixation spricht man, wenn das Auge über einen gewissen Zeitraum hinweg nicht bewegt wird [12]. Diese Methode der Interaktion mittels Eye-Tracking untersuchte beispielsweise Jacob [17]. Wenn der Benutzer mit seinem Blick eine vordefinierte Zeit lang eine Stelle fixiert, wird dort vom Computer eine Aktion ausgeführt, beispielsweise ein Mausklick. So wird zwischen einfachen Betrachtungen von Elementen und der Selektion von Elementen unterschieden. Diese Zeitschwelle korrekt zu wählen ist eine der größten Herausforderungen. Es sollte so gut es geht sichergestellt werden, dass das System eine einfache Betrachtung z. B. eines Buttons nicht für eine gewünschte Selektion des Benutzers hält. Dabei spricht man vom Midas-Touch-Problem. Während man mit einer klassischen Maus einfach den Mauszeiger über einem Element platzieren kann, ohne eine versehentliche Interaktion fürchten zu müssen, ist dies bei der fixationsbasierten Interaktion nicht der Fall. Wenn man die für eine Interaktion nötige Fixationszeit zu lange wählt, dauern Interaktionen zu lange und wirken nicht natürlich, während eine zu kurze Fixationsdauer zu vielen Fehleingaben führen kann. Man muss bei diesem Ansatz also zwischen Geschwindigkeit und Genauigkeit abwägen. [31]

Eine Möglichkeit ist es, die Wahl der gewünschten Fixationsdauer dem Benutzer zu überlassen. Dies wurde beispielsweise für die Texteingabe mittels Eye-Tracking getestet. So können Anfänger mit einer längeren Fixationsdauer beginnen, ehe sie sich mit der Anwendung vertraut gemacht haben und somit eine geringere Dauer effektiver ist [25]. Eine weitere Möglichkeit ist es, dass die Fixationsdauer vom Programm selbst dem jeweiligen Benutzer angepasst wird [32]. Allerdings stellten Huckauf und Urbina fest, dass ein solcher adaptiver Ansatz auf Seiten des Benutzers, der Anwendung und des Algorithmus sehr viel Training erfordert, um gewollte von ungewollten Selektionen in Echtzeit unterscheiden zu können [14].

Eine weitere Herausforderung, die Jacob feststellte, ist die Tatsache, dass die Präzision der Interaktion deutlich geringer ausfällt, als es mit der Maus der Fall ist. Sie ist eher vergleichbar mit der Präzision von Toucheingaben. Darüber hinaus muss ein System, das auf fixationsbasierter Interaktion basiert, auch vor der Verwendung immer kalibriert werden. [17]

Jacob testete diese Methode der Interaktion mit einem Interface, bei dem



■ **Abbildung 1** Interface bei der Studie von Jacob. Wenn ein Nutzer ein Schiff im rechten Fenster betrachtet, wird das Schiff ausgewählt und im linken Fenster werden Informationen dazu angezeigt. [30]

der Benutzer auf der rechten Seite des Bildschirms Schiffe angezeigt bekommt (siehe Abbildung 1). Richtet der Nutzer seinen Blick eine gewisse Zeit lang auf ein Schiff, bekommt er in der linken Spalte Attribute dazu angezeigt. Dabei bemerkte der Autor auch, dass eine recht kurze Fixationsdauer gewählt werden kann, wenn das Resultat, das durch eine Fehleingabe entsteht, auf einfache Weise ungeschehen gemacht werden kann, z. B. indem der Nutzer einfach ein anderes Schiff auswählt. Jacob stellte fest, dass dieser Ansatz in Verbindung mit einer Fixationsdauer von 150-250 ms zu guten Resultaten führte. In einer späteren Studie verglichen Jacob und Sibert die fixationsbasierte Interaktion mit der klassischen Interaktion per Maus [30]. Sie stellten fest, dass die Selektion von Elementen per Eye-Tracking schneller ist, als die Selektion mittels Maus.

Bei Anwendungen, in denen Eye-Tracking zur Texteingabe zum Einsatz kommt, werden hauptsächlich Methoden eingesetzt, die auf der Interaktion mittels Fixationen basieren [24]. Auch bei Zeichenanwendungen wird diese Interaktionstechnik oft verwendet (z. B. [13]).

Zusammenfassend lässt sich sagen, dass diese Interaktionsmethode vor allem dann Sinn macht, wenn die jeweilige Anwendung auf diese spezielle Art der Interaktion ausgelegt ist. So kann beispielsweise sichergestellt werden, dass die Elemente auf dem Bildschirm groß genug sind, sodass die limitierte Genauigkeit der Eye-Tracker nicht zum Problem wird.

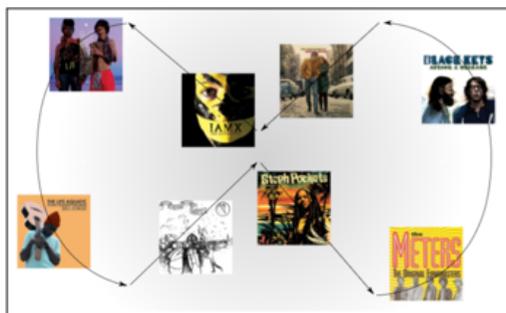
### 3 Interaktion auf Basis von Augenbewegungen

Nachfolgend werden Interaktionstechniken diskutiert, die auf der Basis von Augenbewegungen, wie Smooth Pursuits oder Sakkaden, basieren.

#### 3.1 Smooth-Pursuits-basierte Interaktion

Vidal, Bulling und Gellersen verwenden für ihre Interaktionstechnik *Pursuits* Smooth Pursuits als Basis für Interaktionen [38]. Von Smooth Pursuits spricht man, wenn das Auge einem sich bewegenden Objekt folgt [12]. Vidal et al. wollen mit ihrem Ansatz einigen Nachteilen von fixationsbasierter Interaktion begegnen. Bei *Pursuits* handelt es sich um eine Interaktionstechnik, die Nutzern spontane Interaktionen (ohne Vorwissen, wie das System funktioniert), z. B. bei Public Displays, erlaubt. Auf den Displays werden sich bewegende Objekte dargestellt. Wenn der Nutzer mit seinem Blick einem Objekt folgt, führen seine Augen denselben Bewegungsablauf aus, wie das Objekt, dem er folgt, womit das betrachtete Objekt korrekt identifiziert werden kann. Dies bringt auch den großen Vorteil dieser Interaktionstechnik mit sich, nämlich, dass diese Form der Interaktion somit keinerlei Kalibrierung des Eye-Trackers benötigt, denn der tatsächlich betrachtete Punkt ist bei dieser Methode nicht von Relevanz. Die Verwendung dieser Korrelation bedeutet auch, dass dieser Ansatz mit jedem Eye-Tracking-System und mit jeder Bildschirmgröße verwendet werden kann, da die Blick-Koordinaten nicht in dem selben Bereich wie die Objektkoordinaten liegen müssen. Die wichtigste Anforderung an ein System, das diese Form der Interaktion implementiert, ist, dass das Interface dynamisch sein muss. Während im Bereich der Mensch-Computer-Interaktion sich bewegende Objekte oft als Problem angesehen werden, da die Selektion eines solchen mit der Maus schwierig sein kann [11], ist dies kein Problem für das menschliche Auge. Ein weiterer Vorteil der Smooth-Pursuits-basierten Interaktion im Vergleich zur fixationsbasierten Interaktion ist die Tatsache, dass hier auch sehr kleine Objekte selektiert werden können, da diese Technik nicht positionsbasiert arbeitet. Somit ist es auch kein Problem, wenn die Genauigkeit des Eye-Trackers nicht sehr hoch ist. Um zwischen verschiedenen Objekten auf dem Bildschirm unterscheiden zu können, ist es wichtig, dass die Bewegungsabläufe der einzelnen Objekte so verschieden wie möglich ausfallen, was gleichzeitig auch zu einem Maximum der Anzahl der darstellbaren Objekte führt, wenn man Fehleingaben so gut es geht minimieren möchte. Da die Augenbewegungen mit den Bewegungsabläufen der Bildschirmobjekte korreliert werden, ist es auch wichtig, dass der Bildschirm orthogonal zur Blickrichtung positioniert ist.

Abbildung 2 zeigt das von den Autoren entwickelte *Music Shop Display*, ein Display, das in einem Musikgeschäft angebracht wurde und die Cover



■ **Abbildung 2** *Music Shop Display*, mithilfe dessen Kunden eines Musikgeschäfts mit ihrem Blick Alben zur Wiedergabe einer Hörprobe auswählen können. [38]

der neuesten Musikalben anzeigt. Ein Kunde kann nun mit seinem Blick der Bewegung des gewünschten Albums folgen, was zu der Wiedergabe einer Hörprobe führt.

Auch in anderen Bereichen werden Interaktionstechniken, die auf Smooth Pursuits basieren, eingesetzt. *Orbits* stellt beispielsweise einen Ansatz dar, der es Benutzern einer Smart Watch ermöglicht, diese zu steuern, indem er Interfacelementen, die sich kreisförmig auf dem Bildschirm der Uhr bewegen, für kurze Zeit mit seinem Blick folgt [9].

Zusammenfassend kann man sagen, dass die Smooth-Pursuits-basierte Interaktion im Vergleich zur fixationsbasierten Interaktion einige Probleme löst, wobei anzumerken ist, dass das Midas-Touch-Problem auch hier zu einer Herausforderung werden kann. Darüber hinaus ist im Vergleich zur fixationsbasierten Interaktion hier auch der Aufwand beim Interface- und Interaction Design höher.

### 3.2 Gestenbasierte Interaktion

Der gestenbasierten Interaktion liegen Abfolgen von Sakkaden zugrunde. Bei einer Sakkade handelt es sich um eine schnelle Bewegung der Augen von einem Punkt zum nächsten, so wie es beispielsweise beim Lesen der Fall ist [12]. Der große Vorteil von diesem Ansatz ist die Tatsache, dass diese Interaktionstechnik nicht nur potenziell sehr schnell ist, sondern auch das Potenzial hat, das Midas-Touch-Problem zu vermeiden, wenn die Position des ersten Blickpunkts der Geste nicht von Relevanz ist [31]. Des Weiteren ist die gestenbasierte Interaktion auch besonders für die Interaktion mit kleineren Displays geeignet, da man hier kein Platzproblem, wie bei den zwei zuvor vorgestellten Ansätzen



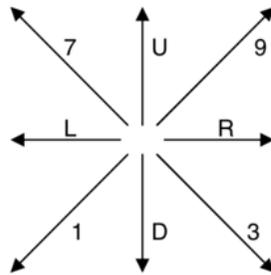
■ **Abbildung 3** Lage der Hot Zones zur Navigation von Webseiten [20].

hat. Wenn man zusätzliche Befehle hinzufügen will, fügt man einfach neue Gesten hinzu, die dann aber natürlich je nach Implementationsform vorher erlernt werden müssen.

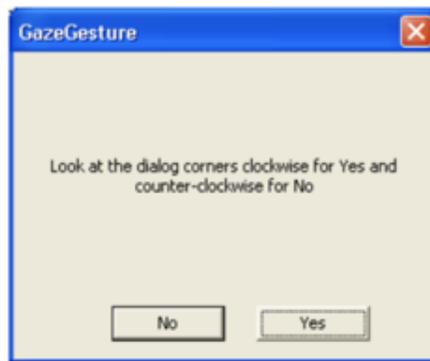
Juang, Jasen, Katrekar, Ahn und Duchowski studierten die Verwendung der Gesten-basierten Interaktion für die Webseitenavigation [20]. Oben, unten, links und rechts des Bildschirms wurden für die Gestensteuerung Hot Zones festgelegt (siehe Abbildung 3). Passierte der Blick der Nutzer eine dieser Zonen, wurde eine bestimmte Aktion ausgeführt. Die obere und untere Zone war nach oben respektive unten scrollen zugewiesen, während die linke und rechte Zone für eine Seite zurück respektive weiter gehen stand. Erst wenn der Nutzer danach wieder auf den Bildschirm blickte, konnte er eine weitere Eingabe durchführen. Des Weiteren konnte eine Hot Zone innerhalb von 100 ms nur ein Mal aktiviert werden um Fehleingaben zu vermeiden. Juang et al. verglichen diese Art der Interaktion mit der klassischen Interaktion per Maus und stellten fest, dass fast alle Probanden mit dieser neuen Interaktionsform ungefähr doppelt so schnell waren, während nur wenige Fehler passierten.

Auch Drewes und Schmidt beschäftigten sich mit der Computersteuerung mittels Gesten [7]. Die zentrale Frage, mit der sie sich beschäftigten, war, ob Personen komplexe Gesten mit ihrem Blick ausführen konnten. Ihr Algorithmus basierte unter anderem auf dem kostenlos erhältlichen *Mouse-Gesture*<sup>1</sup> Plug-In für Firefox. Abbildung 4 zeigt die acht verschiedenen Möglichkeiten für eine Sakkade, die dabei verfügbar sind. Eine Geste wurde definiert als eine bestimmte Abfolge, die aus den in der Abbildung zu sehenden Zeichen besteht. Um

<sup>1</sup> <http://optimoz.mozdev.org/gestures/>



■ **Abbildung 4** Die Namen der acht möglichen Richtungen für eine Sakkade bei Drewes und Schmidt [7].



■ **Abbildung 5** Schließen einer Dialog-Box mittels Gestensteuerung [7].

Gesten besser von den natürlich auftretenden Augenbewegungen unterscheiden zu können, erweiterten Drewes und Schmidt den Algorithmus noch um eine Timeouterkennung und fügten ein neuntes Zeichen, den Doppelpunkt hinzu, der vom Algorithmus generiert wurde, wenn kein anderer Input innerhalb eines festgelegten Zeitfensters erfolgte. Abbildung 5 zeigt eine Aufgabe für die Probanden bei der Studie, welche die Autoren durchführten. Sie sollten eine Dialogbox per Gestensteuerung schließen. Die Zeit, die die Probanden dafür benötigten, war vergleichbar mit der Zeit, die es sie kostete, wenn sie das Dialogfenster auf traditionellem Wege mit der Maus schlossen. Darüber hinaus evaluierten die Autoren auch, welche Gesten beim normalen Surfen im Internet auftreten um festzustellen, welche Gesten nicht verwendet werden sollten. Die Autoren stellten außerdem fest, dass die Gestensteuerung nicht

adäquat für Texteingabe ist, da eine Geste durchschnittlich 1 bis 2 Sekunden dauerte und somit sogar die Fixations-basierte Interaktion schneller ist.

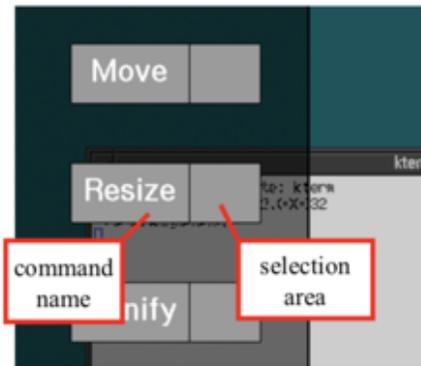
Gestensteuerung findet auch bei Videospielen Einsatz. Istance, Hyrskykari, Immonen, Mansikkamaa und Vickers entwickelten z. B. Gesten, die für die Steuerung von *World of Warcraft* eingesetzt wurden [16].

Der große Vorteil der Gestensteuerung ist die Tatsache, dass dadurch das Midas-Touch-Problem gelöst wird. Darüber hinaus ist eine geringe Genauigkeit des Eye-Tracking-Systems kein Problem, da der Blick nicht zum Zeigen verwendet wird, und es ist auch keine Kalibrierung nötig, da relative Augenbewegungen verwendet werden. Beachtet werden muss jedoch, dass gestenbasierte Interfaces wie auch Smooth-Pursuits-basierte Interfaces mehr Aufwand beim Interactiondesign benötigen.

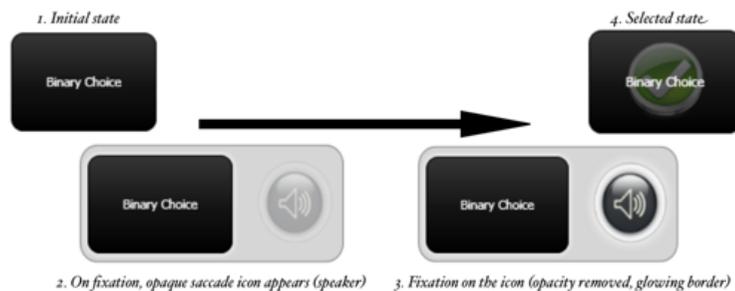
#### 4 Dedizierte Interface-Widgets

Nachfolgend werden zwei Ansätze beschrieben, wie man bereits beim Interface Design darauf achten kann, dass bei der blickbasierten Interaktion gewisse Probleme erst gar nicht auftreten. Bisher wurde beispielsweise bei der fixationsbasierten Interaktion davon ausgegangen, dass sich das Interface im Grunde so verhält, wie man es von der klassischen Interaktion per Maus und Tastatur gewohnt ist. Dass dies nicht immer der Fall sein muss, sieht man z.B. bei der Texteingabetechnik von MacKenzie und Zhang [23], bei der nachdem bereits „th“ eingetippt wurde, „e“ ausgewählt wird, auch wenn man danach eigentlich auf „d“ blickt, weil sich „e“ in der Nähe von „d“ befindet und der Nutzer vermutlich das Wort „the“ und nicht „thd“ schreiben wollte. Auf diese Art und Weise ist es kein großes Problem, wenn die Genauigkeit des Eye-Trackers nicht sehr hoch ist.

Ein weiteres Beispiel stellt die *Quick Glance Selection Method* von Ohno und Mukawa [28] dar, bei der ein Button aus zwei Teilen besteht: einem Teil, der auch den Namen des Interfaceelements enthält und auf das der Nutzer blicken kann, ohne eine Aktivierung fürchten zu müssen und ein Aktivierungsareal, auf das der Nutzer blickt, wenn er den Button betätigen will (siehe Abbildung 6). Somit wird das Midas-Touch-Problem gelöst. Der Nachteil ist, dass ein Button mit dieser Methode mehr Platz auf dem Bildschirm einnimmt. Des Weiteren kann es auch immer noch zu Fehleingaben kommen. Aus diesen Gründen verwendete Tall [37] dynamische Buttons, bei denen das Aktivierungsareal erst nach einer Fixation auf den Button erscheint (siehe Abbildung 7), was allerdings zu einer längeren Selektionszeit führt.



■ **Abbildung 6** Quick Glance Selection Method, bei der ein Button aus zwei Teilen besteht [28].

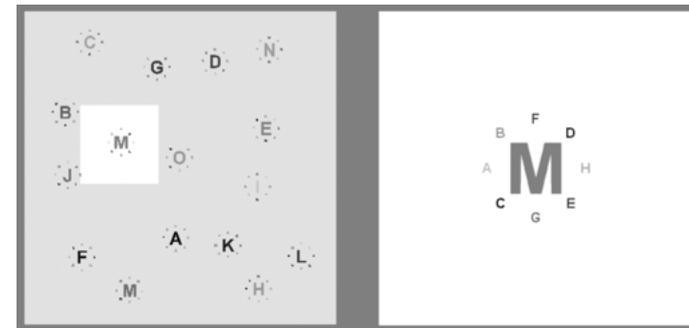


■ **Abbildung 7** Dynamischer Button bei Tall, bei dem das Aktivierungsareal erst nach einem Blick auf den Button erscheint [37].

## 5 Unterstützung bei der Auswahl kleiner Ziele

Zoomen ist eine Möglichkeit um die Auswahl von kleinen Zielen einfacher zu gestalten. Dabei gibt es verschiedene Arten: Man spricht von geometrischem Zoomen, wenn einfach die Größe der Darstellung vergrößert wird, von Fischaugenzoomen, wenn auch Informationen, die sich am Rand befinden, mit leichter Verzerrung sichtbar sind und von semantischem Zoomen, wenn sich beim Zoomen zusätzlich auch die Form bzw. der Kontext, in dem die Informationen präsentiert werden, ändert [34]. Ein Beispiel für diese Art des Zoomens ist Google Maps, wo kleinere Straßen überhaupt erst erscheinen, wenn man die Ansicht zu einem gewissen Grad vergrößert.

**Geometrisches Zoomen** Interfaces, die geometrisches Zoomen umsetzen, sind recht einfach zu implementieren. Ein Beispiel dafür stellt das Interface von Pomplun, Ivanovic, Reingold und Shen dar [29] (siehe Abbildung 8). Während



■ **Abbildung 8** Zoominterface mit zwei Fenstern, bei dem das rechte Fenster eine vergrößerte Version des im linken Fenster hervorgehobenen Areals anzeigt [29].

der Nutzer das linke Fenster, welches die gesamte Menge an Informationen darstellt, inspiziert, folgt dem Blick die ganze Zeit über ein hervorgehobenes Quadrat. Sobald der Nutzer seinen Blick dem rechten Fenster zuwendet, wird dort das zuvor hervorgehobene Areal vergrößert angezeigt. Die Autoren verglichen in ihrer Studie die Blicksteuerung mit der klassischen Steuerung per Maus und stellten fest, dass die Probanden mit Blicksteuerung zwar mehr Vergrößerungsvorgänge auslösten und mehr Fehler machten, die Antwortzeit aber nur minimal länger ausfiel und somit die Blicksteuerung, richtig eingesetzt, eine gute Alternative zu herkömmlichen Kontrollgeräten sein kann.

Skovsgaard, Hansen und Mateo verglichen in ihrer Studie eine neuartige kontinuierliche Zoomauswahltechnik, bei der das Zoomen direkt im betrach-



■ **Abbildung 9** Größer werdende Elemente in einem Listen-Menü [33].

teten Fenster stattfand (also ohne separates Zoomfenster) [18]. Dabei wurde das Zoomfenster dort angezeigt, wo der Nutzer hinblickte und der im Fenster dargestellte Inhalt wurde allmählich vergrößert dargestellt. Sie verglichen ihre neue Interaktionsform (kontinuierliches Zoomen) mit einer traditionellen Selektion in zwei Schritten (diskretes Zoomen) und stellten fest, dass ihre Technik bei Selektionen zwar deutlich schneller, dafür aber ungenauer war.

**Semantisches Zoomen** Špakov und Miniotas entwickelten eine Technik, um Elemente aus einem Listenmenü auszuwählen [33] (siehe Abbildung 9). Wenn ein Nutzer ein Element aus der Liste fixiert, vergrößert sich das Selektionsareal des Elements, um die oft nicht ausreichende Genauigkeit von Eye-Tracking-Systemen zu kompensieren. Die Autoren stellten in einer Studie fest, dass Probanden mit dieser Technik zwar länger brauchten um ein Element zu selektieren, dafür aber auch die Fehlerrate drastisch abnahm.

**Fischaugenzoom** Ashmore, Duchowski und Shoemaker evaluierten in ihrer Studie die Verwendung einer Fischaugenvergrößerung [3]. Das Besondere an dem Ansatz der Autoren ist, dass die Fischaugenvergrößerung erst dann angewendet wird, wenn das System erkennt, dass der Benutzer ein Element fixiert (bis dahin wird das Interface normal, also ohne Vergrößerung dargestellt). Die Studie zeigte, dass diese neuartige Art der Fischaugenvergrößerung, sowohl was die Geschwindigkeit als auch die Genauigkeit angeht, besser abschneidet, als wenn sie die ganze Zeit aktiv gewesen wäre oder wenn gar keine Vergrößerung stattgefunden hätte. Die Fehlerrate fiel unter allen Konditionen vergleichbar aus.

## 6 Eye-Tracking in Kombination mit anderen Input-Modalitäten

Während bisher nur Interaktionstechniken vorgestellt wurden, die Blickzuwendungen als einzige Modalität benutzen, finden sich in diesem Kapitel multimodale Interaktionstechniken.

**Blinzeln, Zwinkern, Pupillengröße** Möchte man Blinzler dazu verwenden, einen Mausklick auszulösen, muss natürlich sichergestellt werden, dass das System absichtliche Blinzler auch als solche erkennt und nicht mit dem natürlichen Blinzeln verwechselt (man blinzelt ungefähr 10 Mal in der Minute [6]). Eine Möglichkeit, dies zu lösen, wäre es, Zwinkern anstatt Blinzler zu verwenden, was aber für einige Personen nicht so leicht ist [31]. Man könnte für Aktivierungen auch längere Blinzler fordern, was dann aber natürlich die Interaktionszeit erhöht. Des Weiteren kann es passieren, dass nach einer längeren Periode, während derer die Augen geschlossen wurden, die aktuelle Fixationsposition verloren geht [14]. Darüber hinaus wird es zum Problem, wenn die Probanden während der Blinzler bzw. während dem Zwinkern ihren Kopf bewegen und man einen kopfgetragenen Eye-Tracker verwendet, da dieser dann verrutschen kann. Ekman, Poikola, Mäkäräinen, Takala und Hämäläinen zeigten, dass sogar die Pupillenerweiterung als Inputmodalität genutzt werden kann [8]. Sie fanden heraus, dass es möglich ist, zu lernen, die Pupillengröße zu kontrollieren, wobei natürlich auch hier zwischen absichtlichen Veränderungen der Pupillengröße und Veränderungen aufgrund von z.B. Helligkeitsschwankungen unterschieden werden muss.

**Tastatur, Maus** Eine weitere Möglichkeit ist, die Eingabe per Tastatur oder Maus als zusätzliche Inputmodalität hinzuzufügen. Kumar, Paepcke und Winograd implementierten in ihrem *EyePoint*-Prototyp Blickzuwendungen mit dem Input einer normalen Tastatur [22]. Zuerst blickt der Nutzer auf das gewünschte Ziel. Dann drückt er eine bestimmte Taste auf der Tastatur, was eine vergrößerte Ansicht des zuvor betrachteten Bereichs auslöst. In dieser Ansicht blickt der Nutzer nun wieder auf das gewünschte Ziel und lässt die Taste los, um die Aktion auszuführen. In einer Studie wurde herausgefunden, dass *EyePoint* zwar für eine reine Selektionsaufgabe langsamer und weniger genau war, als es mit der klassischen Mausinteraktion der Fall war, aber sogar etwas schneller war, wenn die Aufgabe sowohl Selektions- als auch Texteingabelemente enthielt. Es geschahen allerdings in jedem Fall deutlich mehr Fehleingaben bei diesem neuen Ansatz. Zhai, Morimoto und Ihde kombinieren in *MAGIC pointing* die Blickeingabe mit dem Input einer Maus [40]. Bei einem Ansatz wird dabei der Mauscursor automatisch in die Nähe des betrachteten

Objektes verschoben, sodass der Nutzer dann per Maus nur noch die Feinarbeit erledigen muss. Die Autoren stellten fest, dass die Geschwindigkeit hier im Vergleich zu einer reinen Navigation per Maus leicht stieg und die Genauigkeit gleich blieb.

**Gesichtsmuskeln** Surakka, Illi und Isokoski untersuchten, wie man das Runzeln der Stirn in Kombination mit Blickzuwendungen als Mensch-Computer-Interaktionstechnik verwenden kann [36]. Um die Muskelaktivitäten aufzuzeichnen, wurden zwei Elektroden an der Stirn der Probanden befestigt. Die Autoren fanden heraus, dass die klassische Interaktion per Maus für kurze Bewegungsdistanzen zwar schneller war als ihre neue Interaktionsform. Bei mittelgroßen bis großen Distanzen ließen sich allerdings keine signifikanten Unterschiede feststellen. Surakka, Isokoski, Illi und Salminen stellten später fest, dass die Selektion durch Lächeln schneller und genauer ist, als die Selektion durch Stirnrunzeln [35]. Grauman, Betke, Lombardi, Gips und Bradski verglichen die Selektion mittels Blinzeln mit der Selektion durch Anheben der Augenbrauen und fanden heraus, dass erstere Technik eine höhere Erkennungsrate aufwies [10].

**Kopfbewegungen** Adams, Witkowski und Spence verwendeten in ihrer Studie Kopfbewegungen in Kombination mit Augenbewegungen [1]. Sie verwendeten die Distanz zwischen den Augen und dem Bildschirm für das Zoomen. Diesem Ansatz liegt die Tatsache zugrunde, dass Menschen sich von Natur aus näher auf ein Ziel zubewegen, das sie interessiert und sich zurücklehnen, wenn sie sich einen Überblick verschaffen wollen. Eine Aufgabe der Probanden war es, in *Google Earth* mittels kontinuierlichem zoomen und, wenn nötig, korrigierenden Schwenk- und Zoomoperationen, zu einem bestimmten Ziel zu navigieren. Dabei wurden verschiedene Zoomarten miteinander verglichen. Die Probanden brauchten zwar einige Zeit, um sich an diese neue Art der Navigation zu gewöhnen, beim dritten Durchlauf war die Performance allerdings bereits bei allen verglichenen Techniken ähnlich.

**Sprache** Kaur et al. untersuchte Sprache als zusätzliche Inputmodalität für blickgesteuerte Systeme. In einer Studie bekamen die Probanden die Aufgabe, auf ein Objekt auf einem Bildschirm zu blicken, das sie bewegen möchten, dann „Move it“ zu sagen, zum Zielort zu schauen und abschließend „There“ zu sagen [21]. Es wurde herausgefunden, dass die Fixation, welche das Objekt, das bewegt werden soll, am besten identifiziert, durchschnittlich 630 ms bevor das Wort „Move“ ausgesprochen wurde, stattfindet, was verständlich ist, da Nutzer das gewünschte Objekt erst einmal eine Zeit lang betrachten wollen. Miniotas, Špakov, Tugoy und MacKenzie verwendeten Sprache um Ungenauigkeiten

auszugleichen, die auftauchten, wenn die Selektion von Elementen nur blickgesteuert erfolgt und diese Elemente sehr klein bzw. eng beieinanderliegend sind [27]. Zur Selektion kam in der Studie der fixationsbasierte Ansatz zum Einsatz. Um den gemessenen Fixationspunkt herum wurden in einem 100 Pixelradius alle auswählbaren Elemente mit verschiedenen Farben versehen. Der Nutzer konnte dann die Farbe des gewünschten Elementes sagen. Dieser Ansatz erhöhte die Genauigkeit, wobei die Fixationsdauer erhöht werden musste um die gestiegene kognitive Last auszugleichen.

**Gehirn-Computer-Schnittstelle** Zander, Gaertner, Kothe und Volimek untersuchten Blickzuwendungen in Kombination mit einer Gehirn-Computer-Schnittstelle als Input [39]. Diese Art der Interaktion stellt einen neuen Interaktionskanal für Personen dar, die komplett die Kontrolle über alle Muskelaktivitäten verloren haben (Locked-in-Syndrom). Um die Gehirnaktivitäten aufzuzeichnen, verwendeten die Autoren ein EEG, weil die temporale Auflösung recht hoch und die Anwendung vergleichsweise einfach ist. In der Studie, die Zander et al. durchführten, wurde eine zweidimensionale Cursorsteuerung per Eye-Tracking sowie ein von der Gehirn-Computer-Schnittstelle erkennbarer Kontrollgedanke umgesetzt. Das Auge stellt bei Bewegung einen starken Dipol dar und stört somit die Erkennung der Gehirnwellen, die ein wesentlich geringeres Potenzial aufweisen. Der Erkennungsalgorithmus muss also diesen Störfaktor berücksichtigen. Die Autoren verglichen in einer Selektionsaufgabe ihre Inputtechnik mit dem fixationsbasierten Ansatz (mit zwei verschiedenen Aktivierungsdauern). Sie fanden heraus, dass es möglich ist, mit ihrer Interaktionstechnik genauere Selektionen durchzuführen, als wenn man eine kurze Fixationsdauer zur Selektion verwenden würde. Des Weiteren favorisierten auch die Nutzer selbst diesen neuartigen Ansatz und gaben an, von diesem weniger frustriert gewesen zu sein.

## 7 Zusammenfassung

Nachdem nun auf die verschiedenen expliziten Interaktionstechniken mittels Eye-Tracking eingegangen, Lösungsmöglichkeiten für die dabei am häufigsten auftretenden Probleme besprochen sowie Eye-Tracking in Kombination mit anderen Inputmodalitäten erläutert wurde, stellt sich nun natürlich die Frage, ob es eine beste Form der Interaktion mittels Eye-Tracking gibt. Es wurde in der Arbeit klar, dass dem nicht so ist. Jede Interaktionstechnik hat ihre Stärken und Schwächen und somit ein Anwendungsgebiet, in dem sie optimal eingesetzt werden kann. Wenn man Eye-Tracking zur Steuerung einer Anwendung verwenden möchte, muss man sich immer die Frage stellen, welche Art der Blickinteraktion dazu am besten passt. Ist es von größter Wichtigkeit,

dass die Interaktionen so schnell wie möglich ablaufen oder nehme ich eine etwas langsamere Form der Interaktion in Kauf und produziere dabei weniger Fehleingaben? Stecke ich viel Arbeit in das Interface Design um die Interaktion zu optimieren oder bleibe ich bei klassischen Darstellungen?

In dieser Arbeit wurde hauptsächlich auf simple Interaktionen eingegangen, also beispielsweise Point-and-Select-Operationen. Für zukünftige Arbeiten zu diesem Thema wäre es natürlich interessant zu untersuchen, wie Eye-Tracking für komplexere Aufgaben eingesetzt werden kann und natürlich wäre es nicht zuletzt sehr hilfreich, in einer Studie alle vorgestellten Techniken miteinander hinsichtlich Geschwindigkeit, Fehleingaben etc. zu vergleichen.

---

#### Literatur

- 1 Nicholas Adams, Mark Witkowski, and Robert Spence. The inspection of very large images by eye-gaze control. In *Proceedings of the working conference on Advanced visual interfaces*, pages 111–118. ACM, 2008.
- 2 August Ahrens. *Untersuchungen über die Bewegung der Augen beim Schreiben*. C. Boldt, 1891.
- 3 Michael Ashmore, Andrew T Duchowski, and Garth Shoemaker. Efficient eye pointing with a fisheye lens. In *Proceedings of Graphics interface 2005*, pages 203–210. Canadian Human-Computer Communications Society, 2005.
- 4 Edmund B Delabarre. A method of recording eye-movements. *The American Journal of Psychology*, 9(4):572–574, 1898.
- 5 Raymond Dodge and Thomas Sparks Cline. The angle velocity of eye movements. *Psychological Review*, 8(2):145, 1901.
- 6 Michael J Doughty. Further assessment of gender-and blink pattern-related differences in the spontaneous eyeblink activity in primary gaze in young adult humans. *Optometry and Vision Science*, 79(7):439–447, 2002.
- 7 Heiko Drewes and Albrecht Schmidt. Interacting with the computer using gaze gestures. In *IFIP Conference on Human-Computer Interaction*, pages 475–488. Springer, 2007.
- 8 Inger Ekman, Antti Poikola, Meeri Mäkäräinen, Tapio Takala, and Perttu Hämäläinen. Voluntary pupil size change as control in eyes only interaction. In *Proceedings of the 2008 symposium on Eye tracking research & applications*, pages 115–118. ACM, 2008.
- 9 Augusto Esteves, Eduardo Velloso, Andreas Bulling, and Hans Gellersen. Orbits: Gaze interaction for smart watches using smooth pursuit eye movements. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pages 457–466. ACM, 2015.
- 10 Kristen Grauman, Margrit Betke, Jonathan Lombardi, James Gips, and Gary R Bradski. Communication via eye blinks and eyebrow raises: Video-based human-computer interfaces. *Universal Access in the Information Society*, 2(4):359–373, 2003.
- 11 Tyler J Gunn, Pourang Irani, and John Anderson. An evaluation of techniques for selecting moving targets. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems*, pages 3329–3334. ACM, 2009.
- 12 Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost Van de Weijer. *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford, 2011.
- 13 Anthony Hornof, Anna Cavender, and Rob Hoselton. Eyedraw: a system for drawing pictures with eye movements. In *ACM SIGACCESS Accessibility and Computing*, number 77-78, pages 86–93. ACM, 2004.
- 14 Anke Huckauf and Mario H Urbina. On object selection in gaze controlled environments. *Journal of Eye Movement Research*, 2(4), 2008.
- 15 Edmund B Huey. Preliminary experiments in the physiology and psychology of reading. *The American Journal of Psychology*, 9(4):575–586, 1898.
- 16 Howell Istance, Aulikki Hyrskykari, Lauri Immonen, Santtu Mansikkamaa, and Stephen Vickers. Designing gaze gestures for gaming: an investigation of performance. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, pages 323–330. ACM, 2010.
- 17 Robert JK Jacob. What you look at is what you get: eye movement-based interaction techniques. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 11–18. ACM, 1990.
- 18 Henrik Tomra Skovsgaard Hegner Jensen, John Paulin Hansen, and Julio C Mateo. How can tiny buttons be hit using gaze only? In *COGAIN 2008 'Communication, Environment and Mobility Control by Gaze'*, pages 38–42. CTU Publishing House, 2008.
- 19 Rose Johnson, Kenton O'Hara, Abigail Sellen, Claire Cousins, and Antonio Criminisi. Exploring the potential for touchless interaction in image-guided interventional radiology. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3323–3332. ACM, 2011.
- 20 Kevin Juang, Frank Jasen, Akshay Katrekar, Joe Ahn, and Andrew T Duchowski. Use of eye movement gestures for web browsing. *Computer Science Department, Clemson University, Available as early as Jan*, 1:7, 2005.
- 21 Manpreet Kaur, Marilyn Tremaine, Ning Huang, Joseph Wilder, Zoran Gavovski, Frans Flippo, and Chandra Sekhar Mantravadi. Where is it? event synchronization in gaze-speech input systems. In *Proceedings of the 5th international conference on Multimodal interfaces*, pages 151–158. ACM, 2003.
- 22 Manu Kumar, Andreas Paepcke, and Terry Winograd. Eyepoint: practical pointing and selection using gaze and keyboard. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 421–430. ACM, 2007.
- 23 I Scott MacKenzie and Xuang Zhang. Eye typing using word and letter prediction and a fixation algorithm. In *Proceedings of the 2008 symposium on Eye tracking research & applications*, pages 55–58. ACM, 2008.

- 24 Päivi Majaranta. Communication and text entry by gaze. *Gaze interaction and applications of eye tracking: Advances in assistive technologies*, pages 63–77, 2012.
- 25 Päivi Majaranta, Ulla-Kaija Ahola, and Oleg Špakov. Fast gaze typing with an adjustable dwell time. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 357–360. ACM, 2009.
- 26 Päivi Majaranta and Andreas Bulling. Eye tracking and eye-based human–computer interaction. In *Advances in physiological computing*, pages 39–65. Springer, 2014.
- 27 Darius Miniotas, Oleg Špakov, Ivan Tugoy, and I Scott MacKenzie. Speech-augmented eye gaze interaction with small closely spaced targets. In *Proceedings of the 2006 symposium on Eye tracking research & applications*, pages 67–72. ACM, 2006.
- 28 Takehiko Ohno and Naoki Mukawa. Gaze-based interaction for anyone, anytime. In *Proceedings of HCI International 2003*, volume 4, pages 1452–1456. Citeseer, 2003.
- 29 Marc Pomplun, Nada Ivanovic, Eyal M Reingold, and Jiye Shen. Empirical evaluation of a novel gaze-controlled zooming interface. In *Usability Evaluation and Design: Cognitive Engineering, Intelligent Agents and Virtual Reality. Proceedings of the 9th International Conference on Human-Computer Interaction*, 2001.
- 30 Linda E Sibert and Robert JK Jacob. Evaluation of eye gaze interaction. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 281–288. ACM, 2000.
- 31 Henrik Skovsgaard, Kari-Jouko Räihä, and Martin Tall. Computer control by gaze. In *Gaze interaction and applications of eye tracking: Advances in assistive technologies*, pages 78–102. IGI Global, 2012.
- 32 Oleg Špakov and Darius Miniotas. On-line adjustment of dwell time for target selection by gaze. In *Proceedings of the third Nordic conference on Human-computer interaction*, pages 203–206. ACM, 2004.
- 33 Oleg Špakov and Darius Miniotas. Gaze-based selection of standard-size menu items. In *Proceedings of the 7th international conference on Multimodal interfaces*, pages 124–128. ACM, 2005.
- 34 Todd Stephens. A passion for metadata-an interview with todd stephens of bell south. In *Wilshire Conferences*, 2003.
- 35 V Surakka, P Isokoski, M Illi, and K Salminen. Is it better to gaze and frown or gaze and smile when controlling user interfaces. In *Proceedings of HCI International*, volume 2005, 2005.
- 36 Veikko Surakka, Marko Illi, and Poika Isokoski. Gazing and frowning as a new human–computer interaction technique. *ACM Transactions on Applied Perception (TAP)*, 1(1):40–56, 2004.
- 37 Martin Tall. *NeoVisus-Gaze Interaction Interface Components*. PhD thesis, Master Thesis, Department of Cognitive Science, XP002590862, 2008.

- 38 Mélodie Vidal, Andreas Bulling, and Hans Gellersen. Pursuits: spontaneous interaction with displays based on smooth pursuit eye movement and moving targets. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pages 439–448. ACM, 2013.
- 39 Thorsten O Zander, Matti Gaertner, Christian Kothe, and Roman Vilimek. Combining eye gaze input with a brain–computer interface for touchless human–computer interaction. *Intl. Journal of Human–Computer Interaction*, 27(1):38–51, 2010.
- 40 Shumin Zhai, Carlos Morimoto, and Steven Ihde. Manual and gaze input cascaded (magic) pointing. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 246–253. ACM, 1999.

# Das Potential der Potentiale - EEG und Interaktion

Anna Hubert

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

---

## Zusammenfassung

Diese Arbeit untersucht den Einsatz der Elektroenzephalografie (EEG) im Bereich der Mensch-Maschine-Interaktion (MMI) und bietet eine Einführung in die Technologie, verschiedene Einsatzgebiete und eine differenzierte Betrachtung der Vor- und Nachteile der angewandten Methoden. Die EEG wird zu einem großen Teil dazu verwendet, MMI-Elemente zu evaluieren und ein direktes unvoreingenommenes Feedback zu erhalten. Ein weiterer Einsatzbereich sind adaptive Benutzerschnittstellen, die sich dynamisch an den Nutzer und seine Aktionen bzw. Reaktionen anpassen. Ein großer Vorteil der EEG liegt in der Abbildung von Gehirnaktivitäten in Echtzeit, sodass Elektroenzephalogramme (EEG) sofort ausgewertet und verarbeitet werden können. Dadurch ist eine Implementierung von Systemen möglich, die dem Nutzer ohne wahrnehmbare Verzögerung Hilfestellung leisten, beispielsweise im Bereich der Spielentwicklung oder räumlicher Navigation. Da die Durchführung von EEGs im Allgemeinen noch sehr aufwendig ist und Skepsis bezüglich der Aussagekraft der potentiell relevanten Signale herrscht, ist eine weitere Entwicklung der Technologie erforderlich, bis deren Einsatz alltagstauglich wird.

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

**Keywords and phrases** EEG, HCI, BCI, Mensch-Maschine-Interaktion, Interaktionsprobleme

## 1 Einführung in die Elektroenzephalografie

Die Elektroenzephalografie (EEG) ist eine Methode aus dem Bereich der kognitiven Neurowissenschaft, mittels derer die Summe aller elektrischen Aktivitäten im Gehirn gemessen wird. Im Bereich der Medizin wird die EEG hauptsächlich eingesetzt, um Gehirnaktivitäten zu erforschen und Erkrankungen festzustellen. In dieser Arbeit soll jedoch auf den Einsatz der EEG im Bereich der Mensch-Maschine-Interaktion (MMI) eingegangen werden.



© Anna Hubert;

licensed under Creative Commons License CC-BY

Cite as: Anna Hubert. Das Potential der Potentiale - EEG und Interaktion. 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. 4:1–4:18.

Die MMI wählt einen nutzerzentrierten Ansatz, um Produkte zu designen, die den Bedürfnissen des Nutzers größtmöglich gerecht werden [18]. Ebenjene Interaktion zwischen Mensch und Maschine stellt Entwickler von Hardware wie Software dabei vor große Herausforderungen.

Besitzen Systeme, mit denen Menschen arbeiten und interagieren eine hohe Komplexität, kann deren Benutzung sehr anstrengend werden [11]. Verbunden mit hoher Verantwortung, wie z.B. im Bereich der Flugsicherung oder im Bahnbetrieb, müssen komplexe Aufgaben unter großem Druck bewältigt werden. Dies bewirkt eine hohe mentale und kognitive Belastung bei den Nutzern und kann zu längerer Bearbeitungsdauer, einer höheren Fehlerquote oder folgenschweren Fehlern führen [11]. Deshalb ist es erstrebenswert Methoden zu entwickeln, die die kognitive Last der Nutzer überwinden und an Systeme weitergeben, die auf Veränderungen dynamisch reagieren.

Nicht nur komplexe Aufgaben, auch alltägliche Anwendungen, die häufig benutzt werden, sollten möglichst einfach und intuitiv gestaltet werden. So stellt z.B. die Bedienung einer Tastatur für jeden Nutzer eine unterschiedlich große Herausforderung dar. Die Texteingabe hängt stark von der Übung, sowie von den körperlichen Fähigkeiten, ggf. Einschränkungen des Nutzers ab. Ähnlich verhält es sich bei Touchscreens, Mausbedienung oder Spracheingaben [8]. So unterschiedlich die Schnittstellen zwischen Mensch und Maschine sind, so unterschiedlich sind auch die Nutzergruppen, für die Interaktionsmöglichkeiten entwickelt werden.

Für all diese Interaktionsprobleme eröffnen sich durch die Elektroenzephalografie neue Möglichkeiten, Lösungsansätze zu finden.

Diese Arbeit erklärt zunächst die medizinischen Hintergründe der Elektroenzephalografie und gibt eine Einführung zu verschiedenen Arten der EEG-Durchführung, sowie zu technischen Voraussetzungen und Geräten. Im weiteren Verlauf werden unterschiedliche Anwendungsbereiche in der Mensch-Maschine-Interaktion dargestellt und schließlich Herausforderungen beim Einsatz von EEGs diskutiert.

## 2 Medizinischer Hintergrund

Um zu verstehen wie EEGs erstellt und ausgewertet werden können, wird zunächst der Aufbau des menschlichen Gehirns kurz dargelegt. Das Gehirn besteht aus folgenden vier Hauptbereichen:

**Großhirn** Das Großhirn (*cerebrum, telencephalon*) ist in zwei gleichgroße Hälften (*Hemisphären*) unterteilt und von der Großhirnrinde (*cortex cerebri*) umgeben. Die Masse des Großhirns macht die sogenannte Weiße Substanz (*substantia alba*) aus, die überwiegend Nervenfasern und Axone beinhaltet. Der

Kortex besteht aus der Grauen Substanz (*substantia grisea*), zu der hauptsächlich Nervenzellkörper zählen. Jede Hemisphäre kann von vorne nach hinten in Stirnlappen (*Frontallappen*), Schläfenlappen (*Temporallappen*), Scheitellappen (*Parietallappen*) und Hinterhauptlappen (*Okzipitallappen*) eingeteilt werden, die jeweils unterschiedliche Aufgabenbereiche besitzen. Der Kortex ist zusätzlich in drei Felder unterteilt, dem motorischen, sensorischen und assoziativen Feld.

**Zwischenhirn** Das Zwischenhirn (*diencephalon*) filtert wichtige von unwichtigen Informationen der Sinnesorgane und leitet diese zum Großhirn weiter.

**Kleinhirn** Das Kleinhirn (*cerebellum, metencephalon*) lässt sich ebenfalls in zwei von der Kleinhirnrinde umgebene Hemisphären unterteilen und ist für das Gleichgewicht sowie die Steuerung der Bewegungen des Körpers zuständig.

**Nachhirn** Das Nachhirn (*myelencephalon*) stellt lebenswichtige Grundfunktionen des Körpers sicher. Dazu gehören automatische Abläufe, wie Atmung, Herzschlag und Reflexe.

Für ein EEG sind das Großhirn und besonders die Großhirnrinde von Bedeutung, da dort hohe elektrische Aktivität stattfindet. Nervenzellen (*Neuronen*) sind untereinander durch Synapsen verknüpft und tauschen über diese Informationen in Form von elektrischen Potentialen aus. Dabei treten an den Synapsen Spannungsunterschiede auf, die durch entsprechend an der Schädeldecke angelegte Elektroden gemessen werden können.

Es existieren verschiedene Arten von Potentialen, die jedoch einzeln nicht mittels EEG messbar sind. Erst wenn viele tausend Synapsen von parallel verlaufenden Neuronen gleichzeitig aktiviert werden, generieren diese ein extrazelluläres Feldpotential und somit einen ausreichend großen Spannungsunterschied, der aufgezeichnet werden kann.

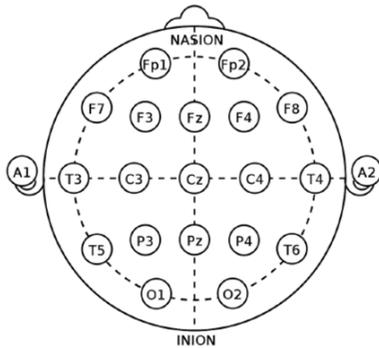
## 3 Methoden und Messgeräte

Die Erstellung eines EEGs ist ein komplexer Vorgang und erfordert ausgefeilte Technik, um aussagekräftige und wiederholbare Ergebnisse zu liefern. Im Folgenden werden gängige Methoden und Messgeräte vorgestellt, die momentan in der Wissenschaft eingesetzt werden.

### 3.1 10-20-System

Um ein auf jede Schädelgröße angepasstes und dennoch vergleichbares Elektroenzephalogramm (EEG) zu erhalten, wird das internationale 10-20-System angewendet, wie in Abbildung 1 dargestellt. Die Schädeldecke wird anhand

zweier Achsen, die von Ohr zu Ohr, sowie von der Nasenwurzel (*Nasion*) zum Hinterhaupt (*Inion*) führen, vermessen und deren Länge als 100 % festgelegt. Diese Achsen werden jeweils in Abstände von 10 Prozent bzw. 20 Prozent unterteilt und so ein Gitter über den Schädel gelegt. Die Elektroden werden nach einem festen Muster auf diesem Gitter positioniert, das je nach Anzahl der Elektroden (16, 32, 64, etc.) verdichtet werden kann. Bei 50 %, dem Schnittpunkt der beiden Achsen, liegt der Vertex (*Cz*), der „höchste“ Punkt des Schädels. *Fz* und *Pz* befinden sich 20 % vor bzw. hinter *Cz* und werden zusammen mit diesem immer als Standardorte angegeben, um Vergleichbarkeit zu garantieren. [14]



■ **Abbildung 1** Darstellung des 10-20-Systems mit Standardorten für Elektroden [6].

### 3.2 Standard-EEG-System

Ein Standard-EEG-System besteht aus einer Elektrodenkappe, die an einen Verstärker angeschlossen ist, (s. Abbildung 2). Die analogen EEG-Signale werden in digitale Signale umgewandelt und an einen Computer weitergeleitet, der die Daten auswertet und darstellt. An den Verstärker ist zusätzlich über einen Trigger das Medium angeschlossen, auf dem Reize dargeboten werden, damit der Zeitpunkt eines ausgelösten Reizes auf dem EEG markiert wird.

Um den elektrischen Widerstand so gering wie möglich zu halten, muss die Kopfhaut sauber sein und gegebenenfalls vor dem Aufsetzen der Kappe gereinigt werden. Zwischen Haut und Elektroden wird ein Elektrolyt, eine elektrischen Strom leitende Flüssigkeit meist in Form eines Gels, gespritzt, ebenfalls um den Widerstand je nach Messgerät auf bis zu  $< 5 \text{ k}\Omega$  zu verringern.

<sup>1</sup> <https://brainvision.com/actichamp.html>



■ **Abbildung 2** Elektrodenkappe, Verstärker und Computer zum Aufzeichnen des EEGs<sup>1</sup>.

### 3.3 Off-the-shelf EEG Headsets

In den letzten Jahren wurden kabellose, tragbare EEG-Messgeräte entwickelt, wie Abbildung 3 zeigt. Sie besitzen eine geringere Anzahl an Elektroden, benötigen weniger Zeit zum Aufsetzen und können ortsunabhängig getragen werden. Die Messergebnisse sind ungenauer als bei Standard-EEG-Systemen, Vi und Subramanian [19] zeigen jedoch, dass für bestimmte Einsatzgebiete, wie der Bereich der Mensch-Maschine-Interaktion eine Messgenauigkeit von 65 - 80 % ausreicht.



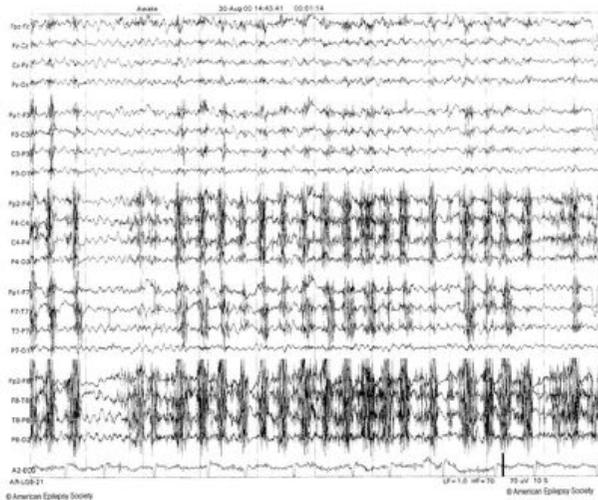
■ **Abbildung 3** Mobiles EEG-Messgerät von *Emotiv*<sup>2</sup>.

#### 4 Erstellen eines EEGs

Ein EEG entsteht, indem der Spannungsunterschied zwischen zwei Elektroden gemessen wird. Man unterscheidet dabei drei Arten der Ableitung:

- Unipolare Ableitung: Spannungsdifferenz zwischen Ableitpunkt und gemeinsamer Referenzelektrode. Diese Methode ist in der Forschung am gängigsten.
- Bipolare Ableitung: Spannungsdifferenz zwischen benachbarten Elektroden.
- Durchschnittsreferenz: Spannungsdifferenz zwischen Ableitpunkten und Mittelwert über alle anderen Elektroden. Diese Methode ist mittels der unipolaren Ableitung berechenbar.

Die wellenförmigen Signale der Elektrode werden auf einzelnen Kanälen untereinander zeitlich synchronisiert dargestellt. So werden ein guter Überblick und eine schnelle zeitliche Zuordnung gewährleistet.



■ **Abbildung 4** EEG mit deutlich sichtbaren Artefakten ausgelöst durch Kauen [3].

Sogenannte **Artefakte** sind Ausschläge im EEG, die von anderen körperlichen Aktionen ausgelöst und aufgezeichnet wurden. Die häufigsten Ursachen für Artefakte sind Blinzeln, Augenbewegungen oder Muskelartefakte, wie Anspannen des Kiefers, Schlucken oder Kauen (s. Abbildung 4). Ein Herausschneiden der Artefakte ist möglich, führt jedoch zu Datenverlust und ist

deshalb zu vermeiden. Regelmäßig wiederkehrende Artefakte wie Puls oder Blinzeln können durch entsprechende Algorithmen herausgerechnet werden [3].

#### 4.1 Frequenzbänder

Bestimmte Potentialfolgen treten, je nachdem in welchem Aktivitätszustand sich das Gehirn befindet, kontinuierlich auf und können verschiedenen Frequenzen zugeordnet werden. Diese oftmals periodisch wiederkehrenden Rhythmen werden Frequenzbänder genannt. Es werden mehrere signifikante Frequenzbänder unterschieden, die im Folgenden nach Höhe der Frequenz aufgelistet sind [14]:

- Delta-Rhythmus (< 4 Hz): Tiefschlaf
- Theta-Rhythmus (4 - 8 Hz): Gedächtnisprozesse
- Alpha-Rhythmus (8 - 13 Hz): Wachzustand mit geschlossenen Augen, sowie geringe kognitive Belastung
- Beta-Rhythmus (13 - 30 Hz): vermutlich motorische Aktivierung
- Gamma-Rhythmus (> 30 Hz): räumlich lokale Aktivität

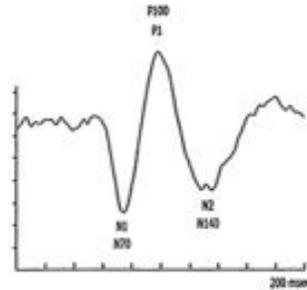
#### 4.2 Ereignis-korreliertes Potential

Ein Ereignis-korreliertes Potential (EKP, *englisch*: event-related potential (ERP)) tritt nach bestimmten Ereignissen auf und kann in verschiedene Abschnitte (Epochen) unterteilt werden, die bestimmte Merkmale haben. Das typische Muster eines EKPs, genannt EKP-Komponente, besteht aus positiven und negativen Deflektionen, d.h. wiederkehrenden Ausschlägen nach oben und unten während eines bestimmten Zeitfensters, dargestellt in Abbildung 5. Die Verzögerung (Latenz) der EKP-Komponente spielt dabei eine große Rolle, um sie zuzuordnen und klassifizieren zu können.

Der Entstehungsprozess eines EKPs ist jedoch noch nicht eindeutig geklärt. Das Additive Modell und in der Literatur weit verbreitete Modell besagt, dass zusätzlich zur Hintergrundaktivität ein Potential aufgezeichnet wird, das einen Reiz auslöst. Diese Signale addieren sich und ergeben das EKP-Muster. Im Gegensatz dazu begründet die Hypothese des „Phase Reset“ die EKP-Muster in einer Neuordnung der Schwingungen im Gehirn in Folge eines Reizes. [16]

#### 4.3 SSVEP

Ein SSVEP (*englisch*: steady-state visually evoked potential) ist ein Potential, das ausgelöst wird, wenn der visuelle Fokus auf ein Element gerichtet wird, das in einer Frequenz höher als 6-8 Hz flimmert. Bestimmte Gehirnsignale passen sich als Folge an diese Frequenz an und sind auf EEGs gut darstellbar [2].



■ **Abbildung 5** Aufzeichnung eines EKP mit erster negativer (N1), erster positiver (P1) und zweiter negativer Deflexion (N2)[1].

Anhand von SSVEPs kann man den Grad der Aufmerksamkeit messen, die Objekten im Raum visuell geschenkt wird. Kombiniert man Elemente, die auf verschiedenen Frequenzen flimmern, können außerdem Augenbewegungen nachverfolgt werden. Dieses implizite Eye Tracking kann auch zur gezielten Steuerung von Objekten verwendet werden.

## 5 EEG und Mensch-Maschine-Interaktion

Folgendes Kapitel gibt nun einen Einblick in verschiedene Anwendungsgebiete von EEGs, in denen Interaktionsprobleme durch EEGs potentiell gelöst werden können.

### 5.1 Evaluation von MMI Elementen

Viele gängige Evaluierungsmethoden bringen Nachteile mit sich, die für die jeweiligen Methoden unumgänglich sind. „Think aloud“-Protokolle beispielsweise, bei denen eine Person während der Benutzung einer Anwendung laut ihre Gedanken mitspricht, können die Person von ihrer eigentlichen Aufgabe ablenken oder verunsichern. Weit verbreitet ist ebenfalls die Bewertung durch Fragebögen, die jedoch nicht zu jedem Zeitpunkt möglich und daher vom Erinnerungsvermögen und der Motivation des Nutzers abhängig ist. Hier bietet der Einsatz der EEG große Vorteile, die vor allem im Bereich der Zeitnutzung liegen. Durch eine gute zeitliche Auflösung von wenigen Millisekunden, werden Daten annähernd in Echtzeit aufgezeichnet und können sofort interpretiert werden. Ein EEG misst somit die direkte und unvoreingenommene Reaktion des Nutzers auf ein MMI Element [19].

Beispielsweise während der Nutzung einer Benutzeroberfläche (*englisch*: user interface, UI), wird ein EEG der Testperson aufgezeichnet. Verhält sich die Interaktion mit der UI anders als erwartet, wird ein **fehlerbezogenes Potential** (*englisch*: error-related potential, ErrP) ausgelöst, das zu den EKPs gehört und somit einem bestimmten Reiz zugeordnet werden kann [7]. Diese Rückmeldung wird genutzt, um zu beurteilen wie intuitiv ein MMI Element ist [8]. Macht ein Nutzer bei der Bedienung einen Fehler und klickt beispielsweise unter zeitlichem Druck auf einen falschen Button, wird zu dem Zeitpunkt, an dem der Nutzer seinen Fehler bewusst oder unterbewusst realisiert, eine **fehlerbezogene Negativität** (*englisch*: error-related negativity, ERN) aufgezeichnet. Dieser Ausschlag lässt sich im EEG ebenfalls gut erkennen und einem Reiz zuordnen. ERNs eignen sich daher sehr gut für einen vielfältigen Einsatz im MMI Bereich [19].

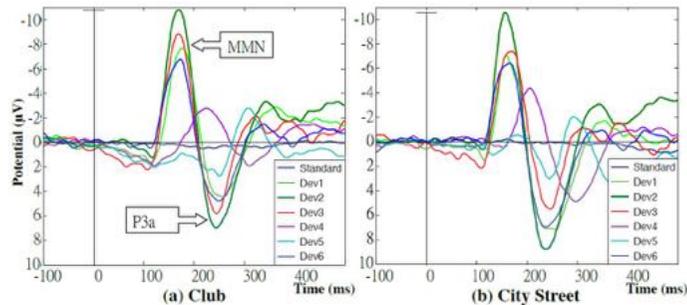
Eine weitere Komponente, mittels derer MMI-Elemente evaluiert werden können, ist der Wert der **kognitiven Last**. Ein EEG misst, in welchen Bereichen des Gehirns große Aktivität stattfindet und macht sichtbar, wenn eine Aufgabe mentale Anstrengung oder Konzentration erfordert [8].

Frey et al. [8] untersuchten in einer Studie die Eignung von Tastatur im Vergleich zu Touchscreen als Schnittstelle, um bestimmte Aufgaben zu lösen. Dabei sollten Probanden eine Figur durch ein auf einem Bildschirm dargestelltes Wegesystem navigieren und Hindernissen ausweichen. Die Bedienung wurde dadurch erschwert, dass Probanden sich räumlich in die Position der Figur versetzen mussten, sodass Richtungswechsel nach rechts oder links aus Sicht der Figur je nach Position im Spiel vom räumlichen „Rechts“ und „Links“ der Probanden abwichen. In mehreren Durchgängen mit verschiedenen Schwierigkeitsstufen wurde die Navigation einmal mit Tastatur und einmal mit Touchscreen getestet und währenddessen ein EEG aufgezeichnet. Es wurde festgestellt, dass die kognitive Last bei der Bedienung der Tastatur von Durchgang zu Durchgang abnahm, die Probanden sich also schneller daran gewöhnten als an die Bedienung über Touchscreen, der für diese Aufgabe demzufolge schlechter geeignet war. Über Tests wie diesen kann man anhand der kognitiven Auslastung die Eignung verschiedener MMI-Elemente für spezielle Aufgaben evaluieren und sie gezielt einsetzen.

Nicht nur visuelle, auch auditive Elemente wie Audio-Benachrichtigungen können anhand eines EEGs evaluiert werden. Hier dient ein EEG dazu, die menschliche Wahrnehmung von Tönen zu untersuchen und auszuwerten. Eine Komponente, die unerwartete bzw. seltene auditive Reize beschreibt, ist die **Mismatch Negativity** (MMN) [13]. Sie zeichnet sich durch einen negativen Ausschlag aus, der 100 - 250 ms nach einem auditiven Reiz, der von den Standardreizen abweicht, ausgelöst wird. Eine weitere Komponente, **P3a** genannt, tritt etwas später als die MMN auf und deutet an, wie viel Aufmerksamkeit ein

auditiver Reiz beansprucht. Übertragen auf Audio-Benachrichtigungen, die gut wahrnehmbar sein sollten, kann man so Töne entwickeln, die signifikante MMN- und P3a-Ausschläge auslösen und demnach besser aus Umgebungsgeräuschen herausgefiltert werden. Lee et al. [13] untersuchten in einer Studie die Eignung von Audio-Benachrichtigungen im Zusammenhang mit Umgebungsgeräuschen. In zwei Testszenarien - einem Café und einem Club - wurden Probanden den jeweiligen Hintergrundgeräuschen in entsprechender Lautstärke ausgesetzt und anschließend in zufällig gewählten Zeitabständen sechs verschiedene Töne abgespielt.

Die erstellten EEGs wurden nach MMN und P3a, wie in Abbildung 6 dargestellt, ausgewertet und zeigten, dass die auditive Wahrnehmung fast unabhängig von mentaler Belastung ist, jedoch stark von den Umgebungsgeräuschen abhängt. Dies legitimiert den Einsatz von Audio-Benachrichtigungen und bietet gute Möglichkeiten zur Verbesserung des Nutzererlebnisses.



■ **Abbildung 6** Gemessene Ausschläge von MMN und P3a nach sechs unterschiedlichen auditiven Reizen mit Hintergrundgeräuschen eines Clubs (a) und eines Cafés (b) [13].

## 5.2 Adaptive Interfaces

Unter Adaptive User Interfaces (AUI) versteht man adaptive Benutzeroberflächen, die sich dynamisch an die Bedürfnisse und Aktionen des Nutzers anpassen. Je nach Vertrautheit des Nutzers mit dem System, kann dieses die Bedienung erleichtern oder eine fortgeschrittene Bedienung ermöglichen.

**Objektauswahl** Betrachtet man EKPs, können EEGs als Unterstützung zur korrekten Auswahl von Objekten dienen. Selektiert ein Nutzer ein falsches Objekt, erkennt das System anhand detektierter ERNs die eigentliche Absicht

des Nutzers und kann die Aktion automatisch rückgängig machen. Dies verhindert Frustration beim Nutzer und spart unter Umständen enorm Zeit, da der Nutzer nicht selbstständig die Auswirkungen seiner Falschwahl beheben muss. Möglicherweise kann so in manchen Situationen der „Zurück“- Button vermieden werden [19].

**Spielentwicklung** Computerspiele erfordern die Interaktion eines Nutzers mit einem System. Über ein EEG könnte das System den Zustand des Spielers überwachen und dadurch den Entscheidungsprozess des Systems beeinflussen. Wenn beispielsweise spielunabhängige Faktoren wie Netzwerkfehler oder zeitliche Verzögerungen den Spielverlauf verändern, erkennt das System die Verwirrung des Spielers anhand ErrPs. Es kann darauf reagieren, indem es den eigentlich erwarteten Zustand annimmt, oder den Spieler entscheidende Sequenzen wiederholen lässt. Dynamische Veränderungen der Schwierigkeit des Spiels sind ebenfalls möglich. Je nach Über- oder Unterforderung des Spielers kann der Spiellevel angepasst werden [19].

**Räumliche Navigation** In der räumlichen Navigation können EEGs als Grundlage für Navigationshilfen eingesetzt werden. Das System erkennt über ERNs die Verwirrung des Nutzers und greift ein, indem es Vorschläge zur weiteren Navigation bietet. Eine Herausforderung dieses Einsatzgebietes ist jedoch, dass sich der Nutzer ständig bewegt. Diese Bewegungen stören ein EEG, werden als Artefakte aufgezeichnet und müssen aufwendig herausgefiltert werden [19].

**Anwendungen mit mehreren Nutzern** In diesen Bereich fallen Spiele mit Multiplayer Modus, aber auch andere Anwendungen, bei denen mehrere Nutzer eine gemeinsame Aufgabe haben. Von jedem Nutzer wird ein EEG aufgezeichnet, das eine etwaige Überforderung erkennen soll. Treten Anzeichen von Überforderung eines Nutzers auf, z.B. hohe kognitive Belastung, werden andere Teammitglieder informiert und können ihn unterstützen [19].

## 5.3 Gehirn-Computer-Schnittstellen

Gehirn-Computer-Schnittstellen (*englisch*: brain-computer-interfaces, BCI) bieten eine weitere Kommunikationsmöglichkeit zwischen Mensch und Maschine, bei der Elemente über das Gehirn gesteuert werden. Abhängig von der Art der auszuwertenden Gehirnaktivitäten erfordert die Steuerung Training oder kann intuitiv verwendet werden. Nach Zander und Kothe [20] lassen sich BCIs in drei Kategorien einteilen:

Bei **aktiven BCIs** erfolgt die Manipulation über die Auswertung der Frequenzbänder und Stimulation verschiedener Gehirnareale. Der Nutzer trainiert entsprechende Gedanken, wie z.B. die Vorstellung den linken oder rechten

Arm zu bewegen. Dieses Muster lernt das System nutzerspezifisch zu erkennen, wertet sie aus und reagiert darauf. Je höher die Konzentration und das Engagement des Nutzers ist, desto bessere Ergebnisse werden erzielt. **Reaktive BCIs** machen sich Potentiale wie EKPs oder SSVEPs zu Nutze, die in Folge eines Reizes ausgelöst werden. Sie benötigen kein Training, da sie mit natürlichen Gehirnaktivitäten arbeiten, die zwar bewusst induziert werden, vom Nutzer jedoch nicht aktiv gesteuert werden können. **Passive BCIs** hingegen überwachen die gesamte kognitive Aktivität und versuchen aus impliziter Information Schlüsse über den emotionalen Zustand und über Absichten des Nutzers zu ziehen.

Aktive BCIs bieten beispielsweise Menschen mit stark eingeschränkten motorischen Fähigkeiten die Möglichkeit, Kontrolle über Objekte zu erlangen und diese explizit zu steuern, wie eine Studie von Philips et al. [17] zeigt. Sie untersuchten die Steuerung eines elektrischen Rollstuhls durch mentale Steuerungsbefehle. Es wurden drei Befehle antrainiert, die kurze mentale Aufmerksamkeit erforderten, um damit drei Richtungen unterscheiden zu können: Vorwärts, Links und Rechts. Diese Befehle wurden über ein EEG erkannt und in einen komplexen Algorithmus eingebaut, der sie zusammen mit unterstützenden Systemen wie Kollisionsvermeidung verarbeitete und letztendlich den Rollstuhl erfolgreich durch einen Korridor steuerte.



■ **Abbildung 7** Die Spielfigur kann durch Fokussieren der entsprechenden Schachbrettfläche gesteuert werden, wenn sie das Gleichgewicht verliert [12].

Lalor et al. [12] verwenden eine reaktive BCI, die SSVEPs auswertet, um einen Avatar in einem 3D-Spiel zu kontrollieren. Die Spielfigur balanciert auf einem Seil und verliert in zufälligen Abständen auf eine zufällig gewählte Seite

das Gleichgewicht. Währenddessen hält sie eine Balancierstange in den Händen, an deren Enden sich jeweils eine Fläche mit Schachbrettmuster befindet, wie in Abbildung 7 dargestellt. Diese Flächen flackern in zwei unterschiedlichen Frequenzen. Verliert die Spielfigur das Gleichgewicht, steuert der Spieler dem entgegen, indem er seinen Fokus auf das entgegengesetzte Schachbrett richtet. Die hierbei gemessenen SSVEPs werden sofort detektiert und in Echtzeit verarbeitet. Diese Technik könnte eingesetzt werden, um Menschen mit eingeschränkten motorischen Fähigkeiten als Kommunikationsmittel zu dienen.

Eine andere Anwendungsmöglichkeit für BCIs liegt im Bereich der Fahrsistenz beispielsweise beim Autofahren. Haufe et al. [10] haben in einer 2D-Fahrsimulation Gefahrenbremsungen nachgestellt, gezeigt in Abbildung 8. Probanden sollten auf einen plötzlichen Bremsvorgang oder Spurwechsel des vorausfahrenden Fahrzeugs reagieren. Nahm der Proband die Aktionen des anderen Fahrzeugs wahr, wurden EKPs ausgelöst, die das System veranlassten sofort eine Gefahrenbremsung einzuleiten. Es zeigte sich, dass die Reaktion des Systems auf EKPs in Echtzeit schneller passierte, als die vollständig ausgeführte Reaktion der Probanden. Das System greift dabei nicht selbstständig in die Steuerung ein, sondern erkennt die Absicht des Fahrers eine Aktion auszuführen und vollzieht diese.



■ **Abbildung 8** 2D-Fahrsimulator mit angeschlossenem EEG um Gefahrenbremsungen einzuleiten [10].

Als weitere Unterstützung dient die Überwachung der mentalen Auslastung während des Fahrens, so dass erkannt werden kann, wie müde ein Fahrer ist. Bei Anzeichen von hoher Müdigkeit kann das System den Fahrer darauf aufmerksam machen oder die Steuerung übernehmen und Unfälle durch Sekundenschlaf vermeiden [10].

## 5.4 Virtuelle Realität

Die Entwicklung von Anwendungen in Virtueller Realität (VR) erlebt einen großen Aufschwung und auch in diesem Bereich kann EEG eingesetzt werden. Eine sehr große Rolle bei VR-Anwendungen spielt die **Präsenz**, die ein Maß dafür ist, wie sehr der Nutzer sich in der virtuellen Umgebung (VE, *englisch*: virtual environment) angekommen fühlt, obwohl sein Körper nicht dort ist. Dieser Wert wird üblicherweise über komplexe Fragebogen ausgewertet, ist jedoch sehr subjektiv und kann selbst bei identischen Testbedingungen von Person zu Person unterschiedlich sein. Mithilfe eines EEGs lässt sich ein direkter Zusammenhang zwischen dem Grad der Aufmerksamkeit, die eine VE erfordert, und der Präsenz des Nutzers in dieser VE ziehen. Es können eindeutige Muster erkannt werden, die sich auf die Präsenz beziehen und VR-Anwendungen können so korrekter und nutzerorientierter bewertet werden. [4]

## 6 Herausforderungen

Ein großer Nachteil von EEGs ist momentan deren mühsame Durchführung. Das Aufsetzen einer Elektrodenkappe kann bis zu einer Stunde dauern, da der Widerstand an jeder Elektrode ausreichend verringert und überprüft werden muss. Dieses Verfahren ist sehr aufwendig und kann zusätzlich für die Versuchsperson unangenehm sein. Die Versuchsperson ist zudem in ihrer Mobilität eingeschränkt und kann sich nicht frei bewegen. Mit Off-the-shelf Headsets wurde ein erster Schritt in Richtung Flexibilität gemacht, jedoch müssen noch einige Probleme wie geringere Messgenauigkeit und Artefakte durch erhöhte Bewegung gelöst werden.

Des Weiteren besitzen EEGs eine schlechte räumliche Auflösung im Bereich von mehreren Zentimetern. Dies bedeutet, dass messbar ist, *ob* eine elektrische Aktivität stattfindet, nicht aber *wo* genau diese sich befindet. So kann man keine Rückschlüsse über Aktivitäten an bestimmten Synapsen ziehen. Das liegt vor allem daran, dass an der Oberfläche der Schädeldecke gemessen wird und die Positionierung der Elektroden nicht exakt dokumentiert ist.

Ein weiterer Nachteil von EEGs liegt in ihrem niedrigen Signal-Rausch-Abstand. Signifikante Signale sind nur schwer zu erkennen, da sie sich nicht deutlich von der allgemeinen elektrischen Aktivität unterscheiden. Besonders bei Kindern ist es schwierig Artefakte durch Muskelbewegungen zu vermeiden. Zuverlässig auswertbare EEGs lassen sich daher nur für eine bestimmte Altersgruppe erstellen. Dies geht einher mit einem weiteren Nachteil, nämlich, dass EEGs abhängig vom körperlichen und mentalen Zustand der Testperson sind. Müdigkeit, der Einfluss von Alkohol oder Medikamenten beeinträchtigen die Messergebnisse des EEGs, die dann nicht verwertbar sind [5, 15].

Trotz des vielseitigen Einsatzes von EEGs im MMI Bereich gibt es gegensätzliche Meinungen über EKPs in der Wissenschaft. Kritiker zweifeln daran, dass EKPs relevante Potentiale sind, die nach einem Ereignis ausgelöst werden und nicht nur die Reaktionszeit auf ein Ereignis darstellen. Für bestimmte Vorgänge im Gehirn existieren bis heute nur Theorien über Entstehung und Wirkung, zu deren Klärung noch einige Forschungsarbeit zu leisten ist. [16]

## 7 Zusammenfassung und Diskussion

Gerade im Bereich der Mensch-Maschine-Interaktion, können schon mit einfachen kostengünstigen EEG-Messgeräten verwertbare Ergebnisse erzielt werden. Dies erleichtert den breiten Einsatz der Elektroenzephalografie bei verschiedenen Arten von Nutzerschnittstellen. Im Moment dienen EEGs zur Unterstützung von Systemen, um möglichst schnell und fehlerfrei mit Nutzern, sowie auf deren Bedürfnisse angepasst zu interagieren. Alle Vorgänge sind dabei Folgen auf Reaktionen der Nutzer. In Zukunft könnten diese Methoden verfeinert und beschleunigt werden, bis vielleicht sogar Vorhersagen über die Reaktion eines Nutzers getroffen werden können und ein negatives Nutzererlebnis schon im Vorfeld vermieden werden kann. Die Algorithmen und Methoden, die für diese Vorgänge entwickelt werden, stehen und fallen allerdings mit der Genauigkeit der Messgeräte. Die Weiterentwicklung der tragbaren Messgeräte nimmt einen großen Bereich in der Forschung ein, sodass der nächste Schritt vom reduzierten, leicht aufsetz- und anpassbaren „Elektrodenkranz“ in Richtung unsichtbare EEGs geht - Elektroden, die ähnlich wie Kopfhörer oder Hörgeräte am, bzw. im Ohr angebracht werden [9].

Entwickelt sich die EEG so weiter, dass sie nahezu unsichtbar wird und in den Alltag Einzug hält, müssen sicherlich neben allen positiven Faktoren auch die ethischen Aspekte, wie das Eindringen in den wohl persönlichsten Bereich der Gedanken, diskutiert werden. Enormes Potential für den Einsatz von EEGs besteht ohne Zweifel im medizinischen Bereich, um Anwendungen zu entwickeln, die Menschen mit eingeschränkten körperlichen Fähigkeiten als Unterstützung dienen.

Im Rahmen dieser Arbeit wurde ein kleiner Einblick in das breite Spektrum der Anwendungsmöglichkeiten gegeben, um einen Zugang zu diesem komplexen Forschungsgebiet zu schaffen und einen Überblick über wegweisende Fortschritte zu geben. Wie sich zeigt, birgt die Elektroenzephalografie in Kombination mit Mensch-Maschine-Interaktionen großes Potential und eröffnet eine neue Art der Kommunikation zwischen Mensch und Maschine. Bis diese Technologie allerdings in den Alltag integriert werden kann, bedarf es noch einiger Weiterentwicklung, vor allem im Bereich der tragbaren EEG-Messgeräte und deren Anwendbarkeit.

## Literatur

- 1 Michael Adamaszek. Ereignis-korrelierte potenziale zur untersuchung cerebraler kognitiver und emotionaler funktionen. *Das Neurophysiologie-Labor*, 38(2):63 – 78, 2016. URL: <http://www.sciencedirect.com/science/article/pii/S1439484716300011>, doi:<https://doi.org/10.1016/j.neulab.2016.03.001>.
- 2 Fabrizio Beverina, Giorgio Palmas, Stefano Silvoni, Francesco Piccione, Silvio Giove, et al. User adaptive bcis: Ssvp and p300 based interfaces. *Psychology Journal*, 1(4):331–354, 2003.
- 3 JW Britton, Lauren C Frey, JL Hopp, P Korb, MZ Koubeissi, WE Lievens, EM Pestana-Knight, and EK St Louis. *Electroencephalography (EEG): An introductory text and atlas of normal and abnormal findings in adults, children, and infants*. American Epilepsy Society, Chicago, 2016.
- 4 Miriam Clemente, Alejandro Rodríguez, Beatriz Rey, and Mariano Alcañiz. Assessment of the influence of navigation control and screen size on the sense of presence in virtual reality using eeg. *Expert Systems with Applications*, 41(4):1584–1592, 2014.
- 5 Eleanor A Curran and Maria J Stokes. Learning to control brain activity: A review of the production and control of eeg components for driving brain–computer interface (bci) systems. *Brain and Cognition*, 51(3):326 – 336, 2003. URL: <http://www.sciencedirect.com/science/article/pii/S0278262603000368>, doi:[https://doi.org/10.1016/S0278-2626\(03\)00036-8](https://doi.org/10.1016/S0278-2626(03)00036-8).
- 6 Public Domain. Electrode locations of international 10-20 system for eeg (electroencephalography) recording. Wikimedia Common, May 2010. Public domain. URL: [https://upload.wikimedia.org/wikipedia/commons/7/70/21\\_electrodes\\_of\\_International\\_10-20\\_system\\_for\\_EEG.svg](https://upload.wikimedia.org/wikipedia/commons/7/70/21_electrodes_of_International_10-20_system_for_EEG.svg).
- 7 P. W. Ferrez and J. del R. Millan. Error-related eeg potentials generated during simulated brain x2013;computer interaction. *IEEE Transactions on Biomedical Engineering*, 55(3):923–929, March 2008. doi:10.1109/TBME.2007.908083.
- 8 Jérémy Frey, Maxime Daniel, Julien Castet, Martin Hachet, and Fabien Lotte. Framework for electroencephalography-based evaluation of user experience. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 2283–2294, New York, NY, USA, 2016. ACM. URL: <http://doi.acm.org/10.1145/2858036.2858525>, doi:10.1145/2858036.2858525.
- 9 Bleichner Martin G., Lundbeck Micha, Selisky Matthias, Minow Falk, Jäger Manuela, Emkes Reiner, Debener Stefan, and De Vos Maarten. Exploring miniaturized eeg electrodes for brain-computer interfaces. an eeg you do not see? *Physiological Reports*, 3(4):e12362, 2015. URL: <https://physoc.onlinelibrary.wiley.com/doi/abs/10.14814/phy2.12362>, arXiv:<https://physoc.onlinelibrary.wiley.com/doi/pdf/10.14814/phy2.12362>, doi:10.14814/phy2.12362.
- 10 Stefan Haufe, Matthias S Treder, Manfred F Gugler, Max Sagebaum, Gabriel Curio, and Benjamin Blankertz. Eeg potentials predict upcoming emergency brakings during simulated driving. *Journal of neural engineering*, 8(5):056001, 2011.
- 11 Naveen Kumar and Jyoti Kumar. Measurement of cognitive load in hci systems using eeg power spectrum: an experimental study. *Procedia Computer Science*, 84:70–78, 2016.
- 12 Edmund C Lalor, Simon P Kelly, Ciarán Finucane, Robert Burke, Ray Smith, Richard B Reilly, and Gary Mcdarby. Steady-state vep-based brain-computer interface control in an immersive 3d gaming environment. *EURASIP Journal on Advances in Signal Processing*, 2005(19):706906, 2005.
- 13 Yi-Chieh Lee, Wen-Chieh Lin, Jung-Tai King, Li-Wei Ko, Yu-Ting Huang, and Fu-Yin Cherng. An eeg-based approach for evaluating audio notifications under ambient sounds. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, pages 3817–3826, New York, NY, USA, 2014. ACM. URL: <http://doi.acm.org/10.1145/2556288.2557076>, doi:10.1145/2556288.2557076.
- 14 Jaakko Malmivuo and Robert Plonsey. *Bioelectromagnetism*. 13. *Electroencephalography*, pages 247–264. 01 1995.
- 15 Andrew Myrden and Tom Chau. Effects of user mental state on eeg-bci performance. *Frontiers in human neuroscience*, 9:308, 2015.
- 16 E. Niedermeyer and F.H.L. da Silva. *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. LWW Doody's all reviewed collection. Lippincott Williams & Wilkins, 2005. URL: <https://books.google.de/books?id=tndqYGPHQdEC>.
- 17 Johan Philips, José del R Millán, Gerolf Vanacker, Eileen Lew, Ferran Galán, Pierre W Ferrez, Hendrik Van Brussel, and Marnix Nuttin. Adaptive shared control of a brain-actuated simulated wheelchair. In *Rehabilitation Robotics, 2007. ICORR 2007. IEEE 10th International Conference on*, pages 408–414. IEEE, 2007.
- 18 Brian Shackel. Human-computer interaction—whence and whither? *Journal of the American society for information science*, 48(11):970–986, 1997.
- 19 Chi Vi and Sriram Subramanian. Detecting error-related negativity for interaction design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 493–502, New York, NY, USA, 2012. ACM. URL: <http://doi.acm.org/10.1145/2207676.2207744>, doi:10.1145/2207676.2207744.
- 20 Thorsten O Zander and Christian Kothe. Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general. *Journal of Neural Engineering*, 8(2):025005, 2011. URL: <http://stacks.iop.org/1741-2552/8/i=2/a=025005>.

# Adaptives Lernen mit Gehirn-Computer-Schnittstellen

Raimund Krämer

Ludwig-Maximilians-Universität München, München, Deutschland  
kraemer.raimund@campus.lmu.de

---

## Zusammenfassung

Für erfolgreiches Lernen ist unter anderem die Wahl des Lernmediums ausschlaggebend. Adaptive Lernsysteme können die Effektivität beim Lernen positiv beeinflussen. Diese Arbeit gibt einen Überblick über Anwendungsfälle von Gehirn-Computer-Schnittstellen in Kombination mit adaptiven Lernsystemen, die deren kognitive Daten auswerten und sich an den mentalen Zustand des Nutzers anpassen. Es zeigt sich, dass Gehirn-Computer-Schnittstellen trotz noch bestehender technischer Einschränkungen großes Potenzial haben Nutzer adaptiver Systeme beim Lernen zu unterstützen.

**2012 ACM Computing Classification** Applied computing-Interactive learning environments

**Keywords and phrases** Adaptives Lernen, Adaptive Benutzeroberfläche, BCI, Gehirn-Computer-Schnittstelle

## 1 Einführung

Erfolgreiches und effizientes Lernen hängt von vielen Faktoren ab. Neben den kognitiven Fähigkeiten und der Motivation des Lernenden wird der Lernerfolg auch von der Wahl des Lernmediums beeinflusst. Laut Yuksel et al. [9], zeigen Metaanalysen positive Ergebnisse beim Einsatz von computergestützten Lernsystemen. *Adaptives Lernen* bezeichnet das Anpassen eines Lernsystems oder -mediums basierend auf dem Kontext oder Informationen über die lernende Person. Diese Informationen können sich auf die Leistung oder das Lerntempo beziehen, oder den psychischen und physischen Zustand des Lernenden beschreiben [9].

Die Verwendung von Gehirn-Computer-Schnittstellen (engl. *brain-computer interface*, BCI) für adaptive Lernsysteme ermöglicht es Daten über den Zustand des Gehirns zu messen und auszuwerten. Aus verschiedenen physiologischen Eigenschaften des Gehirns lässt sich auf bestimmte mentale Zustände schließen.



© Raimund Krämer;

licensed under Creative Commons License CC-BY

Cite as: Raimund Krämer. Adaptives Lernen mit Gehirn-Computer-Schnittstellen. 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. 5:1–5:14.

Solche physiologischen Eigenschaften können beispielsweise der Blutsauerstoffgehalt [4, 9] oder die elektrischen Aktivitäten im Gehirn sein [6], die mittels verschiedener Technologien gemessen werden können.

Die verschiedenen Technologien haben unterschiedliche Genauigkeiten und Reaktionszeiten [3, 4, 8], wodurch sie sich für verschiedene Anwendungen eignen. So kann beispielsweise der Schwierigkeitsgrad eines zu lernenden Musikstücks an die mentale Auslastung angepasst werden [9], oder es wird die Aufmerksamkeit des Lernenden gemessen und daraufhin bestimmte Lerneinheiten zur Wiederholung empfohlen [8].

Diese Arbeit gibt zuerst einen Überblick über häufig verwendete BCI-Technologien (Abschnitt 2) und erklärt grundlegende Begriffe. Dann werden Beispiele von Anwendungsgebieten von Gehirn-Computer-Schnittstellen im Zusammenhang mit adaptiven Systemen anhand von bisherigen Studien aufgezeigt (Abschnitt 3). Anschließend werden die Vor- und Nachteile der verschiedenen Technologien erläutert (Abschnitt 4).

## 2 BCI-Technologien

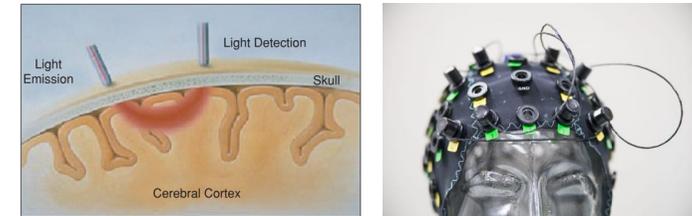
Zur Messung von kognitiven Daten gibt es verschiedene Möglichkeiten. Dieser Abschnitt erklärt wichtige Techniken und Begriffe im Zusammenhang mit BCIs. Unabhängig von der technischen Messmethode werden zwei Arten von Gehirn-Computer-Schnittstellen unterschieden:

**Aktive Gehirn-Computer-Schnittstellen** erlauben dem Nutzer ein System mithilfe von kognitiven Signalen zu steuern. So kann ein Nutzer beispielsweise einen automatischen Rollstuhl mit seinen Gedanken steuern [8], oder durch mentale Kraft einen Mauscursor über den Bildschirm bewegen und dann durch Imagination eines Mausklicks diesen auslösen [2, 8].

**Passive Gehirn-Computer-Schnittstellen** überwachen die kognitive Aktivität des Nutzers, um zusätzliche Informationen zum Kontext zu erlangen, und passen sich an den jeweiligen mentalen Zustand des Nutzers an [2, 8]. Daneben gibt es Kombinationen aus beiden Konzepten [8].

### 2.1 Funktionelle Nahinfrarot-Spektroskopie

Die *funktionelle Nahinfrarot-Spektroskopie* (fNIRS) misst die kognitive Auslastung des Gehirns anhand des Blutsauerstoffgehalts im Gehirn [4]. Bei stärkerer Gehirnaktivität steigen die Durchblutung und der Sauerstoffgehalt des Blutes in den jeweiligen Hirnarealen an. Diese physiologischen Veränderungen können optisch gemessen werden, indem das Gewebe mit Licht nahe am Infrarotspektrum beleuchtet und die Menge der zurück gestreuten Photonen mittels



(a) Die gemessene Intensität des gewebedurchdringenden Lichts hängt von der Aktivität des Hirnareals ab. (Abbildung: Bunce et al. [4])

(b) Beispiel für ein fNIRS-Gerät. Gut erkennbar sind die Lichtemitter und -sensoren. (Abbildung: Universität Trier [1])

■ **Abbildung 1** Funktionelle Nahinfrarot-Spektroskopie (fNIRS)

Lichtsensoren ermittelt wird [2, 4] (siehe Abbildung 1a). Abhängig von der Durchblutung und dem Sauerstoffgehalt werden unterschiedlich viele Photonen vom Gewebe absorbiert bzw. zurück gestreut [4].

fNIRS-Geräte bestehen aus Lichtemittern und -sensoren, die von außen an die Kopfoberfläche angelegt werden. Das Verfahren ist daher nicht-invasiv [4]. Abbildung 1b zeigt ein solches Gerät.

## 2.2 Elektroenzephalographie (EEG)

Bei der *Elektroenzephalographie* wird mittels Elektroden die Gehirnaktivität gemessen, die sich bei der Freisetzung von Neurotransmittern in der Großhirnrinde in Form von elektrischen Ladungen äußert [8]. Die Elektroenzephalographie hat verglichen mit den anderen Methoden eine relativ hohe temporale Auflösung [3, 8] und erlaubt dadurch mehrere zuverlässige Messungen pro Sekunde [3]. Die räumliche Auflösung dagegen ist gering, sodass es schwer ist eine bestimmte Gehirnaktivität dem jeweiligen Teil des Gehirns zuzuordnen [8].

## 3 Anwendungsfälle von Gehirn-Computer-Schnittstellen (BCIs)

### 3.1 Adaptives Lernen

#### 3.1.1 Verbessern der Lerneffektivität

Daniel Szafrin und Bilge Mutlu von der Universität von Wisconsin-Madison haben ein System [8] entwickelt, das die Effektivität des Lernens mit digitalen Lernplattformen erhöht. Das als *ARTFuL* (kurz für *Adaptive Review Technology for Flipped Learning*) bezeichnete System [8] überwacht die Aufmerksamkeit



■ **Abbildung 2** Eine Probandin im Experiment von Szafor und Mutlu. Das System überwacht ihre Aufmerksamkeit während einer Lektion in Kunstgeschichte. (Abbildung: Szafor, Mutlu [8])

der lernenden Person beim Ansehen von Videolektionen. Es eignet sich deshalb als Ergänzung für die als “umgedrehter Unterricht” (engl. *flipped classroom*) bekannte Unterrichtsmethode, bei der die Schüler als Hausaufgabe selbstbestimmt Onlinelektionen ansehen und bei der Anwesenheit im Unterricht, unter Betreuung durch die Lehrkraft, üben und Verständnisfragen klären.

Abbildung 2 zeigt das System in Aktion. Ein Elektroenzephalograph auf dem Kopf der Probandin misst ihre Gehirnaktivität. Basierend auf den Messungen werden ihr anschließend Lektionen zum erneuten Ansehen empfohlen, bei denen eine verringerte Aufmerksamkeit festgestellt wurde.

Neben Gehirn-Computer-Schnittstellen wären auch andere Techniken verwendbar, um die Aufmerksamkeit von Lernenden zu überwachen. Szafor und Mutlu nennen beispielsweise die Blickerfassung (*Okulographie*) und Veränderungen in der Herzfrequenz. Die Elektroenzephalographie (EEG) wurde laut Szafor und Mutlu gewählt, da vorangegangene Studien gezeigt haben, dass die Elektroenzephalographie sowohl zur Messung der Aufmerksamkeit von Personen als auch zur Einschätzung des Schwierigkeitsgrades des Lernmaterials genutzt werden kann.

Die Autoren Szafor und Mutlu stellten zwei Hypothesen auf, die sie in der Studie untersuchten [8]:

- Hypothese 1: Das Wiederansehen einer Lektion, bei der die Aufmerksamkeit beim ersten Ansehen niedrig war, verbessert das Lernergebnis. Falls die Aufmerksamkeit beim ersten Ansehen bereits hoch war, verbessert sich das Lernergebnis nicht bei erneutem Ansehen der Lektion.
- Hypothese 2: Werden die Lektionen, bei denen die Aufmerksamkeit beim ersten Ansehen niedrig war, erneut angesehen, ist das Lernergebnis genauso gut wie beim erneuten Ansehen aller Lektionen, wird aber in kürzerer Zeit erreicht. Umgekehrt erreicht das Wiederholen nur der Lektionen, die beim ersten Mal aufmerksam angesehen wurden, nicht dasselbe Lernergebnis wie eine Wiederholung aller Lektionen.

In einem Laborexperiment wurde die Effektivität des adaptiven Systems untersucht. Um die Hypothesen zu untersuchen, wurde das adaptive System mit drei Varianten des Systems verglichen, die keine Wiederholung, eine Wiederholung aller Lektionen beziehungsweise eine Wiederholung der bereits mit hoher Aufmerksamkeit angesehenen Lektionen bieten. Die Testpersonen wurden in vier gleich große Gruppen mit jeweils gleicher Anzahl an Männern und Frauen aufgeteilt. Die Lektionen wurden aus dem Bereich der Kunstgeschichte gewählt, um ein realistisches Lernszenario zu bieten, mit dessen Inhalten die Teilnehmenden mit hoher Wahrscheinlichkeit noch nicht vertraut waren. [8]

Die Studie von Szafor und Mutlu ergab, dass bei Nutzung des adaptiven Systems und Wiederansehen der Lektionen, bei denen die Aufmerksamkeit am geringsten war, das Lernen deutlich effektiver ist, als bei Testpersonen, denen keine Lektionen zur Wiederholung empfohlen wurden. Bei Nutzung des adaptiven Systems erinnerten sich die Testpersonen unabhängig vom Geschlecht 29 % besser an die gelernten Informationen als ohne Wiederholungen. Testpersonen aus der Gruppe, die die falschen Lektionen wiederholte, nämlich die, die schon beim ersten Mal mit hoher Aufmerksamkeit angesehen wurden, schnitten annähernd gleich gut ab wie die Gruppe ohne Wiederholungen. Somit konnten beide Aspekte der ersten Hypothese bestätigt werden. [8]

Das adaptive System war genauso effektiv wie ein Wiederholen des kompletten Stoffes. Dieses Ergebnis wurde aber mit 75 % geringerem Zeitaufwand für Wiederholungen von Lektionen erreicht. Somit wurde der erste Teil von Hypothese 2 bestätigt. Der zweite Teil der zweiten Hypothese wurde nicht vollständig bestätigt. Das Experiment ergab, dass sich der Lerneffekt bei Testpersonen, die die Lektionen wiederholten, die zuvor mit hoher Aufmerksamkeit angesehen wurden, kaum vom Lerneffekt bei vollständiger Wiederholung unterschied, während Hypothese 2 in diesem Bereich einen signifikanten Unterschied voraussagte. [8]



■ **Abbildung 3** Bei sinkender Aufmerksamkeit des Kindes projiziert das System *FOCUS* eine Übung auf das Buch. Dadurch wird die Aufmerksamkeit zurücklerlangt. (Abbildung: Huang et al. [6])

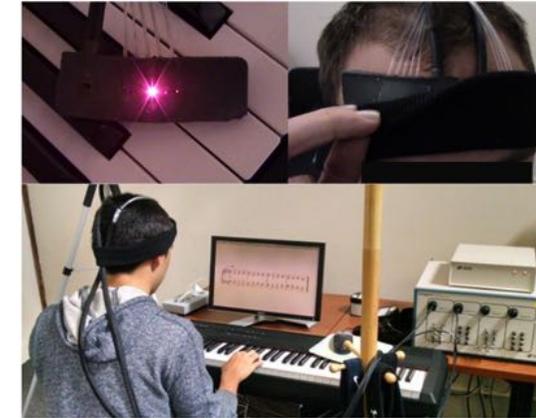
### 3.1.2 Steigern der Leseeffizienz

Einen ähnlichen Ansatz verfolgen Huang et al. mit ihrem System *FOCUS* [6], das die Aufmerksamkeit von Kindern beim Lesen überwacht und beim Feststellen geringer Aufmerksamkeit in einen Übungsmodus wechselt. Die Übungen werden im Kontext des Leseinhalts gewählt, wodurch die Aufmerksamkeit wieder auf das Buch gelenkt werden soll.

Wie ARTFuL nutzt auch *FOCUS* die Elektroenzephalographie (EEG) als Gehirn-Computer-Schnittstelle. Das BCI-Gerät wird während des Lesens getragen (Abbildung 3). Das Buch ist speziell für diesen Zweck entwickelt, damit die Lese- und Übungseinheiten aufeinander abgestimmt sind. Sinkt die Aufmerksamkeit, wird mit einem kleinen Projektor eine Übung auf die dafür vorgesehenen Bereiche des Buches projiziert. In einer Lektion über die Jahreszeiten kann eine solche Übung beispielsweise sein, dass das Kind durch Konzentration der Aufmerksamkeit auf eine virtuelle Blume diese dazu bringen soll, ihre Blüte zu öffnen. [6]

Das System nutzt zwei Aufmerksamkeitsschwellwerte, für niedrige beziehungsweise hohe Aufmerksamkeit. Das System wechselt in den Übungsmodus, wenn die gemessene Aufmerksamkeit für 6 Sekunden unterhalb des unteren Schwellwerts liegt. Die Übung gilt als absolviert, wenn die Aufmerksamkeit für 6 Sekunden oberhalb des höheren Schwellwerts liegt. [6]

Das System wurde im Laborexperiment mit 24 Grundschulern und -schülerinnen getestet. Huang et al. konnten zeigen, dass die im Kontext des Leseinhalts gewählten Übungen die Aufmerksamkeit von Kindern beim Lesen erhöhen und damit die Leseeffizienz verbessern. Ob die Aufmerksam-



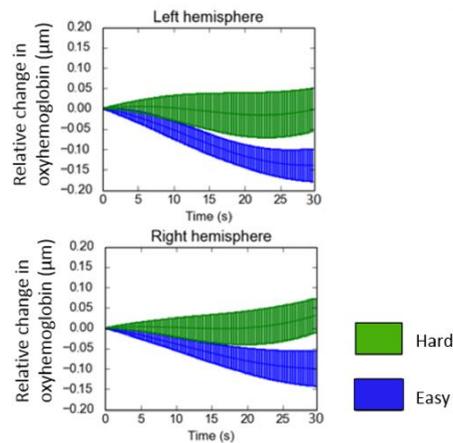
■ **Abbildung 4** Oben links: fNIRS-Sensor. Oben rechts: Trageposition des Sensors. Unten: Der experimentelle Aufbau. Sobald das Musikstück gemeistert ist, wird der Schwierigkeitsgrad automatisch erhöht. Dazu wird die mentale Auslastung des Lernenden überwacht. [9] (Abbildung: Yuksel et al. [9])

keitsübungen einen ähnlichen Effekt haben, wenn sie nicht mit dem Gelesenen zusammenhängen, wurde in der Studie nicht untersucht. [6]

### 3.1.3 Anpassung des Schwierigkeitsgrads

Anders als die beiden vorherigen Systeme gingen Yuksel et al. [9] bei der Entwicklung ihres Systems *BACH* vor, indem sie nicht die Aufmerksamkeit der lernenden Person, sondern deren mentale Auslastung überwachen. *BACH* (kurz für *Brain Automated Chorales*) unterstützt beim Lernen von Musikstücken auf dem Klavier (Abbildung 4). Es nutzt eine fNIRS-Gehirn-Computer-Schnittstelle und passt automatisch den Schwierigkeitsgrad an den Zustand des Lernenden an. Sinkt die mentale Auslastung, deutet dies darauf hin, dass der aktuelle Schwierigkeitsgrad beherrscht wird und Gehirnkapazitäten für einen höheren Schwierigkeitsgrad verfügbar sind. fNIRS eignet sich laut Yuksel et al. für diesen Zweck, da es nicht-invasiv ist und sich damit die kognitive Aktivität in der Großhirnrinde messen lässt.

Das System wurde so entwickelt, dass es den Schwierigkeitsgrad zwar bei geringer mentaler Auslastung erhöht, aber niemals verringert, da die Ursache für eine hohe mentale Auslastung schwer auszumachen ist [9]. Sie könnte bedeuten, dass der Lernende überfordert ist, oder aber, dass er nur *gefordert* ist und gerade aktiv lernt. Laut Yuksel et al. ist die Frage, ob eine hohe mentale Auslastung vorteilhaft für den Lernprozess ist, noch nicht ausreichend



■ **Abbildung 5** Die durchschnittliche Änderung der Menge sauerstoffhaltigen Hämoglobins über die Dauer eines 30-sekündigen Musikstücks. Bei einem schwierigen Musikstück (grün) war die durchschnittliche Änderung deutlich größer als bei einem einfachen Musikstück (blau). (Abbildung: Yuksel et al. [9])

erforscht, und es seien zusätzliche Informationen nötig, wie beispielsweise die aktuellen Emotionen der lernenden Person, um deren mentalen Zustand richtig interpretieren zu können.

Im Experiment wurde bei acht männlichen und acht weiblichen Testpersonen die Gehirnaktivität in der Großhirnrinde beider Gehirnhälften gemessen. Dafür wurde auf jeder Seite der Stirn je eine fNIRS-Sonde mit vier Nahinfrarotlichtquellen und einem Lichtsensor platziert. Die Autoren bestätigten zunächst, dass die gemessenen Daten tatsächlich die mentale Auslastung bei einem schwierigen beziehungsweise einfachen Musikstück repräsentieren. Abbildung 5 zeigt, dass die Menge an sauerstoffhaltigem Hämoglobin bei einem schwierigen Musikstück deutlich höher war als bei einem einfachen Musikstück.

Die Resultate der Studie [9] zeigten, dass die Testpersonen im Experiment deutlich akkurater spielten. Insbesondere spielten sie mehr richtige und weniger falsche Noten und ließen weniger Noten aus, während sie das adaptive System verwendeten. Außerdem gaben sie an, dass ihnen das Lernen des Musikstücks mit dem Lernsystem leichter fiel und sie das Stück anschließend besser beherrschten als beim Lernen ohne *BACH*.



■ **Abbildung 6** Nutzer mit fNIRS-Gerät beim Experiment mit der adaptiven Benutzeroberfläche. (Abbildung: Afergan et al. [2])

### 3.2 Adaptive Benutzeroberflächen

Gehirn-Computer-Schnittstellen können verwendet werden, um alltägliche Aufgaben zu vereinfachen. Kognitive Daten können beispielsweise helfen, die Nutzbarkeit (engl. *Usability*) von grafischen Benutzeroberflächen zu verbessern [2]. Diese adaptiven Benutzeroberflächen könnten aufgrund ihrer Funktionsweise geeignet sein, in Kombination mit adaptiven Lernsystemen eingesetzt zu werden.

Afergan et al. [2] haben ein System entwickelt, das die Elemente einer Benutzeroberfläche automatisch anpasst, wenn der Nutzer gerade mehrere Tätigkeiten gleichzeitig bei hoher mentaler Auslastung ausführt, um das Auswählen der Elemente zu vereinfachen (*Adaptive Interface*). Das System nutzt die funktionelle Nahinfrarot-Spektroskopie für die Messung kognitiver Daten [2] (Abbildung 6).

Während andere Systeme oft bestimmte Schwellwerte haben [2][8][9], ab denen die mentale Auslastung beispielsweise als hoch oder gering angesehen wird, nutzt dieses System eine stetige Messskala, um die Stärke des Multitaskings zu messen [2]. Sie stützen sich dabei auf eine Technik mit dem Namen *Bubble Cursor* [2, 5].

Der *Bubble Cursor* wurde von Tovi Grossman und Ravin Balakrishnan entwickelt [5]. Er passt den Auswahlbereich des Cursors abhängig von der Anzahl der umliegenden auswählbaren Elemente automatisch an. Er löst damit das Problem des *Area Cursors* [5], der die Auswahl von Elementen vereinfachen soll (Abbildung 7a), aber zu Problemen führt, wenn mehrere Ziele

im Auswahlbereich liegen (Abbildung 7b). Daher verändern Grossman und Balakrishnan beim Bubble Cursor die Größe des Auswahlbereichs, damit nur das nächstgelegene Element ausgewählt wird (Abbildung 7c), und verändern zusätzlich die Form des Auswahlbereichs, wenn die Größenanpassung nicht für eine eindeutige Auswahl ausreicht (Abbildung 7d).

Afergan et al. nutzen diese verbesserte Auswahlpräzision des Bubble Cursors, und erweitern die Technik mithilfe der Überwachung von Daten über den mentalen Zustand des Nutzers. Weisen die Daten darauf hin, dass der Nutzer gerade mehrere Tätigkeiten gleichzeitig ausübt (Multitasking), wird die Größe der Elemente in der Benutzeroberfläche erhöht, bevor sie in der Kalkulation für den Bubble-Cursor-Algorithmus verwendet werden. Die visuelle Darstellung der Elemente ändert sich jedoch nicht, anders als beim ursprünglichen Bubble Cursor [2], sondern lediglich die Werte, die für die Berechnungen verwendet werden [2]. Dies wurde bewusst entschieden, damit der Nutzer nicht durch unerwartete Bewegungen im Sichtfeld abgelenkt wird. Bei Benutzeroberflächen, deren Elemente unterschiedliche Prioritäten haben, können sie zusätzlich gewichtet werden, damit der Auswahlbereich zu den Elementen mit höherer Priorität tendiert [2].

Durch die Verwendung einer stetigen Messskala kann die Aktivierungsfläche der Elemente nicht nur zwischen normaler und erweiterter Größe umgeschaltet werden, sondern auch beliebige Größen zwischen der normalen und einer maximalen Größe annehmen [2].

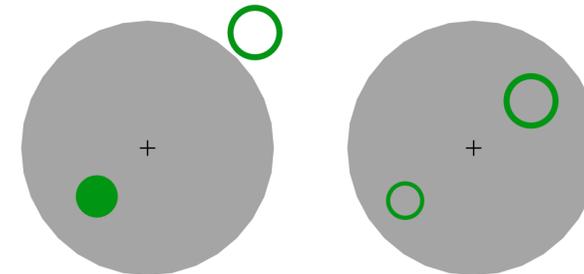
Die Studie zeigt aber auch, dass das Vergrößern von Elementen mit hoher Priorität, wenn der Nutzer nicht mental ausgelastet ist, einen negativen Effekt auf die Auswahlgenauigkeit hat [2]. Ohne die Gehirn-Computer-Schnittstelle wäre die Technik daher wenig brauchbar.

Da adaptive Lernsysteme, bei denen eine grafische Benutzeroberfläche bedient werden muss, Multitasking erfordern, könnte diese Technologie in Kombination mit solchen verwendet werden. Durch die Vereinfachung der Bedienung bei festgestelltem Multitasking könnte die Konzentration des Nutzers oder der Nutzerin weiterhin auf den Lerninhalt gerichtet bleiben, ohne dass eine starke Ablenkung durch die Bedienung der Oberfläche entsteht.

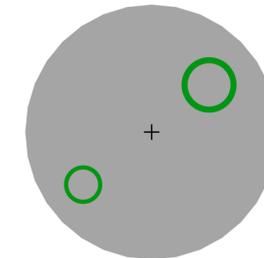
#### 4 Grenzen der Nutzung von BCIs für adaptives Lernen

Die verschiedenen Technologien von Gehirn-Computer-Schnittstellen haben unterschiedliche Eigenschaften. Für jeden Anwendungsfall müssen deshalb Vor- und Nachteile berücksichtigt und die passendste Technologie gewählt werden. Dieser Abschnitt befasst sich mit dieser Problematik und vergleicht die in Abschnitt 2 vorgestellten BCI-Technologien.

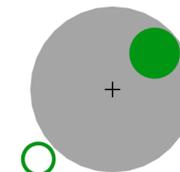
Die in dieser Arbeit vorgestellten BCI-Technologien haben den Vorteil,



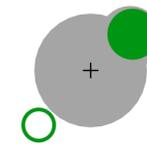
(a) Area Cursor vereinfachen die Auswahl von Elementen in grafischen Benutzeroberflächen [5].



(b) Wenn mehrere Elemente innerhalb der Fläche liegen, wird die Auswahl mit dem Area Cursor schwierig [5].



(c) Der Bubble Cursor passt automatisch seine Größe an, um eine eindeutige Auswahl zu ermöglichen [5].



(d) Der Bubble Cursor passt zusätzlich seine Form an, wenn das Anpassen der Größe nicht für eine eindeutige Auswahl reicht [5].

■ **Abbildung 7** Prinzip des Bubble Cursors, verglichen mit dem Problem des Area Cursors. Grau: Auswahlbereich des Cursors, Grün: auswählbare Elemente (ausgefüllt bedeutet selektiert). Abbildungen: Nach Grossman und Balakrishnan [5].

dass sie nicht-invasiv sind [4, 6]. Daher eignen sie sich besonders für Anwendungsfälle, in denen das Gerät kurzfristig an- oder abgelegt werden kann. Es existieren auch Gehirn-Computer-Schnittstellen, die chirurgisch dauerhaft in die Großhirnrinde implantiert werden [7]. Diese dürften aber aus Gründen der Nutzerfreundlichkeit in den hier vorgestellten Bereichen, besonders im Bereich der Bildung und bei adaptiven Benutzeroberflächen, wenig praktikabel und eher als aktive Gehirn-Computer-Schnittstelle für Menschen mit chronischer Behinderung geeignet sein, wie in dem in Abschnitt 2 erwähnten Fall der Rollstuhlsteuerung.

Ein häufig anzutreffender Begriff im Zusammenhang mit Gehirn-Computer-Schnittstellen ist der der räumlichen beziehungsweise temporalen Auflösung [4]. Die Elektroenzephalographie hat eine hohe temporale Auflösung, wodurch es möglich ist, pro Sekunde mehrere präzise Messungen durchzuführen und die gemessenen Daten mit den auslösenden Reizen in Verbindung zu bringen [3, 8]. Sie eignet sich dadurch für Anwendungsfälle, die geringe oder echtzeitnahe Verzögerungen benötigen. fNIRS dagegen hat eine Verzögerung von 5 bis 7 Sekunden, da es etwas dauert, bis bei mentaler Anstrengung die Durchblutung im Gehirn steigt [9]. Bei den in Kapitel 3 vorgestellten Anwendungsfällen ist dies kein Problem, die Messmethode eignet sich deshalb aber nicht für Echtzeitanwendungen.

Ein weiteres Problem ist das schwierige Schlussfolgern von physiologischen Messungen auf psychologische Zustände. Bei ARTFuL wird davon ausgegangen, dass ein Nutzer oder eine Nutzerin des Systems am ehesten eine Lektion wiederholen sollte, bei der die Aufmerksamkeit am niedrigsten war. Laut Szafrin und Mutlu muss dies aber nicht grundsätzlich zutreffen. Beispielsweise könnte die Ursache für die geringe Aufmerksamkeit sein, dass die lernende Person mit dem vermittelten Stoff bereits vertraut und daher gelangweilt ist, was eine Wiederholung überflüssig macht [8]. Zudem könnte ein Wiederholen einer bereits bekannten Lektion aufgrund geringer gemessener Aufmerksamkeit diesen Effekt weiter verstärken, da der Nutzer oder die Nutzerin weiterhin von der Lektion gelangweilt wäre, sodass das System eine weitere Wiederholung empfehlen würde.

## 5 Zusammenfassung und Fazit

Erfolgreiches und effizientes Lernen hängt unter anderem vom verwendeten Lernmedium ab. Diese Arbeit hat gezeigt, dass bereits verschiedene adaptive Lernsysteme entwickelt wurden, die die Effektivität und Effizienz des Lernens verbessern können. Es wurden die Technologien *funktionelle Nahinfrarotspektroskopie* und *Elektroenzephalographie* einführend erklärt.

Mit ARTFuL wurde ein System vorgestellt, das dem Nutzer oder der Nut-

zerin Lektionen zur Wiederholung empfiehlt, wenn es anhand der gemessenen Gehirnaktivität eine verringerte Aufmerksamkeit feststellt. Mit *FOCUS* wurde ein Beispiel für ein adaptives Lernsystem gegeben, das die Aufmerksamkeit von Kindern beim Lesen verbessern kann. Am Beispiel des Systems *BACH* wurde gezeigt, dass adaptive Lernsysteme den Schwierigkeitsgrad des Lernstoffs dynamisch an den mentalen Zustand des oder der Lernenden anpassen können. Das Konzept von adaptiven Benutzeroberflächen wurde vorgestellt, und es wurde erklärt, warum diese in Kombination mit adaptiven Lernsystemen eingesetzt werden könnten.

Es zeigt sich, dass Gehirn-Computer-Schnittstellen trotz dieser vielversprechenden Entwicklungen noch starke technische Einschränkungen aufweisen. Besonders die geringe räumliche oder temporale Auflösung der jeweiligen Technologien können je nach Anwendung problematisch sein. Dazu kommt, dass beispielsweise die geringe temporale Auflösung der funktionellen Nahinfrarotspektroskopie mehr physiologisch als technisch bedingt ist. Für einige dieser Probleme könnten mit der Zeit technische Verbesserungen zu erwarten sein, oder ganz neue Technologien mit vorteilhaften Eigenschaften erscheinen. Bis Gehirn-Computer-Schnittstellen, in Kombination mit adaptiven Lernsystemen, zur Standardausstattung von Schulen und Schülern gehören dürfte es aber wohl noch etwas dauern.

---

### Literatur

- 1 Uni Trier: fNIRS. <https://web.archive.org/web/20180612122728/https://www.uni-trier.de/index.php?id=59878>. Zugegriffen am 12.06.2018.
- 2 Daniel Afergan, Tomoki Shibata, Samuel W. Hincks, Evan M. Peck, Beste F. Yuksel, Remco Chang, and Robert J.K. Jacob. Brain-based target expansion. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14, pages 583–593, New York, NY, USA, 2014. ACM. URL: <http://doi.acm.org/10.1145/2642918.2647414>, doi:10.1145/2642918.2647414.
- 3 Chris Berka, Daniel J. Levendowski, Michelle N. Lumicao, Alan Yau, Gene Davis, Vladimir T. Zivkovic, Richard E. Olmstead, Patrice D. Tremoulet, and Patrick L. Craven. Eeg correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, Space, and Environmental Medicine*, 78(5):B231–B244, 2007. URL: <https://www.ingentaconnect.com/content/asma/ase/2007/00000078/a00105s1/art00032>.
- 4 S. C. Bunce, M. Izzetoglu, K. Izzetoglu, B. Onaral, and K. Pourrezaei. Functional near-infrared spectroscopy. *IEEE Engineering in Medicine and Biology Magazine*, 25(4):54–62, July 2006. doi:10.1109/MEMB.2006.1657788.

- 5 Tovi Grossman and Ravin Balakrishnan. The bubble cursor: Enhancing target acquisition by dynamic resizing of the cursor's activation area. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, pages 281–290, New York, NY, USA, 2005. ACM. URL: <http://doi.acm.org/10.1145/1054972.1055012>, doi:10.1145/1054972.1055012.
- 6 Jin Huang, Chun Yu, Yuntao Wang, Yuhang Zhao, Siqi Liu, Chou Mo, Jie Liu, Lie Zhang, and Yuanchun Shi. Focus: Enhancing children's engagement in reading by using contextual bci training sessions. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, pages 1905–1908, New York, NY, USA, 2014. ACM. URL: <http://doi.acm.org/10.1145/2556288.2557339>, doi:10.1145/2556288.2557339.
- 7 D. R. Kipke, R. J. Vetter, J. C. Williams, and J. F. Hetke. Silicon-substrate intracortical microelectrode arrays for long-term recording of neuronal spike activity in cerebral cortex. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11(2):151–155, June 2003. doi:10.1109/TNSRE.2003.814443.
- 8 Daniel Szafr and Bilge Mutlu. Artful: Adaptive review technology for flipped learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 1001–1010, New York, NY, USA, 2013. ACM. URL: <http://doi.acm.org/10.1145/2470654.2466128>, doi:10.1145/2470654.2466128.
- 9 Beste F. Yuksel, Kurt B. Oleson, Lane Harrison, Evan M. Peck, Daniel Afergan, Remco Chang, and Robert JK Jacob. Learn piano with bach: An adaptive learning interface that adjusts task difficulty based on brain state. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 5372–5384, New York, NY, USA, 2016. ACM. URL: <http://doi.acm.org/10.1145/2858036.2858388>, doi:10.1145/2858036.2858388.

# Interaktive intelligente Umgebungen

Jennifer Busta

Ludwig-Maximilians-Universität München, München, Deutschland  
[jennifer.busta@campus.lmu.de](mailto:jennifer.busta@campus.lmu.de)

---

## Zusammenfassung

Das Internet der Dinge kann durch sein intelligentes Netzwerk von vernetzten intelligenten Objekten und Diensten eine Verbesserung der Lebensqualität eines jeden Einzelnen ermöglichen. Damit Nutzer die intelligente Umgebung verstehen und akzeptieren, sollte die Beziehung zwischen dem Nutzer und der intelligenten Umgebung auf Vertrauen, Kontrolle und Intuition beruhen. Die Umsetzung von bestimmten Interaktionsmustern bei der Entwicklung intelligenter Umgebungen verbessert das Vertrauen der Nutzer. Diese Interaktionsmuster besagen, dass die intelligente Umgebung eine Interaktion mit dem Nutzer initiieren und beenden sollte, dem Nutzer nach jeder seiner Aktionen Feedback in Bezug auf diese Aktionen geben sollte und gesammelte Daten für den Nutzer zugänglich machen sollte. Zusätzlich erfordert die Neuartigkeit der allgegenwärtigen intelligenten Objekte neue oder angepasste Interaktionstechniken. Durch diese können Benutzer die intelligente Umgebung kontrollieren und intuitiv bedienen. In dieser Forschungsarbeit werden die folgenden Interaktionstechniken erläutert: sogenannte natürliche mobile, proxemische, geräteübergreifende, blickunterstützte und gestenbasierte Interaktion sowie Interaktion mit personalisierten Druckknöpfen und mobilen 3D-Benutzeroberflächen.

**2012 ACM Computing Classification** Human-centered computing → Ubiquitous and mobile computing theory, concepts and paradigms

**Keywords and phrases** Internet der Dinge; Intelligente Objekte; Intelligente Umgebungen; Interaktion.

## 1 Einführung

Eine auf dem Internet basierende Informationsarchitektur ermöglicht den Austausch von Diensten und Gütern zwischen allen Elementen, die mit dem Netzwerk verbunden sind. Das Internet der Dinge (engl. Internet of Things, IoT) bezieht sich auf die vernetzte Verbindung von Alltagsgegenständen, die mit einer Art Intelligenz ausgestattet sind. Während das Internet eine Plattform ist, durch die Geräte elektronisch miteinander kommunizieren und Informationen sowie spezifische Daten austauschen können, ist das IoT eine Evolution des



© Jennifer Busta;

licensed under Creative Commons License CC-BY

Cite as: Jennifer Busta. Interaktive intelligente Umgebungen. 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.6:1–6:18.

Internets. Das bedeutet, dass es eine umfangreichere Verbindungsflexibilität, eine bessere Wahrnehmung von Informationen und verständlichere intelligente Dienste hat. Das IoT kann als eine Kombination von Sensoren und Aktuatoren betrachtet werden, welche digitalisierte Informationen bereitstellen und empfangen und in einem bidirektionalen Netzwerk platzieren. Dieses Netzwerk ist in der Lage, alle Daten zu übertragen, die von den vielen verschiedenen Diensten und Endnutzern verwendet werden sollen. Das Internet der Dinge entstand zwischen 2008 und 2009. 2010 betrug die Anzahl der physischen Alltagsgegenstände und -geräte, die mit dem Internet verbunden waren, etwa 12,5 Milliarden. Bis 2017 ist diese Zahl bereits auf 25 Milliarden gestiegen, was bedeutet, dass jede Person etwa drei intelligente Objekte besitzt. [9]

Die Anzahl der möglichen Anwendungen und Dienste des IoTs ist nicht limitiert und kann die Lebensqualität der Menschen in vieler Hinsicht erleichtern und verbessern. Beispielsweise kann in vernetzten intelligenten Gebäuden das Management und die Speicherung von Energie verbessert werden, da erkannt werden kann, wann Benutzer zu Hause sind oder in welchen Räumen sie sich befinden. Haushaltsgeräte können mithilfe von intelligenten Sensoren und Aktuatoren durch Benutzer oder automatisch gesteuert werden. In intelligenten Städten können Sicherheitsdienste integriert werden und das öffentliche und private Transportwesen kann durch das intelligente Management von Parkdiensten und Verkehr in Echtzeit optimiert werden. Zum Beispiel können Ampeln abhängig vom Verkehr gesteuert werden. Im Bereich der Bildung können virtuelle und physikalische Klassenzimmer miteinander verbunden werden, um das Lernen effizienter und zugänglicher zu machen. Zum Bereich der Verbraucherelektronik zählen Smartphones, intelligente Fernseher, Laptops und Tablets, sowie intelligente Kühlschränke, Waschmaschinen, Trockner und vieles mehr. Diese Geräte miteinander zu verbinden und kommunizieren zu lassen, kann personalisierte Nutzererfahrungen erweitern. So kann beispielsweise über das Smartphone die Waschmaschine gestartet oder die Kühlschranktemperatur abgefragt werden. Die Gesundheit der Menschen kann mithilfe des IoTs durch das Überwachen von chronischen Krankheiten, Aktivitätstracker, Ferndiagnosen, Sportüberwachung, Überwachung von Essgewohnheiten und vieles mehr verbessert werden. Intelligente Autos können den Automobilsektor verändern. Verkehrsflusskontrolle, Informationen über kaputte Teile des Autos, Selbstdiagnosen, Reifendrucküberwachung, intelligentes Energiemanagement und Energiekontrolle sowie GPS-Tracking sind nur einige Beispiele hierfür. [9]

In den folgenden Kapiteln werden zuerst intelligente Objekte und intelligente Umgebungen genauer erläutert. Da die Interaktion zwischen den Nutzern und intelligenten Umgebungen entscheidend für deren soziale Akzeptanz ist, werden anschließend einige allgemeine Interaktionsmuster vorgestellt. Diese Interaktionsmuster sollen das Vertrauen der Nutzer durch das Verbessern

von Verständnis und Kontrolle fördern. Da intelligente Umgebungen sich immer weiter verbreiten [9] und die Neuartigkeit der intelligenten Objekte neue Interaktionsmöglichkeiten bietet, werden anschließend verschiedene Interaktionstechniken aufgezeigt. Zu den Interaktionstechniken gehören die Interaktion mit personalisierten Druckknöpfen, sogenannte natürliche mobile Interaktion, Interaktion mit mobilen 3D-Benutzeroberflächen, proxemische Interaktion, geräteübergreifende Interaktion, blickunterstützte Interaktion sowie gestenbasierte Interaktion über am Kopf befestigte Bildschirme.

## 2 Intelligente Objekte und Umgebungen

Intelligente Objekte werden aufgrund ihrer zusätzlichen Informations- und Kommunikationstechnologie als intelligent bezeichnet. Die Art und Weise, wie die Technologie integriert ist, entscheidet, ob ein Objekt intelligent ist und wie Menschen mit dem Objekt interagieren können. Mit einem intelligenten Objekt sind brauchbare und nützliche Interaktionen möglich. [11]

Verschiedene Eigenschaften können zur Intelligenz eines Objekts beitragen. Die erste Eigenschaft ist Kontextbewusstsein. Darunter versteht man die Fähigkeit eines Objekts, sein Verhalten an sich ändernde Situationen anzupassen. Proaktives Verhalten ist die Eigenschaft eines Objekts, vorherzusehen was passiert und für das eventuell eintretende Ereignis Vorbereitungen zu treffen. Viele intelligente Objekte haben auch verinnerlichtes Wissen. Darunter versteht man die interne Anwendung von wissensbasierten Systemen, insbesondere einer Schlussfolgerungskomponente, die zur Laufzeit neue Fakten erzeugen kann und dadurch die Flexibilität und Adaptierbarkeit des intelligenten Objekts weiter erhöht. Eine weitere Eigenschaft, die zur Intelligenz eines Objekts beiträgt, ist die Produkt-zu-Nutzer-Interaktion. Hier teilt das Objekt dem Nutzer mit, was das Objekt weiß, woher es diese Informationen hat und was es macht. Zusätzlich gibt es die Produkt-zu-Produkt-Interaktion, also die dynamische Ankopplung eines intelligenten Objekts an andere intelligente Objekte. Dies kann hierarchisch erfolgen, wodurch es intelligente Bestandteile oder auch intelligente Umgebungen geben kann [11]. Die Nutzer-zu-Produkt-Interaktion ist eine zielgerichtete Interaktion, bei der der Nutzer ein bestimmtes Ergebnis erreichen will [2]. Diese Interaktion wird vom Benutzer initiiert und das intelligente Objekt oder die intelligente Umgebung kann als Benutzeroberfläche betrachtet werden, in welcher die Aktionen des Benutzers zu System- und Umgebungsantworten führen.

Intelligente Objekte sind also unabhängige Entitäten, die Informationen erkennen und empfangen, selbstständig agieren und in einem bidirektionalen Netzwerk miteinander kommunizieren. Daraus ergibt sich eine aktive Rolle der intelligenten Umgebung. Sie sollte ihr Verhalten also nicht nur an

menschliche Aktionen anpassen, sondern auch aktiv Interaktionen initiieren. Interagierende Akteure in einer intelligenten Umgebung beobachten also die Interaktionsumgebung, andere Teilnehmer und die Situation, interpretieren diese Beobachtungen und schlussfolgern daraus was die zugehörige Aktion ist, welche sie dann ausführen. [2]

Es kann vorkommen, dass Nutzer das Verhalten von intelligenten Umgebungen nicht verstehen oder aufdringlich finden [14]. Sie haben Angst davor, zu viel von Technologien abhängig zu sein und die Kontrolle zu verlieren [8]. Für explizite Interaktionen ist es schwierig, eine intuitive Geräte-Auswahl-Oberfläche zu designen. Für implizite Interaktionen können Benutzer das Verhalten des Systems nicht bestimmen [12]. Dies kann zu einer Überautomatisierung der intelligenten Umgebung führen [12]. Da für implizite Interaktionen die Systemoberflächen fehlen und der Nutzer so nur schwer ein mentales Modell der Umgebung aufbauen kann, fehlt dem Nutzer ein ausreichendes Systembild, durch das er das Verhalten der intelligenten Umgebung nicht versteht [12]. Auch Datenschutz und Sicherheit und das zugehörige Gefühl von Intimität spielen eine entscheidende Rolle bei der Entwicklung von intelligenten Umgebungen [14]. Das Management der Informationen und Aktivitäten muss in einer privaten und sicheren Art und Weise gehandhabt werden [14]. Die Beziehung zwischen den Nutzern und intelligenten Umgebungen ist ein Schlüsselfaktor für die Akzeptanz dieser [14]. Verständnis, Kontrolle und Intuitivität sind die Schlüsselfaktoren, damit Benutzer intelligenten Umgebungen vertrauen [14].

### 3 Interaktionsmuster

Die Umsetzung der nachfolgenden Interaktionsmuster wirkt den gerade beschriebenen Problemen (siehe Kapitel 2) entgegen, und vergrößert somit das Vertrauen der Nutzer. Ein Interaktionsmuster ist das Begrüßungsmuster. Es besagt, dass die intelligente Umgebung ihre Benutzer darüber informieren sollte, ob ein intelligentes Objekt, beziehungsweise eine seiner Anwendungen, verwendet werden kann. Wenn dies nicht möglich ist, sollte ein Grund dafür angegeben werden. Dadurch kann der Benutzer besser nachvollziehen, wie die intelligente Umgebung funktioniert. [14]

Die intelligente Umgebung sollte die Konversation mit den Benutzern initiieren. Damit einhergehend erfordert das Konversationsmuster, dass eine Konversation auch mit einem Feedback an die Benutzer enden sollte, sodass diese sicherstellen können, dass die gesammelten Informationen kontrolliert und aktualisiert wurden. Alle Fragen und Antworten des Konversationsprozesses zwischen dem Nutzer und der intelligenten Umgebung sollten in einem zugänglichen und überprüfbaren Dokument erscheinen. Dadurch kann der Benutzer die Funktionsweise der intelligenten Umgebung besser nachvollziehen.

[14]

Ein weiteres Interaktionsmuster ist das Aktions-Reaktions-Muster, welches besagt, dass die intelligente Umgebung dem Benutzer nach einer Aktion des Benutzers ein Feedback über die Anwendung in Bezug auf diese Aktion geben sollte. Auch dieses Interaktionsmuster trägt zum Verständnis der intelligenten Umgebung bei. [14]

Wird das Abschiedsmuster umgesetzt, so fragt die intelligente Umgebung den Benutzer am Ende einer Interaktion, ob die gesammelten Informationen gespeichert oder gelöscht werden sollen. Unabhängig von der Antwort des Benutzers sollte die intelligente Umgebung den Benutzer darüber informieren, ob die Daten gespeichert oder gelöscht wurden und was der Status dieses Prozesses ist. Dadurch versteht der Nutzer, dass eine Anwendung sicher beendet wurde. [14]

Intelligente Objekte und ihre Anwendungen analysieren die Aktionen und Aktivitäten der Nutzer, damit sich die intelligente Umgebung anpassen und verbessern kann. Dies ist keine direkte Benutzer-Umgebung-Interaktion. Gerade deswegen sollte jede Informationsverarbeitung den Nutzern mittels passenden Feedbacks mitgeteilt werden. Dieser Prozess wird als Explorationsmuster bezeichnet. So kann der Benutzer auch das unsichtbare Verhalten der Umgebung nachvollziehen. [14]

## 4 Interaktionstechniken

Neben den verschiedenen Interaktionsmustern, die bei der Entwicklung von intelligenten Umgebungen umgesetzt werden sollten, gibt es verschiedene Möglichkeiten, mit intelligenten Umgebungen zu interagieren. Für diese Möglichkeiten wurden in den letzten Jahren viele Forschungsarbeiten veröffentlicht. Im Folgenden werden einige davon beschrieben. Die jeweilige Forschungsarbeit ist dabei als Beispiel für die beschriebene Interaktionstechnik zu betrachten. Das Hauptaugenmerk der Beschreibungen liegt auf der Bedienbarkeit und Verständlichkeit der Interaktionstechnik in Bezug auf den Nutzer.

### 4.1 Interaktion mit personalisierten Druckknöpfen

Müller et al. [5] entwickelten als Schnellverfahren für die Interaktion mit intelligenten Geräten personalisierte vom Benutzer getragene Buttons (engl. Personalized User-Carried Single Button Interface, PUCSBI).

Sie entwickelten einen Prototyp in einer Büroumgebung für eine Kaffeemaschine, die mit einem PUCSBI bedient werden kann. Abbildung 1a zeigt den Aufbau des Prototyps. Ein kleiner Berührungsbildschirm neben der Kaffeemaschine fordert den Benutzer auf den Button zu drücken (siehe Abbildung 1b). Drückt der Benutzer den Button, so zeigt der Berührungsbildschirm die



■ **Abbildung 1** Kaffeemaschine, die mit einem personalisierten Druckknopf und einem Berührungsbildschirm bedient werden kann: a) Aufbau des Prototyps; b) Benutzer wird aufgefordert auf den Druckknopf zu drücken; c) Ein benutzerkonfiguriertes Kaffeerezept wird ausgewählt [5]

für diesen Benutzer gespeicherten Rezepte (siehe Abbildung 1c). Nach drei Sekunden wird automatisch das Standardrezept ausgegeben. Vorher kann der Benutzer auch eines seiner anderen konfigurierten Rezepte auswählen. Die Rezepte kann der Benutzer auf einer Webseite bearbeiten. Somit wird das komplexe Konfigurieren der Kaffeemaschine auf einen einzigen expliziten Interaktionsschritt, den Knopfdruck, reduziert. Die Benutzer müssen sich mit dem PUCSBI auch nicht mehr in eine Liste eintragen, um den Kaffee zu bezahlen, da dies automatisch geregelt wird.

Die Benutzer gaben positives Feedback für den Prototypen, da das System schneller und einfacher zu benutzen ist als das Vorherige. Das Prinzip eines PUCSBI kann auch auf andere intelligente Geräte übertragen werden. Das Zielgerät kann durch minimale Distanz, Zeigen oder die räumliche Beziehung zwischen Benutzer und Gerät ausgewählt werden. Im nächsten Kapitel werden diese Auswahlmöglichkeiten für mobile Geräte genauer erläutert.

## 4.2 Natürliche mobile Interaktion

Rukzio et al. untersuchten in [8] sogenannte natürliche mobile Interaktionstechniken. Diese beziehen sich auf die mobile Interaktion zwischen einem Nutzer, einem mobilen Gerät und einem intelligenten Objekt in der realen Welt. Die betrachteten Interaktionstechniken sind Anfassen, Zeigen und Scannen.

Für das Anfassen muss das mobile Gerät in Kontakt mit dem intelligenten Objekt gebracht werden, mit dem der Nutzer interagieren möchte. Dabei muss der Nutzer bereits wissen, dass das Objekt die Anfassen-Fähigkeit hat. Beim Zeigen wählt der Nutzer das intelligente Objekt, mit dem er interagieren möchte, aus, indem er mit dem mobilen Gerät darauf zielt. Für das Scannen werden mit dem mobilen Gerät alle intelligenten Objekte, die in der Nähe sind, gesucht und dem Nutzer über das mobile Gerät in Form einer Liste präsentiert.

In einem Fragebogen gaben Nutzer generell ein positives Feedback für

die Interaktion mit mobilen Geräten, da diese bereits weit verbreitet und Nutzer mit ihnen vertraut sind. Benutzer gaben jedoch auch an, dass sie kein ausreichendes Vertrauen in die Sicherheit von Handy-Technologien haben.

Die Nutzer sahen Anwendungsmöglichkeiten für die natürliche mobile Interaktion mit intelligenten Umgebungen darin, Statusinformationen von einem intelligenten Objekt zu bekommen oder mit intelligenten Objekten aus der Entfernung interagieren zu können.

Abhängig vom Kontext des Benutzers werden Anfassen, Zeigen oder Scannen als Interaktionstechnik bevorzugt. Anfassen wird bevorzugt, wenn das intelligente Objekt in der Nähe ist. Durch Anfassen ist der Auswahlprozess des intelligenten Objekts, mit dem interagiert werden soll, sehr genau und es kommt auch nicht zu Doppeldeutigkeiten, wenn die intelligenten Objekte sehr klein und nahe beieinander sind. Anfassen wird als intuitiv, sicher, fehlerresistent und schnell erachtet. Jedoch ist für diese Interaktion räumliche Nähe erforderlich. Zeigen hingegen wird bevorzugt, wenn sich der Benutzer nicht in der Nähe des Objekts, jedoch in dessen Sichtfeld befindet. Zeigen erfordert daher wenig körperlichen Aufwand und ist einfach zu benutzen. Zielgeräte können intuitiv und schnell erfasst werden und die Benutzeroberfläche hierfür kann einfach gestaltet sein. Jedoch ist ein gewisser kognitiver Aufwand erforderlich und das intelligente Objekt muss sich im Sichtfeld befinden. Es kann außerdem zu Doppeldeutigkeiten im Auswahlprozess kommen, wenn die intelligenten Objekte klein oder nahe beieinander sind. Scannen wird bevorzugt, wenn eine räumliche Distanz überbrückt werden soll. Hierfür ist keine räumliche Nähe erforderlich und das Objekt muss sich nicht im Sichtfeld befinden. Scannen erfordert den geringsten körperlichen Aufwand. Dem Benutzer wird eine Liste mit allen verfügbaren intelligenten Objekten gezeigt, aus welcher der Benutzer das gewünschte Objekt auswählen kann. Dieser Auswahlprozess ist jedoch zeitaufwändiger als bei den beiden zuerst genannten Interaktionstechniken. Außerdem werden dem Nutzer auch Informationen angezeigt, die gegebenenfalls unwichtig für die Situation sind. Scannen erfordert den höchsten kognitiven Aufwand. Natürliche mobile Interaktion wird oft bei der Interaktion mit intelligenten Umgebungen verwendet und wird in den folgenden Kapiteln immer wieder erwähnt.

## 4.3 Interaktion mit mobilen 3D-Benutzeroberflächen

Shirehjini et al. [12] haben eine 3D-basierte Benutzeroberfläche auf mobilen Geräten für die Interaktion mit intelligenten Umgebungen entwickelt. Für das Interaktionsmodell wurde als Metapher der Raum verwendet. Die 3D-basierte Benutzeroberfläche visualisiert also die Umgebung mit allen ihren Objekten und erzeugt somit eine direkte Korrespondenz zwischen den physikalischen Objekten und ihren 3D-Repräsentationen auf der Benutzeroberfläche. Diese

ermöglicht es den Benutzern, auf ein Gerät zuzugreifen, als wäre es in der realen Welt. Es kann also direkt innerhalb der Szene manipuliert werden. Benötigt werden also nur die Position, Orientierung und Form des intelligenten Objekts, aber keine technischen Details wie zum Beispiel seine IP-Adresse.

Durch die 3D-Visualisierung wird dem Benutzer ein Systembild für die intelligente Umgebung gezeigt, welches das Verständnis für die Umgebung und ihre Vorgänge erleichtert. Mit der mobilen Oberfläche kann der Benutzer die Menge und die Art und Weise, in welcher die Umgebung automatisiert wird, bestimmen. Davon profitieren sowohl implizite als auch explizite Interaktionen.



■ **Abbildung 2** Benutzeroberfläche auf dem mobilen Gerät [12]

Die Benutzeroberfläche wird in Abbildung 2 gezeigt und enthält eine 3D-Visualisierung des Raums, eine Dokumentenansicht und eine Aktionsansicht. Die Dokumentenansicht dient hier als einziger Zugriffspunkt für die persönlichen Medien des Benutzers. Eine mögliche Aktion des Benutzers wäre das Ziehen eines Dokuments auf ein intelligentes Objekt, was das Dokument auf das Objekt übertragen würde. Weitere mögliche Aktionen sind in der Aktionsansicht dargestellt.

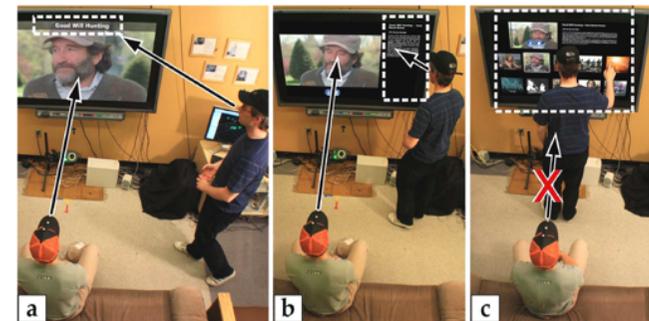
Die Evaluation hat gezeigt, dass das System sowohl für Einsteiger als auch für Fortgeschrittene einfach zu verwenden ist. Dies liegt vor allem daran, dass es Funktionen anbietet, um Aufgaben effizient umzusetzen, verständliche Metaphern verwendet, wenig Lernzeit benötigt und so gestaltet ist, dass Benutzer sich an einmalig gelernte Dinge gut wieder erinnern können.

#### 4.4 Proxemische Interaktion

Proxemische Interaktion basiert darauf, dass intelligente Objekte Position, Identität, Bewegung und Orientierung von Personen und anderen intelligenten Objekten im Raum kennen und ihr Verhalten dementsprechend anpassen [1].

Ballandat et al. [1] haben eine proxemische Medienspieleranwendung mit einem interaktiven an einer Wand befestigten Bildschirm entwickelt. Hierfür wurden verschiedene Aspekte von Proxemik ausgenutzt.

Das System kann sowohl implizite als auch explizite Interaktionen ausführen. Beispielsweise wird die Medienspieleranwendung aktiviert, sobald eine Person den Raum betritt, was einer impliziten Interaktion entspricht. Explizite Interaktionen sind dadurch möglich, dass Benutzer den interaktiven Bildschirm anfassen können, um Eingaben zu tätigen. Über Distanz und Orientierung hinweg können explizite Interaktionen auch mit einem Mobiltelefon ausgelöst werden. Interaktionen können aber auch anhand kontinuierlicher Bewegungen oder durch Bewegungen von Personen und Objekten in und aus diskreten proxemischen Regionen heraus ausgelöst werden. So wird auf dem Bildschirm beispielsweise immer mehr Inhalt angezeigt, je näher der Benutzer dem Bildschirm kommt.



■ **Abbildung 3** Vermitteln zwischen verschiedenen Personen: a) Eintretende Person sieht Basisinformationen wie den Filmtitel; b) Kommt diese Person näher, so sieht sie eine Inhaltszusammenfassung des Films; c) Steht diese Person vor dem Bildschirm, so erhält sie die volle Kontrolle [1]

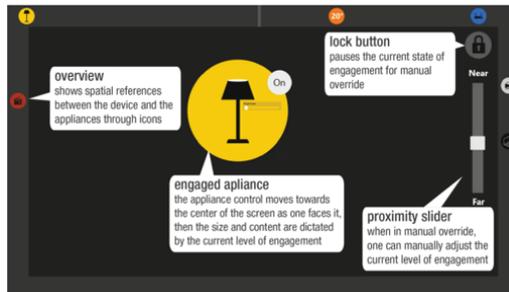
Auch gleichzeitige Interaktionen mehrerer Personen können durch Proxemik verarbeitet werden. Schaut sich eine Person ein Video mit dem interaktiven Bildschirm an, während eine andere Person den Raum betritt, so wird automatisch der Filmtitel angezeigt. Dadurch weiß die zweite Person sofort, welcher Film gespielt wird (siehe Abbildung 3a). Kommt die zweite Person dem Bildschirm näher, wird zusätzlich eine Filmbeschreibung angezeigt (siehe Abbildung 3b). Sobald diese Person direkt vor dem Bildschirm steht, wird ihr die komplette Kontrolle über den Bildschirm übertragen (siehe Abbildung 3c).

Durch Proxemik kann auch die direkte Aufmerksamkeit von Personen gegenüber anderen Personen oder Geräten interpretiert werden. So wird ein Video automatisch pausiert, wenn die Person, die das Video ansieht, beginnt

ein Buch zu lesen, zu telefonieren oder sich mit einer anderen Person zu unterhalten.

Für die Medienspieleranwendung wurde keine Nutzerstudie durchgeführt, dafür aber Anwendungsszenarien vorgestellt, die die Relevanz von proxemischen Interaktionen verdeutlichen.

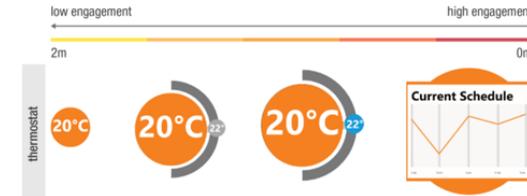
Ledo et al. [4] haben Proxemik für das Design von Fernbedienungen für ubiquitäre Computersysteme verwendet. Das Ziel war es Proxemik auszunutzen, um in einer intelligenten Umgebung kontrollierbare Anwendungen zu finden, eine bestimmte Anwendung auszuwählen, den Status der Anwendung zu betrachten und diese Anwendung zu kontrollieren. Dabei wurde die räumliche Beziehung zwischen der von einer Person in der Hand gehaltenen Fernbedienung und allen umliegenden Anwendungen ausgenutzt, um eine dynamische Steuerungsschnittstelle zu erstellen.



■ **Abbildung 4** Benutzeroberfläche der digitalen Fernbedienung von Ledo et al. auf dem Tablet [4]

Ledo et al. verwendeten als digitale Fernbedienung ein Tablet. Die Benutzeroberfläche, sichtbar in Abbildung 4, zeigt eine Übersicht der verfügbaren Anwendungen im Raum. Die Anwendungen sind am Bildschirmrand entsprechend ihrer räumlichen Position dargestellt. Da die Platzierung der Anwendungen einen Bezug zu den physikalischen Objekten im Raum haben, ist die Anwendungsfundung und -auswahl einfach. Abhängig von der Richtung, in welcher das Tablet gehalten wird, wird eine bestimmte Anwendung selektiert und in der Bildschirmmitte angezeigt. Wendet man das Tablet einer anderen Anwendung zu, wird diese selektiert.

Der Detaillierungsgrad der ausgewählten Anwendung passt sich abhängig von der Distanz zum intelligenten Objekt an. Dies wird in Abbildung 5 gezeigt. Je kleiner die Entfernung zwischen dem Tablet und dem intelligenten Objekt ist, desto mehr Informationen und Optionen werden dem Benutzer angezeigt.



■ **Abbildung 5** Verhalten der Thermostat-Anwendung auf dem Tablet, abhängig von der Distanz [4]

Dadurch sind wesentliche Status-Elemente der Anwendungen verfügbar, aber auch komplexe Operationen der Anwendungen können durchgeführt werden. Eine Sperrtaste ermöglicht es, die räumliche Interaktion zu pausieren und manuell zu überschreiben. Dadurch bleibt eine bestimmte Anwendung ausgewählt, auch wenn der Benutzer seine räumliche Position ändert. Mit einem Proximitätsschieberegler kann der angezeigte Detaillierungsgrad einer Anwendung im Sperrzustand angepasst werden, ohne dass die räumliche Position verändert werden muss.

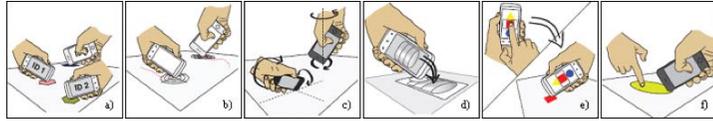
Die proxemikbewusste Fernbedienung baut auf den vorhandenen Interaktionsparadigmen wie Zeigen, Anfassen oder Scannen (siehe Kapitel 4.2) auf. Zeigen entspricht dabei einer Orientierung des Tablets im Raum, Anfassen entspricht einem proxemischen Abstand von null und Scannen wird verwendet, um die Übersicht der verfügbaren Anwendungen am Bildschirmrand zu erstellen.

Auch Ledo et al. führten keine Nutzerstudie durch, sondern stellen verschiedene Anwendungsszenarien vor und bieten dadurch eine generalisierte konzeptionelle Struktur für die Anwendung proxemikbewusster Bedienungselemente.

#### 4.5 Geräteübergreifende Interaktion

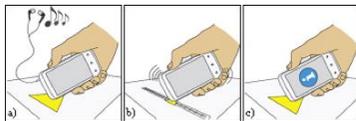
Als geräteübergreifende Interaktion wird eine Interaktion verstanden, welche mindestens zwei intelligente Objekte einbezieht. Schmidt et al. [10] beschäftigten sich mit der direkten Manipulation über Gerätegrenzen hinweg und nutzten für die Interaktion mobile Geräte und interaktive Oberflächen. Dabei wird das mobile Gerät für die Eingabe auf der interaktiven Oberfläche in stiftartiger Weise verwendet.

Bei Berührungseignissen können verschiedene Daten des mobilen Geräts mit übertragen werden. Abbildung 6 zeigt die verschiedenen Eingabeattribute. Da mobile Geräte einen Identifikator haben, kann zwischen verschiedenen Benutzern unterschieden werden (siehe Abbildung 6a). Auch der Ort der



■ **Abbildung 6** Eingabeattribute die mit einem Berührungseignis verbunden werden können: a) Identifikator, b) Ort, c) Orientierung, d) Datenkontext, e) Auswahl, f) Mehrfach-Berührung [10]

Berührung wird erkannt (siehe Abbildung 6b). Die Orientierung des mobilen Geräts bestimmt, welcher Teil des Geräts die interaktive Oberfläche berührt. So kann für jede Ecke des mobilen Geräts eine andere Funktion definiert werden. Auch für die kontinuierliche Rotation um die Achse des mobilen Geräts kann eine Funktion definiert werden (siehe Abbildung 6c). Auf mobilen Geräten wird eine große Menge persönlicher Daten gespeichert. Dieser Datenkontext kann bei einer Berührungsaktion mit der interaktiven Oberfläche übertragen werden (siehe Abbildung 6d). Benutzer können Berührungseignisse parametrisieren, indem sie auf dem mobilen Gerät bestimmte Optionen auswählen. Diese Auswahl wird dann auf die interaktive Oberfläche übertragen (siehe Abbildung 6e). Da zwischen den Berührungseignissen eines Fingers und eines mobilen Geräts unterschieden werden kann, können auch bimanuelle Interaktionen ausgeführt werden (siehe Abbildung 6f).



■ **Abbildung 7** Reaktionen auf ein Berührungseignis: a) Audiofeedback, b) haptisches Feedback, c) visuelles Feedback [10]

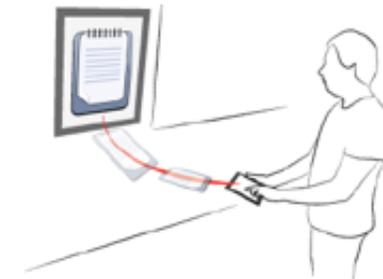
Auf die Berührungseignisse kann die interaktive Oberfläche durch Audiofeedback (siehe Abbildung 7a), haptisches Feedback (siehe Abbildung 7b) oder visuelles Feedback (siehe Abbildung 7c) reagieren.

Aus den genannten Eingabe- und Ausgabewerten ergeben sich einige Anwendungsfälle für geräteübergreifende Interaktionen. Zum einen können Daten vom Telefon auf die interaktive Oberfläche übertragen werden, damit diese mit mehreren Personen geteilt werden können. Auch kann das mobile Gerät zur Personalisierung einer von mehreren Personen genutzten interaktiven Oberfläche verwendet werden. Eine zusammengesetzte Benutzeroberfläche kann entstehen, indem Kommandomenüs und Toolpaletten auf dem mobilen Gerät angezeigt werden und somit die interaktive Oberfläche komplett und ohne Ver-

deckungen sichtbar ist. Auch zur Authentifizierung kann eine Berührung mit einem mobilen Gerät verwendet werden. Auf geteilten interaktiven Oberflächen können Benutzer individuelles Feedback erhalten. Die Eingabeausdruckskraft dieses Interaktionsstils erzeugt mehr Möglichkeiten im Vergleich zur direkten Interaktion mit einer interaktiven Oberfläche mit dem Finger.

Die Vorteile dieses Interaktionsstils sind, dass die Interaktion natürlich ist, keine extra Hardware erforderlich ist und die Eingabe mithilfe eines mobilen Geräts im Vergleich zur Eingabe mit einem Finger überlegen ist, da zusätzliche Eigenschaften ausgenutzt werden können. Ein möglicher Nachteil ist, dass Nutzer sich Sorgen darüber machen könnten, das mobile Gerät oder die interaktive Oberfläche bei der Berührung zu beschädigen. Von den Testnutzern wurde dies jedoch nicht als kritisch empfunden.

#### 4.6 Blickunterstützte Interaktion



■ **Abbildung 8** Benutzer überträgt Daten von einem entfernten Bildschirm auf sein persönliches Gerät und vice versa [13]

Unter blickunterstützter Interaktion versteht man ein Interaktionskonzept, bei dem der Blick des Benutzers überwacht wird und die Art der Interaktion mitbestimmt. Turner et al. [13] entwickelten ein Interaktionskonzept für die Übertragung von Inhalten zwischen öffentlichen Bildschirmen und persönlichen Geräten, basierend auf Blick und Berührung des Benutzers (siehe Abbildung 8). Der Blick des Benutzers und die Szene wurden mit zwei am Kopf befestigten Kameras überwacht.

Augenziehen bedeutet in diesem Kontext, Inhalte von öffentlichen auf persönliche Geräte zu bewegen und Augenschieben das Gegenteil, nämlich Inhalte von persönlichen auf öffentliche Geräte zu übertragen. Um Augenziehen und Augenschieben durchzuführen, wurden drei Interaktionstechniken entwickelt: Augenausschneiden und -einfügen (engl. Eye Cut & Paste, ECP), Augenziehen und -loslassen (engl. Eye Drag & Drop, EDD) und Augengerufen und

-verteilen (engl. Eye Summon & Cast, ESC).

Für Augenziehen mit ECP schaut der Benutzer ein Objekt auf dem öffentlichen Bildschirm an, tippt dann auf sein Tablet, um das Objekt auszuwählen und aus dem Bildschirm auszuschneiden, schaut anschließend sein Tablet an und tippt noch einmal, um das Objekt auf dem Tablet einzufügen. Für Augenschieben funktioniert ECP auf die gleiche Weise, nur dass der Benutzer die Interaktion auf dem persönlichen Gerät startet.

EDD für Augenziehen funktioniert so, dass der Benutzer wieder zuerst das Objekt auf dem öffentlichen Bildschirm anschaut. Statt dann auf das Tablet zu tippen, wird das Berührungseignis auf dem Tablet gehalten, um das angeschaute Objekt auszuwählen. Solange der Benutzer den Finger auf dem Tablet hält, kann das ausgewählte Objekt sichtbar blickgesteuert hin- und herbewegt werden. Um das Objekt auf dem Tablet einzufügen, muss der Benutzer eine Stelle auf dem Tablet anschauen und danach die Berührung loslassen. Analog funktioniert EDD auch für Augenschieben. Hierbei startet die Interaktion auf dem Tablet.

Bei ESC sieht der Benutzer für das Herbeirufen zuerst ein Objekt auf dem öffentlichen Bildschirm an. Danach wischt er auf dem Tablet nach unten, um das Objekt an der Stelle auf dem Tablet einzufügen, auf der die Wischbewegung durchgeführt wurde. Für das Verteilen schaut der Benutzer die Zielposition auf dem öffentlichen Bildschirm an und wischt auf dem Tablet auf dem Objekt nach oben, was das Objekt auf dem öffentlichen Bildschirm einfügt.

Die Benutzerstudie hat gezeigt, dass ECP und EDD ähnlich gut funktionieren, EDD jedoch bevorzugt wird. ESC ist die schnellste Interaktionstechnik, wird von den Testnutzern aber abgelehnt, da sie eine komplexere Hand-Augen-Koordination erfordert.

#### 4.7 Gestenbasierte Interaktion über am Kopf befestigte Bildschirme

Eine weitere Interaktionstechnik ist die gestenbasierte Interaktion mit einem optischen am Kopf befestigten Bildschirm (engl. Optical Head Mounted Display, OHMD) und einer am Kopf befestigten Kamera. OHMDs sind transparente Brillen, auf welche computergenerierte Bilder projiziert werden. Dadurch können die intelligenten Objekte in der urbanen Umgebung mithilfe von virtuellen Objekten, welche die physikalischen intelligenten Objekte ergänzen, manipuliert werden. Benutzer können die virtuellen Objekte und deren zugehörige Informationen sehen, indem sie die realen Objekte direkt anschauen. Da die Hände des Benutzers frei sind, können Interaktionen mit den virtuellen Objekten über Gesten in der Luft erfolgen. Das Überwachen der Hände erfolgt mit einer am Kopf befestigten Kamera. [6]

Kollee et al. [3] untersuchten gestenbasierte Interaktionen mit intelligenten Umgebungen in Bezug auf egozentrisches Tracking. Egozentrisches Tracking ermöglicht es Nutzern, mit jeder Art von verbundenen Geräten räumliche Interaktionen auszuführen, ohne dass diese Geräte weiter instrumentiert werden müssen. Die Gesten werden von der Kamera in Bezug auf den Nutzer getrackt. Gesten ergänzen hierbei die Eingabemöglichkeiten von OHMDs.

In [3] war der Anwendungsfall ein intelligentes Büro mit verschiedenen Bildschirmen, an welchen dynamisch Bilder mithilfe des OHMDs nebeneinander angezeigt werden sollten. Nötige Eingabeschritte hierfür waren das Auswählen des Bildschirms, welcher das gesuchte Bild enthält, das Auswählen dieses Bildes und das Auswählen des Bildschirms, zu welchem das Bild transferiert werden soll.



■ **Abbildung 9** Kopfgesten [3]



■ **Abbildung 10** Hand-Tracking [3]



■ **Abbildung 11** Handgesten [3]

Die gestenbasierten Interaktionen waren Kopfgesten, Hand-Tracking und Handgesten. Für Kopfgesten schaut der Benutzer den Zielbildschirm oder das Zielbild an und nickt, wie dargestellt in Abbildung 9, mit dem Kopf, um diese Aktion zu bestätigen. Hand-Tracking bedeutet, dass die Hände im Verhältnis zum OHMD getrackt werden und die Auswahl über eine Drückbewegung im

Raum erfolgt (siehe Abbildung 10). Entsprechende Handgesten werden in Abbildung 11 gezeigt. Im Gegensatz zum Hand-Tracking wird hier die Stellung der Hände überwacht und die Auswahl erfolgt über eine Greifgeste.

In zwei Studien verglichen Kollee et al. die genannten gestenbasierten Eingabetechniken mit der herkömmlichen Eingabe mithilfe von Anfassen oder mithilfe von Sprache. Es zeigte sich, dass Benutzer Kopfgesten weniger intuitiv, komfortabel und effektiv fanden als die anderen Eingabetechniken. Außerdem werden Hand-Tracking und Handgesten gegenüber den nicht gestenbasierten Eingabetechniken bevorzugt. Hinsichtlich Ausführungszeiten und Fehlerraten lieferten Hand-Tracking, Handgesten und Eingaben mithilfe von Anfassen ähnliche Ergebnisse. Lediglich Spracheingaben funktionierten schlechter als die anderen Techniken. In der Studie war die Listenansicht für die Eingabe durch Anfassen sehr effektiv gestaltet und enthielt nur wenige verfügbare Bildschirme in der Umgebung. Für längere Listen kann es sein, dass die gestenbasierte Interaktion doch überlegen wäre. Ein großer Vorteil von egozentrischem Tracking ist die Möglichkeit, für große Bildschirme die Interaktionen von vielen Benutzern gleichzeitig verarbeiten zu können, da jeder Benutzer mit seinem eigenen Messgerät ausgestattet sein kann.

Park et al. [6] beschäftigten sich mit der Verwendung eines OHMDs, um Regeln zwischen verschiedenen intelligenten Objekten in der Umgebung anzusehen und zu definieren. Beispielsweise könnte ein Benutzer eine Regel definieren wollen, die besagt, dass sich die Intensität der Deckenbeleuchtung automatisch abhängig von einem Wert, welcher von einem Umgebungslichtsensor gelesen wird, ändert.

Benutzer können Regeln erstellen, entfernen und abbilden. Für das Erstellen einer Regel greifen Benutzer die virtuelle Repräsentation eines realen Objekts und ziehen diese auf die virtuelle Repräsentation eines anderen realen Objekts. Zwischen diesen Objekten erscheint dann eine virtuelle Verbindungslinie. Basieren die intelligenten Objekte auf unterschiedlichen Datentypen, so erscheint ein Dialogfenster, in welchem eine Typkonvertierung definiert werden kann. Da für bestimmte Datentypen keine Konvertierung möglich ist, können nicht zwischen allen intelligenten Objekten Regeln erstellt werden. Regeln können entfernt werden, indem der Benutzer die virtuelle Repräsentation eines verbundenen intelligenten Objekts greift und in der leeren Umgebung loslässt. Um eine Regel abzubilden, muss der Benutzer die virtuelle Repräsentation eines verbundenen intelligenten Objekts greifen und auf eine andere virtuelle Repräsentation eines Objekts ziehen.

Die Teststudie hat ergeben, dass die Gestenerkennung nicht vollständig funktioniert und dass das Dialogfenster für die Typkonvertierung mit Gesten in der Luft schwer zu bedienen ist. Außerdem haben OHMDs ein hohes Gewicht

und sind sozial noch nicht akzeptiert. Jedoch hat sich herausgestellt, dass die Greif- und Ziehgesten gut funktionieren, um Verbindungen zu erstellen und den Nutzern bei der Bedienung Spaß bereiten.

## 5 Fazit und Ausblick

In dieser Forschungsarbeit wurde untersucht, was bei der Entwicklung intelligenter Umgebungen zu beachten ist, damit Nutzer sie verstehen und akzeptieren. Ein wichtiger Punkt ist die Transparenz und Verständlichkeit intelligenter Umgebungen. Nutzer sollten diese beeinflussen und ihre Aktionen und Reaktionen nachvollziehen können. Die Art und Weise, wie Interaktionen in intelligenten Umgebungen erfolgen, sollte intuitiv und einfach sein. In dieser Forschungsarbeit werden verschiedene Interaktionstechniken vorgestellt. Abhängig vom Kontext sollte eine passende ausgewählt werden. Weitere Interaktionstechniken könnten beispielsweise auf Elektroenzephalografie (EEG) oder Gehirn-Computer-Schnittstellen (engl. Brain-Computer-Interface, BCI) basieren.

Die momentan am weitesten verbreitete Interaktionstechnik ist vermutlich die sogenannte natürliche mobile Interaktion. Die Verwendung von Smartphones hat in den letzten Jahren enorm zugenommen und Smartphones sind heutzutage allgegenwärtig [7]. Jedoch haben auch die anderen aufgezeigten Interaktionstechniken positives Benutzerfeedback erhalten. Einige von ihnen werden in Zukunft, mit der weiteren Verbreitung der dafür nötigen Technologien, vermutlich auch ihren Einzug in unseren Alltag finden.

---

### Literatur

- 1 Till Ballendat, Nicolai Marquardt, and Saul Greenberg. Proxemic interaction: Designing for a proximity and orientation-aware environment. In *ACM International Conference on Interactive Tabletops and Surfaces - ITS '10*. ACM Press, 2010. doi:10.1145/1936652.1936676.
- 2 Ekaterina Gilman, Oleg Davidyuk, Xiang Su, and Jukka Rieki. Towards interactive smart spaces. *Journal of Ambient Intelligence and Smart Environments*, 5(1):5–22, 2013. doi:10.3233/AIS-120189.
- 3 Barry Kollee, Sven Kratz, and Anthony Dunnigan. Exploring gestural interaction in smart spaces using head mounted devices with ego-centric sensing. In *Proceedings of the 2nd ACM symposium on Spatial user interaction - SUI '14*. ACM Press, 2014. doi:10.1145/2659766.2659781.
- 4 David Ledo, Saul Greenberg, Nicolai Marquardt, and Sebastian Boring. Proxemic-aware controls: Designing remote controls for ubiquitous computing ecologies. In *Proceedings of the 17th International Conference on*

- Human-Computer Interaction with Mobile Devices and Services - MobileHCI '15*. ACM Press, 2015. doi:10.1145/2785830.2785871.
- 5 Florian Müller, Martin Schmitz, Markus Funk, Sebastian Günther, Niloofar Dezfuli, and Max Mühlhäuser. Personalized user-carried single button interfaces as shortcuts for interacting with smart devices. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. ACM Press, 2018. doi:10.1145/3170427.3188661.
  - 6 Heonjin Park, Kyle Koh, Yuri Choi, Han Joo Chae, Jeongin Hwang, and Jinwook Seo. Defining rules among devices in smart environment using an augmented reality headset. In *Proceedings of the Second International Conference on IoT in Urban Space - Urb-IoT '16*. ACM Press, 2016. doi:10.1145/2962735.2962746.
  - 7 Sinead Pugh. Investigating the relationship between smartphone addiction, social anxiety, self-esteem, age & gender. Bachelors final year project, Dublin Business School, 2017.
  - 8 Enrico Rukzio, Karin Leichtenstern, Vic Callaghan, Paul Holleis, Albrecht Schmidt, and Jeannette Chin. An experimental comparison of physical mobile interaction techniques: Touching, pointing and scanning. In *Proceedings of the 8th international conference on Ubiquitous Computing - UbiComp'06*, pages 87–104. Springer Berlin Heidelberg, 2006. doi:10.1007/11853565\_6.
  - 9 Jordi Salazar and Santiago Silvestre. *Internet of Things*. European Virtual Learning Platform for Electrical and Information Engineering, 2017.
  - 10 Dominik Schmidt, Julian Seifert, Enrico Rukzio, and Hans Gellersen. A cross-device interaction style for mobiles and surfaces. In *Proceedings of the Designing Interactive Systems Conference on - DIS '12*. ACM Press, 2012. doi:10.1145/2317956.2318005.
  - 11 Daniel Schreiber, Kris Luyten, Max Mühlhäuser, Oliver Brdiczka, and Melanie Hartman. Introduction to the special issue on interaction with smart objects. *ACM Transactions on Interactive Intelligent Systems*, 3(2):1–4, jul 2013. doi:10.1145/2499474.2499475.
  - 12 Ali Asghar Nazari Shirehjini and Azin Semsar. Human interaction with IoT-based smart environments. *Multimedia Tools and Applications*, 76(11):13343–13365, jul 2016. doi:10.1007/s11042-016-3697-3.
  - 13 Jayson Turner, Jason Alexander, Andreas Bulling, Dominik Schmidt, and Hans Gellersen. Eye pull, eye push: Moving objects between large screens and personal devices with gaze and touch. In *Human-Computer Interaction - INTERACT 2013*, pages 170–186. Springer Berlin Heidelberg, 2013. doi:10.1007/978-3-642-40480-1\_11.
  - 14 Mario Vega-Barbas, Ivan Pau, Juan Carlos Augusto, and Fernando Seoane. Interaction patterns for smart spaces: A confident interaction design solution for pervasive sensitive IoT services. *IEEE Access*, 6:1126–1136, 2018. doi:10.1109/access.2017.2777999.

# Entwicklung eines Prozesses für Voice User Interface Design

Sabine Christina Kaupp

Ludwig-Maximilians-Universität München, München, Deutschland  
sabine.kaupp@campus.lmu.de

---

## Zusammenfassung

Sprachsteuerung, oder Voice, wird von Experten seit längerer Zeit ein großes Marktwachstum vorausgesagt. Mittlerweile gibt es PAs, die ganzheitlich auf Sprache ausgelegt sind, sogenannte Sprache-Zuerst (Voice-First) Geräte. Für diese gibt es auch immer mehr Applikationen. Hiermit wird eine neue Form von Applikation geschaffen, die sich von der für bildschirmbasierte Geräte in einigen Punkten unterscheidet. Dadurch unterscheiden sich auch zum Teil deren Designmethoden. Dieses Paper will eine Grundlage für Designer sprachgesteuerter Nutzeroberflächen (Voice User Interfaces, VUIs) bilden, indem es nach einer kurzen historischen und technischen Einleitung einen Leitfaden abbildet, der Schritt für Schritt durch den Designprozess führt. Dieser entwickelte Leitfaden wird im Anschluss angewandt, um beispielhaft eine sprachgesteuerte Applikation zu designen. Das Paper diskutiert Einschränkungen und zeigt, wie überprüft werden kann ob Sprache für eine Applikation die geeignete Interaktionsmethode ist. Ebenso zeigt es Eigenschaften und Eigenheiten von VUIs auf, die ein(e) Designer(in) im Hinterkopf behalten sollte. Somit wird ein konkretes Praxisbeispiel geschaffen, das als Orientierung für einen VUI-Designprozesses dienen kann.

**2012 ACM Computing Classification** Human-centered computing — Natural language interfaces

**Keywords and phrases** Voice User Interfaces; VUIs; Voice User Interface Design; Personal Assistants; Sprachsteuerung; Voice Design Methoden; Voice Design Prozess

## 1 Einführung

Im Jahr 2018 erschienen viele Artikel mit Überschriften wie 'Voice-First Devices are the next big thing – Here's why' [13], 'Voice First: The Future of Interactions' [29] oder 'Sprachsteuerung im Alltag: Ohne geht es nicht mehr' [15]. Es gibt Kurse, Tutorials und Blogartikel für werdende VUI-Designer(innen) (vgl. z.B. [3], [17] oder [7]). Dieses Paper will einen Überblick schaffen über die bisher



© Sabine Christina Kaupp;

licensed under Creative Commons License CC-BY

Cite as: Sabine Christina Kaupp. Entwicklung eines Prozesses für Voice User Interface Design. 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. 7:1–7:28.

zusammengetragenen Designgrundlagen für VUIs und deren Anwendbarkeit, es entwickelt eine Anleitung für den Designprozess, zeigt Einschränkungen auf und wendet den entwickelten Prozess an einem Beispiel an. Es hat zum Ziel, eine Grundlage zu schaffen für aufstrebende VUI-Designer(innen) indem es ihnen ein mit Beispielen unterlegtes Schema an die Hand gibt. Das folgende Kapitel leitet über eine kurze Historie von Spracherkennung in das technische Hintergrundwissen und schafft somit eine Basis auf der im nächsten Kapitel die Designgrundlagen für VUIs aufbauen können. In Kapitel 4 werden diese Grundlagen an einem konkreten Beispiel angewandt, worauf eine Diskussion und ein kurzer Ausblick des Themas folgt.

## 2 Historisches und technisches Wissen zu Voice User Interface Design

### 2.1 Kurze Geschichte der Sprachsteuerung

Die ersten kommerziell verfügbaren Applikationen von Sprachsteuerung gab es in den neunziger Jahren. Diese mussten vor dem Benutzen trainiert werden und kosteten um die 700 US Dollar. 1996 kam VAL auf den Markt, eine Spracherkennungssoftware die Anrufe entgegennehmen konnte und ab diesem Zeitpunkt bei immer mehr Firmen als erster Kontakt im Kundendienst eingesetzt wurde [23]. Diese „Telefonsysteme“ sind unter dem Term *Interaktive Sprach-Antwort-Systeme* (Interactive Voice Response Systems, IVRs) die offiziellen Vorgänger von VUIs [22]. 2001 lag die Spracherkennungs-Akkuratesse bei 80 Prozent und stagnierte auf diesem Niveau für die nächsten zehn Jahre [23]. 2010 fügte Google die Funktion *Personalisierte Erkennung* zu allen sprachgesteuerten Suchen auf Android Handys hinzu, um durch die aufgezeichneten Suchanfragen ein akkurateres Sprachmodell zu erstellen. Seit dieser Zeit benutzt Google diese Daten, um Voice als Interaktionsmethode stetig zu verbessern [23]. Ein Jahr später installierte Apple 2011 erstmals seine sprachgesteuerte Assistentin *Siri* als Betaversion auf dem iPhone 4s [8]. Sechs Jahre danach, 2017, lag die Spracherkennungs-Akkuratesse bei 95 Prozent [16]. Zum Vergleich: Es wird angenommen, Menschen verstehen untereinander mit etwas weniger als 95 Prozent Akkuratesse [27]. Die Haushalte im Besitz eines Amazon Echo stiegen in zwei Jahren zwischen 2015 und 2017 von unter zwei bis auf über elf Millionen in den USA alleine [16]. Trendanalysen zeigen, dass Voice beginnt, klassisches Eintippen bei Online-Suchanfragen abzulösen [16]. Sprechen ist schneller [26] und intuitiver, außerdem bleiben die Hände frei [22]. Die meisten PAs sind bezahlbar (vgl. z.B. der Amazon Echo für ca. 90 Euro bzw. der Echo Dot für ca. 40 Euro [1]) oder sogar kostenlos (vgl. z.B. Siri auf allen iPhones seit dem 4s). Auch auf Entwicklerseite ist Voice mittlerweile ein zugängliches und bezahlbares Medium. Die Anzahl an verfügbaren Applikationen stieg im letzten

Jahr (2017) von etwa 1.000 auf ca. 30.000 [16]. Ebenso gibt es für Entwickler eine Reihe an benutzbaren, bezahlbaren APIs. Einige der meistgenutzten sind die Google Speech API, Amazon Lex und Amazon Polly, die alle im Jahr 2016 eingeführt wurden [27]. Google Speech beispielsweise war in der Beta-Version kostenlos und kostet jetzt 0,006 Dollar pro 15 Sekunden [5]. Zusammengefasst bedeutet das, Voice als Technologie ist

1. akkurat genug, um eine normale Konversation mit einem technischen Gerät zu führen und damit ein angenehmes Nutzererlebnis (User Experience, UX) zu schaffen
2. angenehm, intuitiv und schnell zu benutzen
3. auf Konsumentenseite preisgünstig genug, um massentauglich zu sein
4. auf Entwicklerseite zugänglich, preisgünstig und unter Konsumenten verbreitet genug, um dafür zu entwickeln

All diese Punkte tragen dazu bei, dass Voice Erfolg vorausgesagt wird. Daneben stehen Statistiken, die den Prognosen Recht zu geben scheinen. 2013 gaben 85 Prozent der iOS-Nutzer(innen) an, Siri nicht zu verwenden, 2017 nutzten 40-50 Prozent Sprachsteuerung jeden Tag [21]. Mittlerweile sind laut dem VoiceLabs Report 33 Millionen Voice-First Hardware-Geräte im Umlauf [14]. Prognosen schätzen Spracherkennung auf eine 600-Millionen-Dollar-Industrie im Jahr 2019 [11]. Der folgende Absatz analysiert, wie Voice als Technologie funktioniert und wo man sie am besten anwendet.

### 2.2 Technisches Hintergrundwissen

#### 2.2.1 Generelle technische Funktionalität von Spracherkennung

Die erste Spracherkennungs-Technologie, die ein kommerzieller Erfolg wurde, war ein Kinderspielzeug 1911. Es war ein Hund namens Rex, der bei Erwähnung seines Namens aus seiner kleinen Hundehütte geschossen kam, ein Resultat geschickt eingestellter Magnete, Hebel und der Resonanz des Wortes 'Rex' [6]. Mittlerweile wird Voice allerdings nicht mehr nur benutzt, um eine Reaktion zu erzeugen, sondern um eine Interaktion zu ermöglichen ([6], [19]). Wenn man ein VUI designt arbeitet man an einer Konversation zwischen Mensch und Maschine, in der die Maschine sinnvoll aufnehmen, verarbeiten und verstehen muss, was der Mensch sagt. Daraufhin muss eine sinnvolle Antwort produziert und ultimativ eine Tätigkeit ausgeführt werden. Dafür braucht das Gerät zwei Schlüsselkomponenten: Automatische Spracherkennung (Automatic Speech Recognition, ASR) und Proaktivität. Eine proaktive Maschine wird als Agent bezeichnet (im Kontrast zum Werkzeug, das ohne eigenständige Aktion von Nutzer(inne)n manipuliert wird, um ein Ergebnis zu erzeugen) [19]. Um zu

verstehen was bei ASR im Hintergrund geschieht, lehnt sich dieses Paper an die Systematik von Cohen et al. [6] an (vgl. Abbildung 1).



■ **Abbildung 1** Systematik einer ASR nach Cohen [6]

Bei Aktivierung der Spracherkennung wird eine festgelegte Aktionskette in Gang gesetzt. Hier wird kurz auf jeden Schritt eingegangen: *Endpunktfindung* bedeutet, die Maschine definiert, wann ein an sie gerichteter Satz beginnt und wann er endet. Bei den gerade populären PAs gibt es dazu Aktivierungsworte, sogenannte *Wakewords*, z.B. *Alexa* für den Amazon Echo (Dot), oder *OK Google* für Google Now. Um das Ende eines Satzes zu erkennen wird normalerweise darauf gewartet, dass der/die Nutzer(in) eine bestimmte Dauer (z.B. fünf Sekunden) schweigt. Alles, was ein(e) Nutzer(in) sagt, wird als *Äußerung* (Utterance) bezeichnet. *Eigenschaftszерlegung* bricht diese Äußerungen in eine Folge von Eigenschaftsvektoren runter. Eigenschaftsvektoren repräsentieren die messbaren Eigenschaften einer Äußerung, beispielsweise die Energielevel bestimmter Frequenzen. Normalerweise wird eine Äußerung in mehrere Eigenschaftsvektoren aufgespalten. *Erkennung* benutzt diese Vektoren, um zu bestimmen, was wirklich gesagt wurde. Der PA hat eine vordefinierte Bibliothek an Worten in bestimmten Sprachen, Dialekten und möglichen Betonungen, aus der er das am besten passende Match herausuchen kann [6]. *Natürliches Sprachverständnis* (Natural Language Understanding, NLU) schließlich versteht die von der Erkennung weitergegebenen Wortstränge mit Bedeutung. Hierfür gibt es mehrere Möglichkeiten. Ein häufig verwendeter Ansatz ist beispielsweise, ein (zuerst) leeres Set an benötigten oder erwarteten Kategorien bereitzustellen, das die Nutzer(innen) mit ihren Antworten füllen. Auch hier wird oft mit einer Bibliothek gearbeitet, die in diesem Fall die Richtungen beinhaltet, in die ein Gespräch gehen kann. Ruft ein(e) Nutzer(in) beispielsweise bei einer Airline an, könnte diese sein: Flug buchen, Flugstatus abfragen, bestehende Buchung verändern (Sitzplatz reservieren/ändern, Gepäck dazu buchen), Flug stornieren, etc.. Sagt der/die Nutzer(in) 'Ich möchte nach Berlin.', kann das System durch 'Ich möchte nach...' darauf schließen, dass er/sie aus den oben aufgelisteten Kategorien am wahrscheinlichsten einen Flug buchen will. Um einen Flug zu buchen braucht es eine Handvoll vordefinierter Informationen, wie zum Beispiel den Zielort. Der wird nun auf 'Berlin' gesetzt ([6], [19]). Nach Ende der Äußerung übernimmt der *Dialogmanager* die weiteren Schritte. Im vorangegangenen Beispiel wäre das wahrscheinlich nach zusätzlichen Informationen wie Datum, Abflughafen und Uhrzeit zu fragen, um die noch benötigten Informationslücken zu füllen. Grob arbeitet der Dialogmanager in zwei Ka-

tegorien. Die erste ist Aufgabenausführung, was bedeutet, die identifizierte Aufgabe wird weitergegeben an die entsprechenden Stellen im System. Die zweite ist Antwortgenerierung. Diese besteht entweder aus einer Nachfrage, wenn Äußerungen nicht verstanden wurden oder Informationen fehlen, einer Bestätigung, wenn eine Aufgabe ausgeführt wurde oder einem Vorschlag, wenn das System Äußerungen nach mehrmaligem Nachfragen nicht erkennt und dem/der Nutzer(in) als Lösung verschiedene Möglichkeiten der Interaktion vorschlägt. Zusammenfassend sind hier nochmals die einzelnen Schritte mit den im Hintergrund ablaufenden Aktionen definiert [6] (vgl. Abbildung 2).



■ **Abbildung 2** Ergänzte Systematik einer ASR nach Cohen [6]

Nachdem die generelle Funktionalität beschrieben ist, beleuchtet der nächste Absatz drei Eigenschaften, die in das Repertoire von VUIs eingebaut werden sollten. Ebenso werden hier technische Schwierigkeiten erläutert, die man im Hinterkopf haben muss.

## 2.2.2 Wichtige technische Eigenschaften von Spracherkennung

Es gibt bestimmte Elemente, die eingebaut dazu führen, dass ein ASR angenehmer zu benutzen ist.

**Einmischung** Einmischung (Barge-In) beschreibt das Designen für die Möglichkeit, dass der/die Nutzer(in) das Gerät unterbricht, während es spricht und ist Standard im VUI Design. Die meisten PAs kann man nur mit ihrem Wakeword unterbrechen. Manchmal macht es aber auch Sinn, *Hotwords* einzubauen, die den gleichen Effekt haben, wie zum Beispiel *Next* um beim Vorlesen langer Listen zum nächsten Item zu wechseln ([6], [22]). Um eine reibungslose Einmischung zu garantieren, müssen drei Dinge sichergestellt sein: Erstens sollte ein sofortiger Abbruch der Wiedergabe erfolgen, sobald das Gerät den/die Nutzer(in) wieder sprechen hört. Zweitens sollte die Unterbrechung in Kontext gesetzt werden mit dem gerade erfolgten Gespräch. Drittens sollte bei komplizierten Interaktionen alles vorherig Gesagte gespeichert bleiben. Sagt der /die Nutzer(in): 'Ich will nach Berlin' sollte Berlin als Zielflughafen gespeichert bleiben, während das Gerät nach weiteren Informationen fragt.

**N-Beste Listen und Vertrauensmaße** Sobald zu einer Spracherkennungssoftware gesprochen wird, versucht sie zu interpretieren was gesagt wurde. Dafür verwenden die meisten Geräte eine Strategie, die eine Kombination aus sogenannten N-Besten Listen (N-Best Lists) und Vertrauensmaßen (Confidence Measures) ist. Sagt man zum Beispiel im Gespräch „Vier Meilen“ könnte eine N-Beste Liste so aussehen:

1. Firm eilen
2. Vier Meilen
3. Vieh Meilen

Da das Gerät aus dem Kontext wissen sollte, dass eine Längenangabe gesucht wird, ignoriert es den ersten Treffer und arbeitet mit dem zweiten weiter. Dann, bevor eine Reaktion für den/die Nutzer(in) generiert wird, überprüft es, wie sicher es sich ist, etwas Bestimmtes gehört zu haben (Vertrauensmaß). Unterschreitet die Sicherheit eine bestimmte Schwelle, beispielsweise 85 Prozent, fragt das Gerät nach einer Bestätigung ('Du sagtest Vier Meilen, habe ich das richtig verstanden?'). Unterschreitet die Sicherheit eine zweite Schwelle, beispielsweise 50 Prozent, entsteht ein Fehler. Hier kann eine Fehler-Nachricht programmiert werden ('Das habe ich leider nicht verstanden.'). Die meisten PAs ignorieren dann mittlerweile jedoch, dass überhaupt etwas gesagt wurde. Die Logik dahinter ist, dass die meisten Nutzer(innen) auf das Schweigen von Geräten genauso reagieren wie auf das Schweigen von Mitmenschen: sie wiederholen sich [22] und geben damit dem Gerät eine zweite Chance. Zuletzt ist noch wichtig, eine N-Beste Liste anzupassen. Ist der erste Treffer nicht das richtige, muss er von der Liste gestrichen und mit dem zweiten Item weitergemacht werden, anstatt im Zweifel jedes Mal bei der Antwort auf die Nachfrage des PAs eine neue Liste zu erstellen.

### 2.2.3 Technische Schwierigkeiten von Spracherkennung

Ebenso wie wichtige Eigenschaften gibt es auch einige Schwierigkeiten, mit denen man rechnen muss, wenn man für Sprache designt.

**Lärm** Die meisten PAs sind in der Lage, die Geräusche, die sie selbst machen, auszublenden (z.B. der/die Nutzer(in) fordert den Echo auf, die Musik, die er abspielt, lauter zu machen). Jedes andere Geräusch ist allerdings ein Problem. In sogenannten nicht-sterilen Umgebungen geht die Verständnis-Akkuratesse von PAs von 95 Prozent schnell auf 75-80 Prozent zurück [27].

**Mehrere Sprecher** Das ist vor allem in öffentlichen, aber eher ruhigen Räumen wie Büros ein Problem. Haben zum Beispiel mehrere Nutzer(innen) die gleiche Sprachsteuerung auf ihren Mobiltelefonen, hören immer alle Geräte zu,

wenn ein(e) Nutzer(in) das Wakework sagt, obwohl er/sie nur mit seinem/ihrer eigenen Gerät sprechen wollte. Bei Befehlen wie: 'OK Google, setze eine Erinnerung für ein Meeting mit Julia um 15 Uhr' kann das dazu führen, dass alle Geräte den Befehl ausführen und somit auch viele Kollegen unnötig an das Meeting mit Julia erinnert werden [22].

**Kinder** Kinder sind von PAs schwerer zu verstehen als Erwachsene. Sie haben höhere Stimmen, für die es momentan noch wenig Vergleichsmaterial gibt. Außerdem sprechen vor allem junge Kinder anders als Erwachsene, wiederholen sich, schildern Ereignisse nicht chronologisch oder machen lange Pausen. Wenn PAs nicht wissen, dass sie mit einem Kind sprechen, gibt es meist Verständnisschwierigkeiten [22].

**Namen, Buchstabierung, Alphanumerik und Listen** Selbst Menschen fällt es manchmal schwer, Namen zu verstehen, da es für diese viele Schreibweisen und keinen Kontext gibt. Befiehlt ein(e) Nutzer(in) also, in seinen/ihren Terminen den mit 'Mayer' zu finden, bekommt er/sie höchstwahrscheinlich keine Treffer, da das System nach 'Meier', 'Maier', 'Meyer' oder einer anderen Schreibweise sucht. Auch Nachfragen helfen hier wenig, da sie die Schreibweise selten korrigieren können. Für solche Fälle ist ein Interface, das nur durch Sprache bedienbar ist und keinen visuellen Output liefern kann eher ungeeignet [22]. Das ist auch der Fall, wenn das Ergebnis einer Frage eine lange Liste ist (z.B. 'Alexa, wer sind die zehn reichsten Menschen der Welt') oder ein Video (z.B. 'Hey Siri, wie schäle ich einen Granatapfel?'). Ebenso sprechen PAs Zahlen immer automatisch als ganze Zahl aus. 112 gibt also hundertzweölf und nicht eins, eins, zwei. Will man Zahlen als einzelne Ziffern gesprochen haben muss das einprogrammiert werden.

**Privatsphäre** Damit Nutzer(innen) jederzeit mit PAs sprechen können, müssen diese auch jederzeit passiv zuhören. Das bedeutet, dass sie viele Stunden Gespräch mitbekommen, die nicht für sie bestimmt sind. Hier ist die Versuchung groß, das Material zu Trainingszwecken zu verwenden oder zu speichern. Beides ist strafbar, muss also unbedingt vermieden werden [22].

Nach dem kurzen Einblick in die Geschichte und mit einem Verständnis für die technischen Funktionalität und potentiellen Einschränkungen von Sprachsteuerung beschreibt der folgende Abschnitt grundlegende Design-Guidelines, die man beachten sollte.

### 3 Designprinzipien für Sprachsteuerung

Die folgenden Methoden fokussieren sich auf das Design von Voice-First Applikationen für PAs, und nehmen an, dass die gesamte Konversation und damit auch der Output nur verbal erfolgt. Sie unterstützen klassisches Konversationsdesign, nehmen also an, dass der/die Nutzer(in) sich mit seinem/ihrer PA genauso unterhalten möchte wie mit Freunden oder Bekannten. Um die Interaktion zwischen Nutzer(in) und PA dementsprechend zu gestalten, sollten Designer einige wichtige Faktoren beachten, die zum Teil schon vor diesem Kapitel angesprochen wurden. Am natürlichsten ist es, dem PA die Fähigkeit zu geben, Gesagtes in Kontext zu setzen. Wird zum Beispiel gefragt 'Wer ist der reichste Mann der Welt?' sollte das System bei der Nachfrage: 'Und wo lebt er?' wissen, dass es sich bei 'er' um den Mann handelt, den es gerade genannt hat. Das ist aber nur möglich, wenn das Gerät eine echte Konversation erlaubt. Bei den meisten PAs hingegen muss bisher allerdings jede Frage erneut mit dem Wakeword beginnen, weil die Beantwortung der ersten Frage im Verständnis des PAs das Gespräch für beendet erklärt. Dadurch gibt es auch keinen Kontext, weil jede Frage eine neue, abgeschlossene Handlung ist. Diese Art der Interaktion, bei der Nutzer(in) und PA durch sogenannte *one-turn tasks* kommunizieren, definiert man als *Command-and-Control* Ansatz [22]. Im nächsten Absatz wird erläutert, wie man bestimmen kann, ob Voice für eine bestimmte Anwendung das geeignete Interface ist. Danach werden die Methoden und Werkzeuge erklärt, die im VUI-Design eingesetzt werden und die Systematik hinter der Fehler-Handhabung dargestellt.

#### 3.1 Sinnvolle Einsatzumgebungen von Sprachsteuerung

Dieser Absatz soll ein Gefühl vermitteln, wann der Einsatz von Voice als Steuerung tatsächlich sinnvoll ist. Googles Voice Design Team hat eine Liste von Eigenschaften ausgearbeitet, die eine sprachbasierte Applikation zumindest zum Teil erfüllen sollte. Die Eigenschaften lassen sich grob um vier Vorteile von Konversation gruppieren, die auch oben schon erwähnt wurden:

1. Konversation ist intuitiv
2. Konversation spart Nutzer(inne)n mehr Zeit und Aufwand als ein bildschirmbasiertes User Interface
3. Konversation ermöglicht Multi-Tasking
4. Konversation ermöglicht freies Sprechen

Abbildung 3 nach Google [20] stellt dar, wie sie die oberen vier Punkte in Eigenschaften einer Applikation übersetzt haben. Je mehr der Kästchen angekreuzt werden können, desto besser ist der Fit für eine Sprachsteuerung. Das impliziert aber ebenso, dass es Situationen gibt, in denen Sprache nicht

das richtige Medium ist, um eine Applikation zu steuern. Sensible oder private Daten sind damit zum Beispiel schwieriger zu verwalten, da die Nutzer(innen) sich unwohl fühlen, bestimmte Informationen laut auszusprechen. Auch Aufgaben wie Essen bestellen können über Voice kompliziert sein, wenn man nicht genau weiß, was man will. Ein PA wie der Amazon Echo hat keinen Bildschirm,

Kreuzen Sie jede Aussage an, die auf Ihre Funktion zutrifft

<input type="checkbox"/> Nutzer haben schon Gespräche von Mensch zu Mensch über dieses Thema	<input type="checkbox"/> Nutzer müssen viel klicken, um Aufgaben auf einem Bildschirm auszuführen
<input type="checkbox"/> Die Interaktion ist kurz, mit möglichst wenig „Hin-und-Her“-Dialog	<input type="checkbox"/> Nutzer müssen mehrere Applikationen navigieren zur Ausführung am Bildschirm
<input type="checkbox"/> Nutzer können die Aufgabe ausführen während sie multitasken	<input type="checkbox"/> Die Funktion ist schwierig oder aufwendig zu finden
<input type="checkbox"/> Nutzer können diese Aufgabe ausführen, wenn Hände und Augen beschäftigt sind	<input type="checkbox"/> Nutzer fühlen sich wohl, über dieses Thema zu sprechen

■ **Abbildung 3** Checkliste zur Bestimmung der Eignung von Sprachsteuerung einer Applikation nach Google [20]

auf dem er Ergebnisse anzeigen kann. Wenn der/die Nutzer(in) sich also von der Menüvielfalt inspirieren lassen will, müsste er/sie sich von Alexa lange Listen an möglichen Restaurants und deren Gerichten vorlesen lassen. Das ist weder intuitiv noch schnell und zudem problematisch, da der Mensch im Durchschnitt nur etwa vier Elemente in seinem Kurzzeitgedächtnis behalten kann [24]. Sobald Alexa Nutzer(inne)n mehr als drei bis vier Elemente präsentiert, überstrapaziert sie meist deren Merkfähigkeit. Ist also das Ergebnis einer Frage an PAs eine Liste mit mehr als vier Items, sollte das Gerät den/die Nutzer(in) darauf hinweisen und ihm/ihr als Output eine priorisierte Liste von drei bis vier Items geben, mit der Option, in einem zweiten Anlauf mehr zu hören. Langwierige Aufgaben wie die Neuanmeldung eines Kontos können ebenfalls zum Problem werden, wenn das Gerät mitten in einem einstündigen Prozesses etwas nicht versteht. Unten wird genauer darauf eingegangen, wie man mit Fehlern dieser Art umgeht. Trotzdem sollte man als Designer(in) nicht riskieren, dass ein Fall auftreten kann, in dem der/die Nutzer(in) möglicherweise gezwungen ist, den Prozess nochmals von vorne an einem anderen Gerät zu starten. Auch kann beim oben genannten Beispiel das Problem auftreten, viel mit Ziffern oder Eigennamen arbeiten zu müssen. Wie schon beschrieben gibt es meist viele Schreibweisen eines Namens, und ohne verfügbaren Bildschirm ist es vielleicht nötig, den/die Nutzer(in) häufig buchstabieren zu lassen, um keine Fehler zu machen. Buchstabieren selbst ist aber wieder schwierig für PAs, da sie Buchstaben wie 'P' und 'B' ohne Kontext schwer auseinanderhalten können. Zusammengefasst bedeutet das, Sprachsteuerung ist für Applikationen ungeeignet, die

1. Private oder sensible Daten behandeln

2. Längere Listen oder Aufzählungen als Ergebnis haben können
3. Grafischen Output brauchen
4. Einen langen und möglicherweise komplizierten Prozess betreuen (z.B. Abschließen einer Versicherung)
5. Viele alphanumerische Folgen und/oder Eigennamen behandeln

Als Faustregel kann festgehalten werden, dass Voice nur dann Sinn macht, wenn dadurch das Ausführen einer Applikation leichter gestaltet werden kann, als mit einer anderen Alternative.

### 3.2 Methoden und Werkzeuge

Die grundlegenden Methoden im VUI-Design sind mehr suggerierte, aufeinander folgende Arbeitsschritte als feste Frameworks. Diese sollten angepasst werden, je nachdem was genau designt wird.

**1. Schreiben von Beispieldialogen** Beispieldialoge sollten der Anfang des Designprozesses sein. Dafür greift sich der/die Designer(in) einen sogenannten *Happy Path* aus den möglichen Gesprächssituationen zwischen PA und Nutzer(in) heraus. Ein Happy Path ist eine Schlüsselfunktion der Applikation. Ein Beispiel für eine Kino-Applikation wäre: 'Hey Google, was kommt heute im Kino?' und die darauf folgende Interaktion des Kartenkaufens für einen der Filme. Der Dialog wird wie ein Filmskript niedergeschrieben, es wird genau spezifiziert, was Nutzer(in) und PA zueinander sagen. Das schließt verschiedene Pfade eines Gesprächs, wie zum Beispiel unterschiedliche Reaktionen basierend auf einer 'Ja'- oder 'Nein'-Antwort, mit ein. Die Antworten oder Aussagen des PAs heißen *Prompts*. Nach dem Schreibprozess empfiehlt es sich, das Geschriebene laut vorzulesen, im Idealfall mit einer zweiten Person. Manchmal klingt etwas in der Konversation anders als auf dem Papier. Das Durchspielen lässt einen ebenso darauf achten, ob bestimmte Prompts zu lang für ein normales Gespräch sind [20]. Eine Alternative dazu, direkt mit dem Schreiben zu beginnen, ist ein tatsächliches Gespräch über einen Happy Path mit einem anderen Menschen. Hier würde der/die Designer(in) eine(n) Bekannte(n)/Freund(in) instruieren, was er/sie machen will (z.B. nach Kinofilmen fragen) und selbst in die Rolle des PAs schlüpfen. Das Gespräch wird geführt, aufgezeichnet, transkribiert und darauf aufbauend dann das Skript erstellt [20].

**2. Erstellen von Flow-Diagrammen** Flowdiagramme bilden Gesprächsfüsse ab. Jede Applikation sollte ein solches Flow-Diagramm zugrunde liegen haben. Es wird aus den Beispieldialogen erstellt und ist holistisch, beinhaltet also jeden möglichen Pfad, den ein Gespräch innerhalb einer Applikation zwischen Nutzer(in) und PA nehmen kann. Damit stellt es sicher, dass der/die Designer(in)

sich dediziert mit allen möglichen Gesprächsverläufen auseinandersetzt und verbildlicht auch fehlerhafte Antworten oder Situationen. Diese können entstehen, wenn Nutzer(innen) nicht weiterwissen und das Gespräch abbrechen bzw. den PA um Hilfe bitten oder wenn das Gerät etwas nicht versteht oder nicht ausführen kann. Hier muss nicht jeder Ausdruck von Verwirrung, Problemen oder Fehlern aufgelistet werden, den Nutzer(innen) sagen können, aber alle unterschiedlichen Fälle sollten einer Kategorie zugeordnet werden können ([20], [22]). Ebenso in das Design von Flows eingearbeitet werden sollten Bestätigungen seitens der PAs. Es gibt drei Möglichkeiten einer Bestätigung für Sprache: Explizit ('Ich habe verstanden, dass Du einen Wecker auf 8:00 Uhr stellen willst. Ist das korrekt?'), implizit ('Ich stelle Deinen Wecker auf 8:00 Uhr') und nonverbal (der Wecker wird gestellt, aber der/die Nutzer(in) bekommt kein verbales Feedback). Nonverbale Bestätigungen sind vor allem gebräuchlich bei Reaktionen, die der/die Nutzer(in) direkt sehen oder hören kann ('Mach das Licht in der Küche an'). Andernfalls sind implizite Bestätigungen die am häufigsten verwendete Methode. Explizite Bestätigungen erzeugen einen weiteren Arbeitsschritt auf Nutzerseite, da er/sie das Gesagte immer nochmal bestätigen muss. Sie werden verwendet, wenn ein potentieller Fehler größeren Schaden anrichten könnte (z.B. einen Flug nach Schwerin statt nach Berlin buchen) [6], [22]. Ein Orientierungsbeispiel für ein Flow-Diagramm findet sich im Anhang (Abbildungen 4 - 8). Darauf wird auch noch einmal genauer in Kapitel 4.3 eingegangen.

**3. Einbauen von dialogorientierten Elementen** Dialogorientierte Elemente sind Füller, die nicht zum Erfolg eines Gesprächs beitragen, aber die Konversation natürlicher und angenehmer machen. Das können unter anderem Zeitangaben sein (z.B. 'Die Hälfte hast Du schon!'), Anerkennungen ('Danke', 'Alles klar!', 'Entschuldige!') oder Feedback ('Gut gemacht!'). Meist wird beim Flow-Design aber erst einmal nicht auf diese Füller geachtet, sondern eine sinnvolle Konversation entworfen. Um gute und sinnvolle Platzierungen für Füller zu finden, bietet sich nochmals ein Durchspielen des Dialogs mit einer zweiten Person an. Hierbei sollte man dann gezielt darauf achten, was gesagt wird, und die verwendeten Füller niederschreiben [20], [22].

### 3.3 Fehler-Handhabung

Der angemessene Umgang mit Fehlern ist ein wichtiger Punkt im Interface Design. Ist es einem/einer Nutzer(in) nicht möglich, sein/ihr Vorhaben erfolgreich abzuschließen, wird er/sie das Gerät oder die Applikation möglicherweise nicht mehr nutzen. Bei Sprache gibt es drei Arten von Fehlern, für die man designen sollte:

**1. Keine Eingabe** Hier hat der/die Nutzer(in) entweder nichts gesagt, etwas nicht laut genug gesagt oder die Sicherheit für das Gehörte rutscht unter die zweite Schwelle des Vertrauensmaßes. Wie oben schon erwähnt, ist eine übliche Reaktion von PAs, darauf zu warten, ob sich der/die Nutzer(in) wiederholt. Ist das Gespräch allerdings eine komplexere Aktion, wie der Kauf von Kinokarten, und bricht in der Mitte ab, sollte das System nachfragen. Es gibt drei Stufen des Eingabefehlers. Bei Stufe eins geht das System davon aus, dass es etwas einfach nicht verstanden hat und stellt (idealerweise etwas umformuliert) die gleiche Frage noch einmal. Stufe zwei ist ein weiteres (umformuliertes) Wiederholen der Frage. Stufe drei schließlich ist der höfliche Abbruch der Konversation und Aktion, um den/die Nutzer(in) nicht zu belästigen ('Es scheint als wolltest Du gerade keine Kinokarten mehr. Falls Du sie später doch brauchst, stehe ich Dir gerne wieder zur Verfügung!') [20].

**2. Keine Übereinstimmung** Keine Übereinstimmung bedeutet, das System hat zwar etwas gehört, das Gesagte stimmt aber nicht mit dem überein, was es erwartet hat bzw. die Sicherheit etwas gehört zu haben rutscht unter die erste Schwelle des Vertrauensmaßes. Auch hier gibt es drei Stufen. Stufe eins ist identisch zu der beim Eingabefehler. Bei Stufe zwei schlägt das System dem/der Nutzer(in) mehrere Dinge vor, die er/sie sagen kann, um das Gespräch und damit die gewünschte Aktion erfolgreich abzuschließen. Führen Stufe eins und zwei nicht zum Ergebnis, ist Stufe drei wieder der Abbruch des Gesprächs durch das System. Im Idealfall schlägt der PA eine Möglichkeit vor, die Aktion anders abzuschließen und entschuldigt sich für sein „Versagen“ ('Entschuldigung, das scheint über meine Fähigkeiten hinaus zu gehen. Vielleicht möchtest Du stattdessen unsere Webseite besuchen unter...'). Generell sollten Nutzer(innen) weder bei Übereinstimmungs- noch bei Eingabeschwierigkeiten öfter als dreimal hintereinander eine Fehlermeldung bekommen [20].

**3. Systemfehler** Systemfehler entstehen, wenn der/die Nutzer(in) korrekt verstanden wurde, der PA aber an der Ausführung der Aktion intern oder extern gehindert wird. Intern bedeutet meist einen Fehler im Programm oder eine Lücke im Design, extern einen Fehler eines anderen Programms. Will der/die Nutzer(in) zum Beispiel ein Flugticket von Lufthansa kaufen, deren Webseite ist aber nicht erreichbar oder der gewünschte Flug ausverkauft, muss der PA auch dafür eine Antwort haben. Diese enthält im Idealfall eine Entschuldigung, eine nachvollziehbare Erklärung und hilfreiche nächste Schritte ('Entschuldigung, ich kann Lufthansa gerade nicht erreichen. Soll ich es in ein paar Minuten nochmal versuchen oder andere Fluganbieter überprüfen?') [20].

### 3.3.1 Universalien für Voice-Design

Einige Design-Grundsätze sind immer im Hinterkopf zu behalten, wenn man ein VUI entwirft.

**1. Die Nutzer(innen) sind nie schuld** Wenn ein Fehler auftritt, sollte man nicht den/die Nutzer(in) verantwortlich machen. Das ist vor allem relevant für leistungsbewertende Systeme wie Sportapplikationen oder VUIs in Autos. Merkt ein System, dass der/die Nutzer(in) zu schnell fährt, sollte es nicht sagen: 'Du fährst zu schnell', sondern elegant formulieren: 'Die Straßenverhältnisse hier sind sehr schwierig, schnelles Fahren erschwert die Bedingungen'. Nutzer bauen dadurch mehr Sympathie auf und hören verstärkt auf Empfehlungen des Systems [18].

**2. Erst- und erfahrene Nutzer(innen)** Die meisten PAs haben leicht abgewandelte Begrüßungstexte für neue, sporadische und regelmäßige Nutzer(innen). Beim ersten Mal ist der Prompt etwas länger und erklärt vielleicht Einiges, bei regelmäßigen Nutzer(inne)n ist es oft nur ein 'Willkommen zurück!'. Zwischen sporadischen und regelmäßigen Nutzer(inne)n wird meist unterschieden, indem das System prüft, wann eine Applikation zum letzten Mal benutzt wurde.

**3. Hilfe** PAs unterstützen ihre Nutzer(innen) mit vielen Dingen. Im VUI-Design sollte man sich deshalb bewusst sein, dass manche Nutzer(innen) ihre PAs fragen, was sie können und eine Aufstufung ihrer Fähigkeiten erwarten, generell oder applikationsspezifisch. Die Frage kann sehr viele Variationen haben, von 'Hilfe!' über 'Was kannst du so?' bis 'Was soll ich jetzt machen?'. Der PA sollte alle diese Varianten erkennen und passende Antworten dafür bereithalten.

**4. Wartezeit** Manchmal brauchen Nutzer(innen) Zeit, bis sie Fragen des PAs beantworten können. Will das System zum Beispiel eine Flugnummer wissen, sollte im Vorhinein berechnet worden sein wie viele Sekunden/Minuten das Gerät wartet, bis es im Zweifel nochmal nachfragt. Genauso sollte der PA Antworten wie 'Moment noch!' oder 'Gleich hab ichs!' erkennen und darauf antworten ('Kein Problem, lass Dir Zeit.'), um dem/der Nutzer(in) nicht das Gefühl von einem ungeduldig wartenden Assistenten zu geben und ihn/sie dadurch in eine Stresssituation zu bringen [22].

**5. Mehrdeutigkeit** Manchmal sind Nutzeranfragen nicht eindeutig. Hört ein PA zum Beispiel: 'Spiele Yesterday' weiß er nicht, ob es sich um das Lied oder den Film handelt und selbst wenn er statistisch gesehen auf das Lied schließt, weiß er nicht welche Version gemeint ist. Hört er 'Ruf Julia an', weiß er nicht

welche Julia, und ob er sie auf dem Handy oder dem Festnetz anrufen soll. In solchen Situationen sind Nachfragen kritisch, die die Informationen, die schon erhalten wurden, speichern und nur mit den noch benötigten Dingen auffüllen. PAs sollten nie raten. Einmal bestätigen oder verneinen ist für den/die Nutzer(in) weniger Aufwand als eine Korrektur ([20], [22]).

**6. Dauer** Ein VUI sollte so designt sein, dass Aufgaben zeiteffizient erledigt werden. Das bedeutet, ein PA sollte seine(n) Nutzer(in) zum Beispiel immer nach seiner/ihrer gesamten Adresse fragen, nie erst nach der Straße, dann der Hausnummer, dann der Stadt, usw. [22]. Der zweite Aspekt von Dauer ist die Geschwindigkeit, mit der der PA spricht. Im Idealfall ist diese durch den/die Nutzer(in) einstellbar. Der dritte Aspekt ist die Länge der Prompts. Zuhören ist für Menschen anstrengend. Wir hören sehr viel langsamer als wir lesen [25] und können, wie oben angesprochen, nur wenige Elemente auf einmal im Kopf behalten. Daher sollte der/die Designer(in) darauf achten, Prompts so kurz wie möglich zu gestalten.

**7. Persönlichkeit** Nutzer(innen) können die Sprache wählen, in der sie sich mit ihrem PA unterhalten wollen. Bald können sie aber auch die Persönlichkeit wählen, die ein PA verkörpern soll. Google will 2018 sechs neue, englische Stimmen für seinen Assistenten einführen [4]. Die jetzigen populären PAs, Amazons Alexa, Googles OK Google, Microsofts Cortana und Apples Siri, haben überwiegend weibliche Stimmen, wofür es viel Theorie gibt. Unter anderem soll es einfacher sein, eine weibliche Stimme zu finden, die alle Nutzer(innen) mögen, als eine männliche, denn das menschliche Gehirn sei darauf trainiert, weibliche Stimmen zu mögen [9]. Kritiker behaupten, es hinge damit zusammen, dass Menschen eher Frauen in leicht unterwürfigen, administrativen Rollen wie der eines PAs erwarten [10]. Eine Studie des Bremen Ambient Assisted Living Lab (BAALL) widerspricht dem jedoch. Hier bevorzugten die Befragten überwiegend eine männliche Stimme [2]. Es scheint keine klare Richtung zu geben, weswegen es Nutzer(inne)n möglich sein sollte, die Persönlichkeit ihres PAs selbst auszuwählen, bezogen auf Alter, Geschlecht, Akzent, Dialekt und möglichen anderen Variationen.

#### 4 Konkretes Designbeispiel einer Applikation für Sprachsteuerung

Nach der Theorie der letzten Kapitel widmet sich dieses dem praktischen Design einer Applikation. Ich designe für Amazons Alexa, da die dazugehörigen Geräte, Echo und Echo Dot, zusammen über zwei Drittel des Marktanteils für PAs bilden [12].

#### 4.1 Hintergrundrecherche

Die Applikationen für Amazons Alexa heißen Skills. Zuerst durchsuchte ich die Skill Bibliothek (das Pendant zum Apple oder Play Store) von Amazon Deutschland, um einen relevanten Anwendungsfall zu finden. Nach einem kleinen Brainstorming und etwas Recherche entschied ich mich für eine Skill, die Nutzer(inne)n während der Arbeit zu mehr Konzentration verhilft. Die Umsetzung erfolgt durch sehr kurze Meditation und Akupressur, beides Tätigkeiten, die gut durch Sprache zu beschreiben und unterstützen sind. Des Weiteren kann die Skill durch im Hintergrund erklingende, ruhige Musik begleitet werden, was sich wiederum gut für ein VUI eignet. Von den oben aufgeführten Punkten aus dem Google-Test zur Anwendbarkeit einer Sprachsteuerung konnten alle bis auf den ersten angekreuzt werden, wenn auch mit Einschränkungen. Nutzer(innen) können die Skill zwar im Multi-Tasking ausführen, sollten es aber idealerweise bei einer Konzentrationsübung nicht. Das gleiche gilt für Benutzung, wenn Augen oder Hände beschäftigt sind. Die Nutzung von mehreren Apps und die vielleicht schwierige Auffindbarkeit bezieht sich auf die Schleifeneigenschaft der Skill, die es Nutzer(inne)n ermöglicht, die Übung in bestimmten, frei wählbaren Zeitabschnitten zu wiederholen. Diese Wiederholungen werden sonst oft erzeugt, in dem sich Nutzer(innen) selbst einen Wecker stellen und die Applikation immer wieder aufrufen. Es wurde noch eine kurze Hintergrundrecherche durchgeführt, um geeignete, konzentrationssteigernde Übungen zu finden. Ebenso wurde das Gefundene auf Umsetzbarkeit geprüft. Bequemlichkeit (Convenience) ist ein sehr wichtiger Faktor im generellen User Interface Design, und somit auch im VUI-Design (vgl. zB [28], [30]). Es gibt beispielsweise einige Akupressurpunkte an den Füßen, aber dazu müssten die Nutzer(innen) während der Übung die Schuhe ausziehen. Die Skill soll aber überall einsetzbar sein, wo lange gearbeitet wird, was Büros oder Bibliotheken einschließt, also entschied ich mich gegen diese Punkte. Genauso darf die Skill nicht zu lange dauern. Je länger die Skill ist, desto mehr Aufwand braucht es, sie ins Arbeiten zu integrieren, da sie ja Zeit benötigt, in der man arbeiten könnte. Deswegen habe ich mich entschieden, die Skill höchstens drei Minuten dauern zu lassen. Mit diesem Wissen im Hinterkopf wurde das Gefundene durchgesehen und zwei kurze Meditationen und drei Akupressurpunkte ausgewählt.

#### 4.2 Beispiel-Dialog schreiben

Wie oben beschrieben, war der erste Schritt das Schreiben eines Beispieldialogs. Dieser Dialog beschreibt, was geschieht, wenn Nutzer(in) und Alexa entlang des Happy Paths der Konzentrationsübung interagieren, von der Öffnung der Applikation bis zum Ende der Übung. Hier entstand auch der Name der

Skill: Konzentrationsminuten (KM). Er beschreibt in einem Wort, was die Skill macht. Das ist wichtig, vor allem in Anlehnung an die oben erwähnte Bequemlichkeit. Wenn vom Titel nicht erkennbar ist, wofür es sich handelt, nimmt der/die Nutzer(in) im Zweifel lieber eine andere Applikation, anstatt sich meine Beschreibung durchzulesen. Da meine Skill wenig Nutzereingaben benötigt, sieht der Beispieldialog zunächst unkompliziert aus. Hier soll gleich auf Kapitel 3.3.1 verwiesen werden: Erst- und erfahrene Nutzer(innen). Wird die Skill häufiger benutzt, sollte der/die Nutzer(in) einen anderen Begrüßungstext bekommen als beim ersten Öffnen. Hier gibt es drei Eröffnungsversionen, eine für die erste, eine für sporadische und eine für regelmäßige Nutzung:

Nutzer(in): 'Alexa, starte Konzentrationsminuten.'

Alexa: 'Konzentrationsminuten wird gestartet.'

Applikation an Erstnutzer(in): 'Herzlich Willkommen bei Konzentrationsminuten. Schön Dich zu sehen! Unsere Übung dauert drei Minuten und wird Dich drei Punkte an Deinem Körper massieren lassen. Am Anfang hörst Du die Einsteigerversion, die diese Punkte beschreibt. Benutzt Du uns öfter, wirst Du die Kennerversion hören, wobei Du immer wechseln kannst. Wenn Du willst, kann ich Dich auch mit einer Schleife zum Beispiel alle 90 Minuten beim Konzentrieren unterstützen.'

Applikation an sporadische(n) Nutzer(in): 'Schön Dich wieder zu sehen! Du warst schon eine Zeit nicht mehr hier. Möchtest Du die Einsteiger- oder die Kennerversion unserer 3-minütigen Konzentrationspause?'

Applikation an regelmäßige(n) Nutzer(in): 'Willkommen zurück!'

Danach folgt der Monolog der Skill während der Konzentrationsübung. Dieser findet sich im Anhang (vgl. Abbildung 9). Der Monolog und der Anfangsdialog wurden dann mehrmals mit verschiedenen Personen aus meinem Freundes- und Kollegenkreis durchgesprochen, um ihn möglichst kurz, verständlich und natürlich zu gestalten.

### 4.3 Flow-Diagramm erstellen

Aus den gewonnenen Dialogen wurde ein Flow-Diagramm erstellt. Dieses enthält alle möglichen Gesprächsrichtungen, inklusive Hilfefunktion und impliziten Bestätigungen. Das Diagramm findet sich ebenfalls in der Appendix (vgl. Abbildungen 4 - 8). Es besteht aus vier Szenen, wobei jede Szene den Flow einer Hauptfunktion verwaltet. In diesem Flow-Diagramm gibt es die

1. Globale Skill, die das Starten der Skill verwaltet
2. Einsteigerversion, für Erst- oder unregelmäßige Nutzer(innen)

3. Kennerversion, für regelmäßige Nutzer(innen)

4. Hilfefunktion

Alle Funktionen sind miteinander verbunden und werden durch bestimmte Zustände oder Nutzereingaben aktiviert. Es wird stark empfohlen, sich das Diagramm genau anzusehen. Es zwingt den/die Designer(in), sich detailliert Gedanken zu machen, an welcher Stelle was passiert oder passieren könnte und wie die Skill darauf reagieren sollte. Die Gesprächspfade von oben werden alle mit definitiven Anfangs- und Endpunkten abgebildet. Durch die Beschaffenheit eines Flow-Diagramms können Aktionen nicht alleine stehen, sondern müssen immer sinnvoll eingebunden werden. Somit entsteht durch das Designen eines Flows ein rundes Bild der Skill, in das alles eingebunden ist.

### 4.4 Einbauen von dialogorientierten Elementen

Sobald der Flow steht, kann man als Designer(in) mehr ins Detail gehen und die Prompts als echte Dialogantworten konzipieren. Fragt ein(e) Nutzer(in) beispielsweise noch einmal nach, wo genau sich ein Konzentrationspunkt befindet, beginnt Alexa mit: 'Gerne!' oder 'Alles klar!'. Wie schon erwähnt, machen diese Elemente ein Gespräch nicht gehaltvoller, aber angenehmer. Um gute Füller zu finden sprach ich den Flow nochmals mit zwei Freunden durch, auch um final sicherzustellen, dass nichts gestellt oder hölzern klingt.

### 4.5 Fehler-Handhabung und Universalien

Die Fehler-Handhabung wird in diesem praktischen Designbeispiel nicht mehr behandelt. Sie stellt den nächsten Schritt nach dem Fertigstellen des Flow-Diagramms dar und involviert das Entwerfen von Reaktionen, wenn Alexa etwas nicht versteht. Von den Universalien wurden in diesem Beispiel dediziert (2), (3), (4) und (6) umgesetzt. Auf *Erst- und erfahrene Nutzer(innen)* und *Hilfe* wurde oben bereits eingegangen, eventuell benötigte *Wartezeit* ist umgesetzt, indem der/die Nutzer(in) Pausen oder Prompt-Wiederholungen verlangen kann und auf *Dauer* der Prompts wurde von Anfang an geachtet. (1), (5) und (7) sind hier nicht anwendbar, da es für *Nutzerschuld* und *Mehrdeutigkeit* keine Prompts gab, die dieses Thema hätten aufnehmen können, und das Wechseln der *Persönlichkeit* auf Alexa noch nicht vollständig möglich ist.

## 5 Diskussion

Sprachsteuerung als Interaktionsmethode für bestimmte Applikationen wirkt derzeit sehr vielversprechend. Trotzdem muss man klar sehen, dass sie nicht für alles anwendbar ist. In der hier ausgeführten Konzentrationsskill gab es Schwierigkeiten beim Designen der Prompts, wenn es darum ging, bestimmte Punkte

am Körper zu finden. Zum Vergleich: diese sind mit Hilfe eines Bildes innerhalb einer Sekunde und ohne Probleme und Zweifel lokalisiert. Trotzdem ist es angenehm, eine solche Übung rein stimmbasiert durchzuführen. Da es nichts zu sehen gibt, kann der/die Nutzer(in) die Augen entspannt geschlossen halten. Weiterhin ist es leichter und schneller zu sagen: „Alexa, starte Konzentrationsminuten“, als die Applikation auf dem Mobiltelefon zu starten. Vor allem bei der Arbeit entspricht das Entsperren des Telefons manchmal der Überschreitung einer Barriere, da dann Applikationen wie Facebook den/die Nutzer(in) in Versuchung führen könnten. Beim Sprachbefehl ist das nicht vorhanden. Sprachbasierte Applikationen sind (idealerweise) sehr nutzer(innen)freundlich, da der Designprozess selbst die Designer(innen) bis zu einem gewissen Grad zwingt, ihre Kreation immer wieder durch Andere validieren zu lassen. Das Durchspielen der Dialoge mit Außenstehenden gibt gleichzeitig ja immer die Möglichkeit für konstruktives Feedback an der Applikation. Außerdem verhindert die Beschaffenheit des Flowdiagramms selbst, dass Elemente alleine stehen und nicht sinnvoll eingebunden werden.

## 6 Einschränkungen

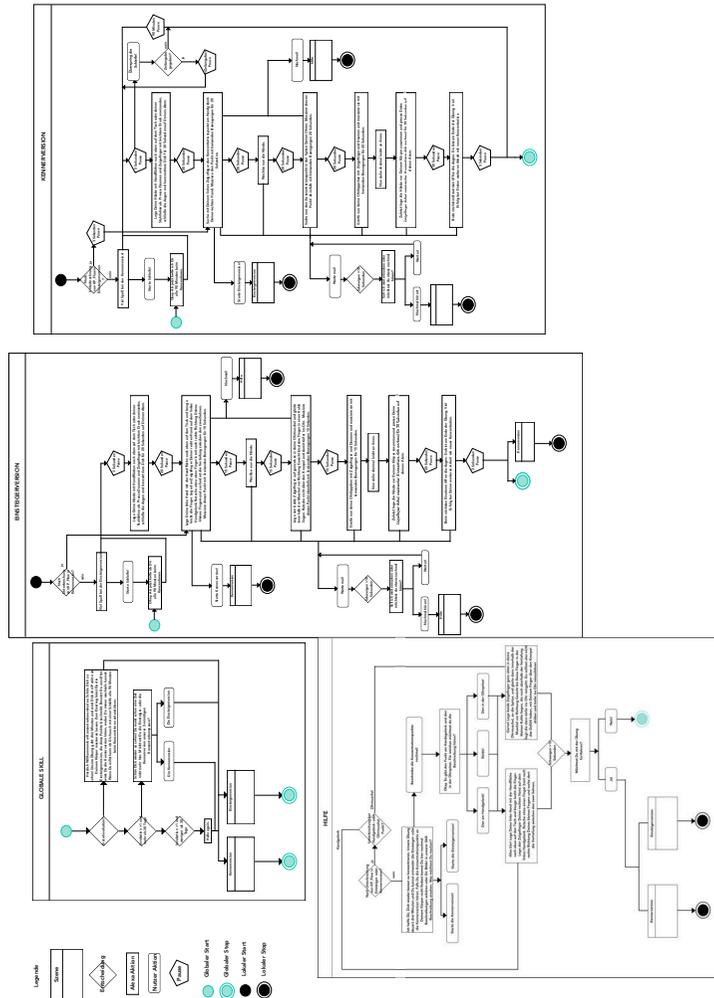
Diese Arbeit erhebt keinen Anspruch darauf, den einzig richtigen Designprozess für VUIs entworfen zu haben. Sie ist mehr wie eine Anleitung zu lesen, mit einem praktischen Anwendungsbeispiel zur Orientierung. Die Fehler-Handhabung sowie die technische Umsetzung werden nicht mehr im Detail behandelt. Ich selbst und mein Umfeld sind aus der Millennial-Generation, die Skill wurde also in dieser Umgebung entwickelt und getestet. Ist die Zielgruppe eine andere, hält das Beispiel trotzdem, muss aber angepasst werden, vor allem im Gesprächsstil und der Detailbearbeitung (siehe z.B. Abschnitt 2.2.3).

## 7 Zusammenfassung und zukünftige Arbeit

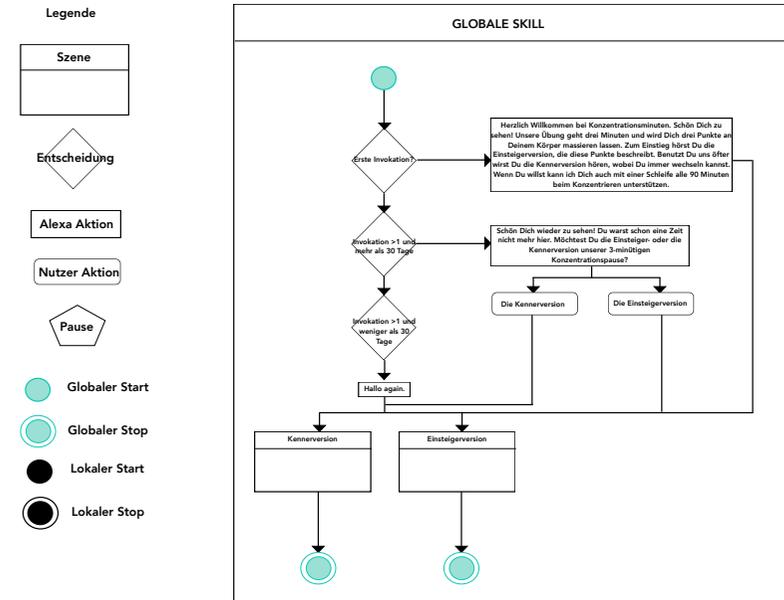
Die Arbeit gibt einen Überblick über das Designen für VUIs, mit Methoden, die man anwenden sollte und Punkten, auf die man achten muss. Sie behandelt kurz die Geschichte und technische Funktionalität von VUIs und zeigt in diesem Zug auch deren Eigenheiten und technische Schwierigkeiten auf. Sie entwirft einen Designprozess für VUIs und weist dabei auf spezifische Eigenschaften von Sprache hin, die besondere Beachtung erfordern. Sie behandelt, wann es sinnvoll ist, eine sprachbasierte Applikation zu entwerfen, erstellt eine Liste von Faktoren, auf die man achten muss und diskutiert, wie man mit Fehlern umgeht. Abschließend designt sie exemplarisch einen Beispieldialog und ein Flow-Diagramm für eine sprachgesteuerte Applikation auf Amazons Alexa und zeigt potentielle Einschränkungen auf. Ein nächster Schritt sollte auch

die Fehler-Handhabung und technische Umsetzung exemplarisch darstellen. Außerdem können weitere wissenschaftliche Arbeiten zum Thema VUI-Design die Qualität des ausgearbeiteten Prozesses verbessern und sicherstellen, dass graduell mehr und mehr Punkte zusammengetragen werden, die beim Designen und Umsetzen helfen können. Generell erwarte ich einen weiteren Anstieg der Popularität von sprachgesteuerten Systemen und damit verbunden auch eine Ausweitung der wissenschaftlichen Arbeiten zu diesem Thema. Mein Paper soll als Anfangspunkt solcher Arbeiten dienen. Es strebt an, eine Grundlage für konkretere Forschung zum Design für sprachgesteuerte Systeme zu sein.

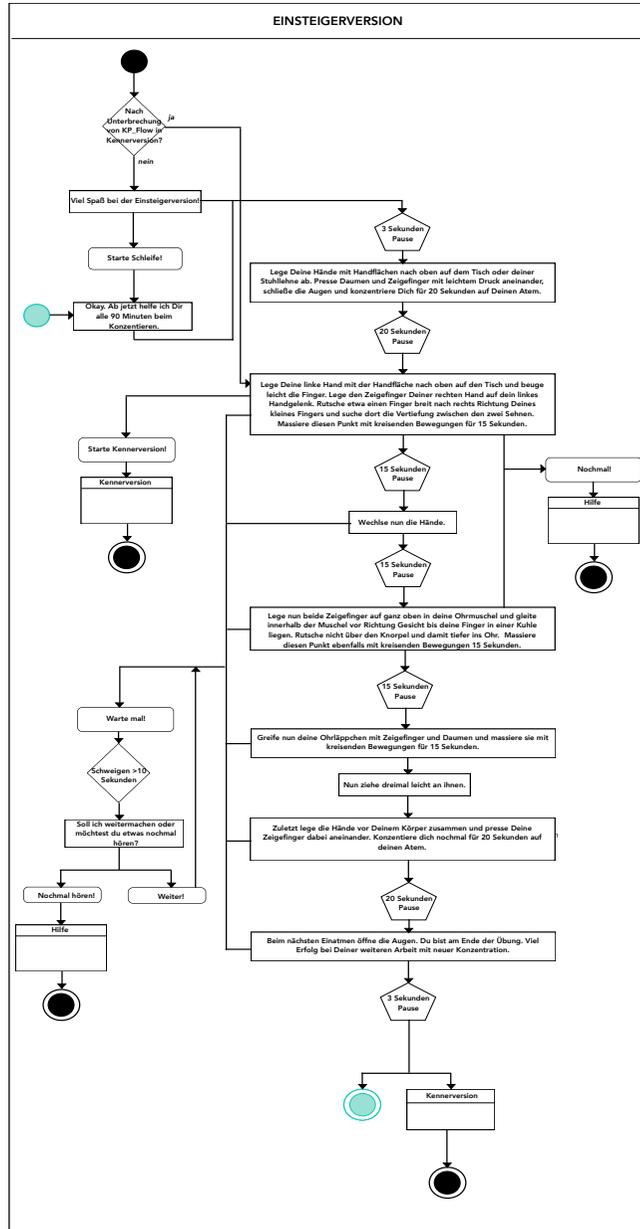
**A** Appendix



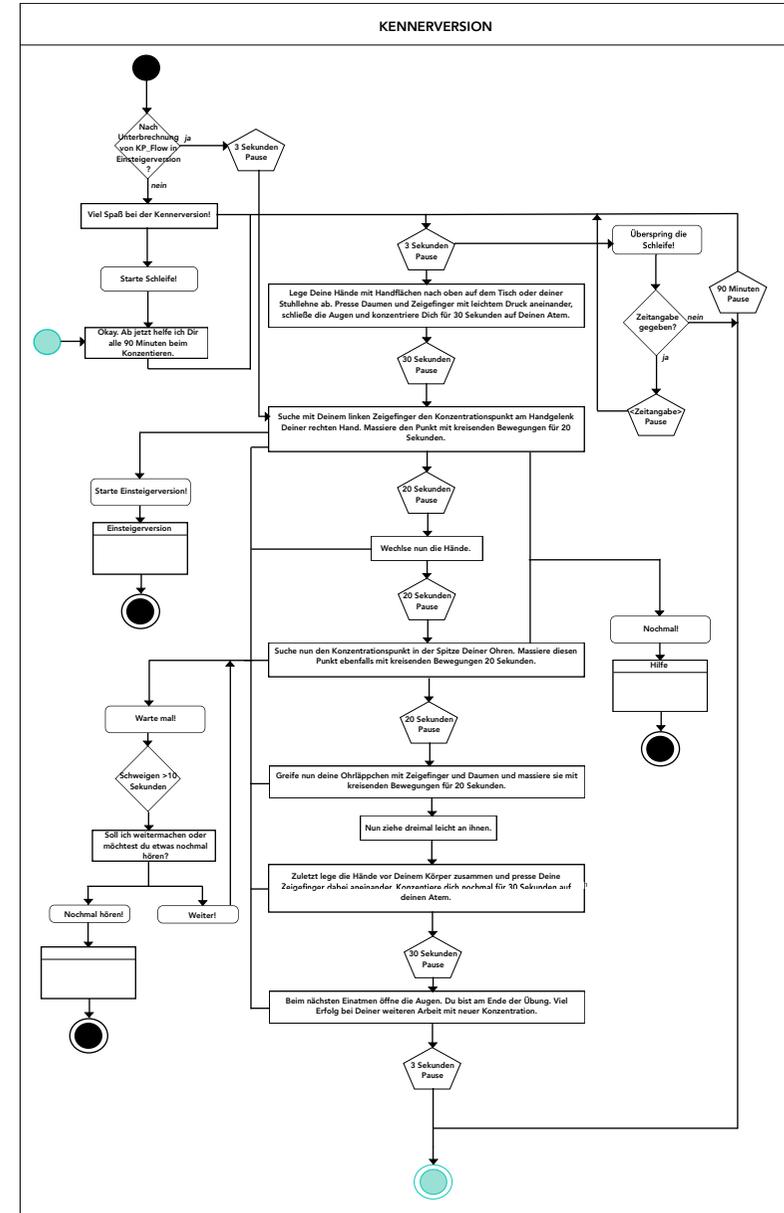
**Abbildung 4 KM:** Überblick des Flowdiagramms. Detailansichten der einzelnen Diagramme finden sich auf den Folgeseiten



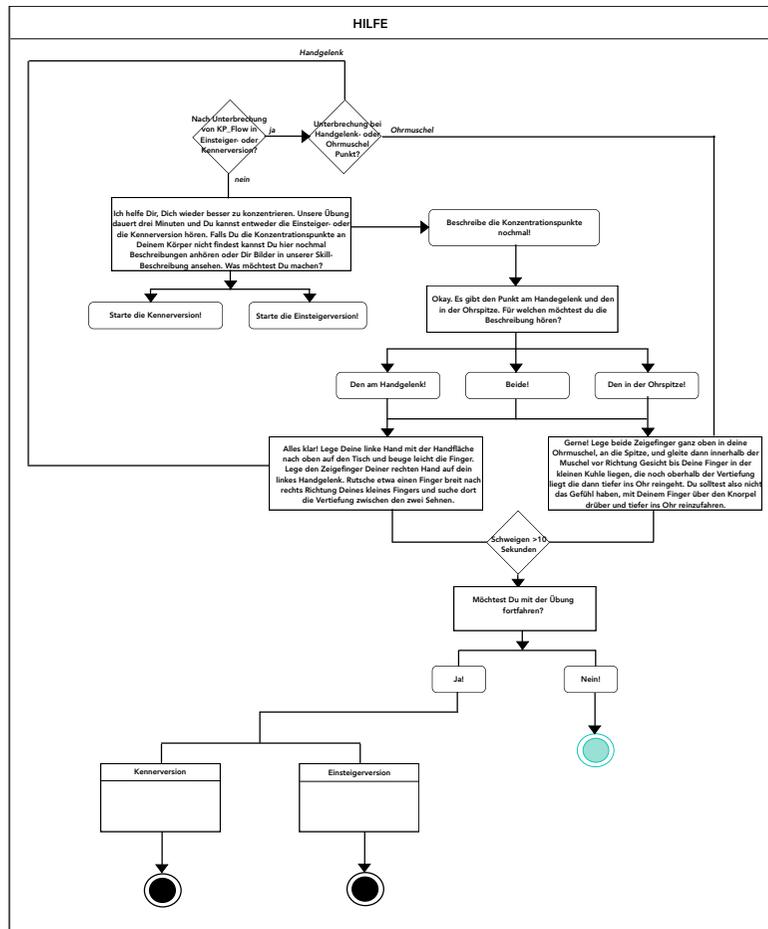
**Abbildung 5 KM:** Globale Skill und Legende



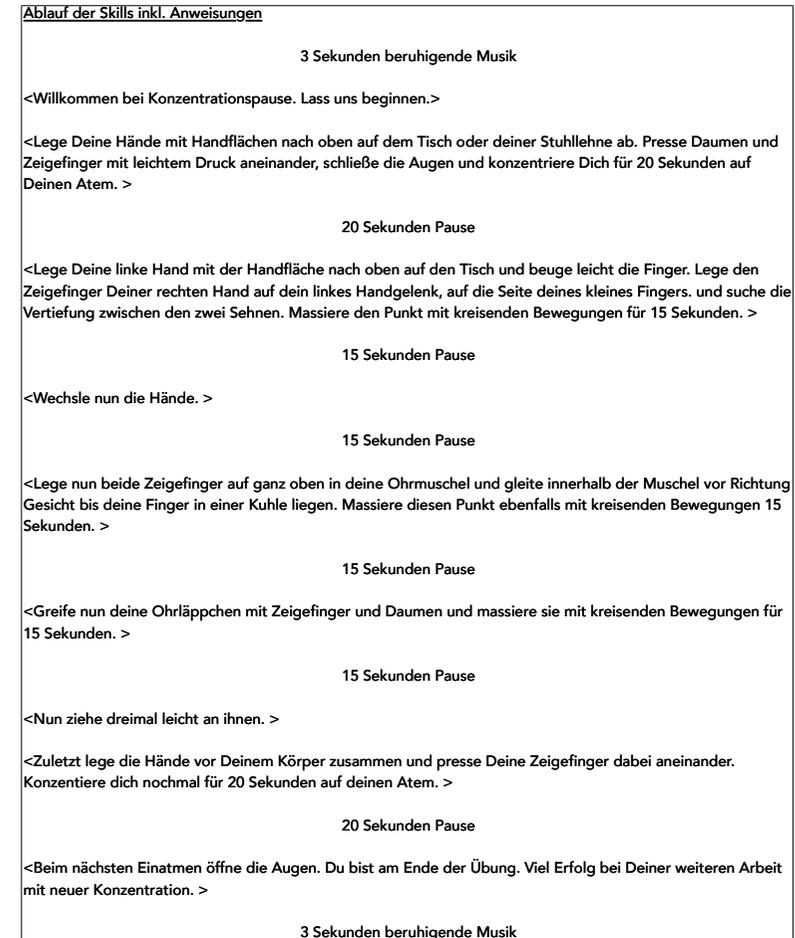
■ **Abbildung 6** KM: Einstiegsersion



■ **Abbildung 7** KM: Kennerversion



■ **Abbildung 8** KM: Hilfe



■ **Abbildung 9** KM: Instruktionstext

## Literatur

- 1 Amazon. Amazon echo, 2018. Retrieved June 10, 2018 from. URL: [https://www.amazon.de/s/ref=sr\\_nr\\_n\\_1?fst=as%3Aoff&rh=n%3A10983902031%2Ck%3Aecho&keywords=echo&ie=UTF8&qid=1526930029&rnid=1703609031](https://www.amazon.de/s/ref=sr_nr_n_1?fst=as%3Aoff&rh=n%3A10983902031%2Ck%3Aecho&keywords=echo&ie=UTF8&qid=1526930029&rnid=1703609031).
- 2 Dimitra Anastasiou, Cui Jian, and Desislava Zhekova. Speech and gesture interaction in an ambient assisted living lab. In *Proceedings of the 1st Workshop on Speech and Multimodal Interaction in Assistive Environments*, pages 18–27. Association for Computational Linguistics, 2012.
- 3 CareerFoundry. Voice user interface design, 2018. URL: <https://careerfoundry.com/en/courses/voice-user-interface-design-with-amazon-alexa/>.
- 4 Ashley Carman. Here are the six new voices coming to google assistant, including john legend. 2018. URL: <https://www.theverge.com/2018/5/8/17332014/google-assistant-voice-john-legend>.
- 5 Google Cloud. Google cloud speech-to-text api preisliste, 2018. Retrieved June 10, 2018 from. URL: <https://cloud.google.com/speech-to-text/pricing>.
- 6 Michael H Cohen, James P Cohen, Michael Harris und Giangola, and Jennifer Balogh. *Voice user interface design*. Addison-Wesley Professional, 2004.
- 7 Amazon Developers. Voice design 101: An introduction, 2017. URL: <https://www.youtube.com/watch?v=XdUooJoTKcA>.
- 8 Luke Dormehl. Today in apple history: Siri debuts on iphone 4s. 2017. URL: <https://www.cultofmac.com/447783/today-in-apple-history-siri-makes-its-public-debut-on-iphone-4s/>.
- 9 Brandon Griggs. Why computer voices are mostly female. 2011. URL: <https://edition.cnn.com/2011/10/21/tech/innovation/female-computer-voices/index.html>.
- 10 Charles Hannon. Gender and status in voice user interfaces. *interactions*, 23(3):34–37, 2016.
- 11 Mike Jeffs. Ok google, siri, alexa, cortana; can you tell me some stats on voice search? 2018. URL: <https://www.branded3.com/blog/google-voice-search-stats-growth-trends/>.
- 12 Bret Kinsella. Amazon echo maintains large market share lead in u.s. smart speaker user base. 2018. URL: <https://www.voicebot.ai/2018/03/08/amazon-echo-maintains-large-market-share-lead-u-s-smart-speaker-user-base/>.
- 13 Ilker Koksalsal. Voice-first devices are the next big thing - here's why. 2018. URL: <https://www.forbes.com/sites/ilkerkoksalsal/2018/02/01/voice-first-devices-are-the-next-big-thing-heres-why/#2d3756916873>.
- 14 Voice Labs. The 2017 voice report, 2017. URL: [https://s3-us-west-1.amazonaws.com/voicelabs/report/vl-voice-report-exec-summary\\_final.pdf](https://s3-us-west-1.amazonaws.com/voicelabs/report/vl-voice-report-exec-summary_final.pdf).
- 15 Sascha Lobo. Sprachsteuerung im alltag: Ohne geht es nicht mehr. 2018. URL: <http://www.spiegel.de/netzwelt/gadgets/sprachsteuerung-im-alltag-ohne-geht-es-nicht-mehr-kolumne-a-1187056.html>.
- 16 Mary Meeker. Internet trends 2017. code conference, may 31, 2017. rancho palos verdes, 2017. URL: <http://www.kpcb.com/internet-trends>.
- 17 Ditte Mortensen. How to design voice user interfaces. 2018. URL: <https://www.interaction-design.org/literature/article/how-to-design-voice-user-interfaces>.
- 18 Clifford Nass and Scott Brave. *Wired for speech: How voice activates and advances the human-computer relationship*. MIT press, 2005.
- 19 W Nguyen, Anh Thi Thuc und Wobcke. *An Agent-Based Approach to Dialogue Management in Personal Assistants*. University of New South Wales, 2006.
- 20 Actions on Google. Conversation design, 2018. URL: <https://designguidelines.withgoogle.com/conversation/#welcome-designing-actions-on-google>.
- 21 Neil Patel. Nie mehr tippen: Wie die sprachsuche die seo-landschaft beeinflussen wird, 2016. URL: <https://neilpatel.com/de/blog/nie-mehr-tippen-wie-die-sprachsuche-die-seo-landschaft-beeinflussen-wird/>.
- 22 Cathy Pearl. *Designing Voice User Interfaces: Principles of Conversational Experiences*. O'Reilly Media, Inc., 2016.
- 23 Melanie Pinola. Speech recognition through the decades: How we ended up with siri. 2011. URL: [https://www.pcworld.com/article/243060/speech\\_recognition\\_through\\_the\\_decades\\_how\\_we\\_ended\\_up\\_with\\_siri.html?page=2](https://www.pcworld.com/article/243060/speech_recognition_through_the_decades_how_we_ended_up_with_siri.html?page=2).
- 24 Jeffrey N Rouder, Richard D Morey, Nelson Cowan, Christopher E Zwilling, Candice C Morey, and Michael S Pratte. An assessment of fixed-capacity models of visual working memory. *Proceedings of the National Academy of Sciences*, 105(16):5975–5979, 2008.
- 25 Sandra Schuster. Voice control um jeden preis? theoretische und praktische Grundlagen für erfolgreiche sprachsteuerungs-angebote aus user experience-sicht. In Henning Brau, Andreas Lehmann, Kostanija Petrovic, and Matthias C. Schroeder, editors, *Tagungsband UPI3*, pages 226–231, Stuttgart, 2013. German UPA e.V.
- 26 Aarti Shahani. Voice recognition software finally beats humans at typing, study finds. 2016. URL: <https://www.npr.org/sections/alltechconsidered/2016/08/24/491156218/voice-recognition-software-finally-beats-humans-at-typing-study-finds>.

- 27 Savina van der Straten. Voice is the next big thing: Why now? 2017. URL: <https://medium.com/point-nine-news/voice-is-the-next-big-thing-913b9bbf9016>.
- 28 Ari Weissman. Convenience. 2012. URL: <https://uxmag.com/articles/convenience>.
- 29 Kathryn Whitenon. Voice first: The future of interaction? 2017. URL: <https://www.nngroup.com/articles/voice-first/>.
- 30 Think with Google. Mobile app ux principles: Improving user experience and optimising conversion. 2015. URL: <https://www.thinkwithgoogle.com/intl/en-gb/advertising-channels/mobile/mobile-app-ux-principles-improving-user-experience/>.

# Nutzbare Sicherheit: Erkenntnisse aus dem Bereich der Authentifizierung

Sebastian Halder

Ludwig-Maximilians-Universität München, München, Deutschland  
halder@cip.ifi.lmu.de

---

## Zusammenfassung

Obwohl die frühen Arbeiten zum Thema nutzbare Sicherheit inzwischen über 20 Jahre alt sind, ist weiterhin kein Standard für die Entwicklung von anwenderfreundlichen Sicherheitslösungen in Sicht. Vor allem im Bereich der Authentifizierungsverfahren gerät der menschliche Aspekt oft in den Hintergrund, weil hohe Sicherheitsanforderungen und leichte Bedienbarkeit oft im Konflikt zu stehen scheinen [3]. Wie aktuelle Beispiele [10] jedoch zeigen, werden Sicherheitspolizen von Anwendern nicht unbedingt so umgesetzt, wie ursprünglich geplant. Dadurch entstehen neue Sicherheitsprobleme, die hauptsächlich aus umständlicher Bedienung und mangelnder Nutzerfreundlichkeit resultieren.

In dieser Arbeit werden unterschiedliche Nutzbarkeitsprobleme aus dem Bereich der Authentifizierungsverfahren untersucht. Dazu wird vor allem auf vorangegangene Arbeiten und Studien zurückgegriffen, um eine umfangreiche Liste an Design Prinzipien zu erstellen, die dabei helfen sollen sichere und nutzbare Systeme zu entwickeln.

**2012 ACM Computing Classification** Security and privacy → Usability in security and privacy

**Keywords and phrases** Sicherheit; Nutzbarkeit; Mensch-Maschine-Interaktion; Authentifizierung.

## 1 Einleitung

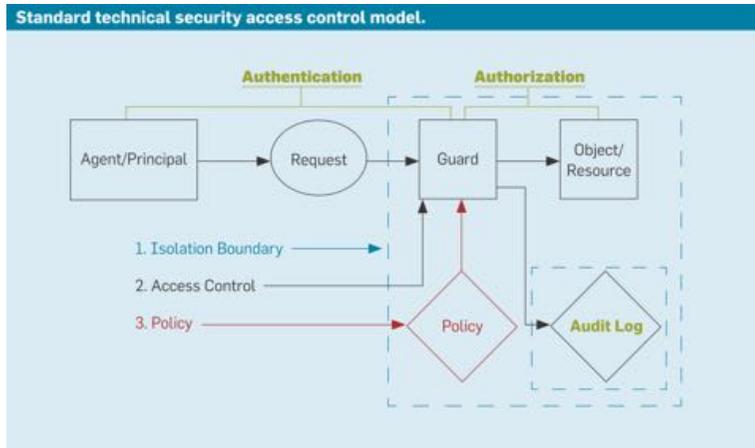
Nutzbare Sicherheit (engl. Usable Security) befasst sich mit Problemen im Bereich der IT-Sicherheit, die vor allem daraus resultieren, dass gängige Sicherheitssoftware und -lösungen für den durchschnittlichen Anwender verständlich oder schwer zu bedienen sind. Als Folge davon hebeln Anwender viele Eigenschaften von Sicherheitssoftware bewusst oder unbewusst aus und schaffen damit neue Sicherheitsprobleme. Das wohl prominenteste Beispiel dafür ist die vergangene Generation an Passworrichtlinien, die vor allem durch Zusammensetzungsregeln und regelmäßig erzwungene Passwortwechsel



© Sebastian Halder;

licensed under Creative Commons License CC-BY

Cite as: Sebastian Halder. Nutzbare Sicherheit: Erkenntnisse aus dem Bereich der Authentifizierung. 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. 8:1–8:14.



■ **Abbildung 1** Zugriffskontrollmodell (access control model): Ein Client versucht auf ein Objekt oder eine Ressource zuzugreifen und wird aufgefordert sich zu authentifizieren. Nach erfolgreicher Authentifizierung seitens des Access Control Managers (hier „Guard“) wird im Regelwerk nachgeschlagen, ob der Client die Autorisierung besitzt, auf diese Ressource zuzugreifen. Die Interaktion zwischen Nutzer und Access Control Manager wird in einem Zugriffsprotokoll aufgezeichnet. Quelle:[7]

charakterisiert werden können. Trotz guter Absichten haben diese Regeln nicht dazu geführt, dass die Mehrzahl der Nutzer sicherere Passwörter verwendet [10], sondern womöglich das Gegenteil erreicht.

Der Aufbau der Arbeit ist wie folgt: In Abschnitt 2 werden Hintergrundinformationen zu Authentifizierungsverfahren und Zugriffskontrolle im Allgemeinen geliefert. Abschnitt 3 beschäftigt sich mit Fallbeispielen aus unterschiedlichen Bereichen der Authentifizierungsverfahren. In Abschnitt 4 werden aus den zuvor analysierten Fallbeispielen und weiteren Quellen allgemeine Design Prinzipien zur Entwicklung von nutzbarer Sicherheit abgeleitet.

## 2 Hintergrund

Im folgenden Abschnitt soll ein allgemeiner Hintergrund über Zugriffskontrolle und gängige Authentifizierungsverfahren vermittelt werden.

### 2.1 Zugriffskontrollmodell

Das übliche technische Modell zur Beschränkung von Zugriff auf Ressourcen oder Geräte ist das Zugriffskontrollmodell [7] wie es in Abbildung 1 dargestellt

ist. Im ersten Schritt wird dabei eine Authentifizierung des Nutzers durchgeführt. Unter Authentifizierung versteht man den Prozess in dem die Identität einer Person oder eines Geräts überprüft wird. Anschließend erhält der Nutzer Zugriff auf Ressourcen (z.B. Daten, Geräte, Anwendungen, etc.), die ihm anhand eines vordefinierten Regelwerks zur Verfügung gestellt werden. Dies wird als Autorisierung bezeichnet. Zusätzlich wird von der Kontrollinstanz ein Zugriffsprotokoll geführt, um eventuelle Konfigurationsfehler oder Angriffe nachzuvollziehen zu können.

## 2.2 Authentifizierungsverfahren

Allgemein lassen sich Authentifizierungsverfahren in drei Kategorien unterteilen: Wissen, Besitz und inhärente Merkmale [9].

### 2.2.1 Wissen

Verbreitete wissensbasierte Authentifizierungsverfahren sind die Eingabe von Benutzernamen und Passwörtern sowie die Verwendung von PINs [9]. Die Authentifizierung basiert dabei auf geheimem Wissen, welches, im Idealfall, nur den Personen zur Verfügung steht, die sich mit eben dieser Identität authentifizieren sollen. Die Schwierigkeit bei dieser Art der Authentifizierung ist der Umstand, dass Wissen sowohl vergessen, als auch bewusst oder unbewusst weitergegeben, sowie erraten werden kann. Auf passwortbasierte Authentifizierung im Speziellen wird in Abschnitt 3.2 genauer eingegangen. Weitere Beispiele für wissensbasierte Authentifizierung sind Sicherheitsfragen oder Geheimfragen, wie sie teilweise zur Passwortwiederherstellung eingesetzt werden, und Sperrmuster für Smartphones.

### 2.2.2 Besitz

Authentifizierung durch Besitz findet beispielsweise bei Chipkarten (Smartcards) und Magnetstreifenkarten, Zertifikaten (z.B. TLS), Einmalpasswörtern oder herkömmlichen Schlüsseln statt. Dabei wird der Nutzer anhand eines Objektes identifiziert, das sich in seinem Besitz befindet. Gleichzeitig bedeutet dies, dass jede Person, die sich im Besitz dieses Objekts oder einer Kopie dessen befindet automatisch die selbe Identität teilt. Dieser Umstand spielt sowohl gewollt (z.B. gemeinsamer Schlüssel, Schlüsselweitergabe), als auch ungewollt (z.B. Diebstahl) eine Rolle. Weitere Herausforderungen bei dieser Art der Authentifizierung sind Verwaltung und Aufbewahrung, sowie Verlust eines „Schlüssels“.

### 2.2.3 Inhärente Merkmale / Biometrie

Zu inhärenten Merkmalen zählen biometrische Merkmale wie Fingerabdrücke, Gesichts- und Iriserkennung, sowie Handschrift (z.B. Unterschrift), Gestik und Tippverhalten. Inhärente Merkmale sind an eine Person gebunden und, im Idealfall, nicht auf andere Personen übertragbar oder reproduzierbar. Allerdings sind sie in der Regel auch nicht veränderlich und öffentlich sichtbar.

## 3 Fallbeispiele

Im Folgenden sollen Fallbeispiele aus unterschiedlichen Bereichen der Nutzer-Authentifizierung vorgestellt werden. Ziel davon ist es, aus Fehlern und Erfolgen der unterschiedlichen Systeme zu lernen und diese im nächsten Kapitel zu verallgemeinerten Lösungsansätzen zusammenzufassen.

### 3.1 Tipp-Biometrie für Touchscreens

Im Bereich der Smartphones und anderen Mobilfunkgeräten gehören Sperrmuster und PINs zu den meistverwendeten der Nutzer-Authentifizierungsverfahren [9]. Dass in diesem Bereich trotzdem noch viel Raum für Verbesserung bleibt, zeigen Buschek et al. [5] mit ihrer Arbeit zu Tipp-Biometrie für Touchscreens.

Tastendruck Biometrie an sich ist kein neues Verfahren und wurde im Laufe der letzten drei Jahrzehnte von unterschiedlichen Wissenschaftlern untersucht [2]. Buschek et al. übertragen diese Art von Biometrie auf die Interaktion der Nutzer mit den Touchscreens von Smartphones. Hierzu wurden unterschiedliche Merkmale beim Schreibverhalten der Nutzer in drei verschiedenen Handpositionen identifiziert. Die Nutzung von räumlichen Merkmalen zusätzlich zu den üblichen temporalen Merkmalen sorgt außerdem für eine Verbesserung der Fehlerraten. Besonders hervorgehoben werden Fortschritte und Erkenntnisse im Bereich der Nutzbarkeit. Dazu gehören Techniken um unterschiedlichen Handpositionen von Nutzern zu erlauben, sowie mit natürlichen Änderungen im Schreibstil der Nutzer umzugehen.

#### 3.1.1 Erkenntnisse

Verfahren zur Tipp-Biometrie sind in der Lage bestehende Authentifizierungsverfahren wie Passwörter und PINs um eine zusätzliche Sicherheitskomponente zu erweitern und bedeuten im Idealfall keinen Mehraufwand für den Anwender. Dennoch machen Buschek et al. deutlich, dass gängige Studien die Genauigkeit ihrer Ergebnisse überschätzen, weil ein Großteil der Publikationen Daten für Training und Evaluation in einer einzigen Sitzung erheben [5].

### 3.2 Passwort Richtlinien

In diesem Abschnitt sollen Richtlinien zur Erstellung von sicheren Passwörtern in den letzten Jahren verglichen werden. Besonderer Fokus liegt dabei auf den Publikationen des „US National Institute of Standards and Technology“ (NIST), das in regelmäßigen Abständen Leitsätze zu sicherheitsrelevanten Themen veröffentlicht. Diese Leitsätze sind zwar primär für öffentliche Einrichtungen in den USA bestimmt, haben aber seit geraumer Zeit auch großen Einfluss auf den privaten Sektor [10]. Die neuesten NIST Richtlinien zum Thema Passwörter finden sich in den „Digital Identity Guidelines: Authentication and Lifecycle Management“ (SP 800-63B [6]), veröffentlicht Juni 2017, und unterscheiden sich vor allem im Bezug auf Nutzbarkeit deutlich von der damit abgelösten „Electronic Authentication Guideline“ aus dem Jahr 2013.

#### 3.2.1 NIST Richtlinien

In der zuletzt 2013 überarbeiteten „Electronic Authentication Guideline“ [4] (Erstveröffentlichung 2003) lag das Hauptaugenmerk für die Passwortwahl auf der Bewertung der Passwortstärke anhand von Entropie-Analyse (Shannon Entropie)[10, 4]. Daraus resultierend wurden Kompositionsregeln für Passwörter empfohlen, die eine Mindestanzahl von bestimmten Zeichen wie beispielsweise Groß- und Kleinbuchstaben, Zahlen und Sonderzeichen fordern. Des Weiteren sollen Passwörter nach Ablauf einer vordefinierten Zeitspanne ablaufen und vom Anwender erneuert werden müssen. Diese Ansätze bergen unterschiedliche Probleme. Wie Weir et al. [10] zeigen, ist das verwendete Konzept von Passwort-Entropie kein geeignetes Maß um die Sicherheit von Passwörtern zu beschreiben. Darüber hinaus erlauben gängige Kompositionsregeln trotzdem die Erstellung von sehr unsicheren Passwörtern. Ein Umstand der durch den verpflichtenden, regelmäßigen Passwortwechsel noch verschlimmert wird. Sasse und Flechais [8] argumentieren, dass besonders unregelmäßig genutzte Passwörter Nutzer dazu motivieren gängige Sicherheitsrichtlinien zu missachten. Dies führe zu Passwortnotizen in Büroräumen und der Erstellung von möglichst einfachen Passwörtern, die trotzdem den jeweiligen Kompositionsregeln entsprechen (z.B. *Password1!* [10]).

In deutlichem Gegensatz dazu wird in einer Arbeit von Grassi et al. [6] die Verwendung von langen, einfach zu merkenden Passwörtern, z.B. in Form von Passphrasen, empfohlen. Außerdem soll auf Passwortwechsel ohne speziellen Anlass (z.B. Datenpanne, etc.) verzichtet werden. Zur Verbesserung der Sicherheit sollen zusätzlich keine Kennworthinweise eingesetzt und eingegebene Passwörter bei der Erstellung gegen existierende Listen von bekannten/schlechten Passwörtern verglichen werden. Darüber hinaus wird mehr Augenmerk auf die sichere Speicherung von Passwörtern (z.B. starke Hash-Algorithmen,

mehrere Salts) gelegt und die Verwendung von Zwei-Faktor-Authentifizierung empfohlen. Außerdem soll die Verwendung von Passwortmanagern vereinfacht werden, indem Loginmasken standardmäßig das Kopieren und Einfügen von Passwörtern erlauben.

### 3.2.2 Erkenntnisse

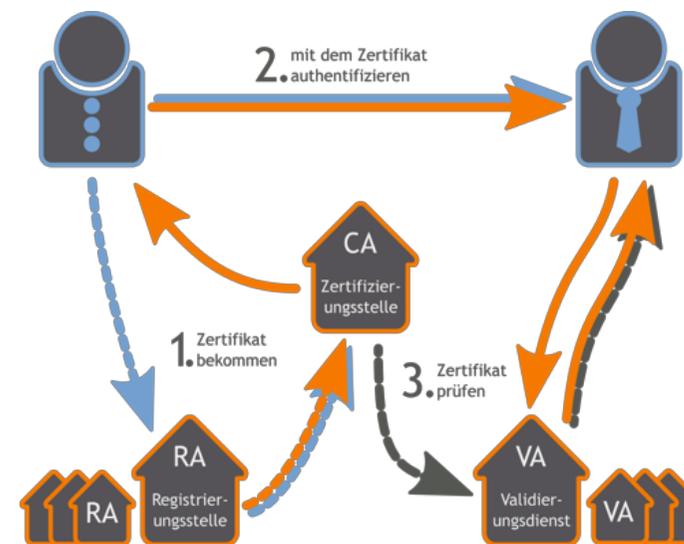
Die deutlichsten Veränderungen bei den oben betrachteten Richtlinien beziehen sich nicht auf die Erstellung von möglichst sicheren und komplexen Passwörtern, sondern darauf, die Nutzerfreundlichkeit soweit zu verbessern, dass Anwender in der Lage sind Passwörter zu verwenden, die sowohl vergleichbar sicher, als auch einfach zu merken sind. Ursprung des Problems ist der von Lampson [7] beschriebene Konflikt zwischen Sicherheitsfunktionen und dem Arbeitsfluss des Anwenders. Die Einrichtung von Sicherheitsfunktionen hat damit eine wirtschaftliche Komponente, bei der es darum geht deren Kosten und Nutzen abzuwägen und festzustellen, ob der damit verbundene Mehraufwand für die Nutzer den zu erwartenden Sicherheitsgewinn rechtfertigt. Außerdem darf der Umstand nicht unberücksichtigt bleiben, dass ein theoretisch sicheres System nicht unbedingt praktisch sicher ist, wenn Anwender Sicherheitsfunktionen bewusst oder unbewusst außer Kraft setzen, weil sie der eigentlichen Arbeit im Wege stehen [8].

## 3.3 Public-Key-Infrastruktur

In diesem Abschnitt wird der Einsatz einer Public-Key-Infrastruktur (PKI) und damit verbundene Probleme, wie von Balfanz et al. [1] durchgeführt, beschrieben. In einer PKI werden digitale Zertifikate zur Nutzerauthentifizierung eingesetzt (vgl. Abbildung 2). In dem hier beschriebenen Fall soll die Sicherheit des WLAN Netzes in einem Forschungszentrum durch Einsatz einer PKI verbessert werden.

### 3.3.1 Klassischer Ansatz

Im ersten Ansatz wurden Nutzer dazu aufgefordert die Netzwerkkonfiguration und Installation des Zertifikats für ihre eigenen Maschinen selbst zu übernehmen. Hierfür wurde eine genaue Anleitung bereitgestellt, die die unterschiedlichen Schritte (insgesamt 38), von der Konfiguration des Netzwerkadapters bis zur Anforderung und Installation eines von der lokalen Zertifizierungsstelle (Certificate Authority, CA) ausgestellten Zertifikats, dokumentiert. Trotz Dokumentation und dem Umstand, dass viele der Anwender im Bereich der Informatik tätig waren, wurde im Rahmen einer Nutzerbefragung festgestellt, dass der durchschnittliche Zeitaufwand, der zu Einrichtung benötigt wurde,



■ **Abbildung 2** Modell einer Public-Key-Infrastruktur: Nach Authentifizierung bei der Registrierungsstelle wird dem Benutzer von der Zertifizierungsstelle ein Zertifikat ausgestellt. Dieses kann der Anwender anschließend dazu verwenden, sich gegenüber anderer Benutzern oder Geräte zu authentifizieren. Die Echtheit des Zertifikats und damit die Identität des Anwenders kann durch Nutzung des Validierungsdienstes überprüft werden. Bildquelle: [https://de.wikipedia.org/wiki/Datei:Orange\\_blue\\_public\\_key\\_infrastructure\\_2\\_de.svg](https://de.wikipedia.org/wiki/Datei:Orange_blue_public_key_infrastructure_2_de.svg)

mehr als zwei Stunden betrug. Darüber hinaus erklärten die meisten der Befragten, dass es ihnen in diesem Zeitrahmen nicht gelungen ist, den Installationsprozess im Detail zu verstehen und sie damit im Störfall auch selbst keine Fehlerdiagnose durchführen könnten. Diese Feststellungen decken sich auch mit früheren Studien, wie beispielsweise von Whitten und Tygar [11], die im speziellen Fall von PGP argumentieren, dass das System nicht nutzbar genug sei um effektive Sicherheit für den Anwender bereitzustellen.

### 3.3.2 Verbesserter Ansatz

Um die Nutzbarkeitsprobleme aus dem vorherigen Ansatz zu beheben, wurde das System einer Neuplanung unterworfen und die manuelle Einrichtung durch eine automatische Installation via Software ersetzt. Darüber hinaus wurde die Ausstellung der Zertifikate ausschließlich durch Interaktion mit einer „Enrollment Station“ durchgeführt, die als Registrierungsstelle diente. Diese Maschine

stellt zuerst eine verschlüsselte ad hoc Drahtlosnetzwerkverbindung (WLAN) zum betreffenden Endnutzengerät her und tauscht darüber Informationen zur Erstellung und anschließend das erstellte Zertifikat selbst aus. Die „Enrollment Station“ befand sich in einem abgeschlossenen Raum, der nur nach Vorzeigen des Mitarbeiterausweises betreten werden konnte. Außerdem wurde die Nutzerinteraktion mit der Registrierungsstelle via Gestensteuerung implementiert. Die durchschnittlich benötigte Zeit einen neuen Rechner mit dem WLAN zu verbinden wurde mit dieser Lösung auf unter zwei Minuten reduziert.

### 3.3.3 Erkenntnisse

Anstatt die bestehende Lösung schrittweise zu verbessern, haben Balfanz et al. sich dafür entschieden von Grund auf neu zu beginnen und ein System zu entwickeln das von Anfang an großen Fokus auf die Nutzbarkeit legt. Durch die Platzierung der Registrierungsstelle („Enrollment Station“) in einem abgeschlossenen Raum wurde automatisch eine Verknüpfung zwischen physikalischer Sicherheit und digitaler Sicherheit hergestellt und dadurch ein intuitives Vertrauensmodell geschaffen [1]. Bei der Entwicklung der unterschiedlichen Ansätze wurde jeweils Rückmeldung von den Anwendern über die Nutzbarkeit des Systems zu eingeholt, um etwaige Mängel erkennen und beheben zu können. Durch Einsatz von Gestensteuerung war es den Anwendern möglich intuitiv Sicherheitsentscheidungen zu treffen, indem sie der „Enrollment Station“ auf diese Weise zu verstehen gaben, welches Gerät sie mit dem Netzwerk verbinden wollten.

## 4 Designprinzipien

Basierend auf vorangegangenen Arbeiten und teilweise aus den obigen Fallbeispielen abgeleitet, werden im Folgenden mehrere Designprinzipien vorgestellt, die die Nutzbarkeit im Bereich der Nutzerauthentifizierung verbessern sollen. Die vorgestellten Lösungsansätze stellen dabei keine vollständige Liste an Richtlinien dar, sondern versuchen in verallgemeinerter Form Hinweise zur Entwicklung von benutzerfreundlichen und effektiven Systemen zu vermitteln.

### 4.1 Nutzbarkeit von Anfang an einplanen

Nutzbare Sicherheit lässt sich in der Regel nicht nachträglich für ein bestehendes System entwickeln. Oder wie Balfanz et al. [1] es ausdrücken, ist es nicht die Lösung ein verwirrendes System um erklärende Dialogfenster zu erweitern. Viele fundamentale Designprinzipien, die die Nutzbarkeit des Systems beeinflussen, müssen zu Beginn des Entwicklungsprozesses getroffen werden [1]. Diese Designphilosophie lässt sich sowohl am Beispiel der NIST Richtlinien,

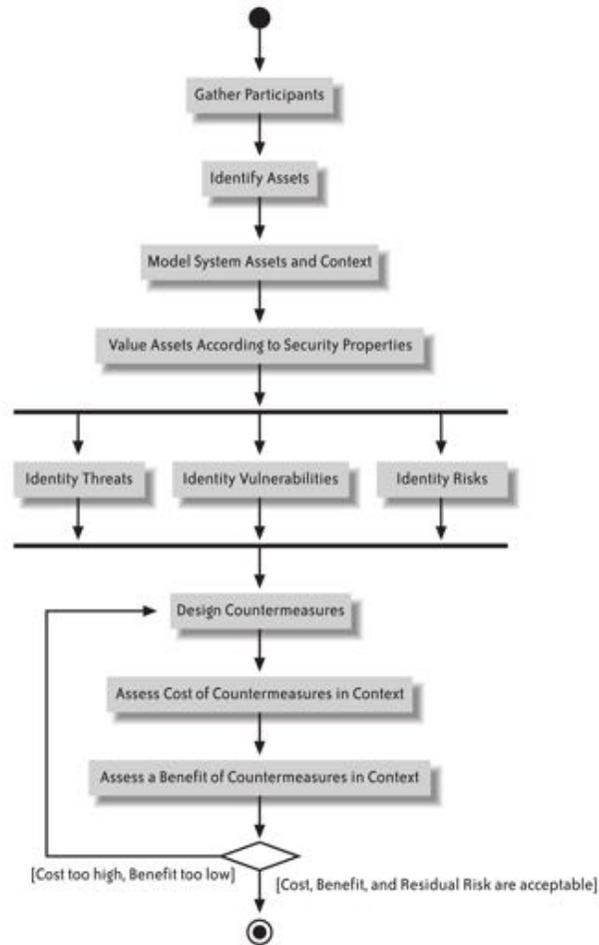
als auch beim vorgestellten Beispiel der Public-Key-Infrastruktur erkennen. In beiden Fällen wurden im ersten Ansatz zuallererst Sicherheitsinteressen umgesetzt. Dabei wurde davon ausgegangen, dass Anwender sich an die gegebenen Systeme anpassen würden. Wie Weir et al. [10] zeigen, hat dies allerdings nicht zur erwünschten Verbesserung der Passwortsicherheit geführt. Dementsprechend wird in den aktuellen Richtlinien von Grund auf mehr Augenmerk auf Nutzerfreundlichkeit und Bedienbarkeit auf Anwenderseite gelegt.

### 4.2 Arbeitsfluss erhalten

Einer der Kernpunkte aus Yees [12] Designprinzipien für Benutzeroberflächen ist das Prinzip vom „Weg des geringsten Widerstands“. Dabei argumentiert Yee, dass im Designprozess so gut wie möglich erreicht werden muss, dass bei der Interaktion mit den Sicherheitsfunktionen der intuitive Ansatz aus Anwendersicht auch gleichzeitig dem sichersten Ansatz entspricht. Sichtbar ist dieses Problem auch im Fall der Passwortrichtlinien beschrieben in Abschnitt 3.2. Anwender, die sich von regelmäßigen Passwortwechseln oder restriktiven Kompositionsregeln beeinträchtigt fühlen, tendieren dazu, Sicherheitsregeln (z.B. keine Passwortnotizzettel am Arbeitsplatz) bewusst außer Kraft zu setzen [8]. In anderen Fällen werden Passwörter gewählt, die zwar einprägsam, aber nicht sicher sind [8, 10]. Diese Problematik ist Teil der wirtschaftlichen Komponente von IT-Sicherheit. Hier gilt es, Kosten und Nutzen der Systeme abzuwägen und dabei die Nutzerinteressen im Bezug auf Arbeitsaufwand und etwaige Unannehmlichkeiten im Auge zu behalten. „Sicherheitsaufgaben müssen so ausgelegt sein, dass sie Produktionsaufgaben unterstützen“ [8]. Das bedeutet insbesondere, dass der Arbeitsfluss der Anwender nicht unverhältnismäßig gestört werden darf, um zu vermeiden, dass Benutzer Sicherheitsfunktionen deaktivieren oder umgehen.

### 4.3 Feedback einholen

Sasse und Flechais [8] stellen mit ihrem AEGIS Modell einen Prozess vor um nutzbare Sicherheitssysteme zu entwerfen. Ein Hauptbestandteil dabei stellt die Interaktion mit unterschiedlichen Interessenvertretern wie beispielsweise Anwendern, Mitarbeitern und Managern dar. Dies dient dazu, Mängel beim Systemdesign aufdecken zu können und gleichzeitig das Bewusstsein für die Sicherheitsproblematik zu steigern. Diese Interaktion wird vor allem in den ersten beiden Schritten des AEGIS Prozesses fokussiert, in denen Teilnehmer und Kapital analysiert werden (vgl. Aktivitätsdiagramm Abbildung 3). Auch Balfanz et al. [1] betonen die Wichtigkeit von Benutzerfeedback. Die in Abschnitt 3.3 beschriebenen Nutzbarkeitsprobleme stammen vor allem daher, dass die Sicherheitsforscher beim ersten Ansatz davon überzeugt waren, ein einfach



■ **Abbildung 3** AEGIS Aktivitätsdiagramm [8]: Der AEGIS Prozess ist ein iterativer Ansatz zur Entwicklung von nutzbaren Systemen im Bereich der IT-Sicherheit. Ein besonderer Fokus liegt dabei auf der Einbeziehung der beteiligten Parteien (Anwender, Mitarbeiter, Manager, etc.) in den Entwicklungsprozess (vgl. Schritt 1 & 2). Außerdem werden iterativ Kosten und Nutzen aller Maßnahmen des geplanten Sicherheitssystems geprüft. Dazu zählen neben finanziellen Kosten, auch Arbeits- und Organisationsaufwand für die Anwender.

zu verwendendes System konzipiert zu haben, nicht zu allerletzst deswegen, weil es dem üblichen Vorgehen entsprach. Da sie selbst genug Erfahrung mit der Verwendung von Public-Key Infrastrukturen hatten, konnten sie selbst im ersten Ansatz Rechner problemlos und in kurzer Zeit mit dem Netzwerk verbinden. Dass ihre Anwender diese Erfahrung nicht teilten, wurde erst in der Nutzerbefragung offensichtlich. Darüber hinaus, so erklären Balfanz et al., seien Nutzbarkeitsprobleme nicht unbedingt am Erfolg oder Misserfolg des Systems zu erkennen. Im Fall der Standard-PKI wären Anwender sehr wohl in der Lage gewesen, das Drahtlosnetzwerk der Forschungseinrichtung zu benutzen, wenn auch unter sehr hohem Zeitaufwand.

#### 4.4 Sicherheitsmechanismen sichtbar machen

Wie Balfanz et al. [1] beschreiben, sollen Sicherheitsmechanismen nicht nur in den unteren Anwendungsschichten sichtbar sein. Stattdessen soll es dem Benutzer ermöglicht werden intuitiv Sicherheitsentscheidungen zu treffen, indem seine Absichten im natürlichen Umgang mit den Geräten umgesetzt werden. Im in Abschnitt 3.3 vorgestellten Fall wurde dies zum Beispiel durch den Einsatz von Gestensteuerung realisiert. Yee [12] beschreibt ein ähnliches Konzept mit seinen Designprinzipien „Explicit Authority“ und „Clarity“. Dabei erklärt er, dass Rechte nur durch eine explizite Aktion des Anwenders vergeben werden dürfen und diese Aktion vom Benutzer auch als solche verständlich sein muss. Außerdem müssen dem Anwender die Auswirkungen für jede sicherheitsrelevante Aktion im voraus bekannt gemacht werden. Einfacher ausgedrückt muss dem Anwender klar sein: „Wer wie mit welchen Dingen interagieren kann“ [7].

#### 4.5 Probleme lokal lösen

Balfanz et al. [1] argumentieren, dass Anwendungssicherheit oft scheinbar universelle Lösungen erfordere, die gar nicht existierten. Stattdessen soll Sicherheit lokal geplant sein und damit auf eine Koordination mit einer „größeren Infrastruktur oder Organisation“ verzichtet werden. In ihrem Fall wurde diese Herangehensweise im Design ihrer „Enrollment Station“ umgesetzt, die eine Zertifizierungsstelle, WLAN-Zugangspunkt und gestenbasierte Registrierung in einem Gerät vereint (vgl. Abschnitt 3.3).

## 5 Fazit

Im Rahmen dieser Arbeit wurden Fallbeispiele für nutzbare Sicherheit aus unterschiedlichen Kategorien von Authentifizierungsverfahren betrachtet. Dabei wurden Unterschiede zwischen mehr und weniger benutzerfreundlichen Ansätzen herausgearbeitet und anschließend in verallgemeinerter Form als Vorlage

für die Entwicklung von Sicherheitslösungen präsentiert, die den Anforderungen an nutzbare Sicherheit entsprechen sollen.

Vor allem im Bereich der Passworrichtlinien konnte in den letzten Jahren eine deutliche Veränderung zu mehr Anwenderfreundlichkeit wahrgenommen werden. Dies ist wohl darauf zurückzuführen, dass unterschiedliche Studien und auch aus Angriffen veröffentlichte Passwortlisten deutlich zeigen, welche Konsequenzen Anwender aus regelmäßigen Passwortwechseln und restriktiven Kompositionsregeln gezogen haben. Trotzdem wird Anwendungssicherheit in vielen Fällen immer noch auf eine Art und Weise umgesetzt, die Benutzer in ihrer alltäglichen Interaktion und Arbeit einschränkt. Damit werden Anwender gezwungen eine Entscheidung zwischen besserer Sicherheit oder erhöhtem Arbeitsfluss zu treffen. Arbeiten wie die von Buschek et al. [5] zeigen dennoch die Vielzahl an Möglichkeiten um bestehende oder zukünftige Authentisierungsverfahren zu erweitern und zu verbessern und dabei gleichzeitig auf Nutzbarkeit und Anwenderfreundlichkeit zu achten.

---

#### Literatur

- 1 Dirk Balfanz, Glenn Durfee, Diana K Smetters, and Rebecca E Grinter. In search of usable security: Five lessons from the field. *IEEE Security & Privacy*, 2(5):19–24, 2004.
- 2 Salil P Banerjee and Damon L Woodard. Biometric authentication and identification using keystroke dynamics: A survey. *Journal of Pattern Recognition Research*, 7(1):116–139, 2012.
- 3 Christina Braz and Jean-Marc Robert. Security and usability: the case of the user authentication methods. In *Proceedings of the 18th Conference on l'Interaction Homme-Machine*, pages 199–203. ACM, 2006.
- 4 William E. Burr, Donna F. Dodson, Elaine M. Newton, Ray A. Perlner, W. Timothy Polk, Sarbari Gupta, and Emad A. Nabbus. NIST Special Publication 800-63-2 Electronic Authentication Guideline, August 2013. URL: [http://dx.doi.org/10.6028/NIST.SP.800-63-2\(June2018\)](http://dx.doi.org/10.6028/NIST.SP.800-63-2(June2018)).
- 5 Daniel Buschek, Alexander De Luca, and Florian Alt. Improving accuracy, applicability and usability of keystroke biometrics on mobile touchscreen devices. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1393–1402. ACM, 2015.
- 6 Paul A. Grassi, James L. Fenton, Elaine M. Newton, Ray A. Perlner, Andrew R. Regenscheid, William E. Burr, Justin P. Richer, Naomi B. Lefkovitz, Jamie M. Danker, Yee-Yin Choong, Kristen K. Greene, and Mary F. Theofanos. NIST Special Publication 800-63B Digital Identity Guidelines, June 2017. URL: [https://doi.org/10.6028/NIST.SP.800-63b\(June2018\)](https://doi.org/10.6028/NIST.SP.800-63b(June2018)).
- 7 Butler Lampson. Privacy and security usable security: how to get it. *Communications of the ACM*, 52(11):25–27, 2009.
- 8 M Angela Sasse and Ivan Flechais. Usable security: Why do we need it? how do we get it? O'Reilly, 2005.
- 9 Roland Schlöglhofer and Johannes Sametinger. Secure and usable authentication on mobile devices. In *Proceedings of the 10th International Conference on Advances in Mobile Computing & Multimedia*, pages 257–262. ACM, 2012.
- 10 Matt Weir, Sudhir Aggarwal, Michael Collins, and Henry Stern. Testing metrics for password creation policies by attacking large sets of revealed passwords. In *Proceedings of the 17th ACM conference on Computer and communications security*, pages 162–175. ACM, 2010.
- 11 Alma Whitten and JD Tygar. Usability of security: A case study. Technical report, Carnegie-Mellon University Pittsburgh, 1998.
- 12 Ka-Ping Yee. User interaction design for secure systems. In *International Conference on Information and Communications Security*, pages 278–290. Springer, 2002.

# Reduzierung der kognitiven Belastung durch natürliche Interaktionsprinzipien in VR-Anwendungen

Thomas B. M. Puls

Ludwig-Maximilians-Universität München, München, Deutschland  
t.puls@campus.lmu.de

---

## Zusammenfassung

Natürliche Interaktionsprinzipien wie Gesten- oder auch Sprachsteuerung treten vermehrt in virtuellen Realitätsanwendungen auf, was u.a. an der inzwischen relativ günstigen und genaueren Hardware und der Verbreitung von Sprachassistenten liegt. Dieser Überblick adressiert Usability-Verbesserungen in virtueller Realität (VR) durch den Einsatz von natürlichen (im Gegensatz zu gerätebasierten) Interaktionsprinzipien. Geräte müssen erlernt und verstanden werden. Dies führt zu kognitiver Belastung und kann die Präsenz in VR-Umgebungen negativ beeinflussen. Bereits erfolgte Untersuchungen in dem Bereich der natürlichen Interaktion in virtueller Realität lassen positive Tendenzen in der Verbesserung der Usability erkennen. Einen konkreten Beweis für die Reduzierung der kognitiven Belastung durch natürliche Interaktion zur Optimierung der Usability in VR-Anwendungen liefern aktuelle Forschungsergebnisse (noch) nicht.

**2012 ACM Computing Classification** Human-centered computing — Virtual reality

**Keywords and phrases** NUI, VR, Cognitive Load, Usability

## 1 Einleitung

Virtuelle Realität erlebt einen neuen Aufschwung, was den Einsatz in Arbeit und Unterhaltung angeht [19]. Neue kostengünstige, hochauflösende und für die Verbraucherinnen und Verbraucher verfügbare Head-Mounted-Displays (HMD) bieten ein stark immersives Erlebnis [19]. Wie bei jedem neuen Medium, welches ubiquitär zur Verfügung steht, sollte es leicht zu bedienen sein, also eine zufriedenstellende Usability (Gebrauchstauglichkeit, für eine genaue Spezifikation s.u. Kapitel 3.1) bieten. Es kann nicht mehr davon ausgegangen werden, dass sich die Zielgruppe ausschließlich auf Expertinnen und Experten beschränkt, welche bereits diverse Erfahrungen im Bereich virtuelle Realität



© Thomas B. M. Puls;

licensed under Creative Commons License CC-BY

Cite as: Thomas B. M. Puls. Reduzierung der kognitiven Belastung durch natürliche Interaktionsprinzipien in VR-Anwendungen. 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. 9:1–9:20.

aufweisen. Um eine Optimierung der Usability in virtuellen Umgebungen zu schaffen, sollte sich die Interaktion in virtuellen Umgebungen mit der Interaktion in der realen Welt decken (s.u. Kapitel 4.2) [34]. Gerätebasierte Interaktion, wie bei der HTC Vive oder den Oculus Touch Controllern, bietet zwar die Möglichkeit, innerhalb der virtuellen Umgebung mit Objekten zu interagieren, jedoch bringen Controller auch negative Eigenschaften mit sich, welche in diesem Überblick diskutiert werden. Natürliche Interaktion via Gesten- oder Sprachsteuerung ist natürlich und intuitiv [40]. Aktuelle Studien zeigen auf, dass natürliche Interaktion wie Sprache und Geste allgemein zur Verbesserung der User Experience (Nutzungserfahrung) beitragen können. Dabei ist User Experience als die *Wahrnehmung* und Reaktion einer Person definiert, basierend auf dem (ggf. antizipierten) Gebrauch eines Produkts, Systems oder Service [7]. Wie in diesem Überblick beschrieben, gibt es einen starken Zusammenhang zwischen der User Experience und der Präsenz (s.u. Kapitel 4) in virtuellen Umgebungen. Die Glaubwürdigkeit einer VR-Anwendung lässt sich durch natürliche Interaktionsprinzipien steigern [33]. Unnatürliche Interaktion (z.B. gerätebasiert) muss erst erlernt werden. Jede Art von bewusster kognitiver Aktivität (z.B. das Erlernen einer Interaktionsmetapher) beeinträchtigt die Arbeitsspeicherkapazität des Gehirns, was zu einer Überlast führen kann [36]. Diese Überlast kann sich negativ auf die Präsenz und dadurch negativ auf die User Experience einer VR-Anwendung auswirken.

Whitenton [38] spricht im Zusammenhang mit der Usability einer Website von der Minimierung der kognitiven Last zur Maximierung der Usability. Diese Aussage, erweitert um die Forderung nach natürlicher Interaktion in virtueller Realität, bildet die grundsätzliche Motivation für diesen Überblick: Es soll die These untersucht werden, ob die durch natürliche Interaktionsprinzipien erreichte Reduzierung der kognitiven Belastung zur Optimierung der Usability in VR-Anwendungen führt. Studien, in der Geste und Sprache als natürliche Interaktionsprinzipien in virtuellen Umgebungen eingesetzt wurden, zeigen positiven Tendenzen auf, was die Verbesserung der Usability anbelangt, allerdings wird sich zeigen, dass bislang noch kein hinreichender Beweis für die untersuchte These gefunden wurde.

Dieser Überblick ist wie folgt gegliedert: Kapitel 2 vermittelt grundlegende Prinzipien der Gesten- und Sprachsteuerung. Das darauf folgende Kapitel erklärt die Grundzüge der kognitiven Belastungstheorie und setzt diese in den Zusammenhang mit Usability. Kapitel 4 stellt den Zusammenhang mit der virtuellen Realität dar. Hierbei werden die Terminologien der Immersion und Präsenz veranschaulicht und Heuristiken präsentiert, die es erlauben, Usability in virtuellen Umgebungen zu messen. Des Weiteren werden natürliche Interaktionsprinzipien in virtueller Realität verdeutlicht und bereits durchgeführte Untersuchungen in diesem Kontext analysiert und diskutiert.

Abschließend folgt in Kapitel 5 eine Bewertung der aufgezeigten Ergebnisse und eine Zusammenfassung dieses Überblicks.

## 2 Natürliche Interaktionsprinzipien

Nach Zielke u.a. [40] können NUIs (Natural User Interfaces) dadurch definiert werden, dass Nutzerinnen und Nutzer mit Hilfe natürlicher Interaktionsmetaphern direkt mit dem System ohne zusätzliche Eingabegeräte wie Controller oder Maus interagieren. Die Interaktion mit dem System erfolgt über natürliches menschliches Verhalten und muss diesbezüglich nicht erst erlernt werden [6]. Natürliche Interaktionsprinzipien wie Spracheingabe, Gestensteuerung oder Multi-Touch sind intuitiv und die Interaktion mit der Oberfläche bleibt unsichtbar [40]. Diese Unsichtbarkeit ermöglicht es, durch intrinsische, natürliche und menschliche Aktionen direkt und effizient die Technologie ohne Zwischenschmittstellen zu bedienen. Nach Kaushik u. Jain [14] definieren sich NUIs aber nicht nur durch sensorischen Input wie Geste, Sprache und Berührungen, sondern können auch kontextbezogen Gesichter, Umgebung oder Emotionen und Beziehungen erkennen.

### 2.1 Gestensteuerung

Laut Kaushik u. Jain [14] besteht zwischenmenschliche Interaktion und Kommunikation zu einem Großteil aus Gesten. Bhuiyan u. Picking [2] zeigen auf, dass Kinder, schon lange bevor sie sprechen lernen, mit Hilfe von Gesten kommunizieren. Eine Geste ist eine nonverbale Kommunikation durch den Einsatz verschiedener Körperteile. Es gibt eine Vielzahl an unterschiedlichen Gesten, geprägt durch Kultur oder Kontext. Kommunikation über Gesten kann jedoch auch kontextbasiert isoliert stattfinden, z.B. die Gestikulation bei Gesprächen am Telefon. In Verbindung mit Objekten ermöglichen Gesten eine Vielzahl an Varianten, wie das Zeigen auf Objekte, das Berühren oder Bewegen von Objekten, die Veränderung einer Objektform, das Aktivieren von Objekten oder auch das Übergeben von Objekten an andere Mitmenschen [14]. Kaushik u. Jain unterteilen Gesten deshalb in drei verschiedene Gruppen, nämlich als semiotisch: der Kommunikation von aussagekräftiger Information, als ergotisch: der Veränderung der physischen Welt und Erschaffung von neuen Objekten, und als epistemisch: dem Erlernen der Umwelt durch taktile oder haptische Erkundung. Demzufolge können unsere natürlichen und intuitiven Körperbewegungen oder Gesten als Befehl verstanden werden, eine Benutzeroberfläche zu steuern, oder auch mit intelligenten Umgebungen zu kommunizieren [2]. Solche Benutzeroberflächen können in ihrer Komplexität stark variieren, von der Erkennung weniger symbolischer Gesten, bis hin zur Interpretation vollumfänglicher Gebärdensprache, oder auch von statischer

Handhaltung bis hin zu dynamischen Handbewegungen. Nach Kaushik u. Jain [14] bietet gestenbasiertes Interaktionsdesign mehrere Vorteile gegenüber herkömmlicher Eingabe durch Tastaturbefehle oder Mausklicks. Es bietet eine einfache und intuitive Benutzeroberfläche und damit mehr Raum und Freiheit in der Mensch-Computer-Interaktion. Außerdem werden die kognitiven und psychologischen Aspekte ideal ausgereizt, da die Erwartung erfüllt und die Interaktionsprinzipien intuitiv sind. Folglich wird die Interaktion zwischen Mensch und Computer natürlicher. Anwendungsfälle können hierbei Zeichensprache, Navigationssysteme, Medizin, Spiele-Anwendungen, erweiterte Realitäts- (Augmented Reality (AR)) oder virtuelle Realitäts-Anwendungen (Virtual Reality (VR)) sein [14].

## 2.2 Sprachsteuerung

Natürliche Spracherkennung (Natural language processing (NLP)) ist ein wichtiger Faktor in Bezug auf NUIs und viel Entwicklungsarbeit wurde in diesem Bereich schon geleistet. Als Ergebnis ist die jüngste Verbreitung von Sprachassistenten wie Amazon Echo, Siri, oder Google Home zu nennen [40]. Doch welche Eigenschaften besitzt Sprache überhaupt, und was bedeuten diese in Zusammenhang mit Benutzeroberflächen? Nach Schnelle-Walka und Döweling [28] weist das menschliche Ohr im Vergleich zum menschlichen Auge Defizite in der schnellen Auffassung von Informationen auf [29]. Zudem ist Sprache asymmetrisch, denn Menschen können schneller sprechen als schreiben, jedoch langsamer zuhören als lesen. Des Weiteren ist Sprache unsichtbar, und es ist deshalb essentiell, den Nutzerinnen und Nutzern verständlich zu machen, welche sprachliche Eingabe gemacht werden darf bzw. kann. Zusätzliche Probleme treten dadurch auf, dass Spracheingabe und Spracherkennung auch von einem Menschen nicht vollständig korrekt erfasst werden [29]. Außerdem muss der Möglichkeit Rechnung getragen werden, dass das gleiche Anliegen auf verschiedene Arten und Weisen ausgedrückt werden kann, aufgrund von unterschiedlichem Vokabular, aber auch durch unterschiedliche Kontexte.

Nach McGlashan u. Axling [20] gibt es drei technische Herausforderung an die Sprachinteraktion. Erstens muss gewährleistet werden, dass Sprache erkannt wird. Je größer das Vokabular und die Grammatik einer Sprache, desto höher ist die Chance auf Fehlinterpretation des gesprochenen Wortes im System. Es muss also ein Kompromiss bezüglich der Einschränkungen der Nutzenden gefunden werden, um dennoch eine leicht nutzbare Interaktion zu gewährleisten. Die zweite technische Voraussetzung ist das Sprachverständnis. Wie bereits beschrieben, kann es durch unterschiedliche Kontexte zu verschiedenen Sprachinterpretationen kommen. Das erschwert dem System zu erkennen, auf welches Objekt sich ein Befehl bezieht. Der dritte Punkt stützt sich auf die Interaktionsmetapher. Der Einsatz von Sprache ändert die Beziehung zwi-

schen Computer und Mensch. Sprache benötigt einen Dialogpartner [20]. Wie Kaushik u. Jain [14] aufzeigen, verlangt dieser Dialog zwischen Mensch und Computer, dass die Maschine verbalen als auch nonverbalen Output generieren muss, genau wie bei einer Angesicht-zu-Angesicht-Konversation bei Menschen. Um dies zu gewährleisten, bedarf es einer Art visueller Interpretation auf der Benutzeroberfläche, die es ermöglicht, auch nonverbale Kommunikation sowohl zu erkennen als auch selber zu erzeugen.

## 3 Kognitive Belastung

Im Gegensatz zu natürlichen Interaktionsmetaphern müssen andere, z.B. gerätebasierte Interaktionsprinzipien, erst erlernt werden. Dabei stellt sich die Frage, inwieweit das Erlernen die kognitiven Prozesse des Menschen beeinflusst. Das Grundprinzip der kognitiven Belastungstheorie (Cognitive Load Theory (CLT)) basiert nach Van Merriënboer und Sweller [36] auf der Vorstellung einer begrenzten Arbeitsgedächtniskapazität und einer großen Langzeitgedächtniskapazität. Die Arbeitsgedächtniskapazität bezieht sich auf die temporäre Aufbewahrung von Informationen, die z.B. durch kognitive Aufgaben wie Lesen oder Lernen erzeugt werden [1]. Jede Art von bewusster kognitiver Aktivität beeinträchtigt die Arbeitsspeicherkapazität des Gehirns. Folgt man der CLT, also der Einschränkung durch das *begrenzte* Arbeitsgedächtnis, kommt es dagegen zu keiner Überlast bei Lernenden, und ein gutes Lernergebnis wird ermöglicht [36]. Gemäß Hollender u.a. [9] teilt sich die CLT in drei verschiedene Arten von kognitiver Belastung auf. Die erste Art bezieht sich auf die intrinsische kognitive Belastung (intrinsic cognitive load), sprich dem Interaktivitätslevel des zu erlernenden Elements. Vokabellernen verursacht eine geringe intrinsische kognitive Last, da jedes Wort unabhängig gelernt werden kann, während das Erlernen von Sätzen in einer Fremdsprache ein Beispiel für eine hohe intrinsische kognitive Belastung ist. Die zweite Art ist die extrinsische kognitive Belastung (extraneous cognitive load), welche durch mangelhafte Aufbereitung und Präsentation von Lernmaterial verursacht werden kann. Als Beispiel ist hier die Integration von Informationen aus verschiedenen Informationsquellen zu nennen. Die Information aus der ersten Quelle muss im Arbeitsgedächtnis behalten und mit der Information der anderen Quelle zusammengebracht werden. Als Drittes ist die lernbezogene kognitive Belastung (germane cognitive load) zu nennen, welche aus aktiven Lernschema-Konstruktionsprozessen resultiert und somit für das Lernen von Vorteil ist. Diese drei Arten werden als additiv angesehen, d.h. sie beeinflussen sich gegenseitig. Wird die extrinsische kognitive Belastung bei Nutzerinnen und Nutzern reduziert, führt dies zu mehr Arbeitsplatzspeicherkapazität, welche für die lernbezogenen kognitiven Belastungen genutzt werden kann [9].

### 3.1 Auswirkungen von kognitiver Belastung auf die Usability

Kognitive Belastung steht im starken Zusammenhang mit dem Lernen. Doch wie wirkt sich kognitive Überlastung auf die Usability einer Benutzeroberfläche aus? Einfach gesagt, drückt der Begriff Usability die Gebrauchstauglichkeit aus, d.h. die Art und Weise, wie man mit einer Benutzeroberfläche interagiert. Eine genauere Spezifikation dieses Begriffs liefert die Internationale Organisation für Normung (International Organization for Standardization (ISO)). Hier wird Usability als der Umfang verstanden, in dem ein Produkt durch bestimmte Nutzerinnen und Nutzer genutzt wird, um bestimmte Ziele effektiv, effizient und zufrieden in einem bestimmten Nutzungskontext zu erreichen [11]. Man sollte jedoch beachten, dass sich das Usability-Level nur im Zusammenhang mit der spezifischen Aufgabe und der spezifischen Zielgruppe messen lässt. Untersucht man Usability-Richtlinien, stellt man fest, dass die Literatur diverse Ansätze präsentiert. Nielsen [22] z.B. definiert fünf Usability-Ziele, die ein System erreichen soll: Lernfähigkeit, Einprägsamkeit, Effizienz, niedrige Fehlerrate und Zufriedenheit. Lernfähigkeit bezieht sich auf die unerfahrenen Nutzerinnen und Nutzer eines Systems, die innerhalb kürzester Zeit einen gewissen geeigneten Kompetenzstand und Level erreichen sollen. Mit der Einprägsamkeit zielt Nielsen auf die Leichtigkeit ab, mit der Nutzerinnen und Nutzer ein System wieder benutzen können, nachdem sie es bereits erlernt haben. Effizienz zielt auf den Lernaufwand ab, der aufgebracht werden muss, um eine bestimmte Aufgabe zu bewerkstelligen. Fehler innerhalb eines Systems, welche durch inkorrekte Eingabe zustande kommen, müssen minimiert werden. Wenn solche Fehler auftreten, muss zumindest ein leichter Weg zum Ausgangszustand gewährleistet sein. Wie häufig und gern ein System benutzt wird, hängt von der Zufriedenheit ab, die umso wichtiger wird, je weiter Systeme außerhalb des Arbeitskontextes stehen [9].

Oviatt [23] bringt die kognitive Belastung in Zusammenhang mit der Usability von Benutzeroberflächen. Um nutzerzentrierte Oberflächen zu gestalten, muss die kognitive Belastung zur Steigerung der Leistung minimiert werden, indem auf überflüssige Ablenkung verzichtet wird. Dies gilt auch für ablenkende und zusätzliche Funktionen, die die Aufmerksamkeit der Nutzerinnen und Nutzer auf sich ziehen. Schmutz u.a. [27] untersuchten die kognitive Belastung bei der Interaktion mit Online-Shops. Ihre Studie ergab, dass durch eine hohe kognitive Belastung (gemessen durch den NASA-TLX-Index [8]) eine höhere Bewältigungszeit der gestellten Aufgabe (hier: das Suchen von Büchern), eine höhere Fehlerrate und eine generell schlechtere Zufriedenheit mit dem Online-Shop gemessen wurde.

### 3.2 Reduzierung der kognitiven Belastung

Setzt man nun diese Grundüberlegungen von Usability in den Zusammenhang mit kognitiver Belastung, muss man nach van Nimwegen u.a. [37] die Speicherlast der Nutzerinnen und Nutzer minimieren, um eine Gebrauchstauglichkeit der Systeme zu gewährleisten. In diesem Kontext spricht man von Off-Loading (Entlastung). Wie weiter oben beschrieben, muss das Arbeitsgedächtnis entlastet werden, um eine maximale Anzahl kognitiver Ressourcen zu gewährleisten, damit die spezifische Aufgabe erfolgreich erledigt werden kann. Informationen werden hierbei kontextbezogen dargestellt, indem inaktive Funktionen ausgeblendet oder deaktiviert werden und Hilfestellungen, in Form von interaktiven Assistenten, auf der Benutzeroberfläche angeboten werden. Dadurch wird die kognitive Belastung reduziert, indem die Anzahl an Auswahlmöglichkeiten eingeschränkt und permanentes Feedback zurückgespielt wird. Mayer und Moreno [18] präsentieren in ihrer Studie neun Wege, wie sich die kognitive Belastung im Zusammenhang mit multimedialem Lernen reduzieren lässt. Beim Off-Loading (Entlasten) wird der Überlastkanal von wesentlichen Verarbeitungsanforderungen frei gehalten, indem z.B. Wörter als erzählte Geschichte präsentiert werden. Das Segmentieren bezieht sich auf das Aufteilen von Wörtern und Bildern. Den Lernenden wird eine Zeitscheibe gewährt, beide Systeme gleichwertig zu erfassen. Wenn das Segmentieren nicht möglich ist, sollte das Pretraining (Vortraining) angeboten werden, welches Vorabinstruktionen bereitstellt, die die Bearbeitung beider Überlastkanäle ermöglicht. Des Weiteren wird beim Weeding (Aussortieren) überflüssige Information beseitigt. Dies ist jedoch nicht immer möglich. So zeigt das Signaling (Signalisieren) Hinweise auf, wie das bereitgestellte Material zu verstehen ist. Außerdem soll es z.B. vermieden werden, Lernenden eine Animation in einem Browserfenster zusammen mit einem konkurrierenden Bildschirmtext in einem anderen Browserfenster anzubieten. Wichtig ist es auch, auf redundante Inhalte zu verzichten. Als Letztes sollte beim Individualizing (Individualisierung) sicher gestellt werden, dass alle Lernenden die Fähigkeit besitzen, die aufgenommenen Informationen gleichwertig im Gedächtnis zu behalten, da es Unterschiede bei räumlich und nicht räumlich lernenden Personen gibt.

## 4 NUI in VR

Die vorherigen Kapitel gaben einen Überblick über natürliche Interaktionsprinzipien, sowie die Grundlagen von kognitiver Belastung, auch im Zusammenhang mit Usability. Die folgenden Abschnitte umfassen den Kontext der virtuellen Realität. Zuerst werden die beiden Grundbausteine von virtueller Realität, die Immersion und Präsenz erläutert, und anschließend in Bezug zur Usability in VR-Anwendungen gebracht. Danach wird der bisherige Stand des Einsatzes

von Gesten- und Sprachsteuerung in virtuellen Umgebungen beschrieben und abschließend bereits geleistete Studien zu diesem Thema kurz diskutiert.

#### 4.1 Immersion und Präsenz

Laut Slater [30] gibt es eine große Verwechslungsgefahr, wenn es um die Immersion in virtuellen Umgebungen geht. Immersion wird in vielen Fällen mit Präsenz (s.u.) gleichgesetzt, doch ist diese etwas anderes. Immersion quantifiziert die objektive und sensorische Genauigkeit einer VR-Anwendung [4, 30]. Mit anderen Worten hängt Immersion ausschließlich von der Genauigkeit der Rendering-Software, der Anzeige-Technologie und des Trackings (Bewegungsverfolgung) ab. Immersion ist messbar und abhängig von objektiven Faktoren wie Sichtfeld, Gesichtsfeld, Bildschirmgröße, Bildschirmauflösung, Stereoskopie, der Bewegungsverfolgung des Kopfes, einer realistischen Beleuchtung, der Bildrate und der Aktualisierungsrate.

Präsenz hingegen ist die menschliche Reaktion auf die Immersion [30]. Steuer [12] grenzt Präsenz von Technologien wie HMDs ab. Für ihn ist die Präsenz das zentrale Ziel der virtuellen Realität. Präsenz ist die eigene Erfahrung der physischen Umgebung. Sie bezieht sich jedoch nicht auf die Umgebung, wie sie in der physischen Welt existiert, sondern auf die Wahrnehmung dieser Welt, die durch automatische und kontrollierte mentale Prozesse vermittelt wird. Wahrnehmung bezieht sich in der virtuellen Welt auf zwei Umgebungen gleichzeitig, nämlich der physischen Umgebung, in der Personen tatsächlich existieren, und der virtuellen Umgebung, welche über ein Medium wie ein HMD vermittelt wird. Wenn man sich in dieser durch das Medium vermittelten virtuellen Realität gegenwärtiger fühlt als in der tatsächlichen physischen Welt, spricht Steuer [12] von Telepräsenz. Präsenz bezieht sich also nach ihm auf die natürliche (mentale) Wahrnehmung einer Umgebung und Telepräsenz auf die durch Kommunikationsmedien vermittelte Wahrnehmung einer Umgebung. Steuer [12] definiert eine virtuelle Realität als reale oder simulierte Umgebung, in welcher ein Wahrnehmender Telepräsenz erlebt.

Um die Begriffe *Immersion* und *Präsenz* (nach Steuer *Telepräsenz* [12]) zu veranschaulichen, vergleichen Slater u.a. [31, 30] diese beiden Terminologien mit der Farbenlehre. Eine Farbe kann man objektiv anhand ihrer Wellenlänge definieren, so wie Immersion anhand von gewissen Parametern auch definiert werden kann. Jedoch ist die Wahrnehmung und emotionale Reaktion auf eine Farbe eine subjektive und menschliche Erfahrung, wie die Präsenz in virtuellen Umgebungen. Wie beschrieben, ist Präsenz eine menschliche Reaktion auf Immersion [30]. Handelnde können eine unterschiedliche Wahrnehmung der virtuellen Umgebung im selben immersiven System haben. Die Autorinnen und Autoren [31] sprechen von gemeldeter Präsenz, also der unmittelbaren subjektiven Wahrnehmung innerhalb der virtuellen Realität. Für sie ist Interaktion

ein Teil dieser Wahrnehmung, besonders die Interaktion mit dem gesamten Körper. Die Anforderung an die virtuelle Realität, also so zu reagieren, als ob sie echt wäre, beinhaltet natürliche menschliche Interaktion. Lässt sich also die Präsenz in VR-Anwendungen durch natürliche Interaktion steigern?

#### 4.2 Usability in VR

Um schnelle und erfolgreiche Lösungen für Usability-Probleme in VR zu finden, definieren Sutcliffe u. Gault [34] 12 heuristische Auswertungskriterien. Dabei liegt ihr Fokus auf der unterschiedlichen Natur von virtuellen Umgebungen, insbesondere dem Verlangen nach intuitiver Interaktion und dem Gefühl nach Präsenz, um die Realität von VR-Anwendungen zu steigern. Die Interaktion in virtuellen Umgebungen sollte sich mit der Interaktion in der natürlichen Welt decken. Idealerweise ist es Nutzerinnen und Nutzern nicht bewusst, dass sie sich in der virtuellen Realität befinden. Das Verhalten der virtuellen Umgebung, sowie deren Objekte, sollte der Erwartung an die wirkliche Welt entsprechen. Das Agieren und Erforschen der virtuellen Umgebung muss so natürlich wie möglich sein und darf die natürliche Körperhaltung einer Person nicht einschränken. Zudem leidet ohne haptisches Feedback zwangsläufig die Natürlichkeit der VR-Anwendung. Um Motion-Sickness (Bewegungskrankheit) entgegen zu wirken, muss die Reaktionszeit zwischen Bewegung der agierenden Person im virtuellen Raum und der Aktualisierungszeit des VR-Systems so gering wie möglich sein. Die visuelle Präsentation der virtuellen Umgebung sollte den Erwartungen der Handelnden entsprechen. Die Interaktion mit Objekten sollte unmittelbar dargestellt werden und mit den Gesetzen der Physik in Einklang stehen. Zu jeden Zeitpunkt muss klar sein, wo sich die Nutzerinnen und Nutzer augenblicklich in der virtuellen Welt befinden und wie sie wieder zu ihrer Ausgangspositionen zurückkehren können. Die Eintritts- bzw. Austrittspunkte der virtuellen Welt sollen klar kommuniziert werden. Spezielle unnatürliche Interaktionsprinzipien aufgrund von Designkompromissen sollten den Nutzerinnen und Nutzern vor dem Eintritt in die virtuelle Realität erläutert werden. Aktive Objekte sollen hervorstechen und selbsterklärend sein. Verlangt die VR-Anwendung einen Wechsel von Nutzer- zu Systemsteuerung, muss diese kommuniziert und eine Konvention dafür angelegt werden. Die Präsenz muss so natürlich wie möglich sein [34].

Nach Sutcliffe u.a. [33] beziehen sich die Heuristiken einerseits auf die natürliche Interaktion in virtuellen Umgebungen, um den Erwartungen an eine reale Welt gerecht zu werden. Andererseits jedoch auch auf unmittelbares Feedback für ausgeführte Aktionen, Navigationsprinzipien sowie den Umgang mit aktiven Objekten. Diese Autoren fügen zu den oben genannten 12 Heuristiken noch weitere hinzu, um die Präsenz, Glaubwürdigkeit und Führung in virtuellen Anwendungen zu verbessern. Dies führt ihrer Meinung nach zu einer besseren

User Experience. Die Glaubwürdigkeit bzw. Echtheit von VR-Anwendungen kann durch natürliche Interaktion erhöht werden. Objektmanipulation und -steuerung, die sich auf die zu erledigende Aufgabe beziehen, müssen natürliche und selbsterklärende Interaktionsmetaphern nutzen.

### 4.3 NUI-Prinzipien in VR

Folgt man dem aktuellen Stand der Forschung, lässt sich feststellen, dass NUI-Prinzipien keine Seltenheit mehr in VR-Anwendungen sind. Jedoch ist das Darstellungssystem ein wichtiger Faktor. Wie weiter oben beschrieben, werden häufig HMDs verwendet, da sie kostengünstiger und weniger sperrig als CAVE-Systeme sind [16]. CAVE (Cave Automatic Virtual Environment, Vgl. Abbildung 1), ist ein begehbare Würfel, dessen Darstellungsbildschirme die Handelnden umringen. Während sich die Nutzerinnen und Nutzer innerhalb der Würfelgrenzen bewegen, wird die korrekte Perspektive und die Stereoprojektion der Umgebung dargestellt [5]. VR-Systeme wie die HTC Vive oder auch die Oculus Rift setzen auf HMDs in Kombination mit Controllern (Vive Controller, Oculus Touch Controller). Controller bringen jedoch die negativen Eigenschaften mit sich, dass sie eine oder beide Hände blockieren und immer vorhanden sein müssen. Insbesondere müssen die Interaktionsprinzipien erst einmal erlernt und verstanden werden [24]. Um diesen negativen Aspekten entgegenzutreten, wird immer häufiger Gestensteuerung als natürliches Interaktionsprinzip in VR-Anwendungen eingesetzt. Die hierfür benötigten Bewegungserkennungs-Kameras werden immer kostengünstiger. Zum Beispiel erfasst die Microsoft Kinect Kamera den dreidimensionalen Raum und den Körper einer Person ohne zusätzliches handbedientes Gerät. Sie erkennt Hände, sowie die Interaktion der Hände mit Objekten, und manipuliert die Objekte anhand der Handinteraktion [15]. Als weiteres Beispiel ist die Leap Motion Kamera zu nennen, welche Gestensteuerung durch Erfassen von Handverfolgungen ermöglicht. Diese beiden Geräte können den Einsatz von Controllern in VR-Anwendungen überflüssig machen und somit eine natürliche und intuitive Interaktion mit der Technologie gewährleisten [24]. Wie Pfeuffer u.a. [24] aufzeigen, wurden verschiedene Möglichkeiten der Gestensteuerung untersucht. Eine Möglichkeit ist die virtuelle Hand. Hierbei werden die realen Hände im virtuellen Raum abgebildet und damit eine intuitive Bedienmöglichkeit mit virtuellen Objekten ermöglicht. Jedoch wird man dabei auf die Reichweite der eigenen Hände im virtuellen Raum eingeschränkt. Um eine größere Reichweite zu ermöglichen, gibt es auch das Prinzip des virtuellen Zeigers. Dieser erlaubt durch Fingerzeig, weiter entfernte Objekte auszuwählen und diese dann zu bewegen und zu manipulieren etc. Ungenauigkeiten können hierbei jedoch durch Handzucken oder dem Heisenberg-Effekt [3] entstehen (Unsicherheit zwischen Zeige- und Auswahlgeste).



■ **Abbildung 1** Beispiel eines CAVE-Systems. Die Person wird von Wänden umgeben, auf denen dreidimensionale Bilder projiziert werden [17].

Um Sprachinteraktion in den Kontext der virtuellen Realität zu stellen, hat nach McGlashan u. Axling [20] Sprache drei entscheidende Vorteile gegenüber controllerbasierter Interaktion. Sprache bietet die Möglichkeit, Interaktionsbefehle zu äußern, während die Hände als auch die Augen freibleiben. Das Verändern oder Manipulieren von Objekten in virtuellen Umgebungen (Transport, Ansichtsänderung, das Erstellen oder Löschen von Objekten) kann stattfinden, ohne dass weitere Prozesse durchgeführt werden müssen. Dementsprechend können mehrere Aktionen gleichzeitig stattfinden. Dies ist von besonderem Vorteil, wenn bereits andere Aufgaben ausgeführt werden. Beim virtuellen (und realen) Autofahren werden z.B. die Hände und Augen gleichzeitig beschäftigt. Sprache kann dann benutzt werden, um das Radio, Navigationssystem oder andere Funktionen zu bedienen [20]. Ein weiterer Vorteil ist die Objektmanipulation außerhalb des sichtbaren Fokus. Der wichtigste Faktor ist jedoch, dass man mit der Sprache als Interaktionsmittel vertraut ist [20]. Es ist zu beachten, dass die Gesetze der natürlichen Welt sich oftmals nicht mit den entsprechenden Gesetzen der virtuellen Welt decken. Zum gegenwärtigen Zeitpunkt ist auch Sprache in der direkten Konversation mit einem Menschen anders zu verstehen als mit einer Maschine [20].

### 4.4 NUI-Interaktionen in VR-Anwendungen - einige Beispiele

Gestenbasierte Interaktion in VR-Anwendungen findet sich in zahlreichen Untersuchungen wieder. Die zunächst folgenden Studien sind positive Beispiele einer natürlichen Interaktion mittels Gesten in VR-Anwendungen. Trotz vielversprechender Ergebnisse beinhalten diese Studien auch Hindernisse und

Misserfolge. So entwickelten Tumkor u.a. [35] mit Hilfe zweier Microsoft Kinect Kameras ein Handfassungssystem, welches die Position und Orientierung der Hände ermittelt. In einer NutzerInnenstudie zeigten sie auf, dass einige Aufgaben mit Hilfe von Gestensteuerung tatsächlich einfacher zu bewältigen waren. Jedoch stellten sie fest, dass Gestensteuerung durch herkömmliche Maus- und Keyboard-Interaktionsphilosophien erschwert wurde. Ihrer Meinung nach sollten Anwendungen, die auf Gestensteuerung setzen, ausschließlich auf NUI-Prinzipien basieren, um Verwirrung zu vermeiden. Pfeuffer u.a. [24] haben die Kombination aus Blick- und Gestensteuerung in einer NutzerInnenstudie untersucht. Wie die Autoren beschreiben, ist Blicksteuerung kein natürliches Interaktionsprinzip, bringt jedoch den Vorteil, dass auch weit entfernte Objekte im virtuellen Raum punktgenau angesteuert werden können. Die Gestensteuerung vermeidet das Erlernen neuer Interaktionstechniken, da diese aus der realen Welt bekannt sind. Blicksteuerung wirkt jedoch der Natürlichkeit entgegen, da die natürliche Reichweite zum Objekt, die Armlänge, nicht gegeben ist. Ihre informelle NutzerInnenstudie erbrachte jedoch positives Feedback, was die Usability anging. Als hinderlich erwies sich noch die aktuelle Ungenauigkeit der Augen- sowie der Gestenerfassungssysteme. In einer weiteren Studie [13] interagierten Nutzerinnen und Nutzer, mittels Oculus Rift in Kombination mit der Leap Motion Kamera, mit medizinischen Bildern in einer virtuellen Umgebung. Durch wenige, einfache, intuitive und natürliche Einhand-Gesten wurden die 3D-Objekte präzise im virtuellen Raum gedreht und bewegt. Die Autoren zeigen auf, dass man auch anhand von relativ günstigen Systemen natürliche Interaktion bieten kann. MacAllister u.a. [16] setzten auf der Studie von Karolczak und Klepaczko [13] an und untersuchten die Integration von kostengünstiger Hardware in kommerzieller 3D Software. Die Autoren erstellten einen funktionsfähigen Prototyp, der auf Oculus Rift in Verbindung mit der Leap Motion Kamera setzt (Vgl. Abbildung 2). Anhand eines Oberflächentests stellten sie jedoch fest, dass es bei der Leap Motion Kamera zu Problemen in Genauigkeit und/oder Zuverlässigkeit kommen kann. Während des Tests war die Gestenerkennung unzureichend. Als Fazit führten die Autoren an, dass die Leap Motion Kamera zur damaligen Zeit nicht für eine integrierte Softwarelösung geeignet war.

In der derzeitigen Forschung wird die sprachbasierte Interaktion in VR noch wenig, wenn überhaupt, behandelt. Die Gründe könnten hierfür in den Herausforderungen der Genauigkeit der Spracherkennung, dem Verständnis des sprachlichen Inputs im Kontext der Mensch-Computer-Interaktion und der Entwicklung einer angemessenen Interaktionsmetapher liegen [20]. Pick u.a. [26] verglichen in einer Studie die Leistung und die User Experience von existierenden hierarchischen Tortenmenüs (pie-menu) mit Spracherkennungsmenüs. Dazu nutzten die Autoren die existierende Fabrikationsanwendung



■ **Abbildung 2** NutzerInnenstudie von MacAllister u.a. [16] in der die Testpersonen mittels Gesten, anhand von Leap Motion Bewegungsverfolgung, in der virtuellen Umgebung interagieren, welche über ein HMD (hier Oculus Rift) vermittelt wird.

"flapAssist"[25] (Vgl. Abbildung 3). In einer CAVE-artigen virtuellen Umgebung führten Nutzerinnen und Nutzer definierte Aufgaben durch. Die sprachbasierte Interaktion erfolgte durch eine Zeigegeste und ein einfaches Verb-Objekt Satzmodell. Anstelle des Objektname wurde der Sprachbefehl auf *this* oder *that* mit unterstützender Zeigegeste reduziert. Die Autoren untersuchten ihre drei Hypothesen, dass Sprache schneller ist als das Tortenmenü, dass Sprach- und Tortenmenüs sich nicht in der Fehlerquote unterscheiden und dass die User Experience bei der sprachbasierter Interaktion besser ist. Das Ergebnis belegte die erste Hypothese, das Ausführen der Aufgaben mit Sprache gelang erheblich schneller. Jedoch konnten die beiden anderen Thesen nicht belegt werden, da die Autoren keine signifikanten Unterschiede in der Fehlerquote als auch der User Experience feststellten [26]. Mirzaei u.a. [21] stellten in ihrer Studie zum Thema multimodaler Navigation und Interaktion fest, dass der Einsatz von natürlicher Sprachsteuerung und gestenbasierter Navigation weniger Motion-Sickness hervorbringt als gerätebasierte Navigation. Ebenso sei gestenbasierte Navigation effizienter als gerätebasierte Navigation. In ihrer Studie untersuchten sie die sogenannte Speak-to-VR (Spracheingabe in VR) Benutzeroberfläche. Hierbei entwickelten sie zweierlei Navigationsprinzipien. Die erste unterstützte natürliches Gehen in der virtuellen Umgebung, mit Hilfe der Xbox Kinect Kamera. Als Vergleichsmedium wurde der sogenannte *Flystick* eingesetzt, ein elektromagnetisches Gerät, das es erlaubt, sich in der künstlichen Umgebung zu bewegen. Als zusätzliches Interaktionsprinzip für die Objektmanipulation wurde natürliche Spracheingabe eingesetzt. Die Testergebnisse zeigten einen signifikant höheren Grad von Motion-Sickness und Orientierungslosigkeit bei der gerätebasierten Navigation mit Hilfe des

Flysticks, verglichen mit der natürlichen gestenbasierten Navigation [21].



■ **Abbildung 3** In der VR-Anwendung "flapAssist" führen Testpersonen Objektmanipulation durch Sprachsteuerung in Kombination mit Pointerzeiger aus [25].

## 5 Bewertung und Zusammenfassung

Dieser Überblick wurde durch die Aussage von Whitenton [38] motiviert, dass durch die Reduzierung der kognitiven Belastung die Usability maximiert werden kann. Diese Aussage wurde hier durch die Forderung nach natürlichen Interaktionsprinzipien in virtuellen Umgebungen erweitert. Nach aktuellem Forschungsstand kann solch eine grundlegende These jedoch nicht vollständig belegt werden. Einige positive Tendenzen lassen sich dennoch ableiten. Bei natürlichen Benutzeroberflächen erfolgt die Interaktion über natürliches menschliches Verhalten, wie Gesten-, oder Sprachsteuerung und muss deshalb nicht erst erlernt werden. Natürliche Benutzeroberflächen basieren auf einer positiven User Experience. Sie benutzen menschliche kreative Fähigkeiten, um die Nutzerinnen und Nutzer innerhalb kürzester Zeit auf ein ExpertInnenlevel zu bringen [39, 22]. Eine Anwendung sollte in Bezug auf die Usability leicht zu erlernen sein. Lernen hängt, wie beschrieben, stark von der kognitiven Belastung ab [36, 1, 9]. Wie Hollender u.a. [9] aufzeigen, folgt einer Reduzierung von extrinsischer kognitiver Last mehr Platz im Arbeitsspeicher und somit höheres Lernpotential. Reduziert man also den Überlastkanal durch intuitive und selbsterklärende Interaktionsmöglichkeiten, werden wesentliche Verarbeitungsanforderungen frei gehalten. Mayer und Moreno [18] sprechen in diesem Kontext von Off-Loading, um die kognitive Belastung der Nutzenden zu reduzieren. Des Weiteren wirken sich natürliche Gesten positiv auf die

Aufgabenbewältigung aus. Viele Menschen benutzen Gesten in ihrer Kommunikation, was die kognitiven Prozesse unterstützt [32]. In einer Studie wiesen Steed u.a. [32] nach, dass auch in der virtuellen Realität Körperbewegungen und Gesten eine wichtige Rolle spielen, um kognitive Prozesse besser zu erfassen. Hostetter u.a. [10] wiesen auf, dass Personen umso mehr Gesten benutzten, je komplexere Aufgaben sie beschreiben müssen, und zeigten, dass diese Gesten die kognitive Belastung reduzieren. Lässt sich also behaupten, dass man mit Hilfe von natürlichen Interaktionsmöglichkeiten die kognitive Last reduziert und sich somit mehr Platz für die Präsenz in virtuellen Umgebungen schafft? Präsenz ist die subjektive Wahrnehmung der virtuellen Umgebung [31]. Um nach Sutcliffe u.a. [33] die Präsenz und damit auch die User Experience in virtuellen Anwendungen zu steigern, kann die Glaubwürdigkeit bzw. Echtheit von VR-Anwendungen durch natürliche Interaktion erhöht werden. Um die Usability in VR-Anwendungen zu steigern, sollte sich die Interaktion mit der Interaktion in der realen Welt decken [34]. Hier unterscheiden sich die Forschungsergebnisse in bereits durchgeführten Untersuchungen. Zum einen zeigen sich positive Aspekte der gestenbasierten Interaktion, was die Usability gegenüber gerätebasierter Interaktion angeht [35, 24, 13, 16]. Jedoch ist der aktuelle Stand der Technik, was die Kameras für das Bewegungs-Tracking angeht, noch unzureichend. Vergleicht man die Interaktion mittels Spracheingabe, lässt sich nach den in diesem Überblick aufgezeigten Ergebnissen keine Verbesserung der Usability feststellen. Vielmehr ergaben die Studien von Pick u.a. [25] keinerlei Unterschiede in der User Experience bzgl. gerätebasierter Interaktion gegenüber sprachbasierter Interaktion. Dennoch erkannten die Autoren, dass die erteilten Aufgaben schneller erledigt wurden [25], während Mirzaei u.a. [21] eine Reduzierung der Motion-Sickness und eine bessere Orientierung im virtuellen Raum feststellten. Führt man nun diese Ergebnisse zusammen, lässt sich feststellen, dass durch natürliche Interaktion die kognitive Belastung reduziert werden kann. Zudem wird durch natürliche Interaktion die Usability der VR-Anwendungen verbessert. Einen direkten Beweis für die Ausgangsthese, dass eine Reduzierung der kognitiven Belastung durch natürliche Interaktionsprinzipien zur Optimierung der Usability in virtuellen Realitätsanwendungen führt, lässt sich allerdings (noch) nicht finden.

### Literatur

- 1 Alan Baddeley. Working memory. *Science*, 255(5044):556–559, 1992.
- 2 Moniruzzaman Bhuiyan and Rich Picking. Gesture-controlled user interfaces, what have we done and what's next? Technical report, Centre for Applied Internet Research (CAIR), Wrexham, UK, 2009. URL: [http://www.newi.ac.uk/computing/research/pubs/SEIN\\_BP.pdf](http://www.newi.ac.uk/computing/research/pubs/SEIN_BP.pdf).

- 3 Doug Bowman, Chadwick Wingrave, Joshua Campbell, and Vinh Ly. Using pinch gloves (tm) for both natural and abstract interaction techniques in virtual environments. 2001.
- 4 Doug A Bowman and Ryan P McMahan. Virtual reality: how much immersion is enough? *Computer*, 40(7), 2007.
- 5 Carolina Cruz-Neira, Daniel J. Sandin, and Thomas A. DeFanti. Surround-screen projection-based virtual reality: The design and implementation of the cave. In *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '93, pages 135–142, New York, NY, USA, 1993. ACM. URL: <http://doi.acm.org/10.1145/166117.166134>, doi:10.1145/166117.166134.
- 6 Christianne Falcao, Ana Catarina Lemos, and Marcelo Soares. Evaluation of natural user interface: A usability study based on the leap motion device. *Procedia Manufacturing*, Procedia Manufacturing, Procedia Manufacturing.
- 7 ISO FDis. 9241-210: 2009. ergonomics of human system interaction-part 210: Human-centered design for interactive systems (formerly known as 13407). *International Organization for Standardization (ISO)*. Switzerland, 2009.
- 8 Sandra G Hart and Lowell E Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, pages 139–183. Elsevier, 1988.
- 9 Nina Hollender, Cristian Hofmann, Michael Deneke, and Bernhard Schmitz. Integrating cognitive load theory and concepts of human–computer interaction. *Computers in Human Behavior*, 26(6):1278 – 1288, 2010. Online Interactivity: Role of Technology in Behavior Change. URL: <http://www.sciencedirect.com/science/article/pii/S0747563210001718>, doi:<https://doi.org/10.1016/j.chb.2010.05.031>.
- 10 Autumn B Hostetter, Martha W Alibali, and Sotaro Kita. I see it in my hands' eye: Representational gestures reflect conceptual demands. *Language and Cognitive Processes*, 22(3):313–336, 2007.
- 11 ISO. ISO 9241-11:1998 Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability. Technical report, International Organization for Standardization, 1998. URL: <http://www.userfocus.co.uk/resources/iso9241/part11.html>.
- 12 Steuer Jonathan. Defining virtual reality: Dimensions determining telepresence. *Journal of Communication*, 42(4):73–93, 1992. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1460-2466.1992.tb00812.x>, arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1460-2466.1992.tb00812.x>, doi:10.1111/j.1460-2466.1992.tb00812.x.
- 13 Krzysztof Karolczak and Artur Klepaczko. A stereoscopic viewer of the results of vessel segmentation in 3d magnetic resonance angiography images. In *Proceedings of the International Conference on Signal Processing: Al-*

- gorithms, Architectures, Arrangements, and Applications (SPA2014)*, pages 160–165, 2014.
- 14 Dr Kaushik, Rashmi Jain, et al. Natural user interfaces: Trend in virtual interaction. *arXiv preprint arXiv:1405.0101*, 2014.
- 15 Dr Manju Kaushik and Rashmi Jain. Gesture based interaction nui: an overview. *arXiv preprint arXiv:1404.2364*, 2014.
- 16 Anastacia MacAllister, Tsung-Pin Yeh, and Eliot Winer. Implementing native support for oculus and leap motion in a commercial engineering visualization and analysis platform. *Electronic Imaging*, 2016(4):1–11, 2016.
- 17 Siddhesh Manjrekar, Shubhrika Sandilya, Deesha Bhosale, Sravanthi Kanchi, Adwait Pitkar, and Mayur Gondhalekar. Cave: An emerging immersive technology—a review. In *Computer Modelling and Simulation (UKSim), 2014 UKSim-AMSS 16th International Conference on*, pages 131–136. IEEE, 2014.
- 18 Richard E Mayer and Roxana Moreno. Nine ways to reduce cognitive load in multimedia learning. *Educational psychologist*, 38(1):43–52, 2003.
- 19 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*, pages 2143–2152. ACM, 2015.
- 20 Scott McGlashan and Tomas Axling. A speech interface to virtual environments. In *Proc., International Workshop on Speech and Computers*, 1996.
- 21 M. A. Mirzaei, J.-R. Chardonnet, F. Mérienne, and A. Genty. Navigation and interaction in a real-scale digital mock-up using natural language and user gesture. In *Proceedings of the 2014 Virtual Reality International Conference*, VRIC '14, pages 28:1–28:4, New York, NY, USA, 2014. ACM. URL: <http://doi.acm.org/10.1145/2617841.2620716>, doi:10.1145/2617841.2620716.
- 22 Jakob Nielsen. *Usability engineering*. Elsevier, 1994.
- 23 Sharon Oviatt. Human-centered design meets cognitive load theory: designing interfaces that help people think. In *Proceedings of the 14th ACM international conference on Multimedia*, pages 871–880. ACM, 2006.
- 24 Ken Pfeuffer, Benedikt Mayer, Diako Mardanbegi, and Hans Gellersen. Gaze + pinch interaction in virtual reality. In *Proceedings of the 5th Symposium on Spatial User Interaction*, SUI '17, pages 99–108, New York, NY, USA, 2017. ACM. URL: <http://doi.acm.org/10.1145/3131277.3132180>, doi:10.1145/3131277.3132180.
- 25 S Pick, S Gebhardt, K Kreisköther, R Reinhard, H Voet, C Büscher, and T Kuhlen. Advanced virtual reality and visualization support for factory layout planning. *Proc. of „Entwerfen Entwickeln Erleben—EEE2014*, pages 187–198, 2014.
- 26 Sebastian Pick, Andrew S Puika, and Torsten W Kuhlen. Comparison of a speech-based and a pie-menu-based interaction metaphor for application

- control. In *Pick, Sebastian and Puika, Andrew S and Kuhlen, Torsten W*, pages 381–382. IEEE, 2017.
- 27 Peter Schmutz, Silvia Heinz, Yolanda Métrailler, and Klaus Opwis. Cognitive load in ecommerce applications: measurement and effects on user satisfaction. *Advances in Human-Computer Interaction*, 2009:3, 2009.
  - 28 Dirk Schnelle-Walka and Sebastian Döweling. Speech augmented multitouch interaction patterns. In *Proceedings of the 16th European Conference on Pattern Languages of Programs*, page 8. ACM, 2012.
  - 29 Sandra Schuster. *Voice Control um jeden Preis? Theoretische und praktische Grundlagen für erfolgreiche Sprachsteuerungs-Angebote aus User Experience-Sicht*. German UPA eV, 2013.
  - 30 Mel Slater. A note on presence terminology. *Presence connect*, 3(3):1–5, 2003.
  - 31 Mel Slater, Beau Lotto, Maria Marta Arnold, and Maria V Sanchez-Vives. How we experience immersive virtual environments: the concept of presence and its measurement. *Anuario de Psicología*, 40(2), 2009.
  - 32 Anthony Steed, Ye Pan, Fiona Zisch, and William Steptoe. The impact of a self-avatar on cognitive load in immersive virtual reality. In *Virtual Reality (VR), 2016 IEEE*, pages 67–76. IEEE, 2016.
  - 33 A. G. Sutcliffe, C. Poullis, A. Gregoriades, I. Katsouri, A. Tzanavari, and K. Herakleous. Reflecting on the design process for virtual reality applications. *International Journal of Human-Computer Interaction*, 0(0):1–12, 2018. URL: <https://doi.org/10.1080/10447318.2018.1443898>, arXiv:<https://doi.org/10.1080/10447318.2018.1443898>, doi:10.1080/10447318.2018.1443898.
  - 34 Alistair Sutcliffe and Brian Gault. Heuristic evaluation of virtual reality applications. *Interacting with Computers*, 16(4):831–849, 2004. URL: <http://dx.doi.org/10.1016/j.intcom.2004.05.001>, arXiv:[/oup/backfile/content\\_public/journal/iwc/16/4/10.1016\\_j.intcom.2004.05.001/3/iwc16-0831.pdf](http://dx.doi.org/10.1016/j.intcom.2004.05.001), doi:10.1016/j.intcom.2004.05.001.
  - 35 Serdar Tumkor, Sven K Esche, and Constantin Chassapis. Hand gestures in cad systems. In *ASME 2013 International Mechanical Engineering Congress and Exposition*, pages 1–9. American Society of Mechanical Engineers, 2013.
  - 36 Jeroen JG Van Merriënboer and John Sweller. Cognitive load theory and complex learning: Recent developments and future directions. *Educational psychology review*, 17(2):147–177, 2005.
  - 37 Christof van Nimwegen, Herre van Oostendorp, Daniel Burgos, and Rob Koper. Does an interface with less assistance provoke more thoughtful behavior? In *Proceedings of the 7th International Conference on Learning Sciences, ICLS '06*, pages 785–791. International Society of the Learning Sciences, 2006. URL: <http://dl.acm.org/citation.cfm?id=1150034.1150148>.
  - 38 Kathryn Whitenton. Minimize cognitive load to maximize usability. *Nielsen Norman Group*, 2013.
  - 39 Daniel Wigdor and Dennis Wixon. *Brave NUI world: designing natural user interfaces for touch and gesture*. Elsevier, 2011.
  - 40 Marjorie A Zielke, Djakhangir Zakhidov, Gary Hardee, Leonard Evans, Sean Lenox, Nick Orr, Dylan Fino, and Gautham Mathialagan. Developing virtual patients with vr/ar for a natural user interface in medical teaching. In *Serious Games and Applications for Health (SeGAH), 2017 IEEE 5th International Conference on*, pages 1–8. IEEE, 2017.

# "Ich sehe mich, also bin ich"

## Selbstwahrnehmung und Avatardesign in Virtueller Realität

Nadine Sarah Schüler

Ludwig-Maximilians-Universität München, München, Deutschland  
N.Schueler@campus.lmu.de

---

### Zusammenfassung

Welchen Einfluss hat mein digitales Ich auf mich und meine Umgebung in einer Welt in der man alles sein kann? Im Laufe dieser Arbeit werden einige der Aspekte beleuchtet, die einen Avatar in der Virtuellen Realität ausmachen und welche Einflüsse eine Veränderung dieser auf einen selbst als Nutzer haben kann. Vom Aussehen, über Stil und Bewegungen bis hin zu den eigenen unbewussten Assoziationen haben fast alle Faktoren der Avatargestaltung Einfluss auf die Präsenz und das damit verbundene Verhalten einer Person in der Virtuellen Realität. Um dies deutlich zu machen werden im folgenden einige beispielhafte Experimente vorgestellt, bei denen uns unser Gehirn auf die eine oder andere Art verblüffen wird.

**2012 ACM Computing Classification** CSS – Human-Centered Computing – Human Computer Interaction (HCI) – Interaction Paradigms – Virtual Reality

**Keywords and phrases** Virtuelle Realität, Selbstwahrnehmung, Avatar, Immersion, Präsenz.

## 1 Willkommen in einer neuen Welt

Schon lange leben wir nicht mehr nur in einer Welt. Unsere Realität wurde in den letzten zehn bis 20 Jahren durch unzählige Soziale Netzwerke, Videospiele und andere Plattformen im Internet erweitert und auch wenn dies keine „Welten“ im physischen Sinne sind, so entwickeln Menschen doch oft schnell ein gewisses Gefühl der Vertrautheit und Freiheit, wenn sie mit Hilfe von Technik „abtauchen“.

In diesem Paper geht es um die Grundlage des Eintauchens in fremde, virtuelle Welten: Den eigenen Avatar. Ein Avatar gibt uns in der virtuellen Umgebung eine Gestalt und Interaktionsmöglichkeiten. Wir identifizieren uns mit ihm und genau wie mit unserem eigenen Körper wollen wir ihn vor Schaden



© Nadine Sarah Schüler;

licensed under Creative Commons License CC-BY

Cite as: Nadine Sarah Schüler. "Ich sehe mich, also bin ich"

Selbstwahrnehmung und Avatardesign in Virtueller Realität. 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. 10:1–10:18.

und unangenehmen Situationen beschützen. Insbesondere in Virtueller Realität, die eine wesentlich realistischere Illusion einer künstlichen Welt erzeugt spielt dabei die Gestaltung eines Avatars eine zentrale und oft unterschätzte Rolle. Auf Faktoren, wie das Aussehen, die Größe und Form sowie weitere Eigenschaften wird im Laufe dieses Papers eingegangen.

## 1.1 Virtuelle Realität

Der Begriff der Virtuellen Realität, so wie wir ihn heute kennen, ist bereits seit den 1960er Jahren bekannt [2]. Damals wurden mithilfe der frühen Computern und gigantischen Apparaturen das erste virtuelle Objekt in das reale Bild gelegt. Die sogenannte erweiterte Realität (engl.: Augmented Reality) ist geboren. Es folgen in den folgenden Jahren eine ständige Entwicklung der Technik. Insbesondere die steigende Rechenleistung und das immer kompakter werden der technischen Bauteile sorgen für eine „Virtuelle Realität zum Mitnehmen“. Waren bei den ersten Versuchen noch an der Decke montierte, schwere Gebilde aus Metall nötig, so kann man heute die Virtuelle Realität mit einem Stück Pappkarton und seinem Mobiltelefon erleben, wie zum Beispiel Google-Cardboard [3] zeigt.

Für VR ist wichtig, dass es eine computergenerierte, simulierte Umgebung ist, die auf den Benutzer reagiert und dessen Bewegungen und Aktionen umsetzen kann. Dabei ist es oft das Ziel eine möglichst glaubhafte Illusion für den Benutzer zu erzeugen, sodass dieser sich als Teil der simulierten Welt fühlt.

## 2 Bin da, wer noch – Präsenz und Immersion

Damit der Benutzer so gut es geht in diese simulierte Welt eintauchen kann, wurden in den letzten Jahren erstaunliche Fortschritte in den Bereichen der Hardware gemacht. Displays werden immer hoch-auflösender, die Steuerung wird intuitiver und die Welten werden durch komplexe Algorithmen abwechslungsreicher und natürlicher. Aber auch einige Entdeckungen in der Psychologie und der Verhaltensforschung sorgen für ein immer stärkeres Gefühl wirklich in diese virtuelle Realität einzutauchen. Eine gute Immersion beschreibt also eine sehr gute Art des „Eintauchens“ in die virtuelle Welt, zum Beispiel durch realistisch klingenden Ton von allen Seiten oder ein scharfes Display. Der Grad der Immersion hängt also davon ab, wie viele Faktoren und äußere Einflüsse der realen Welt mit dem Benutzer in die virtuelle Welt übertragen werden können. Je mehr Eindrücke der Nutzer von der virtuellen Welt, anstatt der realen empfängt, umso höher die Immersion.

Ist die Immersion gelungen und der Benutzer fühlt sich als Teil der neuen Welt, spricht man von Präsenz [4]. Die Präsenz misst sich insbesondere in der

VR vor allem durch Interaktionsmöglichkeiten mit der Umgebung. In einer Welt präsent zu sein bedeutet auch, dass die Welt auf einen reagiert. Sich selber oder Gegenstände bewegen zu können, oder mit anderen Individuen aus dieser Welt interagieren zu können sind zentrale Bestandteile für das Schaffen von Präsenz.

## 3 Das neue Ich – Der Avatar

Oft konzentriert man sich beim Erschaffen neuer Welten auf das möglichst realistische Nachbauen der Umgebung des Benutzers, mit zum Beispiel simuliertem Wind und aufwendigsten Animationen, und ganz sicher tragen diese auch sehr zum Erlebnis in der virtuellen Realität bei, aber mindestens genau so wichtig ist etwas, das man eigentlich gar nicht direkt sieht: Sich selbst.

Die eigene Spielfigur, die man durch die Welt steuert hat erheblichen Einfluss auf das Wohlbefinden des Nutzers und damit auf die Präsenz. Ein so-genannter Avatar muss her. Wer dabei jetzt zuerst am James Camerons Fantasyfilm [1] mit den blauen Kreaturen aus dem Jahr 2009 denkt, liegt gar nicht mal so falsch. Ein Avatar ist eine steuerbare, virtuelle Kunstfigur, die den Benutzer in der jeweiligen Welt repräsentiert [10], ganz ähnlich also zu den Geschehnissen im oben genannten Film, in dem ein Mensch einen künstlich gezüchteten Körper einer Alienrasse „fernsteuert“. Ganz so weit ist der heutige Stand der Technik dann allerdings doch noch nicht und im Bereich der VR beschreibt der Begriff ein meist vollständig dreidimensional modelliertes Individuum, das den Nutzer oder Spieler in der jeweiligen Welt repräsentiert.

Damit ein Avatar in der virtuellen Umgebung interagieren kann sind einige grundlegende Funktionen nötig. Vieles davon erscheint im ersten Moment offensichtlich, aber in den folgenden Abschnitten wird deutlich werden, welchen gewaltigen Einfluss auf die Interaktionen mit der Umgebung diese Faktoren haben. Es werden nun mehrere Beispiele präsentiert, bei denen bestimmte Eigenschaften und Merkmale der Avatare verändert wurden und was dies zur Folge hatte.

### 3.1 Anderes Aussehen

Das erste, was einem bei dem Wort Avatargestaltung einfällt ist wohl das Aussehen. Bereits zu Beginn der Entwicklung von Virtueller Realität gab es Experimente zum dem Thema, ob es einen Unterschied macht, wie ein Avatar aussieht. Schnell wurde festgestellt, dass unser Gehirn eine erstaunliche Fähigkeit besitzt, fremde Körper als den eigenen anzusehen. Der so-genannte „Proteus Effekt“ tritt in Kraft [17]. Der nach dem griechischen Gott der Verwandlung benannte Effekt beschreibt das Phänomen, dass sich das Verhalten

einer Person in einer virtuellen Welt mit den Eigenschaften und Charakteristiken seines Avatars verändert. Diese Veränderungen basieren in der Regel auf Erwartungen und Vorurteilen, wie sich ein Individuum mit den verwendeten Charakteristiken benimmt und agiert. Weil der Benutzer denkt, dass andere in der gleichen virtuellen Welt diese Erwartungen ebenfalls an ihn haben, ändert sich tatsächlich das eigene Verhalten oft unterbewusst. Ganz allgemein ist zum Beispiel mittlerweile bekannt, dass attraktive, große Avatare eher ein positiv gestimmtes, professionelles Auftreten fördern [8].

Aber nicht nur das eigene Auftreten, welches ja sowieso stimmungsabhängig ist, wird verändert, auch Motorik sowie die Sicht, die wir auf andere Individuen oder sogar Gegenstände haben, kann beeinflusst werden.

### 3.1.1 „Das sind nicht meine Hände“ – Akzeptanz

Unsere Hände sind das wohl wichtigste Interaktionswerkzeug, das wir besitzen. Umso wichtiger ist ihre Darstellung in virtueller Realität. Wenn wir unsere Hände nicht als unsere eigenen wahrnehmen, stellt dies eine erhebliche Verletzung der Präsenz, und damit dem allgemeinen Wohlbefinden in einer künstlichen Welt dar. In dem Experiment der Universität Stuttgart aus dem Jahre 2017 [15] wurde genau dieses Problem untersucht. Es wurden insgesamt sechs verschiedene Handmodelle entwickelt und an 28 Studienteilnehmern getestet, ob es signifikante Unterschiede in der Präsenz oder der Akzeptanz der Hände als die eigenen gab. Die Modelle lassen sich in zwei Gruppen einteilen, realistische Hände und künstliche Modelle, siehe Abbildung 1.

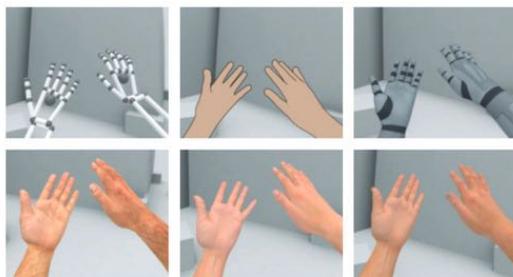


Figure 2. 1<sup>st</sup> row: screenshots of non-human hands: abstract, cartoon, robot; 2<sup>nd</sup> row: human hands: male, female, androgynous

■ **Abbildung 1** Die verschiedenen Handmodelle. Oben die künstlichen Modelle als abstrakte Skelettform, Zeichentrick und Roboter, unten die realistischen Modelle mit Männer-, Frauen- und Geschlechtslosen Händen (aus Paper [15] entnommen)

Die drei realistischen Handmodelle stellen behaarte, kräftige Männerhände dar, ein Paar gepflegte Frauenhände und ein möglichst ungeschlechtliches Paar,

das durch die gleichmäßige Vermischung der anderen beiden Modelle generiert wurde. Die drei künstlichen Handpaare sind ein abstraktes Modell, das an ein Skelett erinnert, ein Paar Zeichentrick Hände, ohne sichtbare Kanten und Textur sowie einer schwarzer Umrandung und zu Letzt ein paar Roboter Hände. Alle 28 Testteilnehmer, zur Hälfte männlich, zur Hälfte weiblich, wurden nacheinander virtuell mit allen Händen ausgestattet und sollten drei einfache Aufgaben in virtueller Realität lösen. Dabei wurden Fragebögen ausgefüllt und sich nach Wohlbefinden und den Gedanken der Teilnehmer erkundigt. Dabei ließen sich die Erfahrungen grob in drei große Kategorien einteilen [15]:

1. Abweichung von einer menschlichen Erscheinung
2. Abweichung vom eigenen Geschlecht
3. Abweichung vom eigenen Körper

Kategorie 1 wurde offensichtlich hauptsächlich bei den künstlichen Händen bemängelt. Vergleiche zu Prothesen wurden gemacht und teilweise trat sogar eine Art Orientierungslosigkeit auf, bei dem die Teilnehmer nicht mehr wussten, wie sie ihre Hände halten sollten. Interessanterweise kamen Frauen mit dieser Art der Darstellung wesentlich besser zurecht als Männer, wenn sie die Modelle als Kostüme oder eine Art Handschuh betrachteten. Kategorie 2 betrifft ebenfalls eher den weiblichen Teil der Studienteilnehmer. Während Männer im großen und ganzen Frauenhände zwar als ungewöhnlich, aber nicht unangenehm empfanden, stießen die behaarten Männerhände auf starke Ablehnung und zu Weilen sogar Ekel bei einigen weiblichen Testern. Die wohl interessanteste Kategorie ist die dritte, bei der bemängelt wurde, dass das Aussehen oder die Bewegung der Hände von den eigenen, natürlichen Händen abwich. Frauen mit Frauenhänden zum Beispiel waren irritiert, dass die Computerhände nicht ihre eigenen Eigenschaften, wie zum Beispiel Sommersprossen, widerspiegeln. Männer kritisierten oft, dass die Bewegungen der realistisch modellierten Hände viel stärker von ihren eigenen Abweichen würde, obwohl dies durch den gleichen Versuchsaufbau nicht der Fall war.

Ganz allgemein lassen sich aus diesem Experiment also die folgenden Schlüsse ziehen: Insbesondere bei Frauen ist die Akzeptanz der virtuellen Hände und damit das Wohlbefinden im eigenen Avatar vom Aussehen abhängig. Tatsächlich zeigte dieses Experiment, dass diese sich mit den künstlichen Händen noch am wohlsten fühlten, so lange diese nicht zu abstrakt gehalten waren. Die realistischen Modelle führten bei Frauen im Falle der Männerhände eher zu Abneigung, als auch zu Verwirrung im Falle der anderen beiden Modelle. Bei Männern war es genau andersherum. Mit den natürlichen Händen jegliches Geschlechts wurde das höchste Präsenzgefühl gemessen. Um also eine hohe Präsenz zu erzeugen kommt es bei der Gestaltung des eigenen Avatars tatsächlich primär darauf an, ob wir unseren Körper akzeptieren können und auch rein äußerlich mögen. Geschlechtertausch oder zu abstrakte Darstellungen



■ **Abbildung 2** Links: Der virtuelle Raum mit den weißen Händen als Grundlage (A) und den beiden verschiedenen Avataren. Rechts: Der Versuchsaufbau und Komponenten des Experiments (aus Paper [11] entnommen)

vermindern unsere Identifikationsfähigkeit und damit unser Wohlbefinden in der virtuellen Welt.

### 3.1.2 Das Trommel-Experiment – Motorik und Erwartungen

Das nächste hier vorgestellte Experiment führt die oben beschriebene Idee noch weiter. Hier wurde sich die Frage gestellt, ob der Avatar nicht nur direkten Einfluss auf das Wohlbefinden, sondern sogar Einfluss auf die eigenen Fähigkeiten bzw. die Hemmschwelle diese zu zeigen hat. Im Paper von Kiltner et al. [11] wurden Teilnehmer gebeten, in einem Anzug, der den gesamten Oberkörper in eine virtuelle Welt überträgt zu musizieren, siehe Abbildung 2 (rechts). Der Versuchsaufbau bestand dabei aus einer echten afrikanischen Djembe Trommel, dem VR Kopfteil, welches ebenfalls die Bewegungen des Kopfes misst und den Anzug mit Markierungen um die Bewegungen des Oberkörpers in die virtuelle Welt übertragen zu können. Das Experiment wurde mit 36 Teilnehmern durchgeführt, die alle kaukasischer Herkunft sind. Nachdem allen eine Einführung in das Trommeln und in virtuelle Realität gegeben wurde, begann das Experiment. Zu Beginn fanden sich die Teilnehmer in einem virtuellen Raum wieder, der außer ihnen selbst, einer virtuellen Trommel und dem Stuhl auf dem sie saßen, einen weiteren asiatisch aussehenden jungen Mann zeigte, der ebenfalls auf einer Trommel einen einfachen Grundtakt spielte. Des weiteren stand ein großer Spiegel in dem virtuellen Raum, so dass der Nutzer sich selbst und seinen Mitspieler darin betrachten konnte, siehe Abbildung 2 (links).

Die Aufgabe war zunächst sich mit der Umgebung vertraut zu machen und

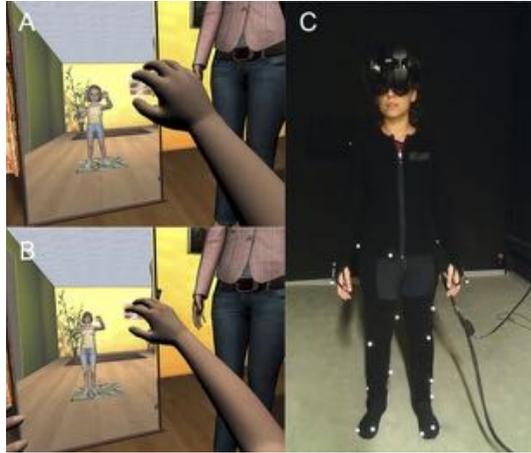


■ **Abbildung 3** Die drei verschiedenen Avatare beim Trommeln aus der Perspektive eines Studienteilnehmers (aus Paper [11] entnommen)

frei drauf los zu trommeln. Die Teilnehmer sahen dabei nur ein paar weiße Hände ohne Textur, die in der Luft schwebten. Ein Körper fehlte. Vier Minuten lang trommelten die Teilnehmer nun frei ohne Körper zu dem Rhythmus, den sie hörten und der in der virtuellen Welt von dem Mitspieler getrommelt wurde. Die Bewegungen, die dabei gemacht wurden dienten im Folgenden als Grundlage um gegebenenfalls auftretende Veränderungen messen zu können. Nach dieser Zeit wurden bei allen Teilnehmern die schwebenden Hände durch einen Avatar ersetzt. Bei der Hälfte der Teilnehmer war dies ein dunkelhäutiger, junger Mann, augenscheinlich afrikanischer Abstammung, der ungezwungen in Jeans und T-Shirt gekleidet war. Die andere Hälfte der Teilnehmer bekam einen hellhäutigen, jungen Mann im Anzug und Krawatte als Avatar zugeteilt, siehe Abbildung 3.

Das Interessante hierbei war nun die Veränderung im Rhythmusgefühl und in der Ausstrahlung der Teilnehmer. Während die Gruppe mit den formal gekleideten Avataren im Vergleich zu den weißen Händen sogar eher noch zurückhaltender trommelten, lies sich bei den Teilnehmern die den dunkelhäutigen Avatar steuerten eine größere Bereitschaft zu Improvisieren feststellen, sie spielten abwechslungsreicher und mit ausholenderen Bewegungen. Je stärker sich dabei eine einzelne Person mit dem virtuellen dunkelhäutigen, ungezwungen gekleideten Körper identifizierte, umso größer der Kontrast zu den eigenen vorher aufgezeichneten Vergleichswerten, während bei dem formell gekleideten, hellhäutigen Avatar genau die gegenteilige Tendenz zu beobachten war.

Das Paper erklärt dieses Phänomen mit der bereits kurz angesprochenen Erwartungshaltung und den Vorurteilen, die wir Menschen haben. Wir erwarten eine bestimmte Reaktion von bestimmten Akteuren in bestimmten Situationen. Die Studienteilnehmer im dunkelhäutigen Avatar gaben an, sich der Situation angemessen repräsentiert zu fühlen, weil wir afrikanische Trommelmusik automatisch mit Menschen dieses Kontinents verbinden. Die formal angezogenen Teilnehmer gaben zwar ein gleiches Maß an Präsenz in der virtuellen Welt an, waren beim Trommeln aber zurückhaltender und fühlten sich unpassend in der gegebenen Situation. Mehr noch, die Studienteilnehmer gaben an, dass sie das Gefühl hatten die virtuelle Person, die als Begleitung mitspielte, erwartete von



■ **Abbildung 4** Links: Die beiden verwendeten Avatare. Einmal mit den Proportionen eines etwa 4 Jahre alten Kindes (A) und einmal mit erwachsenen Proportionen, aber auf die Größe des anderen Modells geschrumpft (B). Rechts: Eine Teilnehmerin im Übertragungsanzug und mit VR-Brille (aus Paper [7] entnommen)

den dunkelhäutigen Avataren mehr Engagement beim Trommeln und bei den hellhäutigen Avataren ein professionelleres Auftreten. Nicht nur die eigenen Erwartungshaltungen beeinflussten also das Trommelspielen der Teilnehmer, sogar die vermeintlichen Erwartungen einer virtuellen Person, mit der kaum interagiert wurde, hatte einen Einfluss auf das Verhalten.

### 3.1.3 Im Körper eines Kindes – Interaktion mit Anderen oder Gegenständen

Doch nicht nur unsere Selbsteinschätzung verändert sich mit unserem Aussehen, auch die Art, wie wir unsere Umgebung wahrnehmen hängt davon ab in welchem Körper wir uns befinden. Unser Körper dient uns unbewusst als Maßstab im täglichen Leben. Das Experiment von Banakou et al. [7] zeigt dies auf erstaunliche Art. In diesem Experiment wurden die Teilnehmer in einen Avatar versetzt, der gerade einmal 90cm groß ist. Einmal wurde dabei der virtuelle Körper so gestaltet, dass dieser auch die Proportionen eines etwa vierjährigen Kindes widerspiegelt, der andere Avatar hatte normal-erwachsene Proportionen und wurde einfach auf die selbe Blickhöhe geschrumpft, siehe Abbildung 4 (A und B).

Die Versuchsteilnehmer trugen Anzüge, die jede ihrer Bewegungen auf den jeweiligen Avatar in Echtzeit übertrugen. Nach der Eingewöhnungszeit

wurden die Teilnehmer dann dazu aufgefordert bestimmte Distanzen mit dem Abstand der Handflächen an zu geben, eine Aufgabe, die vorher bereits ohne den virtuellen Körper mehrmals geprobt und aufgezeichnet wurde.

Erwartet wurde eine Überschätzung der Distanzen, da nun der kleinere Körper als Maßstab dient und wir, wie bereits erwähnt an ihm unsere Welt abmessen. Interessanterweise fiel die Überschätzung bei den Studienteilnehmern im Körper des 4-jährigen Kindes signifikant stärker aus, als bei denen im geschrumpften Körper [12]. Die offensichtlich veränderten Proportionen steigerten also das Verzerrten der Größenverhältnisse noch. Noch interessanter war das folgende Experiment, bei dem mit den Teilnehmern danach ein Test durchgeführt wurde, bei dem sie Worte oder Bilder entweder sich selbst oder anderen zuzuordnen sollten. Die Testteilnehmer im 4-jährigen Avatar assoziierten wesentlich häufiger und vor allem schneller kindliche Begriffe mit sich selbst, während dies bei den „geschrumpften“ Teilnehmern nicht geschah. Die Gründe für dieses Verhalten sind unbekannt. Es ist aber durchaus möglich, dass vergangene Erlebnisse aus der Zeit als die Versuchsteilnehmer selber im Vorschulalter waren, mit der aktuellen Situation im Gehirn assoziiert werden und dadurch alte Erinnerungen unbewusst unser Verhalten beeinflussen.

Für die Gestaltung eines Avatars sind also nicht nur das reine Aussehen, sondern auch die Statur, die Blickhöhe und natürlich die Proportionen relevant, sodass die virtuelle Umwelt auch erwartungsgemäß wahrgenommen werden kann. Größen und Distanzen werden verzerrt, wenn sich unser virtueller Körper verändert, da dieser als Vergleichsobjekt dient und in Kombination mit eigenen Erfahrungen und Erinnerungen kann dies sogar zu einem ganz anderen wahrgenommenen Bild führen, als erwartet oder beabsichtigt.

## 3.2 Andere Bewegung

Neben allen äußeren Gestaltungsmöglichkeiten eines Avatars gibt es noch einen zweiten, nicht weniger wichtigen Bereich zu untersuchen: Die Bewegungen und Animationen. Ein Avatar muss sich bewegen können, damit der Eindruck bestehen bleibt, dass dieser virtuelle Körper der des Nutzers ist. Die Präsenz ist dabei sehr stark abhängig von der Konformität der Bewegungen des Nutzers in der echten Welt mit denen der animierten oder übertragenen Bewegungen des virtuellen Körpers. Bewegt sich der Avatar nur mit Verzögerung oder gar nicht, stört dies nicht nur die Immersion sondern sorgt auch in der Regel dafür, dass die Person sich nicht richtig mit dem virtuellen Körper identifiziert und nicht tief genug in die Welt „eintauchen“ kann. Das oben genannte Experiment mit dem Kinder Avatar wurde beispielsweise einige Zeit später noch einmal wiederholt, diesmal mit asynchronen Bewegungen um die Wichtigkeit dieses Attributes zu beweisen [7].

Ein Avatar sollte sich also möglichst passend bewegen, damit das Gehirn

ihn besser als den eigenen erkennt. Wenn dies gegeben ist, der virtuelle Körper sich aber nicht in der gleichen Position, wie er reale befindet kann dies zu interessanten Illusionen kommen.

### 3.2.1 Die Illusion von Unbequemlichkeit – Körperhaltung

Virtuelle Realität kann für viele verschiedene Anwendungsmöglichkeiten verwendet werden. Eine etwas ungewöhnliche Idee stammt aus dem Jahr 2010 aus dem Journalismus Bereich. Nonny de la Peña et al. beschreiben eine Möglichkeit für Reporter mittels virtueller Realität bestimmte Situationen nach zu stellen, um diesen einen „realistischeren“ Eindruck vom Geschehen zu vermitteln und damit die Qualität der Reportage zu erhöhen [9]. Der Reporter konnte sich, wenn er denn wollte in die gleiche Situation begeben, wie die, über die er schrieb. Auch wenn es vielleicht nicht die erste Wahl eines Reporters sein kann, selbst zum Gefangenen im Gefängnis von Guantánamo zu werden, so wurde doch genau dieses in einem freiwilligen Experiment untersucht. Die für hier relevante Entdeckung dabei, dreht sich um die Körperhaltung der Versuchsteilnehmer. Wie bereits oben beschrieben, ist es für die Präsenz sehr wichtig, dass sich ein virtueller Körper so benimmt, wie wir es von ihm erwarten. Dies tritt allerdings nur in Kraft, wenn auch die Ausgangspositionen gleich sind. Im oben genannten Paper wurde den Versuchsteilnehmern ein Brustgurt umgelegt, um die Atmung und den Herzschlag zu überwachen. Dann wurden sie dazu aufgefordert sich bequem auf einen Stuhl zu setzen und die Hände hinter dem Rücken zu verschränken, siehe Abbildung 5 oben links.

Dann wurden sie mit einer VR-Brille ausgestattet und fanden sich in einer virtuellen Gefängniszelle wieder. Aus dem virtuellen Nebenraum vernahmen sie ein Verhör-Gespräch, dass vorher aus Original Aufzeichnungen nachgesprochen worden war. Wenn die Teilnehmer den Blick nach unten richteten konnten sie ihre, vermeintlich eigenen Füße in einer unangenehmen, gehockten Haltung sehen. Zur Rechten des virtuellen Gefangenen Avatars stand ein großer, raumhoher Spiegel, der den vollständigen Avatar in Gefangenenkleidung und einer so genannten Stress-Position zeigte, vergleiche Abbildung 5 unten rechts. Diese unangenehme Körperhaltung widersprach ganz offensichtlich der tatsächlichen realen Haltung der Versuchsteilnehmer und doch meldeten fast alle nach dem Experiment, ihr Körper habe sich verspannt und unangenehm angefühlt [9, 6]. Durch die erzeugte Präsenz mit Bild und Ton war das Gehirn also der Meinung, der in Stress-Position hockende Avatar sei der Körper des jeweiligen Teilnehmers und sendete somit automatisch Signale von Anstrengung, Unwohlsein und zuweilen sogar Angst. Einige Versuchsteilnehmer versuchten sogar unbewusst die Position nachzuahmen, indem sie sich in der realen Welt vorbeugten.

Nicht nur das Aussehen hat also Einfluss auf unser Empfinden, die Immersion und die Präsenz in virtueller Realität, die Körperhaltung kann ebenfalls



■ **Abbildung 5** Oben: Die Person sitzt bequem auf einem Stuhl mit einem Brustgurt zur Atemüberwachung und einer nach außen hin abgedunkelten VR-Brille. Unten: Links der Blick aus der Ersten-Person, wenn der Teilnehmer nach unten schaut, rechts der Blick in einen virtuellen Spiegel, der die unangenehme Körperhaltung des Avatars deutlich zeigt. (Oben: aus Paper [9] entnommen. Unten: Aus Youtube Video ausgeschnittene Szenen abrufbar unter [6])

eine große Rolle auf das Wohlergehen eines Nutzers in anderen Welten haben.

### 3.2.2 Von der Stelle kommen – Techniken zur Fortbewegung

Bisher wurden nur Experimente betrachtet, bei denen der Nutzer still stand oder saß. Oft wurden zwar die Bewegungen des Körpers aufgezeichnet und in die virtuelle Welt übertragen, doch all dies geschah ohne den eigenen Standpunkt, real oder virtuell, zu verändern. Dabei ist es für die Orientierung enorm wichtig, tatsächlich ein Gebiet abzulaufen, damit wir uns ein Bild bzw. eine Karte der Welt machen können.

Das letzte hier vorgestellte Experiment stammt bereits aus dem Jahr 1994. Bereits zu dieser Zeit war der Einfluss der Animationen des Avatars auf die Immersion und vor allem die Präsenz ein wichtiges Thema. Slater, Usoh und Steed von der Universität London untersuchten zwei verschiedene Arten sich in virtueller Realität fort zu bewegen [16]. Die erste und auch heute noch am verbreitetsten genutzte Möglichkeit ist den eigenen Avatar über einen Controller, oder wie hier im Experiment eine Art 3D-Maus zu steuern. Auch erlernte Handgesten, wie Zeigen werden oft verwendet. Diese Methoden haben den Vorteil, dass es sehr einfach und ohne großen Aufwand umsetzbar ist. Der Nachteil ist aber, dass sich zum einen die Hände kaum noch andersweitig nutzen lassen und zum anderen, dass sich die Bewegungen der realen Person, die

weiterhin auf der Stelle steht zu denen des laufenden Avatars unterscheidet, was einen Einschnitt in die Präsenz bedeutet. Slater et al. präsentieren daher die Möglichkeit des „auf der Stelle Laufens“ als Kompromiss. Die Bewegungen sind nicht identisch, aber schon wesentlich näher am tatsächlichen Laufen und wie jeder weiß, der in einem stehenden Zug sitzend einen anderen anfahren gesehen hat, kann man unserem Gehirn sehr leicht vorgaukeln, es würde sich selbst bewegen. Diesen Effekt der Vektion [5] macht sich auch dieses System zu nutze. Mittels eines Anzugs werden die Bewegungen des Nutzers an den Computer übertragen und ein neuronales Netz erlernte daraus die typischen Bewegungen des „auf der Stelle Laufens“. Der Nutzer konnte sich nun also innerhalb des Raumes, indem die Bewegungen aufgezeichnet wurden frei bewegen und an den Grenzen dessen einfach weiter auf der Stelle laufen um in der virtuellen Welt in Blickrichtung voran zu schreiten. Dieses frühe neuronale Netz, zur Erinnerung dies wurde 1994 entwickelt, lernte von verschiedenen Nutzern das Gehverhalten und konnte nach einigem Training mit guter Sicherheit zwischen „auf der Stelle Laufen“ und anderen Aktionen unterscheiden. Im Experiment wurde dies sogar noch auf Treppenstufen und Leitern ausgeweitet, die durch einfache Kombinationen von Bewegungen dargestellt und erkannt wurden. Alle Testteilnehmer waren der Meinung, dass dieses reale Laufen die Präsenz merklich steigerte und fast alle waren ohne großes Training oder ausführliche Erklärungen in der Lage ihren Avatar durch die Welt zu bewegen. Tatsächliches Bewegen steigert also nicht nur die unmittelbare Präsenz in einer virtuellen Umgebung sondern hilft auch, insbesondere bei größeren Welten extrem bei der Orientierung.

### 3.3 Weitere Faktoren im Überblick

In den vorangegangenen Abschnitten wurden verschiedene Teilbereiche der Avatargestaltung und seiner Funktionen beleuchtet und ihre Wichtigkeit anhand von repräsentativen Experimenten anschaulich dargelegt. Natürlich beschränkt sich die Gestaltung eines „virtuellen Ichs“ nicht nur auf das Aussehen und die Bewegungen, aber dies würde den Rahmen dieses Papers sprengen. Es sei an dieser Stelle nur erwähnt, dass Gerüche, Geräusche zum Beispiel von Kleidung oder Schritten auf dem Untergrund, sowie Schatten und weitere Faktoren eine ebenso große Rolle spielen können, wie die bereits oben erwähnten.

Gerade im Mehrbenutzerbetrieb, also einer virtuellen Plattform, in der mehrere Nutzer gleichzeitig und in Echtzeit miteinander agieren können, spielen Eigenschaften, wie eine gelungene Mimik oder Körperhaltung des Avatars sowie ein aus subjektiver Sicht angenehmes Äußeres eine große Rolle. Menschen verändern ihr Verhalten, je nachdem, wie sie selbst sich in ihrem Avatar wahrnehmen. Dies gilt besonders bei interaktiven, sozialen Plattformen im

Auge zu behalten.

## 4 Ich bin Du und Du bist Ich – Einflüsse und Wirkungen auf das Reale Ich

Viele direkte Auswirkungen einer bestimmten Gestaltung eines virtuellen Avatars wurden bereits im Verlauf der vorgestellten Experimente angedeutet. Es gibt allerdings auch einige indirekte Veränderungen, wie zum Beispiel die Wahrnehmung anderer oder die Präferenz zu Entscheidungen, die sonst vielleicht weniger relevant gewesen wären.

Dass wir oft unwillkürlich zurück lächeln, sobald uns jemand anstrahlt und das dies unsere Stimmung heben kann, ist allgemein bekannt. Dafür bedarf es auch keiner virtuellen Realität, diese zwischenmenschlichen Interaktionen sind fester Bestandteil unseres täglichen Lebens. Umso wichtiger ist es, dies in die virtuelle Welt zu übertragen, um ein Gefühl der Vertrautheit zu erschaffen und die künstliche Welt und ihre Bewohner real erscheinen zu lassen. Der bereits mehrfach angesprochene Proteus Effekt geht noch einen Schritt weiter. Der Effekt, dass wir einen anderen, fremden Körper als unseren eigenen ansehen und in ihm genau die gleichen Reaktionen und Bedürfnisse wie im echten Leben ausführen, ist erstaunlich. Umso interessanter sind dann die Reaktionen, die wir haben, wenn sich unser virtueller Körper stark von unserem realen unterscheidet. Studien haben gezeigt, dass Personen in einem übergewichtigen Avatar sich grundsätzlich in der virtuellen Welt weniger bewegen, als solche in normal gewichtigen Avataren [8]. Andersherum genau so, wenn ein virtueller Körper sehr aktiv war, zum Beispiel auf einem Laufband Sport getrieben hat, wurden die Teilnehmer nach der Rückkehr in die reale Welt oft ebenfalls aktiv und bewegten sich mehr. Es ist weiterhin bekannt, dass die Hautfarbe eines Avatars signifikante Auswirkungen auf die eigenen rassistischen Vorurteile haben kann [13]. Menschen wurden mittels eines so genannten IAT (Impliziter Assoziationstest) darauf untersucht, welche Vorurteile sie Menschen mit dunkler Haut entgegen bringen und einige Tage später in virtueller Realität genau in so einen Körper gesteckt. Direkt nach der Erfahrung im anderen Körper wurde der IAT Test wiederholt und es wurde eine deutliche Verminderung der Vorurteile festgestellt, wenn die Testperson in VR einen dunklen Hautton besaß. Solche, die einen hellhäutigen Avatar, einen lilafarbenen Alien oder eine körperlose VR-Erfahrung zugeteilt bekamen, änderten kaum ihre Einstellungen und Sichtweisen.

Neben unserem direkten Handeln in virtueller Realität bleiben Einflüsse daraus also auch noch einige Zeit nach der Immersion zurück. Ein weiteres sehr interessantes Experiment von Rosenberg, Baughman und Bailenson testete dabei eine Gruppe von Teilnehmern auf ihre Hilfsbereitschaft nach einer VR



■ **Abbildung 6** Oben: Eine nebelige virtuelle Welt zum freien Erkunden für Superhelden und Hubschrauberpassagiere gleichermaßen. Unten: Die Versuchsleitung stößt „aus Versehen“ eine Box voller Stifte zu Boden. Die Teilnehmer helfen beim Aufsammeln (B) (aus Paper [14] entnommen)

Immersion [14]. Die Studienteilnehmer wurden dabei in zwei Gruppen eingeteilt, die in der virtuellen Welt unterschiedliche Fähigkeiten erhielten. Während eine Gruppe als normaler Mensch mit ihrem Avatar in einem Hubschrauber als Passagier saßen, wurde der anderen Gruppe ein Superhelden-Avatar zugeteilt, der sich in der virtuellen Welt frei fliegend bewegen konnte. Dann gab es für beide Gruppen gleichermaßen pro Teilnehmer entweder die konkrete Aufgabe ein Medikament von einem Ort zu einem kranken Kind zu bringen, oder eben keinen konkreten Auftrag und die Welt durfte frei erkundet werden. Nach der Immersion wurden die Teilnehmer von einem Eingeweihten befragt, der dabei „aus Versehen“ eine Tasse mit Stiften vom Tisch stieß, siehe Abbildung 6. Beobachtet wurde, dass die Tatsache, ob in virtueller Realität eine konkrete, sogar helfende Aufgabe erfüllt wurde oder nicht, keinerlei Rolle bei der anschließenden Hilfsbereitschaft im Stifte-Einsammeln spielte. Der Fakt aber, ob dem Avatar Superkräfte verliehen wurden oder nicht hatte erheblichen Einfluss. Signifikant mehr Teilnehmer, die vorher durch die Welt fliegen konnten, erklärten sich schneller bereit zu helfen.

## 5 Leben 2.0 – Anwendungsmöglichkeiten

Insbesondere mit den ständigen Diskussionen über „Killerspiele“ und deren Einfluss, sollte es uns doch glücklich stimmen, dass genau die andere Richtung ein riesiges Potential entfaltet. Diese so genannten pro-sozialen Anwendungen oder Spiele [14] können uns also genau so lehren und positive Stimmungen vermitteln, wie im echten Leben. Wir können lernen zu helfen oder anders zu urteilen und tatsächlich ein Stück weit die Welt aus anderen Augen sehen.

Anwendungsbeispiele für Virtuelle Realität im Allgemeinen gibt es mittlerweile unzählige. Von der ständig steigend und fallenden Meinung in der Videospieleindustrie, über Schulen, Lehrmaterial und Simulationen bis hin zur Forschung und möglichen Erkenntnissen der menschlichen Psyche ist VR schon längst ein zentrales Werkzeug geworden um Menschen, Welten und Situationen zu zeigen, die sie sonst nicht hätten erleben können. Ob virtuelle Realität als Flucht vor dem echten Leben angesehen wird oder als eine Erweiterung der Möglichkeiten bleibt jedem selbst überlassen. Situationen, in denen die oben vorgestellten Phänomene aber hilfreich sein können, gibt es sicherlich ebenso viele, wie Meinungen zu diesem. Dreht man beispielsweise die Idee des immersiven Journalismus aus 3.2.1 herum und versucht aus einer unangenehmen Situation, wie einem Langstreckenflug in der zweiten Klasse mittels virtueller Realität bereits einen Liegestuhl am Strand zu machen, könnte dies mit Sicherheit auf große Begeisterung stoßen.

## 6 Der perfekte Avatar – Ein Fazit

Zusammengefasst gibt es also kein „Rezept“ oder eine Bauanleitung um den perfekten Avatar zu erstellen. Ein in der (Video-)Spielewelt häufig verwendete Ansatz ist scheinbar der beste: Jeder sollte sich seinen Avatar selber erstellen können. Das wichtigste ist, dass man sich selber mit diesem Körper identifizieren kann und ihn ansprechend findet. Er sollte der Situation angemessen sein und nicht fehl am Platz wirken. Die Bewegungen und Animationen sollten möglichst naturnah oder zumindest zum Körper passend sein. Ein Roboter-Avatar darf sich also ruhig etwas eckig bewegen, oder blechern anhören, solange dies keine Bewegungsabläufe verzögert. Unser Gehirn ist unglaublich anpassungsfähig was das Übernehmen von Gegenständen oder eben computergenerierten Körpern angeht.

Die Einflüsse und psychologischen Möglichkeiten, die unser Körper in einer frei zu gestaltenden Welt auf uns als Persönlichkeit hat sind enorm und werden in Zukunft wahrscheinlich noch eine wesentlich größere Rolle spielen, wenn die Entwicklung der Technik immer hochauflösendere Bildschirme, immer genauere Tracking-Mechanismen, die Bewegungen aufzeichnen können und immer realistischer aussehendere Welten hervor bringt. In Kombination mit einem

sorgfältig ausgewähltem virtuellen Ich, das uns tatsächlich so repräsentiert, wie wir es uns wünschen steht dann den Zukunftsvisionen von neuen Welten und virtueller Freiheit nur noch wenig im Weg.

---

#### Literatur

- 1 Avatar, the movie - the official website. Aufgerufen am: 06-06-2018. URL: <http://www.avatarmovie.com/index.html>.
- 2 Die geschichte der virtuellen welten. Aufgerufen am: 06-06-2018. URL: <https://www1.wdr.de/fernsehen/quarks/vr-die-geschichte-100.html>.
- 3 Google cardboard vr. Aufgerufen am: 13-06-2018. URL: <https://vr.google.com/cardboard/>.
- 4 Warum präsent nicht gleich immersion ist. Aufgerufen am: 08-06-2018. URL: <https://www.hardwareluxx.de/community/f324/warum-praesenz-nicht-gleich-immersion-ist-1079193.html>.
- 5 Wikipedia eintrag zum thema vektion. Aufgerufen am: 13-06-2018. URL: <https://de.wikipedia.org/wiki/Vektion>.
- 6 [youtubevideo] positive illusions of self in immersive virtual reality. Aufgerufen am: 06-06-2018. URL: <https://www.youtube.com/watch?v=50Nr8fiBvA>.
- 7 Domna Banakou, Raphaela Groten, and Mel Slater. Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes. *Proceedings of the National Academy of Sciences*, 110(31):12846–12851, 2013.
- 8 Amy Cuddy. Meet virtual you: How your vr self influences your real-life self. Aufgerufen am: 10-06-2018. URL: <https://www.thecut.com/2016/02/how-your-vr-self-influences-your-real-life-self.html>.
- 9 Nonny De la Peña, Peggy Weil, Joan Llobera, Elias Giannopoulos, Ausiàs Pomés, Bernhard Spanlang, Doron Friedman, Maria V Sanchez-Vives, and Mel Slater. Immersive journalism: immersive virtual reality for the first-person experience of news. *Presence: Teleoperators and virtual environments*, 19(4):291–301, 2010.
- 10 Max-Planck-Institut für biologische Kybernetik. Wahrnehmen und handeln in virtuellen umgebungen. Aufgerufen am: 09-06-2018. URL: <http://www.kyb.tuebingen.mpg.de/de/forschung/abt/bu/ehemalige-forschungsgruppen-der-abteilung/pave.html>.
- 11 Konstantina Kilteni, Ilias Bergstrom, and Mel Slater. Drumming in immersive virtual reality: the body shapes the way we play. *IEEE transactions on visualization and computer graphics*, 19(4):597–605, 2013.
- 12 Matthew Mientka. Virtual reality changes behavior and perception, as adults experience world through child’s body. Aufgerufen am: 09-06-2018. URL: <https://www.medicaldaily.com/virtual-reality-changes-behavior-and-perception-adults-experience-world-through-childs-body-video>.
- 13 Tabitha C Peck, Sofia Seinfeld, Salvatore M Aglioti, and Mel Slater. Putting yourself in the skin of a black avatar reduces implicit racial bias. *Consciousness and cognition*, 22(3):779–787, 2013.
- 14 Robin S Rosenberg, Shawnee L Baughman, and Jeremy N Bailenson. Virtual superheroes: Using superpowers in virtual reality to encourage prosocial behavior. *PloS one*, 8(1):e55003, 2013.
- 15 Valentin Schwind, Pascal Knierim, Cagri Tasci, Patrick Franczak, Nico Haas, and Niels Henze. These are not my hands!: Effect of gender on the perception of avatar hands in virtual reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 1577–1582. ACM, 2017.
- 16 Mel Slater, Martin Usoh, and Anthony Steed. Taking steps: the influence of a walking technique on presence in virtual reality. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 2(3):201–219, 1995.
- 17 Nick Yee and Jeremy Bailenson. The proteus effect: The effect of transformed self-representation on behavior. *Human communication research*, 33(3):271–290, 2007.

# Präsenz und Messung von Präsenz in Virtual Reality

Chantal Marie Huttenloher

Ludwig-Maximilians-Universität München, München, Deutschland  
c.huttenloher@campus.lmu.de

---

## Zusammenfassung

Virtual Reality (VR) ist ein Begriff, der immer öfter verwendet wird. Virtual Reality Systeme sollen zudem immer realistischer gestaltet werden, jedoch wird VR von jedem Individuum anders wahrgenommen, d.h. das Gefühl „präsent zu sein“ oder „da zu sein“ ist bei jeder Person verschieden stark ausgeprägt. Zudem ist Präsenz ein sehr weitreichender Begriff. Es gibt verschiedene Arten und Komponenten von Präsenz, die wiederum das Gefühl „da zu sein“ verstärken oder abschwächen können. Diese Arten und Komponenten helfen, um den Begriff Präsenz besser verstehen zu können und werden daher in dieser Arbeit erläutert. Wie stark das Gefühl „da zu sein“ in einer virtuellen Umgebung wahrgenommen wird ist subjektiv und zudem gibt es verschiedene Faktoren, die das Gefühl von Präsenz beeinflussen können. Diese Faktoren müssen bei der Messung von Präsenz bedacht werden. Es gibt verschiedene Ansätze und Methoden, um Präsenz zu messen. Die Problematik hierbei ist, dass eine Methode nötig ist, die das Gefühl von Präsenz effektiv, verlässlich und auf jedes Individuum angepasst messen kann. Der Fokus dieser Arbeit liegt auf dem Begriff Präsenz, der Messung von Präsenz und auf den verschiedenen Faktoren, die das Gefühl „da zu sein“ beeinflussen können.

2012 ACM Computing Classification • Human-centered computing→Virtual reality

**Keywords and phrases** Präsenz, Präsenzmessung, VR, Virtual Reality

## 1 Einführung

Ab wann fühlt man sich in einer virtuellen Umgebung präsent? Welche Faktoren führen dazu, dass sich ein Individuum in einer virtuellen Welt präsent fühlt? Präsenz kann auf verschiedene Art und Weise erklärt werden. Wenn wir an Präsenz denken, dann hat jeder seine eigene Vorstellung davon, ab wann ein Gefühl der Präsenz verspürt oder eine Umgebung als so real wahrgenommen wird, sodass man sich physisch in dieser Umgebung präsent fühlt. Dieses Gefühl von Präsenz kann gemessen werden. Dazu gibt es verschiedene Methoden, um



© Chantal Marie Huttenloher;

licensed under Creative Commons License CC-BY

Cite as: Chantal Marie Huttenloher. Präsenz und Messung von Präsenz in Virtual 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. 11:1–11:16.

das Gefühl „da zu sein“ zu messen. Diese Methoden können subjektiv und objektiv sein. Subjektive Methoden sind unter anderem Fragebögen, Interviews, Beobachtungen und psychophysische Messungen. Diese Methoden werden im Abschnitt „Methoden zur Messung von Präsenz“ im Detail erläutert.

Heutzutage ist Virtual Reality ein Begriff, unter dem sich viele Menschen etwas vorstellen können. Das Interesse an VR Systemen scheint zu wachsen. Die HTC Vive<sup>1</sup>, Oculus Rift<sup>2</sup>, Google Daydream<sup>3</sup> und Sony Playstation VR<sup>4</sup> sind alles VR Systeme, deren Verkaufszahlen von 2016 auf 2017 gestiegen sind<sup>5</sup>.

Präsenz ist ein Schlüsselbegriff, wenn es um die Definition von VR geht [13], denn VR Systeme zielen darauf ab, den Benutzer so weit wie möglich in eine virtuelle Welt zu integrieren, sodass der Unterschied zur realen Welt kaum noch wahrgenommen wird. Mit den steigenden Verkaufszahlen von VR Systemen könnte die Bedeutung von Präsenz und daher auch die Bedeutung von Messung von Präsenz steigen. Da VR Systeme immer realistischer gestaltet werden, wird dadurch das Gefühl von Präsenz beeinflusst werden. Daher ist es wichtig, Methoden zur Messung von Präsenz auf VR Systeme abzustimmen.

In dieser Arbeit soll zuerst ein kurzer Überblick über Virtual Reality Systeme gegeben werden, dann auf den Begriff Präsenz und zuletzt auf Messung von Präsenz eingegangen werden, wobei in letzterem Abschnitt vor allem existierende Methoden zur Messung von Präsenz erläutert werden. Am Ende der Arbeit soll zudem ein Ausblick auf die Zukunft von VR und Messung von Präsenz gegeben werden.

## 2 Ein Überblick über Virtual Reality

Bevor Präsenz und Messung von Präsenz im Fokus stehen, soll ein kurzer Überblick über Virtual Reality (VR) gegeben werden. VR ist ein Medium, welches immer mehr an Beliebtheit gewinnt. VR kann als ein Medium gesehen werden, welches den Benutzer vollkommen in eine virtuelle Umgebung involviert.

Für den Begriff Virtual Reality gibt es verschiedene Definitionen. Eine Vielzahl davon bezieht sich bei der Definition des Terms auf ein bestimmtes technologisches System, welches meistens einen Computer involviert, der Echtzeit Animationen erstellen kann, über einen Tracker für bestimmte Positionen und einen Datenhelm (HMD) verfügt. Außerdem ist bei vielen Definitionen für

<sup>1</sup> <https://www.vive.com/de/>

<sup>2</sup> <https://www.oculus.com/>

<sup>3</sup> [https://vr.google.com/intl/de\\_de/daydream/smartphonevr/](https://vr.google.com/intl/de_de/daydream/smartphonevr/)

<sup>4</sup> <https://www.playstation.com/de-de/explore/playstation-vr/>

<sup>5</sup> <https://www.statista.com/statistics/752110/global-vr-headset-sales-by-brand/>

Virtual Reality die Rede von einer elektronisch simulierten Umgebung [13].

Die erste Idee zu einem VR System wurde 1965 von Ivan Sutherland beschrieben. Dabei war es sein Ziel, eine virtuelle Welt zu schaffen, die genauso wie die reale Welt aussieht, sich genauso anfühlt und anhört. Zudem stellte Sutherland sich vor, dass die virtuelle Welt mit dem Menschen auf eine realistische Art und Weise interagieren kann. Jedoch hatte Sutherland nicht nur eine Idee zu VR, sondern er entwarf auch das erste HMD (head mounted display), welches in der Lage war, die Kopfposition und -orientierung auszugeben [5].

Seit dem Jahr 2000 ist ein großer Fortschritt in der Entwicklung von VR Systemen zu beobachten, während Preise für VR Systeme sinken. Besonders die Gaming Industrie hatte Einfluss auf die Entwicklung von VR Systemen<sup>6</sup>.

Wie schon in der Einführung erwähnt, ist Präsenz ein Schlüsselbegriff, um VR zu definieren. Welche Faktoren sind wichtig, um Präsenz in VR zu erreichen? Und welche Faktoren können ein Gefühl von Präsenz begrenzen oder steigern? Sanchez-Vives und Slater schreiben in ihrer Arbeit „From presence to consciousness through virtual reality“ (2005) über eben diese Faktoren. Dabei beschreiben sie, dass z.B. die Größe des Displays einen Einfluss auf das Gefühl von Präsenz haben kann, je nachdem wie groß das Sichtfeld ist (je größer das Sichtfeld, desto wahrscheinlicher ist es, dass das Gefühl von Präsenz steigt). HMDs haben generell ein begrenztes Sichtfeld. Dazu kommt, dass Displays unterschiedliche Auflösungen haben. Ein HMD hat z.B. die Auflösung 640x480 Pixel bei 3,3 cm Diagonale des Displays (per Auge). Ein weiterer Faktor, der Präsenz beeinflusst, ist die Bildfrequenz des Computers, um die virtuelle Umgebung so realistisch wie möglich aussehen zu lassen. Es kann z.B. zum Problem werden, wenn eine virtuelle Szene so komplex ist, dass die Ladezeit des Computers erhöht ist. Dann kann es passieren, dass der Benutzer seinen Kopf dreht und Veränderungen in der Szene nicht schnell genug geladen werden können, sodass das Erlebnis des Benutzers unterbrochen wird [8]. Dies sind nur ein paar ausgewählte Faktoren, die sich auf das Gefühl von Präsenz auswirken können.

Es gibt also verschiedene hardware- sowie softwarebezogene Faktoren, die das Gefühl von Präsenz in einer virtuellen Umgebung beeinflussen können.

Im nächsten Abschnitt wird genauer darauf eingegangen, was Präsenz bedeutet. Zudem wird auf verschiedene Komponenten und Arten von Präsenz eingegangen, um den Begriff verständlicher zu machen.

<sup>6</sup> <https://www.vrs.org.uk/virtual-reality/history.html>

### 3 Definition des Begriffs Präsenz

Präsenz ist ein Begriff, der in vielen verschiedenen Bereichen und in verschiedenen Kontexten genutzt wird. Daher ist anzunehmen, dass Präsenz von einer Menge von unterschiedlichen Definitionen geprägt ist. In diesem Abschnitt soll der Begriff Präsenz anhand von ausgewählten, existierenden Definitionen so klar wie möglich erläutert werden.

#### 3.1 Bedeutung von Präsenz

Matthew Lombard und Theresa Ditton beschäftigen sich in ihrer Arbeit „At the Heart of It All: The Concept of Presence“ (2006)[3] mit dem Thema Präsenz und erläutern mögliche Wege, um Präsenz zu definieren. Dabei handelt es sich um die folgenden Definitionsansätze:

- Präsenz als soziale Vielfalt
- Präsenz als Realismus
- Präsenz als Transportmittel
- Präsenz als Immersion
- Präsenz als sozialer Akteur innerhalb eines Mediums
- Präsenz als Medium, welches wiederum als sozialer Akteur verstanden wird

Um diese sechs verschiedenen Ansätze, Präsenz zu definieren, verständlicher zu machen, soll kurz auf jeden Ansatz eingegangen werden, ohne zu tief ins Detail zu gehen. Schumie et al. (2001) haben zu Lombards und Ditttons Ansätzen je eine prägnante Zusammenfassung erörtert. Dabei beschreiben sie z.B. „Präsenz als soziale Vielfalt“ als ein Medium, welches mit der Interaktion durch andere Personen als warm, sensitiv und personenbezogen wahrgenommen wird.

„Präsenz als Realismus“ bedeutet, dass das Medium zu einem gewissen Ausmaß als realistisch wahrgenommen werden kann. Wobei „Präsenz als Transportmittel“ sich auf das Gefühl, an einem Ort „da zu sein“ oder „zusammen zu sein“ bezieht. „Präsenz als Immersion“ beschreibt, zu welchem Ausmaß unsere Sinne in einer Umgebung angesprochen werden.

Bei den letzten beiden Erklärungen wird Präsenz als eine Art sozialer Akteur beschrieben, wobei dieser Akteur im Falle von „Präsenz als sozialer Akteur innerhalb eines Mediums“ mithilfe eines Mediums mit einer anderen Person interagiert. Im Falle von „Präsenz als Medium (welches als sozialer Akteur dient)“ ist das Medium selbst besagter Akteur [10].

Die oben genannten Ansätze, um Präsenz zu erklären zeigen, dass der Begriff Präsenz sehr weitläufig ist und aus verschiedenen Perspektiven betrachtet werden kann. In den folgenden Abschnitten wird genauer auf die Komponenten und auf die Arten von Präsenz eingegangen, um die bisherigen Definitionsansätze genauer zu veranschaulichen.

#### 3.2 Komponenten von Präsenz

Da der Begriff Präsenz sehr weitläufig ist und viele Definitionen hat, wird in diesem Abschnitt auf die Komponenten, die Präsenz ausmachen, eingegangen. Dadurch soll das Verständnis des Begriffs erhöht werden. Schubert et al. (2001) unterteilt Präsenz in die folgenden drei Komponenten [9]:

- Räumliche Präsenz
- Einbeziehung des Benutzers
- Echtheit der virtuellen Umgebung

Dabei bezieht sich räumliche Präsenz auf das Gefühl, innerhalb einer virtuellen Welt „da zu sein“, sich präsent zu fühlen und durchaus auch zu handeln und nicht nur von außen Dinge innerhalb dieser virtuellen Welt zu beeinflussen. Die Einbeziehung des Benutzers wird als „sich nur auf den virtuellen Raum konzentrieren“ oder als „gefesselt sein von der virtuellen Umgebung“ beschrieben. Die Echtheit der virtuellen Umgebung bezieht sich z.B. auf „wie echt sind Erfahrungen in der virtuellen Welt verglichen mit Erfahrungen in der realen Welt“ oder allgemeiner auf „ist Echtheit als Vergleich zwischen virtueller und realer Welt zu sehen“ [9].

Des Weiteren schreiben Schumie et al. (2001), dass räumliche Präsenz die stärkste Ausprägung von Präsenz sei, da das Gefühl „da zu sein“ ein grundlegender Faktor ist, um Präsenz zu erschaffen. Die anderen beiden Komponenten sind jedoch auch ausschlaggebend, um Aktionen und Wahrnehmungen innerhalb einer virtuellen Umgebung zu beschreiben [9].

#### 3.3 Arten von Präsenz

Zusätzlich zu den Komponenten gibt es auch verschiedene Arten von Präsenz. In diesem Abschnitt wird auf diese Arten eingegangen und erläutert, welche der Arten am bedeutendsten für das Gefühl von Präsenz ist.

Präsenz kann in die folgenden drei Arten untergliedert werden (Riva, Davide, & IJsselstein, 2003):

- Physikalische Präsenz
- Soziale Präsenz
- Kopräsenz

Physikalische Präsenz bezieht sich auf das Gefühl, dass der eigene Körper sich in einem Raum befindet. Soziale Präsenz bezieht sich auf das Gefühl „zusammen zu sein“ und auch darauf, mit einem virtuellen Gegenüber, welches sich durchaus an einem ganz anderen Ort befinden kann, zu interagieren. Bei Kopräsenz handelt es sich um die Schnittstelle zwischen physikalischer und sozialer Präsenz. Dies bedeutet, dass Kopräsenz als das Gefühl „sich zusammen

in einem Raum zu befinden“ beschrieben werden kann, wobei charakteristische Komponenten von physikalischer und sozialer Präsenz vorhanden sind [7].

Die aufgeführten Arten zeigen, dass Präsenz kein einfaches, sondern ein komplexes Konstrukt ist. Daher ist es sicherlich wichtig alle drei Arten von Präsenz in Studien miteinzubeziehen, wobei die physikalische Präsenz als das Gefühl, sich wirklich in einer Umgebung zu befinden am prägnantesten ist.

### 3.4 Immersion und Präsenz

In diesem Abschnitt liegt der Fokus auf Immersion. Des Weiteren soll auf den Zusammenhang zwischen Immersion und Präsenz eingegangen werden. Immersion und Präsenz sind beides Begriffe, die im Bereich Virtual Reality verwendet werden. Der Unterschied zwischen den beiden Begriffen ist nicht auf den ersten Blick klar. Laut Mel Slater sind Immersion und Präsenz logisch trennbar. Er sagt aber auch, dass Immersion und Präsenz stark miteinander in Beziehung stehen [12]. Doug A. Bowman hat Slaters Definitionen von Immersion und Präsenz folgend zusammengefasst [1]:

- Immersion bezieht sich auf den objektiven Grad der sensorischen Genauigkeit, die von einem Virtual Reality System bereitgestellt wird.
- Präsenz bezieht sich auf die subjektive und psychologische Reaktion des Benutzers auf ein Virtual Reality System.

Des Weiteren schreibt Bowman, dass Immersion in einem virtuellen System von der Rendering Software des Systems und von der Technologie der Bildschirme abhängig ist. Außerdem ist Immersion messbar und objektiv, sodass Systeme verschiedene Grade von Immersion vertreten können. Immersion kann als Kombination von mehreren Komponenten gesehen werden, die alle für die Applikation von Vorteil sein können. Somit ist Immersion ein multidimensionales Kontinuum [1].

Dagegen ist Präsenz eine individuelle und kontextabhängige Reaktion des Benutzers auf ein VR System. Zudem steht das Gefühl „da/hier zu sein“ sehr stark in Verbindung mit Präsenz. Dabei ist anzumerken, dass verschiedene Benutzer das gleiche VR System unterschiedlich wahrnehmen können [1].

Die oben genannten Definitionen helfen, um Immersion und Präsenz voneinander unterscheiden zu können. Jedoch soll nun noch auf den Zusammenhang und eventuelle Gemeinsamkeiten von Immersion und Präsenz eingegangen werden. So sagt z.B. Witmer, dass eine virtuelle Umgebung, die einen höheren Grad an Immersion erzielt, auch einen höheren Grad an Präsenz erzeugen kann [15]. Dies zeigt, dass ein Zusammenhang zwischen den beiden Begriffen bestehen muss. Dieser Zusammenhang besteht darin, dass Immersion zu Präsenz führen kann.

Immersion und Präsenz können als Begriffe separat betrachtet werden. Jedoch ist dabei anzumerken, dass Immersion durchaus Präsenz beeinflussen und diese begünstigen kann. Beide Begriffe sind wichtige Bestandteile in virtuellen Systemen.

## 4 Messung von Präsenz in Virtual Reality

Nachdem der Fokus des ersten Teils dieser Arbeit auf Präsenz als solches lag, soll nun der Fokus auf Messung von Präsenz in Virtual Reality gelegt werden. Es gibt verschiedene Methoden, um Präsenz zu messen. Bevor diese Methoden erläutert werden, soll ein grober Überblick über Messung von Präsenz im Allgemeinen gegeben werden.

### 4.1 Bedeutung von Präsenzmessung in VR

Es gibt laut Thomas B. Sheridan ein gewisses Bedürfnis nach einer Messung von Präsenz, welche robust, operational, verlässlich und nützlich ist. Zudem schreibt Sheridan, Präsenz sei subjektiv, ähnlich wie ein „mental model“, wodurch subjektive Berichte als Grundlage für die Messung von Präsenz dienen sollten [11].

Die Messung von Präsenz wird von verschiedenen Faktoren beeinflusst. Darunter sind z.B. Faktoren, wie die Charakteristiken einer virtuellen Umgebung. Darüber hinaus können individuelle Eigenschaften und Fertigkeiten das Präsenzgefühl innerhalb einer virtuellen Umgebung verstärken oder abschwächen. Dadurch ist es wichtig, dass Methoden zu Messung von Präsenz diese Faktoren miteinbeziehen, um das Präsenzgefühl so genau wie möglich zu messen [15].

Wie oben erwähnt, ist ein Ansatz für Präsenzmessung der subjektive Weg, da Präsenz subjektiv wahrgenommen wird. Es gibt jedoch durchaus auch einen objektiven Ansatz für die Messung. Bei diesem Ansatz wird die Reaktion von Benutzern als Indikator dafür, wie stark sich der jeweilige Benutzer in einer virtuellen Umgebung präsent fühlt, gewählt. Zudem kann man zur objektiven Messung des Gefühls „da zu sein“ auch physiologische Indikatoren, wie z.B. Blutdruck, Herzfrequenz, Muskelspannung und Atmung verwenden. Eine weitere Art der objektiven Messung bezieht sich auf das Verhalten von Benutzern. Dazu wird einem Benutzer je ein virtueller und ein realer Hinweis gegeben. Das Ausmaß, mit dem die Person sich mehr auf den virtuellen Hinweis als auf den realen stützt, zeigt, inwiefern Präsenz wahrgenommen wird [4].

Objektive Messungen von Präsenz sind vor allem als Unterstützung von subjektiven Messungen hilfreich. Physiologische und verhaltensbezogene Ansätze zur Messung von Präsenz stellen weitaus weniger detailreiche Angaben

über das Präsenzgefühl bereit als subjektive Ansätze, so kann z.B. ein Zusammenzucken auf eine als real wahrgenommene Umgebung oder auf räumliche Immersion, die von der virtuellen Umgebung geschaffen wird, hinweisen [4].

Des Weiteren schreiben Lombard et al. (2009), dass Methoden mit einem subjektiven Ansatz für Präsenzmessung einfacher durchzuführen seien als Methoden mit einem objektiven Ansatz. Zudem lassen subjektive Methoden eine wesentlichere und genauere Analyse des Präsenzgefühls zu [4].

Nachdem die Unterschiede von objektiven und subjektiven Ansätzen erläutert wurden, soll im nächsten Abschnitt dieser Arbeit auf Methoden zur Messung von Präsenz mithilfe des subjektiven Ansatzes eingegangen werden, da dieser für die Präsenzmessung essentieller zu sein scheint als der objektive Ansatz.

## 4.2 Methoden zur Messung von Präsenz

Um das Gefühl „präsent zu sein“ zu messen gibt es objektive und subjektive Ansätze, wobei der Fokus in dieser Arbeit auf den subjektiven Ansätzen liegt. In diesem Abschnitt sollen daher subjektive Methoden zur Messung von Präsenz vorgestellt werden.

Die subjektive Messung von Präsenz kann laut Joy van Baren (2004) in vier Gruppen eingeteilt werden [14]:

- Fragebögen zu Präsenz
- Qualitative Methoden
- Kontinuierliche Beobachtungen
- Psychophysische Messungen

Dabei ist anzumerken, dass die Fragebögen zu Präsenz weitaus die verbreitetste Art von Methoden zur Messung von Präsenz darstellen [14]. Daher soll in dieser Arbeit besonderer Fokus auf die Fragebögen gelegt werden. Des Weiteren soll auf subjektiv verstärkende Messungen eingegangen werden.

### 4.2.1 Fragebögen zu Präsenz

Bisher wurde eine Vielzahl an verschiedenen Fragebögen zu Präsenz entwickelt. Diese können sich in ihrem Aufbau und inhaltlichem Umfang unterscheiden. Zudem verwenden manche Fragebögen nur ein Element, um Präsenz zu adressieren, wobei in anderen Fragebögen die Komplexität von Präsenz widerspiegelt wird [14]. Daher ist es wichtig, einen allgemeinen Konsens darüber, welche Elemente wichtig für Fragebögen zu Präsenz sind, zu ermitteln. Lessiter et al. (2001) haben bestimmte Kriterien zusammengefasst, die Präsenzfragebögen erfüllen sollten [2]:

- Ein Verständnis von Präsenz sollte nicht erzielt werden, indem man einen Benutzer/in direkt danach fragt, wie präsent sich diese/r gerade fühlt.
- Eine Frage sollte auch wirklich nur aus einer Frage bestehen, nicht aus mehreren.
- Antwortmöglichkeiten sollten konsistent geführt werden.
- Da Präsenz ein multidimensionales Konstrukt ist, sollte dieser Sachverhalt in Fragebögen angedeutet werden und Charakteristiken von Präsenz aufzeigen.
- Fragen sollten keine Andeutungen oder Referenzen auf bestimmte Mediensysteme beinhalten.
- Präsenzmessungen sollten generell an unterschiedlichen Mediensystemen mit verschiedenen medialen Inhalten vorgenommen werden.
- Fragebögen sollten von einer ausreichenden Anzahl an Personen ausgefüllt werden.

Die aufgeführten Kriterien von Lessiter et al. (2001) helfen, um eine Idee davon zu bekommen, wie ein Fragebogen zu Präsenz aussehen sollte. Lombard et al. (2009) zeigen ebenfalls Kriterien auf, die ein Fragebogen zu Präsenz vorweisen sollte. Diese Kriterien unterscheiden sich zu den oben genannten und sind im Folgenden zusammengefasst [4]:

- Ein Fragebogen muss glaubwürdig sein, d.h. dass dieser nach außen und nach innen konsistent sein sollte.
- Ein Fragebogen muss Beweise für seine Richtigkeit aufführen, welche z.B. durch Beziehung zwischen Variablen, die zur Präsenzmessung verwendet werden, geschaffen werden.
- Ein Fragebogen, der gültig und glaubwürdig ist, muss zudem sensitiv sein, d.h. dass die Möglichkeit, zwischen verschiedenen Graden von Präsenz unterscheiden zu können, gegeben sein muss.
- Ein Fragebogen muss gedanklich umfassend sein, um die Multidimensionalität von Präsenz widerspiegeln zu können.
- Ein Fragebogen zu Präsenz muss außerdem ein breites Spektrum an Anwendungsmöglichkeiten anbieten können, dazu zählen z.B. Anwendungen in verschiedenen Umgebungen unter verschiedenen Konditionen.

Dabei ergänzen Lombard et al. (2009), dass die letzten beiden Kriterien schwierig zu erreichen seien, da es nicht ganz geklärt ist, ob Messungen aller verschiedener Typen von Präsenz in jedem beliebigen Kontext auch möglich sind [4]. Diese Kriterien von Lombard et al. (2009) zeigen nicht nur, dass es verschiedene Ansätze gibt, um Fragebögen zu Präsenz zu erstellen, sondern auch, dass es ein gewisses Verlangen nach einem offiziell gültigen Konsens gibt, um Präsenz einheitlich messen zu können.

Fragebögen zu Präsenz bieten gewisse Vorteile. Zu diesen Vorteilen gehört

unter anderem, dass Fragebögen kostengünstig, einfach zu analysieren und zu interpretieren sind. Außerdem stören Fragebögen nicht während einer Studie, da sie immer nach einer abgeschlossenen Etappe einer Studie ausgefüllt werden. Zudem haben viele Fragebögen gezeigt, dass sie gut auf verschiedene Grade von Präsenz anzuwenden sind [14].

Nachdem wir uns mit den Vorteilen von Fragebögen zu Präsenz beschäftigt haben, sollen nun auch Nachteile dieser Fragebögen erläutert werden. So ist z.B. ein Nachteil, dass Fragebögen retrospektiv sind und sich auf das Gedächtnis der Benutzer stützen. Das Problem hier ist, dass das Gedächtnis von Personen meist nur eine bruchstückhafte Reflexion von Erinnerungen beinhaltet und dies kann sich in den Antworten in Fragebögen widerspiegeln. Zudem ist es wahrscheinlich, dass sich Personen an Dinge, die am Ende eines Experimentes passiert sind besser erinnern können, als an Dinge, die am Anfang des Experimentes vonstatten gingen. Außerdem können sich bestimmte Charakteristiken eines Fragebogens, wie z.B. darin enthaltene Hinweise oder Schlagwörter, auf die Antworten der Studienteilnehmer auswirken [14].

Da Fragebögen zu Präsenz innerhalb der subjektiven Methoden zur Messung von Präsenz die am meisten genutzte Methode darstellen, scheinen die Vorteile im Allgemeinen die Nachteile zu überwiegen.

## 4.2.2 Qualitative Methoden

Qualitative Methoden oder Messungen sind ein guter Ansatz, wenn man Informationen gewinnen möchte, die nicht durch Quantifizierung gewonnen werden kann. Darunter fallen z.B. statistische Verfahren. Zu den qualitativen Methoden gehören unter anderem aufgeschriebener Text oder Gedanken, Interviews und ethnographische Ansätze. Nutzung finden diese Methoden oftmals bei Forschung, die darauf abzielt, Hypothesen zu kreieren [14].

Zu qualitativen Methoden gehören z.B. Interviews, Fokusgruppen, ethnographische Beobachtungen, Inhaltsanalysen und Think-Aloud-Methoden [14].

Ein Vorteil von qualitativen Messungen ist, dass die gewonnenen Informationen sehr detailreich sind und sie eine gute Ergänzung zu quantitativen Messungen darstellen [14]. Ein weiterer Vorteil ist, dass Personen in ihren eigenen Worten Erfahrungen beschreiben können und nicht von vorgegebenen Fragen beeinflusst werden [6]. Darüber hinaus können qualitative Methoden Forschern helfen ein tieferes Verständnis für Präsenz zu bekommen. Jedoch ist es schwierig, gewonnene Erkenntnisse auf die Bevölkerung zu generalisieren, da die daraus gewonnenen Daten oftmals wenig zuverlässig und nicht unbedingt stichhaltig sind [4]. Ein weiterer Nachteil ergibt sich aus dem Fakt, dass qualitative Methoden sehr detailreiche Informationen hervorbringen können, denn Datenanalyse ist generell sehr zeitaufwendig, sodass die Menge an Informationen eine Verlangsamung der Analyse bedeuten kann [14].

Qualitative Methoden haben genauso wie Fragebögen zu Präsenz Vor- und Nachteile. Besonders geeignet sind qualitative Methoden für explorative Forschung, bei der sehr viel Wert auf Detailreichtum gelegt wird.

## 4.2.3 Kontinuierliche Beobachtungen

In diesem Abschnitt werden kontinuierliche Beobachtungen (engl. „continuous assessments“) als subjektive Methoden erläutert.

Joy van Baren (2004) beschreibt kontinuierliche Beobachtungen als eine Methode, bei der Studienteilnehmer während dem eigentlichen Experiment Änderungen des Gefühls von Präsenz angeben, sobald eine solche Änderung auftritt. Daher ist einer der Vorteile dieser Methode, dass keine Erinnerungen an das Experiment verloren gehen, da die erlebten Eindrücke sofort geschildert werden. Daraus ergibt sich jedoch auch ein Nachteil dieser Methode, denn dadurch, dass Eindrücke gleich beschrieben werden, wird das Experiment und damit die User Experience unterbrochen. Zudem kann immer nur ein Aspekt zu einem Zeitpunkt von einem Studienteilnehmer beschrieben werden [14].

Kontinuierliche Beobachtungen haben wie andere subjektive Methoden Vor- und Nachteile. Ihre besondere Stärke liegt sicherlich darin, dass es zu wenig Informationsverlust kommt, da Eindrücke sofort durch die jeweilige Person beschrieben werden.

## 4.2.4 Psychophysische Messungen

Psychophysische Messungen werden durch einen Beobachter vorgenommen. Dabei soll der Beobachter eine subjektive Bewertung der Änderung eines Stimulus angeben. Dies kann erreicht werden, indem man Studienteilnehmer dazu auffordert, dem Stimulus, der die wahrgenommenen Präsenz verstärkt oder abschwächt, einen Wert zu geben. Eine andere Art dies zu erreichen ist zu observieren, inwiefern Teilnehmer zwischen zwei Stimuli unterscheiden können. Des Weiteren können Teilnehmer auch danach gefragt werden, das Gefühl von Präsenz auf einen anderen Sachverhalt zu übertragen und damit zu beschreiben [14].

Laut Joy van Baren (2004) werden psychophysische Messungen in Studien nicht häufig verwendet. Er sagt jedoch auch, dass psychophysische Messungen bei mindestens zwei Studien gezeigt haben, dass diese Methode für verschiedene Grade von Präsenz geeignet ist. Außerdem ist diese Methode zur Messung von Präsenz kostengünstig und einfach anzuwenden [14]. Van Baren (2004) schreibt außerdem über Nachteile von psychophysischen Messungen. Darunter z.B. die Tatsache, dass alle subjektiven Methoden gewissermaßen anfällig für individuelle Vorlieben und Vorurteile sind. Zudem sind die Messungen sehr abhängig von den Instruktionen des Studienleiters und ebenso von den

individuellen Interpretationen der Studienteilnehmer [14].

Zu psychophysischen Messungen gehören unter anderem das „Cross-Modality Matching (CMM)“, der Paarvergleich (engl. „Paired comparison“) und der Virtual Reality Turing Test [14].

Für die Messung von Präsenz sind psychophysische Messungen nicht die erste Wahl, da die anderen Methoden (besonders Fragebögen) signifikantere Vorteile aufzeigen können. Trotzdem wird diese Art von Messung manchmal benutzt, um Präsenz kostengünstig und einfach zu messen.

#### 4.2.5 Subjektiv verstärkende Messungen

Als letzte Art der subjektiven Messungen sollen die subjektiv verstärkenden Messungen (engl. „subjective corroborative measures“) erläutert werden.

Indikatoren, die die Subjektivität verstärken, können helfen, die Gültigkeit von Messungen von Präsenz zu unterstützen. Solche Indikatoren sind z.B. Aufmerksamkeit, räumliche Wahrnehmung und Erinnerungen. Beispiele für subjektiv verstärkende Messungen sind z.B. „Breaks in Presence (BIP)“, „Duration Estimation“, „Simulator Sickness Questionnaire (SSQ)“ und „Memory Characteristic Questionnaire (MCQ)“ [14]. Auf letzteres soll kurz genauer eingegangen werden, um subjektiv verstärkende Messungen anhand eines Beispiels genauer zu erklären.

Das MCQ bezieht sich auf physikalische Präsenz und wurde erstellt, um qualitative Unterschiede zwischen Erfahrungen im Gedächtnis festzustellen. Es besteht aus insgesamt 21 Einheiten, welche wiederum Bewertungsskalen beinhalten. Zudem enthält es Beurteilungen über Präsenz, Aufmerksamkeit, Ähnlichkeiten zwischen Umgebungen und über das Sichtfeld [14].

Laut van Baren (2004) wurde das MCQ in einem Experiment genutzt, in dem reale, virtuelle und imaginäre Welten verglichen werden sollte. Zusammengefasst kann man sagen, dass die Verwendung von subjektiv verstärkenden Messungen hilfreich ist, wenn man besonderen Wert auf die Subjektivität eines Experimentes legt.

### 4.3 Bedeutung von Präsenz in verschiedenen Anwendungsbereichen

Wir wissen bereits, dass Präsenz ein Schlüsselbegriff für Virtual Reality ist. Zudem haben wir die Arten von Präsenz und deren Komponenten kennengelernt. Darüber hinaus wurden existierende Methoden zur Messung von Präsenz erläutert. Zum Ende dieser Arbeit sollen nun verschiedene Bereiche, in denen Präsenz von Bedeutung ist, beschrieben werden, um die Wichtigkeit von Präsenz und Messung von Präsenz zu verdeutlichen.

#### 4.3.1 Präsenz und psychologische Therapien

Anwendungsbereiche, die nicht ohne Präsenz auskommen, involvieren die Ausnutzung von virtuellen Umgebungen für z.B. psychologische Therapien (u.a. zum Zwecke der Angstbewältigung). In einer solchen Therapie wird der Patient in eine virtuelle Umgebung gebracht, die die Ängste des Patienten veranschaulicht [8]. Hierbei ist es natürlich von Bedeutung, dass die virtuelle Umgebung so real wie möglich die Ängste des Patienten widerspiegeln kann, ansonsten kann eine Therapie wirkungslos sein. Das bedeutet im Allgemeinen, dass es sehr wichtig ist, dass der Patient sich in der virtuellen Umgebung präsent fühlt. Ängste, die durch den Einsatz von virtuellen Umgebungen therapiert werden können, sind unter anderem [8]:

- Akrophobie
- Soziale Phobie
- Paranoia
- Posttraumatische Belastungsstörung

Des Weiteren kann laut Sanchez und Slater (2005) eine Schmerzreduzierung erzielt werden, wenn Patienten mittels Einsatz von virtuellen Systemen behandelt werden. Die virtuelle Umgebung agiert in solchen Fällen als eine Art Gegenstück zum Schmerz eines Patienten. Die Beweise für Virtual Reality als Medium zur Schmerzlinderung sind jedoch noch nicht im Detail erforscht worden [8].

#### 4.3.2 Präsenz und das Lernen

Nach Witmer et al. (1998) spielt Präsenz eine Rolle für das Lernen und Leistung. Faktoren, die Präsenz beeinflussen, scheinen Lern- und Leistungsfähigkeit zu verbessern. Wenn mit einer virtuellen Umgebung auf natürliche Art und Weise interagiert wird, wird der Grad an Immersion und damit auch Präsenz erhöht. Faktoren, die Immersion verstärken können, wie z.B. die Minimierung von externen Ablenkungsmöglichkeiten oder Erhöhung der Teilnahmemöglichkeiten innerhalb einer virtuellen Umgebung an Events, können auch die Lern- und Leistungsfähigkeit verstärken. Da viele Faktoren, die mit Lernen und Leistung zu tun haben, das Gefühl von Präsenz verstärken können, ist es sehr wahrscheinlich, dass zwischen beiden Bereichen ein Zusammenhang besteht [15].

#### 4.3.3 Präsenz und die Neurowissenschaft

Ein weiterer Fachbereich, welcher sich für die Forschung an Präsenz interessiert, ist die Neurowissenschaft. Forschung an Präsenz spielt sich häufig in

den technologischen Fachbereichen ab, jedoch interessieren sich immer mehr Psychologen für Präsenz [8]. Sanchez und Slater (2005) schreiben, dass Präsenz für die Neurowissenschaft von Bedeutung ist, es jedoch noch nicht viel Literatur über Präsenz in diesem Bereich gibt. Sie fügen dem jedoch hinzu, dass die Forschung an Präsenz für die Neurowissenschaft vorteilhaft ist, denn VR Technologie kann helfen, verschiedene experimentelle Szenarios zu schaffen, die sich speziell auf den Menschen anpassen lassen.

Dies sind nur ausgewählte Bereiche, in denen Präsenz von Vorteil sein kann oder genutzt wird, um ein bestimmtes Ziel, z.B. Therapie von verschiedenen Ängsten, zu erreichen. Sanchez und Slater (2005) ergänzen noch, dass die Forschung an Präsenz ausgeweitet werden sollte, über den Bereich der technologisch orientierten Disziplinen hinweg [8]. Forschung an Präsenz scheint allgemein ein sehr vielversprechendes Gebiet zu sein, welches in den nächsten Jahren sicherlich noch mehr verfolgt werden wird.

## 5 Fazit

Da das Medium Virtual Reality immer mehr an Bedeutung gewinnt, ist es wichtig, Präsenz zu verstehen und eine allgemein gültige und verlässliche Art der Messung für Präsenz zu finden.

Präsenz ist ein multidimensionales Konstrukt, welches subjektiv empfunden wird. Daher muss bei der Messung auf verschiedene Faktoren Rücksicht genommen werden.

Die oben aufgezeigten Methoden zur Messung von Präsenz haben ihre individuellen Vor- und Nachteile, wobei Fragebögen zu Präsenz im Allgemeinen am besten für die Messung von Präsenz geeignet zu sein scheinen. Daher werden sie momentan auch am häufigsten innerhalb von Studien über Präsenz verwendet.

Präsenz zu messen wird immer einen gewissen Nachteil haben, denn das Gefühl „da zu sein“ ist subjektiv und kann daher nur vollkommen durch das Individuum selbst beschrieben und bewertet werden. Daher kann es immer dazu kommen, dass ein Individuum eine Erfahrung in einem virtuellen Raum vorurteilsbehaftet oder nach gewissen Vorlieben bewertet. Daher ist es umso wichtiger, bisherige Methoden zur Messung von Präsenz in virtuellen Umgebungen oder auch ganz allgemein und nach gegebenen Möglichkeiten zu optimieren.

Mit den immer realistischer werdenden Virtual Reality Systemen, wird auch das Gefühl von Präsenz bei vielen Benutzern gesteigert werden können. Dabei spielen eine Vielzahl an Faktoren eine Rolle, wenn es um das Beeinflussen des Gefühls „da zu sein“ geht. So kann z.B. bei Person A ein realistischeres Gefühl in virtuellen Welten erzielt werden durch die Benutzung von verschie-

denen Interaktionsmöglichkeiten wie die Leap Motion<sup>7</sup> im Gegensatz zu der Benutzung von Controllern. Wobei Person B eventuell die Interaktion mit den Controllern bevorzugt und sich dabei viel präsenter fühlt als mit einer anderen Interaktionstechnik. Dabei lässt sich nur wieder betonen, wie subjektiv Präsenz ist. Dieser Faktor sollte immer berücksichtigt werden, wenn man sich mit dem Thema beschäftigt.

Zusammengefasst kann man sagen, dass Präsenz als zentraler Begriff von Virtual Reality in der Entwicklung von VR Systemen von großer Bedeutung ist. Daher sollten ebenso Methoden zur Messung von Präsenz weiter entwickelt werden, damit diese an die neuen Technologien angepasst sind und kein Faktor, der das Gefühl von Präsenz beeinflussen kann, vernachlässigt wird.

## 6 Ausblick

Da Virtual Reality ein Medium ist, dessen Beliebtheit gerade wächst, ist es sehr wahrscheinlich, dass Messungen von Präsenz auch weiterhin an Bedeutung gewinnen werden. Zudem ist es sehr wahrscheinlich, dass die Preise für VR Systeme in der Zukunft sinken werden, da es immer mehr Angebote für VR Systeme geben wird. Dadurch wird der Zugang zu VR Systemen immer mehr Menschen möglich werden<sup>8</sup>. Außerdem ist heutzutage das Programmieren von VR Software für jeden offen zugänglich. Es existiert kostenlose Software, um VR Applikationen auch daheim zu entwickeln (z.B. Unity<sup>9</sup>).

Wie aufgezeigt, gibt es einen Wunsch nach effizienten Messmethoden, wobei existierende Methoden häufig zur Messung von Präsenz genutzt werden. Die verschiedenen Methoden finden je nach Vorteil einen Anwendungsbereich, um Präsenz so genau wie möglich messen zu können.

Wenn in der Zukunft Präsenz in VR durch ein klares Verständnis von Methoden zur Messung von Präsenz noch verbessert werden kann, dann könnte dies genutzt werden, um noch realistischere VR Systeme zu entwickeln. Denn je realistischer ein VR System ist, desto besser kann es dann für Simulationen, Therapien und auch für Lernzwecke genutzt werden.

<sup>7</sup> Leap Motion Website: <https://www.leapmotion.com/en/>

<sup>8</sup> <https://www.vrs.org.uk/virtual-reality/history.html>

<sup>9</sup> <https://unity3d.com/de>

## Literatur

- 1 Doug A Bowman and Ryan P McMahan. Virtual reality: how much immersion is enough? *Computer*, 40(7), 2007.
- 2 Jane Lessiter, Jonathan Freeman, Edmund Keogh, and Jules Davidoff. A cross-media presence questionnaire: The itc-sense of presence inventory. *Presence: Teleoperators & Virtual Environments*, 10(3):282–297, 2001.
- 3 Matthew Lombard and Theresa Ditton. At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3(2):0–0, 1997.
- 4 Matthew Lombard, Theresa B Ditton, and Lisa Weinstein. Measuring presence: the temple presence inventory. In *Proceedings of the 12th Annual International Workshop on Presence*, pages 1–15, 2009.
- 5 Tomasz Mazuryk and Michael Gervautz. Virtual reality-history, applications, technology and future. 1996.
- 6 Lydia MJ Meesters, Wijnand A IJsselsteijn, and Pieter JH Seuntjens. A survey of perceptual evaluations and requirements of three-dimensional tv. *IEEE Transactions on circuits and systems for video technology*, 14(3):381–391, 2004.
- 7 Giuseppe Riva, Fabrizio Davide, and Wijnand A IJsselsteijn. *Being there: Concepts, effects and measurements of user presence in synthetic environments*. Ios Press, 2003.
- 8 Maria V Sanchez-Vives and Mel Slater. From presence to consciousness through virtual reality. *Nature Reviews Neuroscience*, 6(4):332, 2005.
- 9 Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments*, 10(3):266–281, 2001.
- 10 Martijn J Schuemie, Peter Van Der Straaten, Merel Krijn, and Charles APG Van Der Mast. Research on presence in virtual reality: A survey. *CyberPsychology & Behavior*, 4(2):183–201, 2001.
- 11 Thomas B Sheridan. Musings on telepresence and virtual presence. *Presence: Teleoperators & Virtual Environments*, 1(1):120–126, 1992.
- 12 Mel Slater, Vasilis Linakis, Martin Usoh, Rob Kooper, and Gower Street. Immersion, presence, and performance in virtual environments: An experiment with tri-dimensional chess. In *ACM virtual reality software and technology (VRST)*, volume 163, page 72. ACM Press New York, NY, 1996.
- 13 Jonathan Steuer. Defining virtual reality: Dimensions determining telepresence. *Journal of communication*, 42(4):73–93, 1992.
- 14 Joy Van Baren. Measuring presence: A guide to current measurement approaches. *Deliverable of the OmniPres project IST-2001-39237*, 2004.
- 15 Bob G Witmer and Michael J Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7(3):225–240, 1998.

# Freihand basierte Interaktion in Augmented Reality

Duc Anh Le

Ludwig-Maximilians-Universität München, München, Deutschland  
D.Le@campus.lmu.de

## Zusammenfassung

Die Nutzung von Freihand-Gesten in Augmented Reality-Applikationen ist von großer Bedeutung. Es wird nicht nur die Möglichkeit geboten, typische 2D-Eingabeprobleme im dreidimensionalen Raum zu bewältigen, sondern kann auch in risikoreichen Umgebungen wie bei einem Tatort oder einem Operationssaal von großer Notwendigkeit sein. Diese Arbeit befasst sich mit Augmented Reality im Allgemeinen, dem technischen Hintergrund von Freihand Interaktion und dessen Einsatzgebiete. Zudem werden aktuelle Probleme erläutert und mögliche Lösungsansätze geschildert.

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

**Keywords and phrases** Freihand Interaktion, Augmented Reality, Tiefenerkennung

## 1 Einleitung

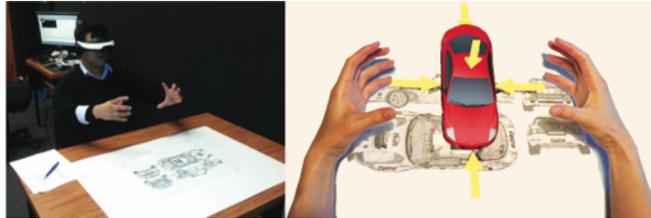
Applikationen auf Basis von Augmented Reality (AR) ermöglichen neue Wege der Interaktion mit der realen Umgebung durch Einblendung von virtuellen Inhalten auf Smartphone oder AR Brille. Diese Inhalte können unter anderem mit der realen Welt über das Display interagieren oder zusätzliche Informationen zur Umgebung anzeigen. Der Nutzer kann diese Inhalte mithilfe von zusätzlichen Steuerungsmöglichkeiten, Fingerberührungen oder sogar durch Neigung des Smartphones [5, 8] bewegen oder ändern. Großer Nachteil bei dieser Art der Interaktion ist jedoch, dass bei mobilem Einsatz bereits eine Hand durch Halten des Smartphones beschränkt wird und dadurch nur noch eine Hand zur Steuerung frei ist. Wenn hierzu noch ein Steuerungsgerät bedient werden muss, erweist sich die Interaktion als sehr umständlich. Auch das Szenario mit Fingerberührungen bringt viele Nachteile mit sich. Unter anderem ist es eine große Herausforderung, dass auf dem ohnehin schon kleinen Bildschirm eines Smartphones durch die Fingereingabe zusätzliche Teile des



© Duc Anh Le;

licensed under Creative Commons License CC-BY

Cite as: 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.



■ **Abbildung 1** Links: Sicht von Außen. Rechts: Sicht innerhalb der AR-Brille [3].

Bildes verdeckt werden und das somit die Nutzerfreundlichkeit einschränkt. Davon abgesehen ist es schwierig, eine zwei-dimensionale Eingabe auf eine drei dimensionale Welt abzubilden. Bei Nutzung einer Daten-Brille, bspw. Microsoft Hololens<sup>1</sup>, wären die Hände grundsätzlich frei, jedoch würden sich auch bei diesem System die Verwendung von Steuerungsgeräten als unpraktisch erweisen. Als Lösung ergibt sich die Nutzung von Gestensteuerung, bei der man mithilfe von tiefererkennenden Kameras die Möglichkeit hat, präzise und konsistent die Bewegungen der Hände zu verfolgen und als Eingabe zu nutzen. Für mobile Geräte gibt es von Vuforia<sup>2</sup> eine Softwareschnittstelle, die die interne Kamera des Smartphones oder Tablets zur Handerkennung verwendet. Wie gut diese Methoden sich im täglichen Gebrauch erweisen, wird im weiteren Verlauf dieser Arbeit beschrieben.

### 1.1 Wann spricht man von Augmented Reality?

Um zu sehen, wann es sich bei der Darstellung einer virtuellen Welt um Augmented Reality handelt, ist es wichtig, den Unterschied zwischen den verschiedenen Realitäten zu verstehen. Im folgenden Abschnitt werden die drei gängigen Realitäten näher erklärt:

**Augmented Reality (AR)** beschreibt die Möglichkeit, zusätzliche grafische oder textbasierte Informationen über ein Display in die reale Welt einzublenden [19].

Funktionsweise: Durch die Nutzung einer Kamera lassen sich visuelle Daten aus der realen Welt extrahieren und über eine Softwareschnittstelle verarbeiten. Diese werden dann mithilfe eines Smartphone Displays oder einer AR-Brille auf die reale Welt visuell überlagert. Dies können z.B. Details zum derzeitigen Standort sein (*siehe Abb. 2*) oder virtuelle Objekte, die sich der Umgebung

<sup>1</sup> <https://www.microsoft.com/de-de/hololens>

<sup>2</sup> <https://www.vuforia.com>



■ **Abbildung 2** Nutzung von AR zur Einblendung von zusätzlichen Informationen.

anpassen und bewegen lassen (*siehe Abb. 1*). Durch AR lässt sich die reale Welt dementsprechend mit zusätzlichen virtuellen Funktionen anreichern, die entweder als Informationsträger dienen oder auch als Unterhaltungsmedium genutzt werden können.

**Virtual Reality (VR)** beschreibt, wenn der Nutzer tatsächlich von der virtuellen Welt umgeben ist und dort mit virtuellen Objekten interagieren kann [19]. Dies ist mithilfe von einer Daten Brille z.B. HTC Vive oder Oculus Rift möglich.

Funktionsweise: Durch Tragen der VR-Brille wird der Nutzer komplett von der realen Welt isoliert und sieht vor sich lediglich das Bild, das auf dem Display der VR-Brille dargestellt wird. Das Bild ist dementsprechend rein virtuell und ermöglicht das vollständige Eintauchen des Nutzers in die virtuelle Welt.

**Mixed Reality (MR)** ist eine Kombination zwischen VR und AR bei der die reale und virtuelle Welt miteinander vereint werden. Die Besonderheit besteht darin, dass physische und digitale Objekte in beiden Welten koexistieren und in Echtzeit veränderlich sind. Um MR zu verstehen ist es wichtig zu wissen das es verschiedene Systeme gibt, die als Mixed Reality gekennzeichnet werden können. Milgram und Kishino [14] haben folgende Klassifizierung aufgestellt:

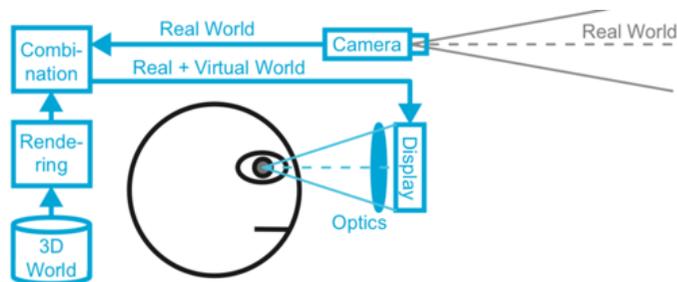
- Manche Systeme sind primär Video basiert und werden durch Computergrafik erweitert, während andere primär auf Computergrafik basieren und durch Videos erweitert werden.
- In manchen Systemen wird die reale Welt direkt angezeigt (durch Luft oder Glas), während in anderen die physischen Objekte gescannt und verarbeitet auf einem Display Gerät wiedergegeben werden.

- In manchen Systemen ist es zwingend erforderlich ein 1:1 Verhältnis der Proportionen vom virtuellen und realen Bild darzustellen während bei anderen Systemen eine Skalierung weniger kritisch oder gar unwichtig ist. Nach Betrachtung der oben genannten Unterschiede lässt sich sagen, dass ein Mixed Reality System sich überwiegend von dessen Definition zu anderen Systemen abgrenzt und der Begriff sich deshalb auf ein weites Spektrum von virtuellen Systemen anwenden lässt. Aus diesem Grund ist die Verwendung des Begriffs MR noch etwas strittig in der Forschung und wird folglich noch viel untersucht.

## 1.2 Wie kann man Augmented Reality darstellen?

Für die Nutzung von AR ist es notwendig, eine digitale Bildschirmfläche zu haben, worauf der virtuelle Inhalt dargestellt werden kann. Dabei kann es sich um gängige Optionen wie z.B. einen Monitor handeln oder für mobile Zwecke dann auch ein Tablet oder Smartphone. Im Folgenden werden verschiedene Darstellungsmöglichkeiten näher beschrieben.

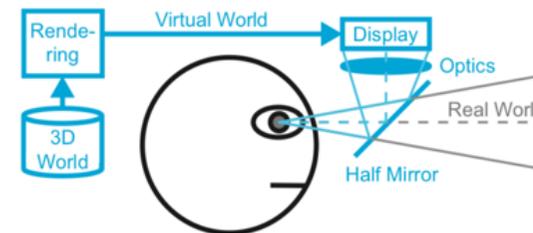
**Mobilgeräte** Nachdem sich die Gesellschaft heute stark in Richtung Portabilität bewegt, entwickelt sich entsprechend auch der technische Markt sehr schnell. Dadurch haben Smartphones und Tablets heutzutage bereits genügend Rechenleistung um durchaus anspruchsvolle und leistungsintensive Grafikberechnungen durchzuführen. Dies ermöglicht das Nutzen von AR-Applikationen zur Echtzeit-Bildverarbeitung, unter anderem Vuforia, einer Softwareschnittstelle zur Erkennung von Freihand-Gesten mithilfe der eingebauten Kamera. Mobilität wird aber nicht nur durch die Nutzung von Mobilgeräten gewährleistet, sondern kann auch mit Daten Brillen bewerkstelligt werden.



■ **Abbildung 3** Funktionsweise von video-basierten AR-Brillen [4].

**Datenbrillen** Datenbrillen, auch sogenannte Head-Mounted-Displays (HMD) werden am Kopf getragen und schränken somit die Nutzung der Hände nicht weiter ein. Man unterscheidet bei Datenbrillen zwischen video-basierter Durchsicht und optischer Durchsicht.

Bei video-basierter Durchsicht (*siehe Abb. 3*) werden die Augen des Nutzers komplett vom Display umschlossen, sodass ein Bild der realen Welt lediglich über eine Kamera übertragen werden könnte. Diese Methode ermöglicht es, das virtuelle und reale Bild ohne weitere Verzögerungen darzustellen und bietet zusätzlich die Option, das Bild nachträglich aufzuhellen oder zu verdunkeln. Nachteil hierbei wäre, dass das reale als auch das virtuelle Bild etwas verzögert auf reale Bewegungen reagiert und es entsprechend zu Bewegungsübelkeit kommen kann. Weiteres Problem ist die Abhängigkeit von einem Computer, der das Bild für das Display verarbeitet, d.h. man ist örtlich gebunden.



■ **Abbildung 4** Funktionsweise von optischen transparenten AR-Brillen [4].

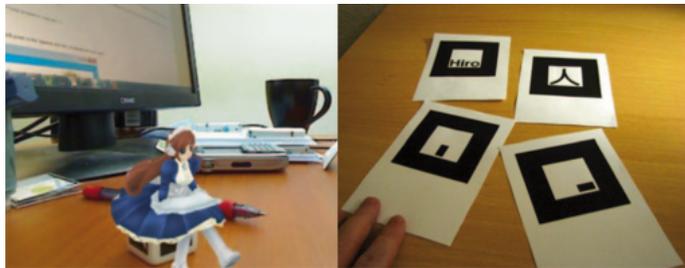
Bei optischer Durchsicht (*siehe Abb. 4*) wird das virtuelle Bild auf einen Halbspiegel projiziert, der für das Auge aber weiterhin durchsichtig bleibt und somit dem Nutzer ermöglicht, die reale Umgebung durch den Halbspiegel zu sehen. Entsprechend besteht ein Vorteil darin, dass das reale Bild nicht zusätzlich verarbeitet werden muss. Führend in diesem Bereich ist unter anderem die Microsoft HoloLens, die bereits viel für Forschung und Entwicklung in AR und MR genutzt wird.

**Projektion** Eine weitere Möglichkeit zur Darstellung von dreidimensionalen Inhalten ist die Nutzung von Holografie. Durch spezielle Nutzung von Lasern, Interferenzen und Lichtbrechung ist es möglich, bei optimaler Beleuchtung ein dreidimensionales Bild in die Luft zu zeichnen [19]. Die Position und Ausrichtung des Bildes verändert sich gemäß dem Betrachtungswinkel des Nutzers, wodurch das Hologramm sehr realistisch erscheint. Angewendet wird das System überwiegend auf Konzertbühnen, bei denen man virtuelle Objekte als Teil der Vorstellung einblendet. Es besteht jedoch ein großes Potential

darin, diese Technik auch für Augmented Reality zu verwenden, da es dem Nutzer ermöglicht, virtuelle Inhalte mit dem bloßen Auge zu sehen.

### 1.3 Interaktionsmöglichkeiten in Augmented Reality

Um mit virtuellen Umgebungen zu interagieren, ist es immer notwendig eine Eingabeschnittstelle zwischen realer und virtueller Welt zu haben. Dies können zusätzliche Eingabegeräte sein, die vom Nutzer bedient werden müssen, oder aber auch Sensoren sein, die Bewegungen erkennen und diese dann in Eingaben umwandeln. Im folgenden Abschnitt werden verschiedene Eingabemethoden näher erläutert.



■ **Abbildung 5** Rechts: Beispiele für QR-Codes als Marker für AR. Links: Virtuelle Figur, die einen Marker als Ankerpunkt für die Darstellung verwendet [19].

#### 1.3.1 Marker basiertes AR

Ein großes Problem der Computergrafik war es bisher, mithilfe von Kameras eine reale Umgebung richtig auszuwerten und zu erkennen, aufgrund der hohen Komplexität der Formen und Materialien, wobei hier von der benötigten Rechenleistung abgesehen wurde [19]. Ein weiteres Problem ist, dass es schwer ist, eine Ankerfläche zu erkennen, auf der man das virtuelle Objekt zeichnen kann. An dieser Stelle kommen Marker ins Spiel (*siehe Abb. 5*), die generell aus einem quadratischen QR-Code<sup>3</sup> bestehen und von einer digitalen Kamera eingelesen werden können. Anhand des hohen Kontrasts des Quadrats und des Winkels der Kanten lässt sich ein 3D-Koordinatensystem erzeugen, in das das virtuelle Objekt dann eingezeichnet wird. Auch heute finden sich noch viele Einsatzgebiete für Marker-Erkennung, da es sehr primitiv in der Handhabung ist z.B. zur Demonstration von Möbeln oder anderen 3D Objekten.

<sup>3</sup> zweidimensionaler, elektronisch lesbarer Code (vergleichbar mit Barcode)

Die Interaktion erfolgt generell durch Fingermanipulation auf dem Display bei Nutzung von mobilen Geräten oder durch gängige Eingabegeräte.

#### 1.3.2 Markerloses AR

Im Vergleich zur Marker-basierten Methode bietet die markerloses AR die Erkennung von statischen Objekten in Echtzeit. Dies ermöglicht die Interaktion mit virtuellen Inhalten durch Manipulation von echten haptischen Objekten [19]. Die Erkennung basiert dabei auf tiefererkennenden Kameras, die durch die zusätzliche Tiefendimension ein Objekt von der Oberfläche unterscheiden können. Diese Methode erweitert die Bandbreite der Interaktion zwischen Mensch und Computer und lässt sich nicht nur auf Objekte anwenden, sondern auch zur Erkennung von Händen, ohne dabei zusätzliche Marker oder tragbare Geräte zu verwenden.

#### 1.3.3 Freihand Interaktion

Durch Nutzung von tiefererkennenden Kameras ist es heutzutage möglich die Hände präzise und konsistent zu erkennen und als Eingabemittel zur Interaktion zu verwenden. Die Verwendung von Mobilgeräten oder herkömmlichen Eingabegeräten setzen das Halten des Gerätes voraus und schränken entsprechend die Interaktion mit den Händen ein. Die Forschung von Freihand basierter Interaktion ergab eine erhöhte Natürlichkeit und eine intuitive als auch effektivere Nutzung von virtuellen Oberflächen [16]. Des Weiteren werden neue Pforten beim Entwickeln von Interaktionsmustern geöffnet, die man auch während einer Bewegung verwenden kann, was bei WIMP<sup>4</sup>-basierter Interaktion nicht möglich ist. Zusätzlich sind Menschen durch ihre tägliche händische Interaktion mit physikalischen Objekten bereits so darauf abgestimmt, sodass viele Aktionen nur minimale visuelle Aufmerksamkeit erfordern [9]. Rückschließend sollte eine optimale Interaktionsoberfläche die Vielfalt der menschlichen Sinne wahrnehmen die aus unseren täglichen Aktivitäten hervortreten. Aus der Studie von Billinghurst et. al. [3] über Gestensteuerung mit AR Oberflächen ergab sich, dass Probanden dazu tendieren für die Manipulation von Objekten in der virtuellen Welt gängige Gesten wie Strecken oder Drücken probieren, wie sie es sonst bei einem echten physischen Objekt machen würde. Auf Nachfrage welche Gesten die Probanden selber entwerfen würden haben die meisten eine Kombination aus symbolischen und metaphorischen Gesten vorgeschlagen, die in aus der echten Welt reflektiert werden [3]. Somit bietet sich die Nutzung

<sup>4</sup> steht für "Windows, Icons, Menus and Pointers" und bezeichnet das Grundkonzept von grafischen Benutzerschnittstellen



■ **Abbildung 6** Marktvolumen von Augmented- und Virtual Reality ab 2016 mit Vorhersage für 2022 [15]

von Händen im Zusammenhang mit Gesten für die Interaktion mit virtuellen Umgebungen an.

## 2 Motivation

AR- und VR-Applikationen sind in der Unterhaltungsbranche stark vertreten. Nimmt man sich in diesem Zusammenhang noch das Marktvolumen (siehe Abb. 6) von VR und AR in den letzten Jahren her, dann sieht man einen starken Anstieg. Dies ist darauf zurückzuführen, dass diese im Laufe der Jahre immer erschwinglicher wurden. Dieser Preisverfall macht es unter anderem auch für Kliniken attraktiver, sich mit dieser Technik näher auseinanderzusetzen, denn genau hier kann man großen klinischen Problemen entgegenreten. Ein großes Risiko besteht unter anderem darin, dass Ärzte im Operationssaal aufgrund der strikten Einhaltung von Sterilität die Geräte nicht selbständig und nur eingeschränkt verwenden können. Dies beinhaltet Aufgaben wie z.B. die Maus zu bedienen oder ein Röntgenbild zu verschieben. All diese Aufgaben werden im Normalfall dann von anderen Ärzten übernommen, was nicht nur die klinischen Kosten erhöht, sondern auch das klinische Risiko [17]. Genau an dieser Stelle ist der Einsatz von Freihand-Interaktion von großer Bedeutung.

## 3 Stand der Technik zur Erkennung von Händen

Eine Hand hat 26 Freiheitsgrade [1], was bedeutet, dass sie im Vergleich zu Objekten mit 6 Freiheitsgraden deutlich schwerer zu erkennen ist. Meistens

werden Kamerasysteme verwendet, bei denen Parameter wie Tiefe, Farbe oder Kontrast ausgelesen werden, um bewegliche Objekte wie die Hände von anderen Umgebungsinformationen zu trennen. Im folgenden Abschnitt werden verschiedene Ansätze und Systeme vorgestellt.

### 3.1 Hardware

Viele Firmen versuchen sich daran, eine Hardwareschnittstelle zu erstellen, die die Interaktion zwischen Mensch und Computer so einfach wie möglich gestaltet. Ein wichtiger Aspekt liegt darin, die Herstellungskosten so niedrig wie möglich zu halten und zu gleich aber die neuste Technik bei der Gesten-/Handerkennung zu verwenden.

Das erste erschwingliche System war die Microsoft Kinect<sup>5</sup>, die als Steuerungsgerät für die Spielekonsole gedacht war. Statt der Nutzung von herkömmlichen Controllern konnte man mithilfe von Körperbewegungen die Oberfläche und Spiele steuern. Durch die integrierte Infrarotkamera konnte man Tiefeninformationen auswerten und somit Personen vor der Kamera in Echtzeit erkennen. Auch Kim et. al. [12] bedient sich einer Infrarottechnik, bei der die Autoren eine IR Kamera an einen am Handgelenk getragenen Gerät befestigt haben. Diese erlaubt es dem System dann, die Fingerbewegung zu erkennen und entsprechende Befehle in einer Applikation auszulösen. Ziel dieses Projekts war es, ein System zu entwerfen, das ohne zusätzliche externe Geräte, d.h. ortsungebunden, verwendet werden konnte. Die Genauigkeit der Erkennung konnte sich bei offenhändigen Gesten sehen lassen, jedoch hatte der Sensor große Probleme bei der Erkennung von überlagernden Fingern. Eines der am weitesten verbreiteten Geräte ist die Leap Motion<sup>6</sup>, die über eine herkömmliche USB-Schnittstelle mit einem Computer verbunden wird. Die Leap Motion besteht aus drei Infrarotstrahlern und zwei Infrarotkameras, welche das Licht aufnehmen. In Verbund mit deren Merkmal-Erkennungsalgorithmus lässt sich das Skelett einer Hand komplett rekonstruieren und somit für Gestensteuerung verwenden. Vielmehr ermöglicht die Leap Motion auch das Greifen und Rotieren von virtuellen Objekten, ohne dabei durch Überlappen der Finger eingeschränkt zu werden.

### 3.2 Funktionsweise

**Active Appearance Model (AAM) [7]** Beim AAM handelt es sich um eine statistische Methode, bei der mehrere Formdaten und Texturen des Objekts miteinander verglichen und ausgewertet werden. Konkret verwendet AAM

<sup>5</sup> <https://developer.microsoft.com/de-de/windows/kinect>

<sup>6</sup> <https://www.leapmotion.com>

zuvor gesammelte Daten und gleicht sie mit den neuen Daten ab, die bspw. von der Kamera gesammelt werden und kann dann einschätzen, ob es sich um eine Hand handelt oder nicht. Im ersten Schritt werden entsprechend erstmal Formdaten von markanten Punkten des zu erkennenden Objekts gesammelt. Diese werden im zweiten Schritt durch sogenanntes „Image Warping“ normalisiert, sodass sich alle Daten in einem einheitlichen Format befinden. Im letzten Schritt lässt sich dann mit diesen Trainingsdaten ein neues Objekt klassifizieren. Im Paper von Dactu et. al. [6] ergab sich die Nutzung von AAM als sehr erfolgreich im Vergleich zur bewährten Annäherungsfunktion von Viola & Jones [21]. Bei Viola & Jones handelt es sich um einen ähnlichen Maschinen-Lernen-Algorithmus, bei dem aber andere Merkmale des Objekts zur Erkennung gesammelt werden.

**Contour-based Template Matching (CTM) [20]** Der CTM-Algorithmus basiert auf dynamischer Programmierung und kann genutzt werden um eine optimale Lösung für ein Problem zu finden. In diesem Fall kann man den Algorithmus dazu verwenden, die Vorlage einer Hand mit den Bilddaten der Kamera eines Smartphones zu vergleichen um einen optimalen Treffer zu finden. Des Weiteren können zusätzliche Parameter zur Erkennung der Haut hinzugefügt werden um die Treffergenauigkeit zu erhöhen. Diese sind jedoch abhängig von den Lichtverhältnissen und der Hautfarbe [13]. Zhihan et. al [13] haben CTM für ihre Studie mit einer Google Glass verwendet. Die Ergebnisse ergaben, dass der Algorithmus sehr gut für die Erkennung von Handgesten war, unter anderem aufgrund der geringen Fehlerrate.

Generell lässt sich sagen, dass mithilfe von Farbinformation und Tiefeninformation die Erkennung von Handeigenschaften sehr stabil funktioniert, unabhängig von der jeweiligen Umgebung. Dies ermöglicht die Nutzung von Gesten im alltäglichen Gebrauch, da hier eine Fehlfunktion oftmals fatal sein kann. Gängige Gesten wie Zeigen, Greifen und Verschieben werden bspw. bereits in aktuellen Autos zur Steuerung von Oberflächenelementen benutzt.

## 4 Einsatzgebiete von Freihand Interaktion

Die Nutzung von Freihand-Interaktion hat ein sehr breit gefächertes Einsatzgebiet, welches im folgenden Abschnitt näher beschrieben und mit Beispielen belegt wird. Ein großes Problem besteht darin, dass bei üblichen Eingabesystemen eine Berührung mit dem Eingabegerät oder dessen Steuereinheit notwendig ist. Dies ist aber in kritischen Szenarien wie Tatorten oder im Operationssaal aufgrund von Sicherheitsproblemen oder kritischen Umgebungen nicht möglich.

### 4.1 Freihand Interaktion am Tatort

Spezifisch für das Tatortszenario hat Dactu et. al. [6] eine Arbeit verfasst, bei der sie im Detail darauf eingehen, welche Anforderungen erfüllt werden müssen um Freihand-Interaktion mit AR möglich zu machen. Ziel dieser Interaktion ist es vor allem, dass eine Verbindung zu externen Experten ermöglicht wird, die bei der Untersuchung unterstützen können. Die Anforderungen unterteilen sich in folgende Punkte: (1) Um den Tatort so unverändert wie möglich zu halten ist eine markerlose Ausrichtung des virtuellen Inhalts notwendig. (2) Zusätzliche Geräte wie Datenbrillen sollten so leicht wie möglich sein um den Ermittler nicht zusätzlich zu belasten. (3) Die Interaktion muss Freihand erfolgen, sodass die Hände physisch mit dem Tatort interagieren können. (4) Des Weiteren muss eine reibungslose Kommunikation zum externen Ermittler gegeben sein. Folglich ist das einzig erlaubte Zusatzgerät eine Datenbrille. Die Nutzung von Spracheingabe ist nur bedingt möglich aufgrund der oftmals sehr lauten Umgebungsgeräusche bei einem Tatort, wodurch nur eine Freihand-Interaktion in Erwägung gezogen werden kann. Die Evaluierung dieser Studie ergab, dass Selektion und Navigation mit dem AR-System ziemlich erfolgreich war und von heutiger Computergrafik gut bewerkstelligt wurde. Die Auswertung der quantitativen Daten erschloss, dass Probanden sich sehr schnell an den Freihand Interaktionsstil anpassen konnten.

### 4.2 Freihand Interaktion im Operationssaal

Ein weiteres Szenario ist die Nutzung von Freihand-Interaktion im Operationssaal. Viele Operationen setzen eine Interaktion des behandelnden Arztes mit den technischen Geräten voraus, die aber viele Einschränkungen mit sich bringt. Die größte Hürde besteht darin, die Grenze zwischen etwas Sterilem und etwas nicht Sterilem aufrecht zu halten, d.h. sobald ein Arzt seinen Mantel und seine Handschuhe angezogen hat, sollte dieser nicht mehr in Kontakt mit anderen Gegenständen kommen, die die Sterilität beeinflussen könnten. Darunter befinden sich auch sämtliche Eingabegeräte oder Gegenstände, die mit den Händen benutzt werden müssten. Eine gängige Lösung für das genannte Problem ist es, einen Assistenten um Hilfe bei der Bedienung zu bitten, bspw. um ein Ultraschall-Bild zu drehen oder eine Gerätschaft an und auszuschalten. Großer Nachteil bei dieser Lösung ist, dass oftmals viele zusätzliche Assistenten notwendig sind, was die klinischen Kosten erhöht aber vor allem auch das klinische Risiko. Indirekte Manipulation ist insbesondere nicht geeignet für analytische und interpretative Aufgaben. Studien zeigen, dass Ärzte die direkte Kontrolle von Bilddaten benötigen, um mental ein besseres Verständnis für die Operation bzw. das Problem zu bekommen [17]. Abbildung 7 zeigt ein Gestensystem zur Behandlung von Gefäßen, welches bereits am Universitätsklinikum



■ **Abbildung 7** Gestensystem für Gefäßchirurgie am Universitätsklinikum in Cambridge [17]

in Cambridge zum Einsatz kommt.

Die Nutzung von Freihand-Gesten liegt demnach also auf der Hand und ist mit dem heutigen Stand der Technik auch definitiv möglich. Jedoch besteht hier an nächster Stelle das Problem, wie man ein System für eine Operation unter bestimmten Umständen und Erfordernissen entwirft. Ein Problem besteht darin, dass je mehr Gesten das System zulässt, die umso unzuverlässiger Erkennung wird. Zusätzlich ist es aber notwendig, dass das System so viele Funktionen wie möglich mit sich bringt, was gegensätzlich zur Gestenerkennung steht. An dieser Stelle muss also ein Kompromiss zwischen Funktionsumfang und Funktionalität geschlossen werden, was es nicht zu vernachlässigen gilt. Das nächste Problem besteht im Kollaborationsaspekt einer Operation. Mehrere Ärzte arbeiten generell gemeinsam an einer Behandlung und dem entsprechend ist es sinnvoll, dass das System sich von mehreren Ärzten steuern lässt. Diese Übergabe sollte möglichst flüssig ablaufen, sodass keine unnötigen Verzögerungen im Operationsaal entstehen. Eine gängige Lösung ist die Nutzung von unnatürlichen Gestern über dem Kopf um die Steuerung zu übernehmen oder abzugeben, um ein versehentliches Auslösen zu verhindern.

### 4.3 Freihand Interaktion in Spielen

Auch in der Spieleindustrie findet die Nutzung von Augmented Reality immer mehr Einsatz, da es die eindringliche Spielerfahrung verbessert [13]. Des Weiteren spielt auch eine intuitive Art der Interaktion eine große Rolle, da gerade



■ **Abbildung 8** AR Spiele mit Hand und Fußerkennung durch die integrierte Kamera des Handys. Abgebildet ist links ein Jonglierspiel, mittig ein Fußballspiel und rechts ein Klavierspiel [13].

bei Spielen der erste Eindruck ausschlaggebend dafür ist, ob der Spieler es weiterspielen möchte oder nicht. Auch hier treffen Entwickler auf die Probleme bei der Gestaltung von Steuerungsmöglichkeiten auf zweidimensionalen Bildschirmen. Am Beispiel von Smartphones würde bei Fingerberührung das Problem von Verdeckung auftreten, wodurch die Spielerfahrung stark eingeschränkt wird. Die Nutzung von berührungslosen Methoden kristallisiert sich an dieser Stelle als die natürlichste Möglichkeit, welche den Spieler mehr auf das Spiel fokussieren lässt anstatt auf die Steuerung des Spiels selbst. Die Ergebnisse von Zhihan et. al. [13] ergaben, dass die Probanden bereit wären AR Spiele mit Freihand-Interaktion in der Öffentlichkeit zu verwenden. Des Weiteren war die Nutzung der Google Glass sehr intuitiv und komfortabel für die Probanden. *Abb. 8* zeigt die verwendeten AR-Spiele der Studie, die durch Hand und Fußerkennung der integrierten Kamera des Handys gesteuert werden.

### 4.4 Kulturelles Lernen in AR Umgebungen mit Freihand basierter Interaktion

In einer globalisierten Gesellschaft wie heute ist es wichtig die unterschiedlichen Kulturen zu verstehen. Als kulturelles Lernen wird an dieser Stelle der Prozess bezeichnet bei dem man Wissen, Fähigkeiten und Einstellungen erlernt, die notwendig sind um die Kommunikation und Interaktion mit Menschen aus anderen Ländern zu ermöglichen [18]. Folglich ist es essentiell beim kulturelles Lernen auch die entsprechende Sprache des Landes zu erlernen. Gängiger

Weise erfolgen beide Schritte während eines Austauschsemesters, bei dem ein Student oder Schüler in ein fremdes Land reist und dort an interkulturellen Aktivitäten und Sprachkursen teilnimmt. Oftmals trifft aber der Fall zu, dass die potenziellen Lerner nicht die Zeit oder das Geld haben, um für längere Zeit ein fremdes Land zu bereisen. Yang und Liao haben ein System entworfen, bei dem AR genutzt wird um virtuelle kulturelle Inhalte zusätzlich zum Englisch Kurs darzustellen. Die virtuellen Inhalte können dabei mit Freihand Gesten verändert und bewegt werden. Besucht wird der Kurs entweder physisch in einem echten Klassenzimmer oder durch externe Zuschaltung über das Internet. [22]

Das System bietet drei Möglichkeiten der Interaktion an: (1) Interaktion zwischen Schülern und Lehrer - Schüler können ein virtuelles Objekt von der Projektionsfläche des Lehrers durch eine Geste greifen und diese auf ihren eigenen virtuellen Arbeitsplatz ablegen. (2) Interaktion zwischen Schülern und Lehrmaterialien - Schüler können virtuelle Objekte mit Handgesten bewegen, rotieren und skalieren. (3) Interaktion zwischen den Schülern - Schüler können ihre virtuellen Objekte mit ihren Tischnachbarn teilen. [22]

Die Ergebnisse der Studie ergaben, dass die Nutzung von Hand-Gesten als sehr natürlich und intuitiv empfunden wurden. Des Weiteren erhöhte die spielerische Gestaltung der Lehrinhalte den kulturellen Lernerfolg und regte zur zwischenmenschlichen Kommunikation zwischen Lehrer und Schülern an. [22]

## 5 Ausblick

Dank der schnell fortschreitenden Technologie ist das Erkennen von Händen heutzutage keine Hürde mehr und eröffnet deshalb neue Pforten des Interaktionsdesigns im Bereich VR und AR. Die Nutzung von Gestensteuerung bietet nicht nur Lösungswege für altbekannte 2D Fingerberührungsproblemen sondern wird zugleich von den Nutzern als natürlicher und intuitiver wahrgenommen [2]. Auch gibt es Szenarios in denen die Freihand basierte Gestensteuerung gegenüber den traditionellen Eingabemethoden schneller und präziser von Statten geht [10]. Gerade im Falle von direkter Manipulation eines virtuellen 3D Objektes sind Freihand Gesten klar im Vorteil, da der Anwender bereits die nötigen Vorkenntnisse aus der realen Welt mit sich bringt [3].

Des Weiteren bietet es sich für die zukünftige Forschung an die Kombination von Gestensteuerung und Sprachsteuerung näher zu untersuchen. Diese Konstellation scheint sehr vielversprechend für das Entwerfen eines vielseitigen Interaktionsdesigns für aussagekräftige AR Oberflächen.

Gestensteuerung bringt aber nicht nur Vorteile mit sich, sondern stößt auch an Herausforderungen, die nicht vernachlässigt werden sollten. Unter anderem ist es wichtig zu verstehen, welche Gesten für welche Zwecke benutzt werden

können, welche Eingabemechnismen sich miteinander kombinieren lassen und welche Technik für die Erkennung genutzt werden soll [17]. Konkret bedeutet das für Entwickler, dass sie ein klares Verständnis darüber besitzen müssen, wo das System zum Einsatz kommt und inwiefern es Sinn machen würde, ein bereits bestehendes traditionelles Eingabesystem durch ein Neues zu ersetzen. Am Beispiel einer Klinik würde das bedeuten, dass ein Freihand basiertes Interaktionssystem nicht einfach nur das traditionelle System ablösen soll aufgrund der Sterilität, sondern es zusätzlich wichtig ist zu verstehen, was Ärzte mit dem System tatsächlich erreichen möchten [17]. Gerade in solchen kritischen Fällen ist es von großer Wichtigkeit, sich mit allen Parametern im Klaren zu sein, sodass keine verheerenden Fehler passieren.

Ein weiterer Punkt besteht darin, dass Freihand basierte Interaktion nicht geeignet ist für Anwendungen, die längere Manipulationsaufgaben von Objekten beinhalten. Je länger diese Aufgaben stattfinden, desto höher ist die Belastung des Armes und entsprechend auch die körperliche Erschöpfung [2]. Dies muss beim Interaktionsdesign mit berücksichtigt werden.

Letzteres können mit heutiger Erkennungstechnik nicht nur die Hände als Eingabemittel verwendet werden, sondern auch andere Körperteile. Unter anderem hat Jackowski et. al. [11] ein System entwickelt, dass mithilfe von Kopfbewegungen gesteuert wird. Dies soll vor allem körperlich eingeschränkten Menschen die Möglichkeit bieten mit virtuellen Inhalten in Kontakt treten zu können oder auch allgemein als Hilfsmittel für den Alltag genutzt werden können. Des Weiteren sind Kopfgesteuerte Systeme aufgrund der Erkennung ähnlich zu Handgesteuerten System, worin der Vorteil besteht, dass Funktionalitäten unter Umständen übertragbar sind. Jedoch bleibt hier noch viel Forschungsbedarf.

Zusammenfassend öffnen Freihand Gesten neue Möglichkeiten der Interaktion mit virtuellen Umgebungen und Objekten und kommen durch ihre Vielseitigkeit mit vielen Vor- und Nachteilen, die bei der Einbindung in Systeme berücksichtigt werden sollten. Entsprechend bietet dieses Thema viel Freiraum für zukünftige Forschung und wir können gespannt sein, was in den nächsten Jahren auf uns zu kommt.

---

## Literatur

- 1 Irene Albrecht, Jörg Haber, and Hans-Peter Seidel. Construction and animation of anatomically based human hand models. In *Proceedings of the 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, SCA '03, pages 98–109, Aire-la-Ville, Switzerland, Switzerland, 2003. Eurographics Association. URL: <http://dl.acm.org/citation.cfm?id=846276.846290>.

- 2 Huidong Bai, Gun A. Lee, Mukundan Ramakrishnan, and Mark Billinghurst. 3d gesture interaction for handheld augmented reality. In *SIGGRAPH Asia 2014 Mobile Graphics and Interactive Applications*, SA '14, pages 7:1–7:6, New York, NY, USA, 2014. ACM. URL: <http://doi.acm.org/10.1145/2669062.2669073>, doi:10.1145/2669062.2669073.
- 3 M. Billinghurst, T. Piumsomboon, and H. Bai. Hands in space: Gesture interaction with augmented-reality interfaces. *IEEE Computer Graphics and Applications*, 34(1):77–80, Jan 2014. doi:10.1109/MCG.2014.8.
- 4 Andreas Butz. Lecture: Mensch-maschine-interaktion 2. Slides 6: VR + AR primer, 2017.
- 5 Lieu-Hen Chen, Chi-Jr Yu, and Shun-Chin Hsu. A remote chinese chess game using mobile phone augmented reality. In *Proceedings of the 2008 International Conference on Advances in Computer Entertainment Technology*, ACE '08, pages 284–287, New York, NY, USA, 2008. ACM. URL: <http://doi.acm.org/10.1145/1501750.1501817>, doi:10.1145/1501750.1501817.
- 6 Dragos Dancu and Stephan Lukosch. Free-hands interaction in augmented reality. In *Proceedings of the 1st Symposium on Spatial User Interaction*, SUI '13, pages 33–40, New York, NY, USA, 2013. ACM. URL: <http://doi.acm.org/10.1145/2491367.2491370>, doi:10.1145/2491367.2491370.
- 7 G. Hackenberg, R. McCall, and W. Broll. Lightweight palm and finger tracking for real-time 3d gesture control. In *2011 IEEE Virtual Reality Conference*, pages 19–26, March 2011. doi:10.1109/VR.2011.5759431.
- 8 Anders Henrysson, Joe Marshall, and Mark Billinghurst. Experiments in 3d interaction for mobile phone ar. In *Proceedings of the 5th International Conference on Computer Graphics and Interactive Techniques in Australia and Southeast Asia*, GRAPHITE '07, pages 187–194, New York, NY, USA, 2007. ACM. URL: <http://doi.acm.org/10.1145/1321261.1321295>, doi:10.1145/1321261.1321295.
- 9 Valentin Heun, Shunichi Kasahara, and Pattie Maes. Smarter objects: Using ar technology to program physical objects and their interactions. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, pages 2939–2942, New York, NY, USA, 2013. ACM. URL: <http://doi.acm.org/10.1145/2468356.2479579>, doi:10.1145/2468356.2479579.
- 10 Zhanpeng Huang, Weikai Li, and Pan Hui. Ubii: Towards seamless interaction between digital and physical worlds. In *Proceedings of the 23rd ACM International Conference on Multimedia*, MM '15, pages 341–350, New York, NY, USA, 2015. ACM. URL: <http://doi.acm.org/10.1145/2733373.2806266>, doi:10.1145/2733373.2806266.
- 11 Anja Jackowski and Marion Gebhard. Evaluation of hands-free human-robot interaction using a head gesture based interface. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '17, pages 141–142, New York, NY, USA, 2017. ACM. URL: <http://doi.acm.org/10.1145/3029798.3038298>, doi:10.1145/3029798.3038298.
- 12 David Kim, Otmar Hilliges, Shahram Izadi, Alex D. Butler, Jiawen Chen, Iason Oikonomidis, and Patrick Olivier. Digits: Freehand 3d interactions anywhere using a wrist-worn gloveless sensor. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*, UIST '12, pages 167–176, New York, NY, USA, 2012. ACM. URL: <http://doi.acm.org/10.1145/2380116.2380139>, doi:10.1145/2380116.2380139.
- 13 Zhihan Lv, Alaa Halawani, Shengzhong Feng, Shafiq ur Réhman, and Haibo Li. Touch-less interactive augmented reality game on vision-based wearable device. *Personal and Ubiquitous Computing*, 19(3):551–567, Jul 2015. URL: <https://doi.org/10.1007/s00779-015-0844-1>, doi:10.1007/s00779-015-0844-1.
- 14 Paul Milgram and Fumio Kishino. A taxonomy of mixed reality visual displays. vol. E77-D, no. 12:1321–1329, 12 1994.
- 15 Sarah Murray. Worldwide spending on augmented and virtual reality to achieve a five-year cagr of 71.6 <https://www.idc.com/getdoc.jsp?containerId=prUS43860118>, accessed 17.07.2018.
- 16 Tao Ni, Doug A. Bowman, Chris North, and Ryan P. McMahan. Design and evaluation of freehand menu selection interfaces using tilt and pinch gestures. *International Journal of Human-Computer Studies*, 69(9):551–562, 2011. URL: <http://www.sciencedirect.com/science/article/pii/S107158191100053X>, doi:<https://doi.org/10.1016/j.ijhcs.2011.05.001>.
- 17 Kenton O'Hara, Gerardo Gonzalez, Abigail Sellen, Graeme Penney, Andreas Varnavas, Helena Mentis, Antonio Criminisi, Robert Corish, Mark Rouncefield, Neville Dastur, and Tom Carrell. Touchless interaction in surgery. 57:70–77, 01 2014.
- 18 Robert Paige, Helen Jorstad, Laura Siaya, Francine Klein, and Jeanette Colby. 1 culture learning in language education : A review of the literature. 2013.
- 19 Jeff Tang and Jordan Tewell. Emerging human-toy interaction techniques with augmented and mixed reality, 08 2015.
- 20 Shafiq ur Rehman, Abdullah Khan, and Haibo Li. Interactive feet for mobile immersive interaction. In *MobileHCI 2012 : Mobile Vision (MobiVis) – Vision-based Applications and HCI*. MOBIVIS, 2012. URL: <http://mobivis.labs-exit.de/AcceptedPapers.html>.
- 21 Paul Viola and Michael Jones. Robust real-time object detection, 01 2001.
- 22 M. T. Yang and W. C. Liao. Computer-assisted culture learning in an online augmented reality environment based on free-hand gesture interaction. *IEEE Transactions on Learning Technologies*, 7(2):107–117, April 2014. doi:10.1109/TLT.2014.2307297.

# Bildung Augmentieren: Anwendung und Herausforderungen beim Lernen mit Augmented Reality

Dimitri Hein

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

---

## Zusammenfassung

In dieser Arbeit wird ein Überblick über den Einsatz von Augmented Reality im Kontext von Bildung gegeben und bietet somit einen guten Einstiegspunkt in die entsprechende Forschung. Neben der Erklärung von Augmented Reality und modernen Technologien, werden auch durchgeführte Studien von Forschern präsentiert, die wichtige Erkenntnisse für zukünftige Studien und Projekte beinhalten. Darüber hinaus werden mögliche Anwendungsfälle, Vorteile aber auch Herausforderungen vom Einsatz von Augmented Reality beim Lernen diskutiert. Abschließend wird ein Ausblick über die mögliche Relevanz von Augmented Reality in naher Zukunft gegeben.

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

**Keywords and phrases** Augmented Reality, Lernen, Mixed Reality, Bildung

## 1 Einleitung

Im Lauf der letzten zwei bis drei Jahrzehnte wurden verschiedene Augmented Reality (AR) Technologien entwickelt, die in unterschiedlichen Bereichen der Forschung eingesetzt wurden. Durch die Veröffentlichung von teils kommerziellen Lösungen wie Apples ARKit<sup>1</sup> oder der Microsoft HoloLens<sup>2</sup> in den letzten Jahren, haben Forscher und Entwickler Zugang zu erschwinglichen AR-Technologien, um diese auch für Anwendungsfälle außerhalb von Unterhaltung und Wirtschaft einzusetzen. So findet sich ein möglicher weiterer Nutzen von AR für den Einsatz beim Lernen. Obwohl neue Technologien in der Regel nicht unmittelbar nach ihrem Aufkommen in Schulen oder Universitäten eingesetzt

---

<sup>1</sup> <https://developer.apple.com/arkit/>

<sup>2</sup> <https://www.microsoft.com/en-us/hololens>



© Dimitri Hein;

licensed under Creative Commons License CC-BY

Cite as: 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.

werden, deutet die weiter voranschreitende Digitalisierung darauf hin, dass Technologien wie AR oder Virtual Reality (VR) durchaus auch im Bereich von Bildung relevant sein könnten. So wie das Fehlen von Computern in so ziemlich allen Aspekten der Wirtschaft heutzutage undenkbar wäre, könnte man auch für mehr Einsätze von Computern im Kontext von Lernen argumentieren. AR bietet dabei aufgrund der Fusion der echten Welt und digitalen Inhalten spezielle Vorteile für das Lernen und könnte somit die Art wie wir uns Wissen aneignen positiv verändern. In dieser Arbeit werden die bisherigen Erkenntnisse aus der Forschung zusammengefasst und ein Überblick über den Stand der Forschung gegeben. Darüber hinaus wird diskutiert, warum speziell AR in diesem Kontext eine besondere Rolle spielen könnte und welche Herausforderungen und Probleme bestehen, die dem möglichen Produktiveinsatz in Zukunft im Wege stehen könnten.

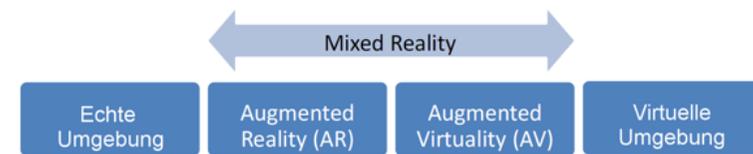
## 2 Verwandte Arbeiten

Seit dem Aufkommen von verschiedenen AR-Technologien wurden zahlreiche Studien durchgeführt, die Vor-, Nachteile und Herausforderungen beim Einsatz von AR beim Lernen untersucht haben. Um allgemeinere Schlussfolgerungen aus den verschiedenen Studien zu ziehen, wurden Literatur-Reviews und Berichte erstellt. Einer der am häufigsten zitierten Berichte über die Anwendungsgebiete und möglichen Vorteile vom Einsatz von AR in Bildung ist die Untersuchung von Azuma aus dem Jahr 1997 [2]. Neben dem allgemeinen Status von AR wurden sechs Bereiche auf Basis von Studien und Projekten wie beispielsweise Medizin identifiziert und auf potenzielle Nutzen untersucht.

In einer weiteren Arbeit aus dem Jahr 2013 beschäftigen sich Wu et al. [36] mit dem Konzept von AR im Kontext von Lernen und allgemeinen Vorteilen als auch speziellen Eigenheiten von Lernmethoden, die auf AR basieren. Diese Arbeit stellt auch Probleme und Limitationen von AR dar, die Forscher in Zukunft helfen sollen bessere Ansätze und Applikationen zu entwerfen. Einer der aktuellsten Berichte ist die Arbeit von Diegmann et al. [13], in der auf Basis von 25 Studien Vorteile von AR in Bezug auf bestimmte Anwendungsfälle identifiziert werden. Darüber hinaus werden Lern- und Lehrkonzepte mit AR erläutert.

## 3 Augmented Reality

In bisher publizierter Literatur wurde AR sowohl in Hinblick auf Konzeption, Technologien als auch der Beziehung zu VR erforscht. So wurden im Laufe der letzten drei Jahrzehnte mehrere Ansätze für die Definition von AR vorgestellt. Milgram et al. beispielsweise, definieren AR basierend auf einem Realität-



■ **Abbildung 1** Realität-Virtualität-Kontinuum basierend auf der Arbeit von Milgram et al. [28].

Virtualität-Kontinuum, wobei auf eine bereits bestehende sehr allgemeine Definition verwiesen wird, die besagt, dass AR durch die Verwendung von simulierten Hinweisen natürliches Feedback<sup>3</sup> erweitert [28]. VR auf der anderen Seite wird häufig durch das immersive Eintauchen in eine synthetische Welt definiert, welche sowohl der echten Welt entsprechen oder auch die Grenzen der physischen Realität sprengen kann [28]. Das Klassifizieren von AR und VR in einem Kontinuum ist demnach sinnvoll, da VR nicht unbedingt als ausschließender Gegensatz zu AR existiert, sondern eine Beziehung zwischen beiden Konzepten besteht. In der Abbildung 1 wird das Realität-Virtualität-Kontinuum dargestellt. Interessant ist hierbei der Begriff "Mixed Reality", der den Bereich zwischen der echten Umgebung und einer virtuellen Umgebung beschreibt. Dieser Bereich beschreibt also jene Umgebungen, in denen sowohl echte als auch simulierte Objekte präsentiert werden [28]. Wenn AR also zum Großteil aus Objekten aus der echten Welt besteht, dann kann Augmented Virtuality (AV) schlussfolgernd als Gegenpol gesehen werden, bei dem zu einer hauptsächlich virtuellen Umgebung echte Objekte (z.B. in Form von Bildern oder Texturen) hinzugefügt werden.

Ein weiterer wichtiger Ansatz in Hinblick auf die zweite eher einschränkende Definition, die von Milgram et al. [28] erwähnt wird ist das Betrachten von AR als Konzept und nicht als konkrete Technologie. So wird jene Definition, die die Verwendung eines transparenten Datenhelms für die freie Sicht auf die reale Welt beinhaltet von Wu et al. [36] als weniger produktiv angesehen, da AR nicht auf eine bestimmte Technologie beschränkt ist und zukünftige Technologien für AR eingesetzt werden könnten. Obwohl AR auf Technologien basieren kann, sollte es konzeptionell nicht nur in diesem Kontext betrachtet werden [36].

<sup>3</sup> Hierbei ist meist die echte physische Realität gemeint, die durch virtuelle Entitäten angereichert wird.

### 3.1 Augmented Reality Technologien

Forschern und Anwendern stehen mittlerweile mehrere verschiedene AR Technologien zur Verfügung, wobei diese grundsätzlich in Hardware- und Software-Aspekten charakterisiert werden können. Bezüglich Positionierung der Darstellung gibt es folgende Kategorien [33] mit Beispielen:

- **Auf dem Kopf tragend:** Microsoft HoloLens
- **In der Hand gehalten:** Smartphone mit Kamera (markerbasierend oder markerlos)<sup>4</sup>
- **Räumlich:** 3D-Hologramme oder Projektionen die statisch innerhalb der Umgebung platziert werden

Heutzutage ist einer der am weitesten verbreiteten Ansätze Inhalte mithilfe von AR anzuzeigen das Überlagern (Superimposing) von virtuellen Objekten mit der realen Umgebung. Dies geschieht entweder mithilfe von Echtzeit-Aufnahmen von Kameras oder das Darstellen von Hologrammen in optischen, durchsichtigen Materialien wie bei der Microsoft HoloLens. Innerhalb der oben aufgezählten Kategorien gibt es verschiedene Technologien, die Forschern und Entwicklern zur Verfügung stehen. Bereits im Jahr 2012 wurden folgende drei Klassifikationen [8] bestimmt, die heute vor allem im mobilen Bereich verbreitet sind (siehe Abbildung 2):

- **Marker-basierte AR:** Für das Darstellen von virtuellen Inhalten wird ein (QR-)Code über Computer-Vision erkannt, sodass die Position und die Ausrichtung präzise ermittelt werden kann. Die virtuellen Inhalte können dann so transformiert werden, dass der Anschein erweckt wird, dass das Objekt sich an der gleichen Stelle wie der Code im Raum befindet.
- **Markerlose AR:** Bei diesem Ansatz werden keine Marker benötigt, stattdessen wird die Umgebung über ein RGB Bild oder sogar komplexere Kamera-Systeme insoweit auf Merkmale untersucht, dass beispielsweise die grobe Geometrie erkannt wird. So kann bestimmt werden wo sich das AR-Gerät relativ zur Umgebung befindet. Mit diesen Informationen können die virtuellen Objekte so transformiert werden, dass es so aussieht als ob sich diese im Raum befinden.
- **Geotagging:** Hierbei werden sogenannte “Geotags” mithilfe von GPS-Koordinaten erstellt und können mit Inhalten wie Videos, Bildern und Texten gekoppelt werden. Ein bekanntes Produkt welches Geotagging unterstützt ist Wikitude<sup>5</sup>.

<sup>4</sup> Mit Markern sind hierbei Bilder wie zum Beispiel QR-Codes gemeint

<sup>5</sup> <https://www.wikitude.com/>



(a) Marker-basierter Ansatz für die Platzierung von Sonnen- und Mondmodellen [17] (b) Markerlose AR für iOS-Geräte durch Apple's ARKit [31] (c) Geo-Tagging Lösung von Wikitude [35]

#### ■ Abbildung 2 Augmented Reality Technologien

Vor allem für moderne Smartphones existieren Software-Frameworks von verschiedenen Anbietern wie zum Beispiel ARKit, ARCore<sup>6</sup> und Vuforia<sup>7</sup>. In einem systematischen Bericht von Fotaris et al. [18] wurden 17 Studien untersucht, wobei Marker-basierte Ansätze 53%, markerlose Ansätze 17% und Standort-basierte Ansätze 24% ausmachten. Zum Zeitpunkt des Reviews wurde die Beliebtheit der Marker-basierten Ansätze mit der Möglichkeit begründet, dass diese eine effizientere und stabilere Erkennung ermöglichten. Mit der Entwicklung von Lösungen wie Apples ARKit oder Vuforia könnte sich dies allerdings in den nächsten Jahren zugunsten von markerlosen Ansätzen entwickeln.

## 4 Lernen mit Augmented Reality

Da Lernen mit AR durch verschiedene Konzepte und Technologien, wie zum Beispiel AR-Bücher, Mixed-Reality Spiele oder computergestützte Instruktion [4, 11, 16] realisiert werden kann, wird in diesem Abschnitt darauf eingegangen welche allgemeinen Ansätze existieren und welche Einsatzgebiete von der bisherigen Forschung identifiziert wurden. Darüber hinaus werden wichtige Faktoren, die das erfolgreiche Lernen vorhersagen besprochen.

### 4.1 Allgemeine Faktoren für gutes und effizientes Lernen

Welche Faktoren und Aspekte sich beim Lernen und Lehren positiv auswirken ist schon ein für sich allein weites Forschungsgebiet und kann aus diesem Grund nicht zur Gänze diskutiert werden, da dies den Rahmen dieser Arbeit

<sup>6</sup> <https://developers.google.com/ar/>

<sup>7</sup> <https://www.vuforia.com/>

sprengen würde. Allerdings können einige Faktoren, die vor Allem in Bezug auf Lernen mit AR besonders relevant sind identifiziert werden. So beschreibt Gee [19], dass Konzepte aus Computerspielen in anderen Umfeldern wie Schulen, Universitäten oder sogar der Arbeitswelt integriert werden und so das Lernen verbessern könnten. Obwohl Gee nicht konkret vom Lernen mit AR spricht, sind viele Konzepte auf AR anwendbar.

Einer der wichtigsten Faktoren beim Lernen ist die Motivation [19], sodass ein Umfeld welches beispielsweise Schüler auf längere Zeit motiviert bessere Lernerfolge einführt. Eine mögliche Definition für Motivation ist dabei die Bereitschaft zum längeren Einsatz zum Lernen in einem neuen Gebiet [14]. In einer Studie von Chiang et al., in der die Effektivität vom Lernen mit AR auf mobilen Geräten untersucht wurde, ist eine signifikant höhere Motivation beobachtet worden [12], sodass dies für AR gegenüber herkömmlichen Lernmethoden im Kontext von Motivation sprechen würde.

Ein weiterer Aspekt beim Lernen in Computerspielen ist die Tatsache, dass gute Spiele notwendige Informationen zur richtigen Zeit zur Verfügung stellen [19]. Das bedeutet, dass Informationen nicht ohne Kontext oder weit entfernt von den Zielen und Absichten der SpielerInnen vermittelt werden. Um zu erzielen, dass sich beispielsweise Schüler erlernte Konzepte für längere Zeit merken und auch außerhalb des Klassenzimmers anwenden, hilft es für das Lernen konkrete Anwendungsfälle und Problemstellungen, die durch die erlernten Konzepte einfacher zu lösen sind, zu integrieren [5].

Eine soziale Komponente des Lernens, die positive Auswirkungen auf das Verständnis von Lernstoffen und dem Lösen von Problemstellungen ist, ist Kollaboration und das Arbeiten in Teams. So lassen sich hier Parallelen zu *Massive Multiplayer Online* Spielen ziehen, in denen mehrere SpielerInnen in Teams mit unterschiedlichen, aber auch überlappenden Fähigkeiten zusammenarbeiten, um bestimmte Ziele zu erreichen [19]. Es wurde in diesem Kontext auch gezeigt, dass Lernansätze die AR mit Zusammenarbeit und sogar Rollenverteilungen kombinieren einen einnehmenden Effekt haben können [15].

Erfolgreiches Lernen ist von mehr als den bisher präsentierten Faktoren abhängig, wie beispielsweise Spaß, eine positive Einstellung und kognitive Auslastung. Darüber hinaus kann argumentiert werden, dass eine effektive Strategie für das Lernen und Lehren der ausschlaggebende Faktor ist [12].

## 4.2 Lern- und Lehrkonzepte mit Augmented Reality

In diesem Abschnitt werden bereits in der Literatur vorgestellte Ansätze für das Lernen mit AR als auch konkretere Lehr-Konzepte vorgestellt. Dabei wird des Weiteren erläutert woraus sich bestimmte Klassifizierungen ergeben, wie zum Beispiel Rückblicke auf durchgeführte Studien und deren Ergebnisse.

### 4.2.1 Mögliche Ansätze für den Einsatz von AR

Durch das Analysieren von bisherigen Studien und Anwendungen können grobe Richtungen für die Entwicklung und Implementierung für das Lernen mit AR klassifiziert werden. Yuen et al. [37] haben so beispielsweise fünf mögliche Richtungen für Anwendungen identifiziert: AR-Bücher, AR-Spiele, Objekt Modellierung, Fähigkeiten-Training und Lernen basierend auf Entdeckung.

**AR-Bücher** AR kann verwendet werden, um herkömmliche Bücher mit AR Merkmalen in Form von 3D-Repräsentationen und interaktiven Erlebnissen zu erweitern. Dies kann beispielsweise durch AR-Marker, die auf die Seiten des Buches gedruckt werden oder auch die Erkennung der bereits beinhalteten “natürlichen Marker” (Bilder) realisiert werden.

**AR-Spiele** Spiele im Allgemeinen können verwendet werden, um Schülern das Verstehen von Konzepten leichter zu gestalten. AR-basierte Spiele können somit die echte Welt mit virtuellen Objekten, mit denen die Schüler interagieren können, anreichern.

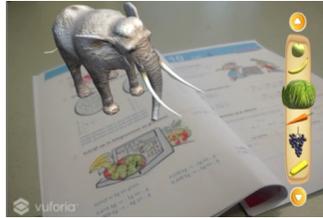
**Objekt Modellierung** AR kann dazu verwendet werden virtuelle 3D-Objekte schnell zu generieren und die Möglichkeit bieten diese zu manipulieren. Dies ist in Anwendungsfällen sinnvoll, in denen sofortiges visuelles Feedback über Ideen und Design hilfreich ist.

**Fähigkeiten-Training** Für das Trainieren von Fähigkeiten in speziellen Umgebungen und dem Lernen durch das Ausführen von Aufgaben hat AR starkes Potential. Solche Trainings können im Kontext von mechanischen Aufgaben wie zum Beispiel Wartungsarbeiten Anwendung finden.

**Lernen basierend auf Entdeckung** Die Verwendung von AR beim Vermitteln von Informationen über Objekte in der echten Welt ermöglicht kreative Ansätze des Anzeigens von Daten über beispielsweise Sehenswürdigkeiten, auch wenn keine realen Informationstafeln vorhanden sind.

Zwischen den genannten Ansätzen gibt es dabei Überschneidungen, wobei darüber hinaus weitere Konzepte wie beispielsweise Lernen durch Zusammenarbeit in mehreren Kategorien implementiert werden könnte. In einer Studie von Dünser et al. [16] wurden AR-Bücher verwendet um 10 Schülerinnen im Alter von 13 bis 15 Jahren Konzepte von Elektromagnetismus beizubringen. Die Ergebnisse der Pilot-Studie deuten dabei darauf hin, dass AR das Potential hat beim Lernen von 3D Konzepten effektiv zu sein.

Ein Beispiel für den Einsatz von Elementen aus Computerspielen ist das AR Spiel “Alien Contact!”, in dem Schüler durch Informationsaustausch und Zusammenarbeit mit anderen verschiedene Puzzles lösen müssen [15]. Ein weiteres AR Spiel basierend auf dem Spiel Go ist “ARGo” [20], welches dazu verwendet wurde, um herauszufinden, ob AR das autodidaktische Lernen unterstützen kann. Der AR-basierte Ansatz wurde dabei mit einem PC-basierten



■ **Abbildung 3** Das AR-basierte Spiel SEE ME ROAR [23], welches Schülern dabei helfen soll mathematische Kenntnisse zu erlernen. Durch das Einscannen der Bücher werden verschiedene Tiere auf den Seiten dargestellt

Ansatz verglichen, wobei physische Interaktion mit dem traditionellen Spiele-Apparat in Kombination mit “ARGo” sowohl die intrinsische Motivation der Teilnehmer als auch eine höhere Entfaltung in Bezug auf Problemlösungen mit sich brachte [20].

Wie AR Systeme durch das Manipulieren von Objekten dabei helfen kann mechanische Konzepte zu verstehen wird von Kaufmann und Meyer mit der Applikation “PhysicsPlayground” demonstriert [21]. Durch dreidimensionale virtuelle Umfelder können sowohl Lernende, als auch Dozenten physikalische Experimente vorbereiten und in Echtzeit simulieren. Kaufmann und Meyer argumentieren, dass dieses System dabei helfen kann mechanische Konzepte besser zu verstehen und sogar Möglichkeiten bietet, die denen in einem realen Physik Labor weit überlegen sind [21].

Fähigkeiten Training mit AR ist in Bereichen sinnvoll in denen beispielsweise mechanische Arbeiten, die meist ähnliche Abläufe haben, durchgeführt werden. Yuen [37] nennt dabei einen Anwendungsfall aus dem Militär, bei dem Entwickler aus dem Computer Graphics und User Interface Lab von der Columbia University ein AR System namens ARMAR entwickeln, welches Mechaniker dabei anleitet Routine-Wartungsarbeiten und Reparaturen in gepanzerten Fahrzeugen durchzuführen.

AR Tools wie zum Beispiel Wikitude erlauben das Lernen basierend auf Entdeckung in realen Umgebungen und können so bei Schulausflügen eingesetzt werden, um Echtzeit-Informationen über Orte und Gebäude auf Smartphones anzuzeigen. Ein möglicher Ansatz wäre dabei, dass Lehrer Aufgaben erteilen, Informationen über Ereignisse und Orte zu sammeln oder sogar selbst relevante Daten (aus dem Lehrplan) erstellen, die die Schüler erkunden können [37].

#### 4.2.2 Lernen basierend auf wissenschaftlicher Erkundung

Chiang et al. definieren Lernen basierend auf (wissenschaftlicher) Erkundung durch studenten- oder schüler-zentrierte Aktivitäten in denen unter Anderem proaktiv konzeptualisiert, exploriert, validiert und diskutiert wird [12]. Im Kontext von Schulen erfolgt diese Art des Lernens typischerweise unter Anleitung eines Lehrers, wobei AR Systeme diese Aufgabe zum gewissen Teil unterstützen könnten. Aufgrund des Fokus auf das selbstständige Erforschen und Lernen entwickeln Schüler so höhergradige kognitive Fähigkeiten als auch Kompetenzen selbstständig zu lernen [24]. In diesem Zusammenhang argumentieren viele Befürworter von authentischem Lernen, dass sinnvolles Lernen anwendbare Kontexte und Situationen erfordert, in denen die erlernten Fähigkeiten eingesetzt werden können [12]. Dies spiegelt das Kontext-Argument aus dem Abschnitt 4.1 wieder. Die Verbindung von AR-Technologie mit dieser Art von Lernmethoden zeigt dabei positive Auswirkungen auf die Lernerfolge von Schülern, wenn reale Kontexte mit digitalen Inhalten zum richtigen Zeitpunkt kombiniert werden [12].

#### 4.3 Anwendungsgebiete

In den vergangenen drei Jahrzehnten haben Forscher bereits potentielle Anwendungsgebiete durch Literatur-Reviews analysiert [2, 37, 18], da es wahrscheinlich ist, dass sich AR als Bestand von Lernmethoden nicht in allen Bereichen eignet. Azuma identifiziert in einem allgemeinen Gutachten zum Stand vom Einsatz von AR sechs potentielle Klassen von Applikationen [2]:

- Medizin
- Produktion und Reparatur
- Visualisierung und Beschriftung
- Roboter Weg-Planung
- Unterhaltung
- Militär (-Flugzeuge)

Es ist dabei wichtig anzumerken, dass Azuma nicht nur Einsatzgebiete im Kontext von Lernen mit AR begutachtet hat, weshalb nicht alle Klassen zum selben Grad für Bildung relevant sind. Dennoch lassen sich Einsatzzwecke wie zum Beispiel in Medizin oder Militär finden, die bereits in Studien und Projekten untersucht wurden [29, 32]. Seit dem Bericht von 1994 hat sich sowohl AR als auch der Einsatz in den jeweiligen Bereichen weiterentwickelt, sodass besonders im Kontext von Schulen teils spezielle Anwendungsfälle gefunden wurden.

Aus einer Meta-Analyse von Saidin et al. [30], in der Studien seit 2007 vom Einsatz von AR zum Lernen untersucht wurden, geht hervor, dass weitere

mögliche Einsatzgebiete in Chemie, Astronomie, Mathematik, Biologie und sogar Geschichte denkbar wären. Wenn man die Anzahl an durchgeführten Studien nach Schulfächern kategorisiert, besteht die Annahme, dass sich vor Allem naturwissenschaftliche Fächer eignen. Dies kann aber natürlich andere Gründe haben wie zum Beispiel die Tatsache, dass die Implementierung von AR in Fächern wie Biologie oder Medizin aufgrund der eindeutigeren visuellen Darstellungen naheliegender ist. Andere Einsatzgebiete wie beispielsweise Sprachen, Geschichte oder Sozialwissenschaften [15] sind deshalb aber nicht ausgeschlossen.

#### 4.4 Zielgruppen

Um effektive Applikationen zu entwerfen, die den Lernprozess von Nutzern unterstützten, ist es oftmals wichtig, die individuellen Zielgruppen und deren Bedürfnisse zu verstehen. Dies wird beispielsweise in der Arbeit von Brederode et al. [4] hervorgehoben, in der es das Ziel war, Kinder mit und ohne physische Behinderungen und Lernschwächen spielerisch durch soziale Interaktionen zusammenzubringen. In diesem speziellen Anwendungsfall wurden so zum Beispiel Interviews mit Lehrern geführt, in denen herausgefunden wurde, dass es Kindern mit Einschränkungen schwer fällt sich spontan an spielerische Aktivitäten mit Kindern ohne größere Einschränkungen anzuschließen. Obwohl dieser Fall speziellere Anforderungen mit sich bringt, die nicht für alle Fächer in Schulen relevant sind, besteht auch dort eine nicht unwesentliche Diskrepanz zwischen Forschern und Kindern [4] in Bezug auf das Design von Applikationen, die zu anderen ebenso wichtigen Anforderungen führen kann. Neben Schulkindern wurden darüber hinaus weitere potentielle Nutzergruppen untersucht, wie aus einer Zusammenfassung von Studien entnommen werden kann [10]. Die Nutzergruppen erstreckten sich dabei über Schulkinder, Schüler aus der Mittel- und Oberstufe, Studenten von Universitäten und sogar Gruppen von Erwachsenen [10]. Literatur-Reviews scheinen aber im Allgemeinen darauf hinzudeuten, dass der Fokus in der bisherigen Forschung auf der Anwendung im Schulunterricht und dem akademischen Umfeld lag [10, 18, 30].

#### 4.5 Vorteile von Augmented Reality Lernmethoden

Welche Vorteile das Integrieren von AR in Lernmethoden im Vergleich zu herkömmlichen Ansätzen mit sich bringt, ist Teil der Forschungsfrage von vielen Studien [12, 16, 15, 29, 20]. Um einen Überblick über die Erkenntnisse der einzelnen Studien zu erlangen, haben unter anderem Diegmann et al. [13] ein ausführliches Literatur-Review über berichtete Vorteile durchgeführt. Dabei wurden 14 unterschiedliche Nutzen gefunden und in sechs Gruppen eingeteilt. Die definierten Gruppen werden in den folgenden Abschnitten besprochen.

#### 4.5.1 Geisteszustand

**Erhöhte Motivation** Wie bereits im Abschnitt 4.1 angesprochen, ist Motivation einer der wichtigsten Faktoren für erfolgreiches Lernen und wird in einer Reihe von Studien als Vorteil genannt. So zum Beispiel in einer Studie zu naturwissenschaftlichen Erkundungsaktivitäten mithilfe von mobilen AR Geräten, in der bei Schülern eine signifikant höhere Motivation zum Lernen befunden wurde [12]. Ein weiteres Beispiel ist in einer Studie zum Lernen des Spieles Go mit der Unterstützung von AR zu finden [20], bei der die Nutzer von einer höheren intrinsischen Motivation zum eigenständigen Lernen berichteten. Darüber hinaus gingen Iwata et al. davon aus, dass eine höhere intrinsische Motivation positive Auswirkungen auf selbständiges Lernen über längere Zeit haben kann [20].

**Erhöhte Aufmerksamkeit** Diegmann et al. identifizierten Indikatoren für eine höhere Aufmerksamkeit unter anderem mit Aussagen, die erhöhtes Interesse von Nutzern beschreiben [13]. Von erhöhter Aufmerksamkeit in Bezug auf den Lernstoff und AR als Lernmethode berichten dabei mehrere Arbeiten [12, 15].

**Erhöhte Konzentration** Dieser Nutzen beschreibt die Konzentration von Nutzern, die eine AR Anwendung verwenden und wird zum Beispiel in der Studie von Iwata et al. im Bezug auf die physische Interaktion im Vergleich zum Lernen an einem PC genannt [20].

**Erhöhte Zufriedenheit** Beschreibt die Zufriedenheit, die Nutzer in Bezug auf ihren Lernfortschritt und den Lernprozess an sich empfinden [13]. Chen et al. maßen in diesem Zusammenhang einen höheren Zufriedenheitsbewertung in einer Pilot-Studie mit 25 Grundschulschülern [7].

#### 4.5.2 Lehr-Konzepte

**Erhöhtes Studenten-zentriertes Lernen** Bei diesem Lernkonzept werden Schüler und Studenten dazu motiviert aktiv und selbstbestimmt zu lernen, wobei diese auch mehr Verantwortung für ihren eigenen Lernfortschritt tragen [13]. AR unterstützt das Studenten-zentrierte Lernen auch insoweit, dass Lehrern mehr Möglichkeiten zur Verfügung stehen, den Unterricht individuell an die Fähigkeiten von Schülern anzupassen [13].

**Verbessertes kollaboratives Lernen** Diegmann et al. erwähnen in ihrer Arbeit drei Studien, in denen Forscher von verbessertem kollaborativen Lernen berichten, welches durch neue Arten der Kommunikation und Kooperation möglich ist [13].

### 4.5.3 Darstellung

Die Kategorie Darstellung beinhaltet drei verschiedene Vorteile: ein erhöhter Detailgrad, verbesserter Zugänglichkeit auf Informationen und mehr Interaktivität. Ein gutes Beispiel für AR, die mehr Details anzeigt als herkömmliche Lernmittel, ist beispielsweise das überlagernde Anzeigen von 3D-Modellen bei AR-Büchern. Abhängig von der Technologie könnten hierbei Bücher und Modelle aus Plastik in einem AR System zusammengeführt werden, anstatt zwei voneinander entkoppelte Lernmittel zu verwenden. Die Verbesserte Zugänglichkeit auf Informationen kann im gleichen Zusammenhang gesehen werden, da es möglich ist Informationen überlagernd gut zu integrieren ohne dass der Lernprozess gehindert wird [20]. AR bietet Lernenden des Weiteren die Möglichkeit mit digitalen Inhalten zu interagieren. Ein sehr gutes Beispiel für ein interaktives AR System ist das "augmented" Stromkreis Exponat von Beheshti et al. [3] bei dem Kinder zusammen mit ihren Eltern die Konzepte von Stromstärke und Widerstand anhand vom eigenständigen Verdrahten von Stromkreis-Komponenten erlernen. Darüber hinaus gibt es noch weitere Studien aus den Bereichen Physik und Medizin [16, 29], in denen erhöhte Interaktion als positiver Faktor beim Lernen mit AR genannt wird.

### 4.5.4 Lerntyp

Diegmann et al. identifizieren in der Lerntyp Kategorie zwei Vorteile, eine verbesserte Lernkurve und erhöhte Kreativität im Vergleich zu traditionellen Ansätzen die AR nicht als Basis haben. Liu berichtet beispielsweise, dass durch die Verwendung von einem AR-unterstützten ubiquitären Lernumfeld namens "HELLO" AR-Nutzer signifikant bessere Ergebnisse bei Tests erzielten verglichen mit Nutzern, die traditionelle Lernmethoden verwendeten [25]. Das Potential von AR kreatives Lernen zu ermöglichen wird darüber hinaus von Chang et al. im Rahmen eines mobilen Führungssystems für ein Kunstmuseum besprochen [6].

### 4.5.5 Verstehen von Inhalten

Wie von Dünser et al. [16] bei einer Studie zu AR-Büchern erwähnt wird, hat AR das Potential räumliche Fähigkeiten von Schülern und Studenten zu verbessern. Eher theoretische oder abstrakte Konzepte können so mithilfe von digitalen und teils animierten 3D-Modellen als auch erhöhter Interaktivität besser verständlich gemacht werden.

In Bezug auf das Beibehalten von angeeignetem Wissen über längere Zeit schlussfolgern Macchiarella et al [26], dass AR einen positiven Effekt hat.

### 4.5.6 Kostenminderung

Leblanc [22] und Martín-Gutiérrez [27] berichten in ihren jeweiligen Arbeiten, dass Kostenminderungen mithilfe von AR-Szenarien im Vergleich zu traditionellem Lernen auf lange Sicht möglich sind. Dies ist zum Beispiel dadurch möglich, dass bei selbstständigen Lernprozessen mit AR weniger Arbeitskräfte notwendig sind. Einmalige Beschaffungskosten sind zwar meist hoch, würden sich aber auf lange Sicht rentieren [22, 13].

## 4.6 Führende Theorien in der AR Forschung für Bildung

In der Zusammenfassung und Analyse von ausgewählten Studien durch Cheng und Tsai [10] geht ein Rahmenkonzept hervor, welches wichtige Ansätze und Vorteile vom Lernen mit AR in übergeordnete Theorien einordnet: Mentale Modelle, Räumliche Kognition, Situations- beziehungsweise Kontext-bedingte Kognition und sozial-konstruktivistisches Lernen.

Mentale Modelle spielen insoweit eine Rolle in der Forschung vom Lernen mit AR, dass sich mentale Modelle von Nutzern beim Verwenden von beispielsweise Büchern im Gegensatz zu AR-unterstützten Büchern unterscheiden.

Wie bereits im Abschnitt 4.5.5 angesprochen wurde, hat AR den Vorteil räumliches Verständnis zu verbessern und ist somit auch für die Forschung von räumlicher Kognition, die sich mit Wissen und Vorstellungen von räumlichen Eigenschaften beschäftigt, relevant [10]. Wie räumliches Wissen angeeignet und über Zeit entwickelt werden kann, könnte somit auch eine Forschungsfrage für das Lernen mit AR darstellen.

Allgemeinere Theorien bezüglich des authentischen Lernens [5] sind ebenso mögliche relevante Forschungsgebiete, da zum Beispiel Orts-basierte AR Applikationen Nutzern durch eigenständiges wissenschaftliches Erkunden ermöglichen im jeweiligen realen Kontext zu lernen [10]. Somit kann die Situations-beziehungsweise Kontext-bedingte Kognition eine nützliche theoretische Perspektive für eine Basis von AR Forschung sein [10].

In einigen Arbeit wurde bereits der Einsatz von AR für kollaborative und soziale Tätigkeiten untersucht [15, 4, 3], sodass sozial-konstruktivistisches Lernen [34], welches eben die sozialen und kollaborativen Eigenschaften vom Lernen herausstellt, eine angemessene Basis bilden könnte [10]. Cheng und Tsai schlussfolgern abschließen, dass diese vier Theorien möglicherweise für die einzigartigen Vorteile, die AR-Technologien bieten, relevant sein und die zukünftige Forschung leiten könnten [10].

## 5 Herausforderungen und Einschränkungen

Trotz der vielen Vorteile und Potentiale, die AR für Schüler, Studenten und Lehrende bietet, gibt es ebenso Herausforderungen und limitierende Faktoren, die es zu beachten gilt. So sollte AR nicht als Allheilmittel für Bildungs-Umfelder gesehen werden, da jede AR Applikation individuell ist und die gefundenen Vorteile nicht in jedem Kontext anwendbar sind [13]. Es ist wichtig, dass jede Applikation konzeptionell durchdacht und sauber implementiert ist, damit Hindernisse wie schlechte Nutzer-Interaktion oder Systemfehler vermieden werden, um den Lernprozess nicht negativ zu beeinflussen.

Das Potential vom selbstständigen und Erkundungs-basierten Lernen wird in dieser Arbeit im Abschnitt 4.2.2 besprochen. Damit Schüler und Studenten von dieser Lernmethode profitieren können, ist es wichtig, dass der Wissens-kontext nicht zu unstrukturiert ist, da sonst der Lernprozess sehr schwierig werden kann [21].

Darüber hinaus gibt es organisatorische Voraussetzungen, die AR-basierte Systeme mit sich bringen wie das Beispiel des Einsatzes einer AR Sandbox in Schulklassen zeigt [1]. Die Autoren empfehlen dabei, dass das Schulpersonal mithilfe einer Trainings-Sitzung mit dem System vertraut gemacht wird bevor der Einsatz stattfindet. Des Weiteren wird darauf hingewiesen, dass solche Projekte eine exzessive Demonstrationsphase erfordern, um den tatsächlichen Nutzen in der praktischen Anwendung nachzuweisen [1]. Für Schüler und Studenten ist ebenso wichtig, sich mit dem AR System vor dem eigentlich Einsatz vertraut zu machen [9], da davon auszugehen ist, dass sie die individuelle Applikation oder sogar AR im Allgemeinen zum ersten mal verwenden.

In dem Review von Fotaris et al. [18] werden fünf Kategorien von Limitierungen von AR im Kontext vom Lernen identifiziert:

- Lehrer können das gleiche System nicht für unterschiedliche schulische Fächer manipulieren
- Die Schüler / Studenten haben zu sehr auf die virtuellen Inhalte geachtet (Faktor Neuheit von AR)
- Lehrer müssen weitere Lernmaterialien entwickeln, die den exklusiven Anforderungen von AR entsprechen
- Komplexe AR-Systeme können eine nicht unwesentliche Lernkurve haben
- Zu kurze Zeiträume für die Messung von der Lernleistung von Schülern und Studenten

In Bezug auf diese Limitierungen werden Lösungsvorschläge gegeben wie beispielsweise das Verbessern von sowohl Hard- als auch Software, dem Erstellen von Richtlinien für das Design von AR-basierten Applikation für das Lernen und die Entwicklung von intuitiver Autorensoftware, damit Lehrende AR Inhalte individuell und ohne Programmierkenntnisse erstellen können [18].

Schlussendlich ist für die Effektivität und den Erfolg von AR Applikationen für das Lernen das Verständnis der Zielgruppen wichtig, wie aus der Studie von Brederode et al. hervorgeht [4]. Vor Allem bei speziellen Anwendungsfällen wie beispielsweise dem Entwurf eines AR-Spiels welches das Zusammenspielen von Kindern mit und ohne Behinderungen fördern soll, ist das Erfassen von Anforderungen, der Definition von klaren Zielen und sogar dem Einbeziehen von Nutzern essenziell. Dadurch können Entwickler Applikationen entwerfen, die konkrete Probleme beim Lernen lösen oder zumindest dabei helfen diese zu mindern.

## 6 Fazit

Aufgrund der Weiterentwicklung von kommerziellen Technologien (mobile AR, Microsoft HoloLens etc.) für die Implementierung von AR Applikationen stehen Forschern mittlerweile eine Reihe von Werkzeugen zur Verfügung, um das Lernen mithilfe von AR zu verbessern. Dennoch werden die meisten Technologien nicht mit einer konkreten Agenda in Bezug auf Bildung entwickelt, sondern eher für den Nutzen in Wirtschaft wie beispielsweise Unterhaltung oder Marketing [37]. Wie aber die in dieser Arbeit präsentierten Studien zeigen, können diese Werkzeuge auch im Rahmen von Bildung eingesetzt werden. AR ist dabei sehr vielseitig, da der Einsatz sowohl hauptsächlich digital aber auch in Kombination mit Objekten aus der echten Welt ablaufen kann. Es ist somit denkbar, dass AR in Zukunft bisherige Lernmethoden oder Werkzeuge nicht wirklich ablöst, sondern vielmehr eine komplementäre Rolle einnimmt, wie man am Beispiel von interaktiven AR-Büchern sehen kann.

Neue Technologien brauchen vor Allem in Schulen oder Universitäten eine gewisse Anlaufzeit, damit gezeigt werden kann, dass diese tatsächlich einen Mehrwert oder Nutzen mit sich bringen. Es wird somit noch einige Jahre oder sogar Jahrzehnte dauern, bis ausreichend Studien und Projekte den Einsatz von AR in Bildung rechtfertigen können. Die bisher erforschten Vorteile, die AR für den Einsatz in Bildung bietet, deuten aber darauf hin, dass AR eine wichtige Rolle in der Zukunft spielen und sogar proaktives und selbständiges Lernen fördern könnte. Schlussendlich ist es durchaus möglich, dass AR in Zukunft eine passende Verknüpfung von der Haptik der Objekte aus der echten Welt und der Vielseitigkeit von digitalen Informationen sein könnte, die die Art wie Menschen lernen verbessern würde.

### Literatur

- 1 Austun Ables. Augmented and Virtual Reality: Discovering Their Uses in Natural Science Classrooms and Beyond. In *Proceedings of the 2017 ACM Annual Conference on SIGUCCS*, pages 61–65. ACM, 2017.

- 2 Ronald T. Azuma. A Survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments*, 6(4):355–385, August 1997. doi:10.1162/pres.1997.6.4.355.
- 3 Elham Beheshti, David Kim, Gabrielle Ecanow, and Michael S Horn. Looking Inside the Wires: Understanding Museum Visitor Learning with an Augmented Circuit Exhibit. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 1583–1594. ACM, 2017.
- 4 Bas Brederode, Panos Markopoulos, Mathieu Gielen, Arnold Vermeeren, and Huib De Ridder. pOwerball: the design of a novel mixed-reality game for children with mixed abilities. In *Proceedings of the 2005 conference on Interaction design and children*, pages 32–39. ACM, 2005.
- 5 John Seely Brown, Allan Collins, and Paul Duguid. Situated cognition and the culture of learning. *Educational researcher*, 18(1):32–42, 1989.
- 6 Kuo-En Chang, Chia-Tzu Chang, Huei-Tse Hou, Yao-Ting Sung, Huei-Lin Chao, and Cheng-Ming Lee. Development and behavioral pattern analysis of a mobile guide system with augmented reality for painting appreciation instruction in an art museum. *Computers & Education*, 71:185–197, 2014.
- 7 Chien-Hsu Chen, Chun-Yen Huang, and Yin-Yu Chou. Integrating Augmented Reality into Blended Learning for Elementary Science Course. In *Proceedings of the 5th International Conference on Information and Education Technology*, pages 68–72. ACM, 2017.
- 8 Chih-Ming Chen and Yen-Nung Tsai. Interactive augmented reality system for enhancing library instruction in elementary schools. *Computers & Education*, 59(2):638–652, 2012.
- 9 Yu-Chien Chen. A study of comparing the use of augmented reality and physical models in chemistry education. In *Proceedings of the 2006 ACM international conference on Virtual reality continuum and its applications*, pages 369–372. ACM, 2006.
- 10 Kun-Hung Cheng and Chin-Chung Tsai. Affordances of augmented reality in science learning: Suggestions for future research. *Journal of Science Education and Technology*, 22(4):449–462, 2013.
- 11 Heien-Kun Chiang, Yin-Yu Chou, Long-Chyr Chang, Chun-Yen Huang, Feng-Lan Kuo, and Hown-Wen Chen. An augmented reality learning space for PC DIY. In *Proceedings of the 2nd Augmented Human International Conference*, page 12. ACM, 2011.
- 12 Tosti HC Chiang, Stephen JH Yang, and Gwo-Jen Hwang. An augmented reality-based mobile learning system to improve students' learning achievements and motivations in natural science inquiry activities. *Journal of Educational Technology & Society*, 17(4):352, 2014.
- 13 Phil Diegmann, Manuel Schmidt-Kraepelin, Sven Van den Eynden, and Dirk Basten. Benefits of Augmented Reality in Educational Environments-A Systematic Literature Review. *Wirtschaftsinformatik*, 3(6):1542–1556, 2015.

- 14 Andrea A DiSessa. *Changing minds: Computers, learning, and literacy*. MIT Press, 2001.
- 15 Matt Dunleavy, Chris Dede, and Rebecca Mitchell. Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning. *Journal of science Education and Technology*, 18(1):7–22, 2009.
- 16 Andreas Dünser, Lawrence Walker, Heather Horner, and Daniel Bentall. Creating interactive physics education books with augmented reality. In *Proceedings of the 24th Australian computer-human interaction conference*, pages 107–114. ACM, 2012.
- 17 Stéphanie Fleck and Gilles Simon. An augmented reality environment for astronomy learning in elementary grades: An exploratory study. In *Proceedings of the 25th Conference on l'Interaction Homme-Machine*, page 14. ACM, 2013.
- 18 Panagiotis Fotaris, Nikolaos Pellas, Ioannis Kazanidis, and Paul Smith. A systematic review of Augmented Reality game-based applications in primary education. In *ECGBL 2017 11th European Conference on Game-Based Learning*, page 181. Academic Conferences and publishing limited, 2017.
- 19 James Paul Gee. What video games have to teach us about learning and literacy. *Computers in Entertainment (CIE)*, 1(1):20–20, 2003.
- 20 Takahiro Iwata, Tetsuo Yamabe, and Tatsuo Nakajima. Augmented reality go: Extending Traditional game play with interactive self-learning support. In *Embedded and Real-Time Computing Systems and Applications (RTC-SA), 2011 IEEE 17th International Conference on*, volume 1, pages 105–114. IEEE, 2011.
- 21 Hannes Kaufmann and Bernd Meyer. *Simulating educational physical experiments in augmented reality*. ACM, 2008.
- 22 Fabien Leblanc, Bradley J Champagne, Knut M Augestad, Paul C Neary, Anthony J Senagore, Clyde N Ellis, and Conor P Delaney. A comparison of human cadaver and augmented reality simulator models for straight laparoscopic colorectal skills acquisition training. *Journal of the American College of Surgeons*, 211(2):250–255, 2010.
- 23 Jingya Li, Erik van der Spek, Jun Hu, and Loe Feijs. SEE ME ROAR: Self-determination Enhanced Engagement for Math Education Relying On Augmented Reality. In *Extended Abstracts Publication of the Annual Symposium on Computer-Human Interaction in Play*, pages 345–351. ACM, 2017.
- 24 Byung-Ro Lim. Challenges and issues in designing inquiry on the Web. *British Journal of Educational Technology*, 35(5):627–643, 2004.
- 25 T-Y Liu. A context-aware ubiquitous learning environment for language listening and speaking. *Journal of Computer Assisted Learning*, 25(6):515–527, 2009.
- 26 Nickolas D Macchiarella, Dahai Liu, Sathya N Gangadharan, Dennis A Vincenzi, and Anthony E Majoros. Augmented reality as a training medium for

- aviation/aerospace application. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 49, pages 2174–2178. SAGE Publications Sage CA: Los Angeles, CA, 2005.
- 27 Jorge Martín-Gutiérrez, Rosa E Navarro, and Montserrat Acosta González. Mixed reality for development of spatial skills of first-year engineering students. In *Frontiers in Education Conference (FIE)*, 2011, pages T2D–1. IEEE, 2011.
- 28 Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. Augmented reality: A class of displays on the reality-virtuality continuum. In *Telemanipulator and telepresence technologies*, volume 2351, pages 282–293. International Society for Optics and Photonics, 1995.
- 29 Christian Moro, Zane Stromberga, Athanasios Raikos, and Allan Stirling. Combining virtual (Oculus Rift & Gear VR) and augmented reality with interactive applications to enhance tertiary medical and biomedical curricula. In *SIGGRAPH ASIA 2016 Symposium on Education: Talks*, page 16. ACM, 2016.
- 30 Nor Farhah Saidin, Noor Dayana Abd Halim, and Noraffandy Yahaya. A review of research on augmented reality in education: advantages and applications. *International education studies*, 8(13):1, 2015.
- 31 Scott Stein. Apple’s first crack at AR looks surprisingly good, 2017. URL: <https://www.cnet.com/news/apples-first-crack-at-ar-looks-surprisingly-good/>.
- 32 Synthetic Training Environment (STE). URL: <http://ict.usc.edu/prototypes/synthetic-training-environment-ste/>.
- 33 DWF Van Krevelen and Ronald Poelman. A survey of augmented reality technologies, applications and limitations. *International journal of virtual reality*, 9(2):1, 2010.
- 34 Lev Semenovich Vygotsky. *Mind in society: The development of higher psychological processes*. Harvard university press, 1980.
- 35 Wikitude. Sales outlet finder, 2018. URL: <https://www.wikitude.com/geo-augmented-reality/>.
- 36 Hsin-Kai Wu, Silvia Wen-Yu Lee, Hsin-Yi Chang, and Jyh-Chong Liang. Current status, opportunities and challenges of augmented reality in education. *Computers & Education*, 62:41–49, March 2013.
- 37 Steve Chi-Yin Yuen, Gallayanee Yaoyuneyong, and Erik Johnson. Augmented reality: An overview and five directions for AR in education. *Journal of Educational Technology Development and Exchange (JETDE)*, 4(1):11, 2011.

# Das Unsichtbare anfassen: Ultrasonic Haptic Feedback.

Verena Schlott

Ludwig-Maximilians-Universität München, München, Deutschland  
Verena.Schlott@campus.lmu.de

---

## Zusammenfassung

Haptisches Ultraschall-Feedback ist eine vielversprechende Methode um gestenbasierte Benutzerschnittstellen mit haptischem Feedback auszustatten. Benutzer können unsichtbare Objekte in der Luft berühren und mit ihnen interagieren, ohne dabei mit einem zusätzlichen Aktuator ausgestattet werden zu müssen. Die berührungslose Rückmeldung löst einige Probleme gestenbasierter Applikationen. Da es sich dabei um eine neuartige und noch nicht weit verbreitete Technologie handelt, gibt es nur wenige Arbeiten, die einen Überblick über die Thematik und mögliche Anwendungen geben, sowie die Grenzen dieses Feedback-Mediums aufzeigen. In diesem Paper wird die Technologie und Funktionsweise Ultraschall-basierter Feedback-Systeme anhand des Prototyps der Firma Ultrahaptics erläutert. Es werden Vorteile aufgezeigt, ein Überblick über bereits existierende Anwendungen gegeben, die darauf basieren, und ein Ausblick auf das Potential sowie weitere Verwendungsmöglichkeiten in der Mensch-Maschine Interaktion gezeigt. Auch die aktuell existierenden Grenzen dieser Feedback-Systeme werden untersucht.

**2012 ACM Computing Classification** Tactile and hand-based interfaces: Haptic devices; Interaction techniques: Gestural input.

**Keywords and phrases** Ultrasonic Haptic Feedback; Ultrahaptics; Touchless Haptic Feedback; Human-Computer-Interaction.

## 1 Einführung

Seit Technologien unser alltägliches Leben bestimmen, wird versucht, die Grenzen der Mensch-Maschine-Interaktion immer weiter aufzubrechen, um noch intuitivere und natürlichere Systeme zu entwickeln. Die primäre Eingabemodalität moderner Endgeräte (Smartphones, Tablets oder Laptops) ist Berührung (engl. „touch“). Die Verwendung menschlicher Gesten (das Tippen virtueller Schaltflächen, das Wischen über den Bildschirm, das Erfassen virtueller Objekte mit zwei Fingern) sind Standard und auf fast jedem technischen Eingabegerät mit berührungssensitiven Bedienoberfläche möglich.



© Verena Schlott;

licensed under Creative Commons License CC-BY

Cite as: Verena Schlott. Das Unsichtbare anfassen: Ultrasonic Haptic Feedback.. 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. 14:1–14:18.

Die Geschicklichkeit und das Feingefühl der menschlichen Finger ermöglichen eine große Auswahl an fein-granularer Manipulation virtueller Objekte und eine subtile Veränderung von Druck. Dies macht „Touch“ zu einer sehr attraktiven Eingabemethode. Um die Effizienz und die Natürlichkeit dieser Eingabe zu verbessern, bieten die meisten Geräte, zusätzlich zum visuellen und auditiven, auch haptisches Feedback. Smartphones reagieren auf Benutzereingaben mit Vibration und können durch eingebaute Drucksensoren entsprechendes haptisches Feedback geben (Apple iPhone 7, Samsung Galaxy S9). Beispielsweise gibt die „Apple Watch“ [9] mit Hilfe der eingebauten „Taptic Engine“ je nach Benachrichtigungsart und Aktion unterschiedliches Feedback durch abweichende Vibrationen. Diese spürt der User auf dem Handgelenk. [11, 20]

Allerdings ist es bei fast allen dieser Geräte nötig, die physischen Bedienelemente oder die Benutzungsoberfläche zu berühren. Dies hat seine Schwachstellen. Eine Problematik ist die Begrenzung der Eingabemöglichkeiten. Nutzer können nicht frei mit dem System interagieren, sondern sind durch die vorgegebenen Gesten beschränkt. Diese sind oftmals umständlich oder unnatürlich. Eine weitere Schwäche ist das Kontaminationsrisiko. Durch das Berühren von Bedienelementen, welche von vielen Menschen angefasst werden (medizinisches Zubehör, Bildschirme in öffentlichen Verkehrsmitteln, Tasten in Aufzügen), kann eine Ansteckung mit einer infektiösen Erkrankung erfolgen. Eine weitere Gefahr ist die Möglichkeit der Verletzung bei Berührung zu heißer oder elektrisch geladener Oberflächen. [20]

Systeme, die Gesten ohne Berührung und ohne zusätzlich getragene Hardware (Handhelds, Sensor-Handschuhe) erkennen können, lösen diese Probleme. Realisiert wird dies meistens durch die Verarbeitung von Video-Bildern. Kameras, welche die Fähigkeit der Tiefenmessung besitzen, können eine hochpräzise dreidimensionale räumliche Darstellung von Gesten bereitstellen [20, 7]. Eines der bekanntesten Beispiele für diese Art der Gestenerkennung ist die Microsoft Kinect [12]. Die Genauigkeit und Nützlichkeit der Erkennung von menschlichen Gesten sind enorm. Das direkte Benutzen der Hände als Eingabegerät ermöglicht eine natürliche und intuitive Interaktion mit Computern. Gesten sind ein mächtiges Kommunikationsmittel und tief in der menschlichen Kommunikation verankert. Objekte können sich mit Leichtigkeit bewegen oder rotieren lassen. [6]

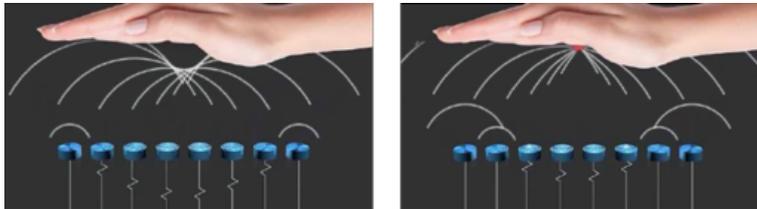
Die meisten berührungslosen Eingabemethoden berücksichtigen allerdings nur die menschlichen Sinne „Sehen“ und „Hören“. Es mangelt ihnen oftmals an taktilem Feedback [22]. Dies limitiert die menschliche Wahrnehmung und die Fähigkeit, ein Gerät effektiv zu kontrollieren. Das Einbeziehen des menschlichen Tastsinns kann nicht nur in Applikationen mit spielerischem Kontext immersive Umgebungen schaffen, sondern hat auch das Potential in sicherheitskritischen Applikationen signifikante Probleme zu lösen (beispielsweise in Fahrzeugen

und Flugzeugen) [20, 3]. Ein System, das eine Reizaufnahme ermöglicht, indem es haptisches aber auch taktiles Feedback erzeugt, ohne dabei eine direkte Interaktion mit einem physischen Eingabegerät oder einer Bedienoberfläche zu erzwingen, wäre daher wünschenswert.

Die Verbindung berührungsloser Gestenerkennung und eines taktilen Rückmeldungssystems kann ein optimales Nutzererlebnis (engl. „User Experience“) generieren und bereits genannte Probleme lösen. Eine Lösung, um haptisches Feedback mitten im Raum zu generieren ohne dass der Nutzer eine physische Oberfläche berühren muss, ist Ultraschall [14]. Es existieren bereits Systeme, welche sich der Ultraschall-Technologie bedienen, um berührbare aber unsichtbare virtuelle Objekte und Texturen zu erzeugen, und vereinen diese mit Bewegungsaufzeichnung der Hand und zusätzlich visuellem Output. In dieser Arbeit werden die Vor- und Nachteile einer Ultraschall-Feedback-Technologie anhand des Systems „Ultrahaptics“ [19] aufgezeigt und zukünftige Anwendungsbereiche in der Mensch-Maschine-Interaktion hierfür beleuchtet. Zunächst soll ein Überblick über die Funktionsweise eines haptischen Ultraschall-Feedback-Systems gegeben werden.

## 2 Haptisches Ultraschall-Feedback

Im Gegensatz zur simplen haptischen Rückmeldung, wie beispielsweise die Vibrations-Funktion in Smartphones, kann ultraschallbasiertes Feedback ein viel größeres Spektrum an Reaktionen liefern. Präzise fokussierte Ultraschallwellen können unsichtbare Sinneseindrücke und dreidimensionale Formen mitten in der Luft erzeugen. Außerdem macht es die gestenbasierte Steuerung noch intuitiver, da eine sofortige Rückmeldung an den Nutzer geliefert werden kann. Haptisches Feedback, das durch Ultraschallwellen ausgelöst wird, kann einen oder mehrere „Brennpunkte“ in der Luft kreieren (siehe Abbildung 1). Wird ein „Brennpunkt“ von der Haut reflektiert, erzeugt die entstandene Kraft eine taktile Wahrnehmung vergleichbar mit einem Windstoß. Benutzer können also die Brennpunkte in der Luft spüren, wenn sie ihre Fingerspitzen darüber bewegen. Fokussierte Ultraschallwellen stimulieren hierbei Rezeptoren des menschlichen Körpers und erzeugen eine taktile Wahrnehmung. Physischer Kontakt, also eine Berührung der Oberfläche, muss dabei nicht hergestellt werden. Gavrilov *et al.* fanden heraus, dass zielgerichtete Ultraschall-Druckwellen von der menschlichen Haut reflektiert werden und dabei Neurorezeptoren stimulieren [7]. Dadurch können verschiedene taktile Empfindungen, wie Tastgefühl, Schmerzempfindung, oder Temperaturwahrnehmung ausgelöst werden [22]. Haptisches Ultraschall-Feedback basiert auf dem Prinzip akustischer Druck-Ausstrahlung. Zielgerichteter Ultraschall erzeugt genügend Kraft um eine kleine Vibration auf der menschlichen Haut auszulösen. Diese werden



■ **Abbildung 1** Fokuspunkt und Bewegungsverlauf von Ultraschallwellen, erzeugt durch ein Ultraschall-Transduktor-Array [19].

durch die genannten Mechanorezeptoren in der Haut wahrgenommen und senden über das zentrale Nervensystem Informationen an das Gehirn. Sie sind für Vibrationen im Bereich zwischen 0.4Hz und 500Hz empfänglich. Der ausgestrahlte Ultraschall kann reguliert werden, um Vibrationen innerhalb des optimalen Frequenzbereichs für die Wahrnehmung auf der menschlichen Haut zu erzeugen. Normalerweise verwenden haptische Ultraschall-Feedback-Systeme eine Frequenz von etwa 40kHz, was über dem Schwellenwert der Aufnahmefähigkeit der Haut liegt. Somit kann ein Nutzer nur den Beginn und das Ende eines solchen fokussierten Ultraschalls wahrnehmen. Um eine Wahrnehmung zu erzeugen, die von den Mechanorezeptoren aufgenommen werden können, wird der gerichtete Ultraschall auf eine niedrigere Frequenz innerhalb dieses wahrnehmbaren Bereichs reguliert. Diese Bandbreite liegt normalerweise zwischen 1Hz und 500Hz. Durch die Veränderung der Ausstrahlungsfrequenz kann auch die Frequenz der Vibrationen auf der Haut verändert werden. Dies wiederum kann dazu genutzt werden, verschiedene taktile Wahrnehmungen zu erzeugen. Indem einzelne Fokus-Punkte im dreidimensionalen Raum (vergleiche Abbildung 1) durch unterschiedliche Frequenz reguliert werden, kann ihnen eine individuelle "Spürbarkeit" verliehen werden. Das bietet die Möglichkeit, unterschiedlich aufgefassten Texturen eine unterschiedliche Bedeutung zuzuweisen und spezifische Informationen zu transportieren. Brennpunkte können mit einer hohen räumlichen Auflösung erzeugt werden und schnell im Bereich oberhalb der Ultraschallquelle bewegt werden. Die Größe beziehungsweise der Fokus der Punkte wird also durch die Wellenlänge des Ultraschalls festgelegt. [22, 3]

Ein komplettes haptisches Ultraschall-System besteht normalerweise aus einer Matrix an Ultraschall-Transduktoren. Das erzeugbare akustische Feld kann kontrolliert werden, indem ein oder mehrere Fokuspunkte in der Luft definiert werden. Jedem Punkt kann ein Wert zugeschrieben werden, welcher einer gewünschten Amplitude an diesem Punkt gleichgestellt wird. Die Transduktoren-Matrix kann dann die festgelegten Werte so einstellen, dass die

fokussierten Punkte sich entsprechend verändern. [3]



■ **Abbildung 2** Links: Erzeugung eines dreidimensionalen, unsichtbaren Objekts. Rechts: Erzeugung eines unsichtbaren Objekts in Form eines Kopfes. [19]

Normalerweise werden die zusammenhängenden Ultraschall-Reihen auf einer flachen Oberfläche aufgestellt. Alle Transduktoren werden in eine Richtung ausgerichtet. Unterschiedliche Frequenzen können dazu verwendet werden, Objekte oder Oberflächen zu generieren, die der Nutzer „anfassen“ und untersuchen kann, indem er seine Hand durch den freien Raum über den Transduktoren bewegt. Ein solches Ultraschall-Array kann berührbare Oberflächen, Linien, Punkte und volumetrische Objekte erzeugen (siehe Abbildung 2). Es eignet sich für gestenbasierte Schnittstellen, welche zusätzlich die Bewegung der Hände aufzeichnen und Gesten erkennen, denn Standardinteraktionen wie „klicken“, „ziehen“ oder „auswählen“ können mit Leichtigkeit simuliert werden. Es wird dem Nutzer ermöglicht, virtuelle Objekte auszuwählen, zu rotieren, zu zeichnen, zu vergrößern oder beliebig zu bewegen und dabei sowohl visuelles als auch taktiles Feedback zu erhalten. Mit einem solchen Aufbau konnte bereits Handschrift übertragen werden. Hoshi kreierte mit Hilfe einer Tiefenkamera (die zum Hand-Tracking verwendet wurde) und zwei Ultraschall-Transduktormatrizen, welche rechts und links auf den Nutzer ausgerichtet waren, eine virtuelle anfassbare Oberfläche – mitten in der Luft [8]. Es ist auch möglich, dass mehrere Benutzer ein Objekt anfassen, oder beliebige Körperteile zur taktilen Wahrnehmung verwendet werden. Ein fokussiertes Ultraschall-Display hat nur eine geringe Latenz und eine ausreichende räumlich-zeitliche Auflösung. Da die Vibration der Haut durch die Bewegung der Hand relativ zu den regulierten statischen Wellen erzeugt wird, tritt kein hörbares Rauschen durch zeitliche Modulation auf. [8, 10, 22, 4]

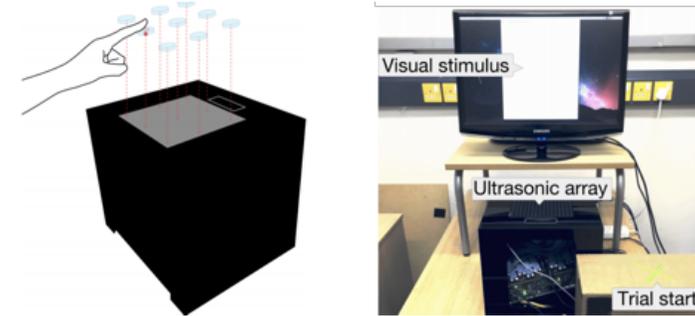


■ **Abbildung 3** TOUCH Development Kit der Firma Ultrahaptics. Erwerbbarer Prototyp (links) und aufgebauter Prototyp mit Gestenerkennungssystem und Software in Verwendung (rechts) [19].

### 3 UltraHaptics: Beispiel eines Ultraschall-Feedback-Systems

Obwohl es bereits viele verschiedene Ansätze zur hardwareseitigen Umsetzung eines solchen haptischen Feedback-Systems gibt (vergleiche [8, 10]) und auch andere Technologien, die nicht auf Ultraschalltechnologien basieren (beispielsweise AIREAL, ein System welches auf strömungsbasierten taktilen Feedback basiert [16]) existieren, fokussiert diese Arbeit auf das System von „Ultrahaptics“ [19]. Die meisten anderen existierenden Installationen sind nur für experimentelle Zwecke konstruiert worden und nicht für die kommerzielle Verwendung ausgelegt. Das Ultraschall-Feedback-System von „Ultrahaptics“, das „TOUCH Development Kit“ [18] (siehe Abbildung 3) wurde bereits in vielen Studien als funktionsfähiger Prototyp verwendet und kann kommerziell erworben werden (vergleiche [1, 14, 21]). In dem Entwickler-Kit inbegriffen sind ein 14x14 Transduktor-Array, ein Leap Motion-Kameramodul und eine entsprechende Software. Mit diesem Development Kit können virtuelle Tasten, Schieberegler und Input-Felder erstellt werden, die der Benutzer fühlen und bedienen kann. Entwickler können die Gestensteuerung erweitern, indem sie eigene Gesteneingaben hinzufügen und neue ultraschallbasierte Elemente kreieren. Es kann außerdem mit einer beliebigen Gestenerkennungs-Software, die 3D-Koordinaten liefert, gearbeitet werden. Das System enthält eine Reihe von Transduktoren, die jeweils eine Ultraschall-Druckwelle von 40kHz übermitteln. Die Druckwellen jedes Transduktors sind entsprechend verzögert, sodass die ausgestrahlte Wellenform an einem Punkt zusammenläuft (siehe Abbildung 1). Dies macht das Drucklevel an diesem Punkt sehr hoch. Wenn eine Testperson ihre Finger über diesen Druck-Fokuspunkt hält, fungiert die Haut als Detektor und ein 200Hz-Signal wird auf dem Finger als Vibration wahrgenommen. Diese

können so übertragen werden, dass dreidimensionale „Objekte“ ertastet werden können. Zusätzlich zum haptischen kann auch visuelles Feedback erhalten werden. Dies muss entsprechend auf die Eingaben durch die Transduktoren abgestimmt werden. [4]



■ **Abbildung 4** Links: Verwendung von haptischem Ultraschall-Feedback zur Leistungsverbesserung. Rechts: Experimentaufbau zur Untersuchung der Geschwindigkeit und Präzision verschiedener Eingabemethoden. [21]

### 4 Vorteile eines Ultraschall-Feedback-Systems

Es wurden bereits einige fundamentale Vorteile von gestenbasierter Steuerung, welches mit Ultraschall-Feedback kombiniert wurde, genannt. Im Folgenden soll auf die Performance-Verbesserung durch haptisches Feedback genauer eingegangen werden. Die Hand als direktes Eingabegerät ist sehr natürlich, denn Gesten sind tief in der menschlichen Kommunikation verankert. Gestenbasierte Steuerung ist somit ein mächtiges Kommunikationsmittel. Objekte können leicht bewegt, gezogen oder rotiert werden. Nutzer werden nicht durch die Einschränkung eines zusätzlichen Eingabegerätes limitiert. Die Effektivität der Kontrolle einer gestenbasierten Schnittstelle wird durch den Einsatz von haptischem Feedback maximiert.

Dies fanden Vo und Brewster in einer Studie heraus [21]. Sie verglichen verschiedene Eingabemethoden hinsichtlich der Geschwindigkeit und Präzision: Eingabe mittels haptischem Feedback, Eingabe mit visueller Rückmeldung, und schließlich die Kombination aus haptischem und visuellem Feedback. Die Aufgabe der Tester war es, einen roten Punkt zu lokalisieren. Für die rein haptische Rückmeldung erzeugte ein Ultrahaptics-Array einen Druckpunkt in der Luft. Zusätzlich wurde für die visuelle Eingabe und die Kombination beider Eingabemodalitäten ein Bild mit einem roten Punkt auf einem dahinter

liegenden Bildschirm an der Zielposition ausgerichtet. Der Experimentaufbau ist in Abbildung 4 einsehbar. Die experimentellen Vergleichsbedingungen wurden vollständig ausbalanciert, um Verzerrungen und Ordnungseffekte so weit wie möglich zu reduzieren.

18 Teilnehmer wurden gebeten, ihren Zeigefinger so genau und so schnell wie möglich an den Zielort zu bewegen. Die Ergebnisse der Studie zeigen, dass die Benutzer, welche nur haptische Rückmeldung erhielten, etwas schneller waren und um 50% genauere Ergebnisse lieferten als mit nur visuellem Feedback. Zusätzlich wurde festgestellt, dass die Größe des virtuellen Objekts verkleinert werden konnte (von 5,5 cm<sup>2</sup> auf 1 cm<sup>2</sup>), ohne dass die Präzision oder Geschwindigkeit wesentlich beeinträchtigt wurde. Die Kombination von haptischer und visueller Rückmeldung ergab eine noch höhere Genauigkeit. Man kann daher annehmen, dass haptisches Feedback Gestensteuerung weiter verbessern kann. [21]



■ **Abbildung 5** Links: Berührungslose Eingabe an einem Bankautomaten. Rechts: Benutzung eines Aufzugs mittels haptischen Feedbacks. [19]

## 5 Anwendungsmöglichkeiten in der Mensch-Maschine-Interaktion

Ultraschall-Feedback befreit die Nutzer von jeglichen taktilen Geräten, welche getragen oder berührt werden müssten. Nutzer können mit ihren bloßen Händen interagieren. Die Befreiung von einer anfassbaren Benutzerschnittstelle bietet entscheidende Vorteile für kritische Mensch-Maschine-Prozesse in medizinischen und industriellen Umgebungen. Es gibt viele Situationen, in denen es sinnvoll wäre, haptische Rückmeldung zu erhalten, bevor eine Bedienoberfläche überhaupt berührt werden muss. Die nicht-invasive gestenbasierte Eingabe in Kombination mit haptischem Feedback kann uns das komfortable und sichere Bedienen von Computern ermöglichen. Eine natürliche und präzise Eingabe wird dadurch ermöglicht. Für Maschinen in öffentlichen Bereichen wie

Selbstbedienungsterminals, Geldautomaten oder Aufzügen ermöglicht diese Technologie eine völlig neue Schnittstelle - hygienisch, sicher und mit reduziertem Bedarf an Reinigung und Wartung (siehe Abbildung 5). Es mindert die Wahrscheinlichkeit, sich mit auf Touch-Displays festsetzenden Bakterien und Viren zu infizieren. Darüber hinaus werden keine Fingerabdrücke beim Bedienen hinterlassen. Dies ist aus sicherheitstechnischen Gründen wünschenswert. Nützlich kann ein solches System aber noch in vielen weiteren Kontexten sein. Ein weiterer sinnvoller Anwendungsfall ist die Verwendung bei eingeschränkter Sicht, beispielsweise beim Autofahren, wenn der Fahrer den Blick nicht von der Straße abwenden sollte. Zudem wäre es sinnvoll, wenn ein Nutzer die Oberfläche nicht berühren möchte, weil seine Hände schmutzig sind, wie dies beim Kochen oder Arbeiten in der Küche der Fall ist (siehe Abbildung 6). Es kann trotzdem bequem im Web gesurft werden [13, 8, 3]. Außerdem ermöglicht haptisches Feedback, welches mitten in der Luft stattfindet, neue Wege zur Kommunikation in der Mensch-Maschine-Interaktion. Es öffnet einen neuen digitalen Informationskanal zusätzlich zum visuellen und auditiven [3].

Im Folgenden werden einige mögliche Anwendungsgebiete und Interaktionsmöglichkeiten im Feld der Mensch-Maschine-Interaktion aufgelistet.



■ **Abbildung 6** Links: Unsichtbarer ultraschallbasierter Bedienknopf im Auto. Rechts: Unsichtbarer ultraschallbasierter Bedienknopf in der Küche.[19]

### 5.1 Ultraschall-Feedback in Multimedia- und VR/AR-Entertainment-Anwendungen

Die virtuelle und die physische Welt verschmelzen immer mehr miteinander. Virtuellen Objekten die Möglichkeit geben, durch den menschlichen Tastsinn tatsächlich fühlbar zu werden, kann diese Verschmelzung vollständig machen. Für innovative Unternehmen im Bereich Gaming und Entertainment bietet dies die Möglichkeit, einzigartige, fesselnde und differenzierende Erlebnisse zu schaffen. Ablart *et al.* haben das Potential haptisch erweiterter Multimedia Erlebnisse untersucht [1]. Sie führten ein Experiment durch, in welchem Kurzfilme um eine berührungslose Feedback-Komponente erweitert wurden. Das

Ergebnis der Studie zeigte, dass haptisches Feedback in der Luft die Fähigkeit hat, die Immersion zu erhöhen, Emotionen zu beeinflussen und die Gesamtqualität von Erlebnissen positiv zu beeinträchtigen - ohne dass der Körper des Betrachters dabei berührt werden muss. Es hat somit Verwendungsmöglichkeit im interaktiven Fernsehen und in Online-Videos.

Die Integration des Tastsinns in die Welt der Virtual und Augmented Reality kann wiederum Nutzererlebnisse schaffen, welche dem Spieler erlauben, wirklich in die virtuelle Welt einzutauchen. In Verknüpfung mit Head-Mounted-Displays kann ein gewisses Bewusstsein in dieser dreidimensionalen Atmosphäre erlangt werden und möglicherweise soziale Verbindungen mit virtuellen Charakteren aufgebaut werden. Aktuell gibt es kaum Möglichkeiten, Nutzern, welche mittels eines kopfgetragenen Displays (engl. Head-Mounted-Displays) in eine virtuelle oder augmentierte Realität eingetaucht sind, unauffälliges taktiles und haptisches Feedback zu liefern, während sie virtuelle Objekte anfassen. Sand *et al.* haben diesem Mangel entgegengewirkt und einen Prototypen eines freischwebenden taktilen Feedback-System für Head-Mounted-Displays gebaut [15]. Das vorgestellte System kann 3D-Benutzerschnittstellen und virtuelle Realität auf eine neue Art und Weise verbessern. Der Prototyp verwendet die Fokuspunkte einer regulierbaren Ultraschall-Transduktoren-Matrix, um unauffälliges taktiles Feedback in der Luft zu generieren. Das Ultraschall-System und ein Hand-Positionssensor sind an der Vorderseite des am Kopf befestigten Virtual-Reality-Displays (Oculus Rift VR-Headset) angebracht. Die Versuchspersonen bevorzugten fast einstimmig die taktile Rückmeldung gegenüber keiner Rückmeldung.

Die Multimediawelt profitiert durch immersive Technologien wie Virtual Reality kombiniert mit der Möglichkeit, Anwendern auch eine körperlich spürbare Reaktion zu liefern. So können sehr persönliche Verbindungen zu Geschichten aufgebaut werden, da direkt teilgenommen, die Rolle von Hauptfiguren übernommen und der Verlauf der Geschichte beeinflusst werden kann. Dies bietet Autoren, Filmemachern und Spieleentwicklern neuartige Möglichkeiten, Konsumenten einzigartige, neuartige und realistische Erfahrungen zu bereiten. Fein abgestimmtes taktiles Feedback ist das fehlende Puzzlestück, welches 4D-Installationen, VR/AR-Erlebnissen, Filmen und Spielen den letzten Rest Immersivität verleihen kann.

## 5.2 Haptisches Marketing und Werbung

Mit der Möglichkeit, haptisches Feedback kommerziell zu nutzen, beginnt eine neue Ära des Markenmanagements. Die Umstellung auf digitale Kanäle ist bereits mit der Entstehung von E-Commerce, Online-Werbung und digitalen Informationsanzeigesystemen (engl. „digital signage“) erfolgt. Die Menschen sind also bereit, mit neuartigen digitalen Systemen zu interagieren.



■ **Abbildung 7** Links: Interaktives Poster „Ready Player One“ (Nutzer können Spielfigur mittels haptischen Feedbacks bewegen. Rechts: Interaktives Poster „Star Wars“ (Nutzer können „Kraftfeld“ auslösen)[19]

Haptisches Feedback hat einen nachweislich signifikanten Einfluss auf die Verweildauer und die Erinnerbarkeit. Dies bietet entsprechende Vorteile für Marketing und Werbung [17]. Man könnte beispielsweise ein Ultraschall-Array auf der Rückseite eines mobilen Endgeräts befestigen, um so medienrelevantes Feedback für Fernsehwerbung zu generieren. Die Ergänzung von Video und Audio mittels haptischen Feedbacks könnte so die Attraktivität mobiler Fernsehübertragung und somit auch die Akzeptanz von Werbung auf diesem Kanal erhöhen [2]. Eine weitere interessante Anwendung, um Ultraschall-Feedback für Werbezwecke zu verwenden, ist die Ausstattung von Filmpostern mit einem solchen Feedback-Medium. Die Firma „Ultrahaptics“ hat bereits mehrere Marketingaktionen mit Hilfe von interaktiven Postern durchgeführt. Ein digitales Poster für den Film „Star Wars Episode VIII“ wurde mit einem haptischem „Kraftfeld“ versehen. Nutzer konnten ihre Hände vor das digitale Filmplakat halten und nahmen einen Kraftstoß, sowie auditives und visuelles Feedback wahr. Ein anderes Plakat für den Film „Ready Player One“ ermöglichte es Betrachtern, eine Spielfigur auf dem Poster durch das Ziehen eines akustisch erzeugten Objekts zu bewegen (siehe Abbildung 7). Solche und weitere kreative Marketing-Aktionen können durch diese Technologie ermöglicht werden und durch ihre Neuartigkeit einen bleibenden Eindruck bei den jeweiligen Zielgruppen hinterlassen.

## 5.3 Übertragung von Emotionen

Menschen sind soziale Wesen und Berührung ist ein fundamentaler Bestandteil der sozialen Kommunikation. Die Übertragung von Emotionen durch Berührung, welche durch ein Ultraschall-Feedback-System erreicht werden kann,

ermöglicht neue Designchancen hinsichtlich der emotionsbasierten Kommunikation. In ihrer Studie haben Obrist et. al. die Kommunikation von Gefühlen mit Hilfe des Ultrahaptics System getestet [14]. Sie fanden heraus, dass durch die Stimulation unterschiedlicher Stellen der Hand unterschiedliche Gefühle vermittelt werden können. Der Bereich um den Daumen, um das Handgelenk und die Handfläche sind Flächen, auf denen Berührungen als positive Emotionen aufgefasst werden. Die Region um den kleinen Finger und die äußeren Bereiche der Handfläche sind im Gegensatz dazu Bereiche um negative Emotionen auszudrücken. Darüber hinaus konnte durch die Studie herausgefunden werden, dass die Ausrichtung der Bewegung ebenfalls Auswirkung auf die emotionale Auffassung hat. Um positive Gefühle zu übermitteln, sollten die vertikale Bewegung in die Richtung der Person gehen. Um negative Gefühle auszudrücken, wird eine Bewegung in Richtung der Fingerspitzen weg von der Handfläche empfohlen. Auch die Stärke des Frequenzbereichs hat einen Einfluss auf die emotionale Auffassung der Berührung. Dies zeigt, dass das Kommunizieren in der Luft auf Basis des Tastsinns nicht willkürlich ist, dass Nutzer so taktile Stimulierungen wahrnehmen können und emotionale Empfindungen übermittelt werden können. Dies könnte beispielsweise für die Kommunikation zwischen Menschen, die sich weit voneinander entfernt aufhalten, nützlich sein. Vor allem für Paare in Fernbeziehungen könnte dies relevant sein [14].

#### 5.4 Ultraschall-Feedback-Systeme im Automotive Bereich

Ein weiterer interessanter und wichtiger Anwendungsbereich sind Fahrzeug-Bedienoberflächen. Die primäre Motivation für die Verwendung von Handgesten für Bedienelemente in einem Fahrzeug basiert auf dem Vorteil, nicht die Augen von der Straße nehmen zu müssen um herkömmliche Einstellung vornehmen zu müssen [6]. Taktiles Feedback ist in neueren Autos bereits ein wichtiger Bestandteil der zur Sicherheit der Fahrzeuginsassen beiträgt. Beispielsweise erhält der Fahrer beim Überqueren der Fahrbahnschwelle eine Reaktion des Lenkrads in Form von Vibration. [20] In Kombination mit Gestenerkennung befreit es von der Ablenkung durch physische Tasten, kleine Touchscreen-Oberflächen oder durch das Betrachten eines Bildschirms. Die Steuerung durch Gesten-Eingabe in der Luft ermöglicht eine intuitive Bedienung des Infotainment-System oder andere Einstellungen im Auto. Taktiles Feedback bestätigt dem Fahrer die Richtigkeit seiner Eingabe, ohne dabei visuelle Aufmerksamkeit vom Fahrer zu fordern (siehe Abbildung 6). [4]

#### 5.5 Ultraschall-Feedback-Systeme im Gesundheitswesen

Auch im Gesundheitswesen liefert eine gestenbasierte Eingabe, die auch ohne direkte Berührung eine spürbare Reaktion zeigt, viele neue Möglichkeiten.

Großes Potential bietet es Menschen mit körperlichen Behinderungen (Tauben, Stumme oder Blinde). Die Sprache der stummen und tauben Menschen - die Gebärdensprache - basiert bereits auf komplexen Gesten. Es würde sich daher anbieten, ein System mit der Funktionalität diese Sprache zu verstehen, auszustatten. Blinde Menschen können vor allem von der taktilen und haptischen Feedback-Fähigkeit profitieren. Sie könnten ohne visuellen Input Objekte erkennen und nur mittels Gesten mit einem Computer interagieren. Möglicherweise könnte dies eine funktionelle Alternative zur Blindenschrift darstellen. Ein Prototyp eines barrierefreien autonomen Fahrzeugs „Accessible Olli“ (hergestellt von Local Motors und IBM Watson Technologies), in welchem die Ultrahaptics-Technologie verbaut ist, um leichter von blinden Menschen verwendet werden zu können, existiert bereits. In Kombination mit Gestenverfolgungs-Software und -Hardware besitzt das Fahrzeug einen virtuellen Halte-Knopf, welcher sich mitten in der Luft befindet. Der Knopf muss nicht von Passagieren lokalisiert werden, sondern macht sich selbst bemerkbar, wenn es nötig ist. Dieser ist viel zugänglicher und erfordert weniger Kraftaufwand als sein herkömmliches physisches Pendant. [5]

Eine sterile Mensch-Maschine-Schnittstelle kann zudem von großer Bedeutung sein. Gestenbasierte Systeme mit haptischem Feedback können Touchscreens ersetzen, die heutzutage in vielen Operationssälen in Krankenhäusern verwendet werden. Ein Chirurg könnte so bei der Bedienung eines medizinischen Informations Geräts die Kontamination mit möglichen Krankheiten oder Infekten vermeiden. Die Feedback-Funktionalität ist auch mit Gummihandschuhen gegeben [17]. Touchscreens müssen außerdem versiegelt und gründlich gereinigt werden, um eine Ansammlung oder Ausbreitung von Verunreinigungen und Infektionen zu verhindern. Dieses Risiko und der zusätzliche Aufwand könnte mit einem feedbackbasierten Hand-Erkennungssystem eingespart werden [6].

Trotz vieler bestehender Anwendungsszenarien ist diese Interaktionsform relativ neu und ungewohnt. Nutzer müssen sich erst noch daran gewöhnen. Die grundsätzlichen Charakteristiken der Wahrnehmung dieses neuen Feedback-Mediums und wie Ultraschall-Berührungen am Besten an einer Schnittstelle zu verwenden sind, muss erst noch richtig verstanden werden. [22]

## 6 Grenzen ultraschallbasierter haptischer Feedback-Systeme

Trotz vieler Anwendungsmöglichkeit in der Mensch-Computer-Interaktion haben das System von Ultrahaptics und generell ultraschallbasierte Feedback-Systeme Grenzen. Wie diese Systeme in verschiedenen Benutzerschnittstellen

bestmöglich verwendet werden können und um die grundsätzlichen Charakteristika der Wahrnehmung dieses neuen Feedback-Mediums vollständig zu verstehen, sind weitere kontrollierte und gründliche Experimente im Bereich der Mensch-Maschine-Interaktion nötig. Im Folgenden werden einige aktuelle Nachteile und verbesserungswürdige Eigenschaften dieser Systeme aufgezählt und erläutert.

- **Mangelnde Skalierbarkeit** Obwohl Prototypen für wissenschaftliche und kommerzielle Zwecke erworben werden können, sind diese nicht günstig. Der Preis eines „TOUCH Development Kit“ liegt beispielsweise bei ca. 2500\$ [18]. Der erhöhte Preis dämpft die schnelle Verbreitung und die Vermehrung weiterer Anwendungsoptionen ein. Bisher sind diese Systeme durch die Notwendigkeit von Strom und einem Computer zur Verarbeitung von Gesten und Steuerung des Systems zudem an einen festen Ort gebunden. Die Integration in bestehende Technologien ist umständlich und sie sind in ihrer Größe nur schwer skalierbar.
- **Verbesserungsfähige User Experience** Da es sich bei Ultraschall-Feedback im Gegensatz zu anderen taktilen Feedback-Systemen, wie zum Beispiel Vibrationsmotoren oder Pin-Arrays, um eine relativ neue Form der Mensch-Maschine-Interaktion handelt, und bisher kaum gut funktionierende Prototypen für die Wissenschaft verfügbar waren, gibt es immer noch einige Herausforderungen bezüglich der User Experience. Dazu gehören nicht nur technische Ansprüche, wie die Maximierung der Reichweite und des Sichtfeldes, sondern insbesondere die Verbesserung der Erkennungsgenauigkeit: das bessere Erkennen von Position und Bewegung, und das richtige Verarbeiten der Benutzerabsicht. [20] Die Gebundenheit an einen Ort und die relativ große Größe führen dazu, dass Feedback nur von einer globalen Stelle und nur in eine bestimmte Richtung ausgestrahlt werden kann. Dies bedeutet eine Limitierung des nutzbaren Interaktionsraums für gestenbasierte Schnittstellen. Ohne diese Einschränkung könnten Nutzer im gesamten offenen Raum und in verschiedenen Richtungen aufgenommen werden. [22]
- **Hardware- und Softwareanforderungen zur Gestenerkennung** Taktiler Feedback erweitert die Genauigkeit der berührungslosen Gestenerkennung. Um Benutzerbefehle optimal zu verarbeiten und immer die richtige Aktion auf eine spezifische Eingabe zu liefern, müssen sowohl bessere und genauere Software-Algorithmen als auch kostengünstigere Kamera- und Sensor-Hardware mit stärkerer Leistung produziert werden. [20]
- **Grobheit von Ultraschall-Objekten** Die räumliche Auflösung ist durch die Ultraschallwellenlänge (derzeit 8,5 mm) beschränkt. Höherfrequenter Ultraschall kann diese Auflösung verbessern, aber der Benutzerbereich kann aufgrund des Dämpfungsfaktors kleiner werden. Weitere Nutzerstudien mit

komplexeren Objekten und dem Vergleich zwischen realen und Ultraschall-Objekten sind notwendig. [10] Ein weiterer Kompromiss eines solchen haptischen Ultraschall-Feedback-Systems ist die Beziehung zwischen der Größe der Matrix an Ultraschall-Transduktoren und dem erzeugten Fokuspunkt. Der Durchmesser dieses Punktes steht im umgekehrten Verhältnis zu der Größe der Transduktoren-Matrix, genauso wie beispielsweise eine größere Linse eine höhere Auflösung erzielt. Die Transduktoren-Matrix ist bei Systemen wie UltraHaptics meist nicht allzu groß, um nicht zu viel Platz einzunehmen. Dies resultiert in einem größeren Durchmesser des Brennpunktes, was wiederum dazu führen kann, dass der Brennpunkt zu groß ist, um einen einzelnen Finger zu stimulieren. [8]

## 7 Fazit und Ausblick

Handgestenerkennung kombiniert mit haptischem Feedback, erzeugt durch Ultraschall, kann in der Zukunft in vielen Bereichen verwendet werden. Die abnehmenden Hardware- und Verarbeitungskosten sind ein bedeutender Faktor, der die Gestenerkennung für eine weit verbreitete Verwendung realisierbar macht. Auch die stetige Verbesserung von Ultraschall-Hardware und -Software zur Gestenerkennung und Verarbeitung wird dazu führen, dass diese Technologie in den nächsten Jahren noch besser funktionieren wird. Um in Zukunft Teil unseres alltäglichen Leben sein zu können, muss dieses Feedback-Medium hinsichtlich der User Experience, der Feingranularität der unsichtbaren dreidimensionalen Objekte und der Skalierbarkeit weiterentwickelt und verbessert werden.

---

### Literatur

- 1 Damien Ablart, Carlos Velasco, and Marianna Obrist. Integrating mid-air haptics into movie experiences. In *Proceedings of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video*, TVX '17, pages 77–84, New York, NY, USA, 2017. ACM. URL: <http://doi.acm.org/10.1145/3077548.3077551>, doi:10.1145/3077548.3077551.
- 2 Jason Alexander, Mark T. Marshall, and Sriram Subramanian. Adding haptic feedback to mobile tv. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '11, pages 1975–1980, New York, NY, USA, 2011. ACM. URL: <http://doi.acm.org/10.1145/1979742.1979899>, doi:10.1145/1979742.1979899.
- 3 A. Carter, Thomas. Perception in a haptic system, 2018.
- 4 Chris Desnoyers and Sarah Pohorecky. Ultrasonic haptic feedback. 2017.

- 5 Alex Driskill-Smith. Accessibleolli | mid-air haptics and accessibility. <https://www.ultrahaptics.com/news/blog/accessible-olli/>. Accessed: 28-05-2018.
- 6 Pragati Garg, Naveen Aggarwal, and Sanjeev Sofat. Vision based hand gesture recognition. *World Academy of Science, Engineering and Technology*, 49(1):972–977, 2009.
- 7 Leonid R. Gavrilov, Grigoryi V. Gersuni, Oleg B. Ilyinski, Efim M. Tsi-rulnikov, and Eugenyi E. Shchekanov. A study of reception with the use of focused ultrasound. i. effects on the skin and deep receptor structures in man. *Brain Research*, 135(2):265 – 277, 1977. URL: <http://www.sciencedirect.com/science/article/pii/0006899377910307>, doi: [https://doi.org/10.1016/0006-8993\(77\)91030-7](https://doi.org/10.1016/0006-8993(77)91030-7).
- 8 T. Hoshi. Development of aerial-input and aerial-tactile-feedback system. In *2011 IEEE World Haptics Conference*, pages 569–573, June 2011. doi: 10.1109/WHC.2011.5945548.
- 9 Apple Inc. Apple watch. <https://www.apple.com/accessibility/watch/>. Accessed: 02-06-2018.
- 10 S. Inoue, Y. Makino, and H. Shinoda. Active touch perception produced by airborne ultrasonic haptic hologram. In *2015 IEEE World Haptics Conference (WHC)*, pages 362–367, June 2015. doi: 10.1109/WHC.2015.7177739.
- 11 F. Kalantari, E. Lank, Y. Rekik, L. Grisoni, and F. Giraud. Determining the haptic feedback position for optimizing the targeting performance on ultrasonic tactile displays. In *2018 IEEE Haptics Symposium (HAPTICS)*, pages 204–209, March 2018. doi: 10.1109/HAPTICS.2018.8357177.
- 12 Microsoft. Kinect für windows. <https://developer.microsoft.com/de-de/windows/kinect>. Accessed: 02-06-2018.
- 13 Yasuaki Monnai, Keisuke Hasegawa, Masahiro Fujiwara, Kazuma Yoshino, Seki Inoue, and Hiroyuki Shinoda. Haptomime: Mid-air haptic interaction with a floating virtual screen. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14, pages 663–667, New York, NY, USA, 2014. ACM. URL: <http://doi.acm.org/10.1145/2642918.2647407>, doi: 10.1145/2642918.2647407.
- 14 Marianna Obrist, Sriram Subramanian, Elia Gatti, Benjamin Long, and Thomas Carter. Emotions mediated through mid-air haptics. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 2053–2062, New York, NY, USA, 2015. ACM. URL: <http://doi.acm.org/10.1145/2702123.2702361>, doi: 10.1145/2702123.2702361.
- 15 Antti Sand, Ismo Rakkolainen, Poika Isokoski, Jari Kangas, Roope Raisamo, and Karri Palovuori. Head-mounted display with mid-air tactile feedback. In *Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology*, VRST '15, pages 51–58, New York, NY, USA, 2015. ACM. URL:

- <http://doi.acm.org/10.1145/2821592.2821593>, doi: 10.1145/2821592.2821593.
- 16 Rajinder Sodhi, Ivan Poupyrev, Matthew Glisson, and Ali Israr. Areal: interactive tactile experiences in free air. *ACM Transactions on Graphics (TOG)*, 32(4):134, 2013. doi: 10.1145/2461912.2462007.
- 17 Ultrahaptics. The magic of mid-air-haptic-feedback. <https://www.ultrahaptics.com/applications/>. Accessed: 28-05-2018.
- 18 Ultrahaptics. Touch development kit. <https://www.ultrahaptics.com/products-programs/touch-development-kit/>. Accessed: 28-05-2018.
- 19 Ultrahaptics. Ultrahaptics official website. <https://www.ultrahaptics.com/>. Accessed: 28-05-2018.
- 20 Ultrahaptics. Gesture recognition and haptic feedback technologies. a whitepaper on touchless interfaces and tactile feedback. 2016. URL: <http://www.eenewsautomotive.com/Learning-center/ultrahaptics-gesture-recognition-and-haptic-feedback-technologies>.
- 21 Dong-Bach Vo and Stephen A Brewster. Touching the invisible: Localizing ultrasonic haptic cues. In *World Haptics Conference (WHC), 2015 IEEE*, pages 368–373. IEEE, 2015.
- 22 Graham Wilson, Thomas Carter, Sriram Subramanian, and Stephen A. Brewster. Perception of ultrasonic haptic feedback on the hand: Localisation and apparent motion. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 1133–1142, New York, NY, USA, 2014. ACM. URL: <http://doi.acm.org/10.1145/2556288.2557033>, doi: 10.1145/2556288.2557033.

# Haptische Rückmeldung in gemischter Realität

Anita C. Baier

Ludwig-Maximilians-Universität, München, Deutschland  
anita.baier@campus.lmu.de

---

## Zusammenfassung

Das Fühlen und Ertasten von Gegenständen ist ein wesentlicher Aspekt des Erlebens einer Realität. Dieser Artikel setzt sich damit auseinander, was der Begriff Haptik umfasst und wie er in die virtuelle und gemischte Realität integriert werden kann. Es werden Schwierigkeiten diskutiert, die speziell bei der Gemischten Realität auftreten können und Beispiele zur Bereitstellung von haptischer Rückmeldung vorgestellt: So kann eine kalte, harte, raue Betonmauer - die nur virtuell existiert - nicht nur gesehen, sondern auch mit Hilfe von Kraftrückmeldung und taktile Rückmeldung haptisch erlebbar werden.

**Keywords and phrases** Gemischte Realität, Haptik, Haptische Rückmeldung, Kraftrückmeldung, taktile Rückmeldung

## 1 Einleitung

Mit der Entwicklung kommerziell vertriebener und günstiger Möglichkeiten zum visuellen Erleben einer Virtuellen oder Gemischten Realität haben simulierte Umgebungen und Gegenstände an Relevanz gewonnen. Die Anwendungsmöglichkeiten reichen dabei vom privaten Gebrauch, wie zum Beispiel Computerspiele zur Vergnügung bis zum professionellen Gebrauch wie Simulationen zum Erlernen komplizierter Reparaturvorgänge einer Industrieanlage. Dabei möchte jeder Nutzer ein möglichst umfassendes Erlebnis haben. Doch ist dafür nicht nur wichtig, was er sieht. Auch die Haptik, das Empfinden wenn er etwas ertastet oder von etwas berührt wird, soll sich so real wie möglich anfühlen.

Aber wie wird etwas fühlbar, was gar nicht physisch existiert? In diesem Artikel werden wir zuerst generell über Haptik und Gemischte Realität sprechen um einen Eindruck zu bekommen, wie diese beiden Dinge zusammenhängen. Es werden die Vorteile der Bereitstellung einer haptischen Rückmeldung vorgestellt, aber auch welche Herausforderung diese mit sich bringt. Dann werden wir Ansätze diskutieren, wie eine haptische Rückmeldung aussehen kann, im



© Anita C. Baier;

licensed under Creative Commons License CC-BY

Cite as: Anita C. Baier. Haptische Rückmeldung in gemischter Realität. In *2nd Seminar on Ubiquitous Interaction (UBIACTION 2018-2)*. Editors: Matthias Hoppe, Jakob Karolus, Thomas Kosch, Pascal Knierim, Albrecht Schmidt. June 28th, 2018. Munich, Germany. pp. 15:1–15:20.

Einzelnen die Krafrückmeldung, die taktile Rückmeldung und die Verwendung physischer Gegenstände, und einige Beispielanwendungen dazu kennenlernen.

## 2 Haptik in der (simulierten) Realität

Spätestens mit der Entwicklung günstiger Systeme wie beispielsweise das Google Cardbord<sup>1</sup> die Virtuelle-Realitäts-Brille Samsung Gear VR<sup>2</sup> und deren kommerzielle Bereitstellung [42] hat das Thema Virtuelle und Gemischte Realität Einzug in das Bewusstsein der breiten Masse erhalten und für die Gerätehersteller an Bedeutung gewonnen. Zum genaueren Verständnis werden wir hier nun näher auf die Begriffe *Virtuelle Realität* (im Folgenden auch *VR* für engl. *Virtual Reality*), *Erweiterte Realität* (im Folgenden auch *AR* für engl. *Augmented Reality*) und *Gemischte Realität* (im Folgenden auch *MR* für engl. *Mixed Reality*) eingehen.

### 2.1 Virtualität und Realität

Bei VR ist der Schlüssel die Bereitstellung einer interaktiven und dreidimensionalen virtuellen Umgebung, die vollständig mit dem Computer generiert ist [22, 8]. Ziel ist es, dass ein Betrachter vollständig in diese Umgebung eintauchen kann, sodass die echte Welt im Bewusstsein des Nutzers in den Hintergrund rückt. Wesentliche Aspekte sind dabei [53, 46]:

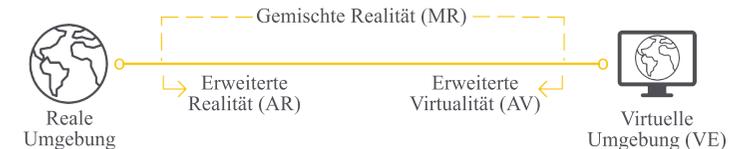
- **Immersion:** Gefühl des *Eintauchens* des Nutzers in die simulierte Realität. Es geht um den technischen Grad des Veränderns seiner Wahrnehmung, so hingehend, dass die simulierte statt der realen Umgebung als wahr empfunden wird.
- **Präsenz:** Gefühl des Nutzers, sich in einer simulierten Umgebung tatsächlich zu befinden. Der Nutzer identifiziert sich beispielsweise mit einem ihn repräsentierenden Avatar und nimmt die simulierte Umgebung als ganzheitliche Welt wahr.
- **Navigation:** Der Nutzer soll durch die virtuelle Umgebung navigieren können und sich darin frei bewegen, aber nicht durch generierte Grenzen hindurchgehen können.
- **Interaktion:** Der Nutzer soll mit seiner Umgebung und darin bereitgestellten Objekten interagieren können.

Diese Punkte spielen auch bei der AR und der MR eine Rolle. Doch zuerst zur AR: auch hier gibt es generierte Dinge, doch muss nicht eine vollständige

<sup>1</sup> <https://vr.google.com/cardboard>, letzter Zugriff: 18. Juli 2018, 18:28 MESZ

<sup>2</sup> <http://www.samsung.com/de/wearables/gear>, letzter Zugriff: 30. Mai 2018, 22:08 MESZ

Umgebung simuliert werden. Während der Betrachter bei der VR komplett in eine virtuelle Welt eingebunden werden soll, geht es bei der AR vorrangig um die Erweiterung der bestehenden Welt die ein Mensch als seine reale, physische Umgebung wahrnimmt. Nach Milgram et. al. werden dieser Welt in Echtzeit computergenerierte Objekte hinzugefügt. Diese können sowohl Gegenstände als auch anderweitige Informationen sein und müssen per Definition interaktiv sein [34, 4]. Die AR kann also einen Schritt vor der VR eingeordnet werden. Die Erweiterung kann weiter verschiedene Ausprägungen haben: Je nach Höhe des Anteils an Realität respektive der Virtualität wird auch von *AV*, engl. *Augmented Virtuality*, zu Deutsch *Erweiterte Virtualität* gesprochen. Milgram et. al. entwarfen deshalb das sogenannte *Realitäts-Virtualitäts-Kontinuum* (im Folgenden auch *VC* für engl. *Virtuality Continuum*) und prägte somit den Ausdruck der MR, welcher als Dachbegriff für die AR und AV bezeichnet werden kann (siehe Abbildung 1).



■ **Abbildung 1** Nach Milgram et. al. umfasst das VC die Ausprägungen der Erweiterung der Realität bzw. Virtualität und reicht dabei bis vor die Erreichung der vollständig realen Umgebung bzw. der vollständig virtuellen Umgebung [34].

Quelle: in Anlehnung an Milgram et. al. [34], 1995

Eine Virtuelle oder Erweiterte Realität kann visuell durch eine Vielzahl an Geräten vermittelt werden. Klassisch vor allem für die VR sind die *Head-Mounted-Displays* zu nennen, eine Anzeige, die direkt am Kopf eines Nutzers befestigt ist und eine Projektion vor seinen Augen oder sogar direkt auf seine Netzhaut erlaubt, oder die sogenannte *CAVE*, bei dem der Betrachter sich in einem Raum befindet, bei dem an bis zu sechs Wänden die virtuelle Umgebung dargestellt wird [22].

### 2.2 Haptik

Der Begriff *Haptik* wurde Ende des 19. Jahrhunderts von Max Dessoir geprägt. Der Berliner Philosoph und Biologe setzte sich in der Dissertation *Über den Hautsinn*<sup>3</sup> umfassend mit dem Tastsinn und der fühlenden Wahrnehmung

<sup>3</sup> Volle Version unter [https://archive.org/details/bub\\_gb\\_IzMaAAAAAYAAJ](https://archive.org/details/bub_gb_IzMaAAAAAYAAJ), letzter Zugriff: 30. Mai 2018, 17:35 MESZ

auseinander. Er führte dabei den Term, in Anlehnung an die Optik und die Akustik, als Ableitung des griechischen Verbs *haptikós*, zu deutsch so viel wie *fähig zu berühren*, ein. Die Haptik umfasst demnach gleichwohl den *Contactsinn* und die *Phelaphesie*, welche sich wiederum aus Berührungs- und Druckempfindung sowie Empfindungen der Muskeln und durch Tasten zusammensetzen ([13], S. 72 ff). Generell ist der Tastsinn keinesfalls nur darauf beschränkt, etwas mit der Hand zu spüren oder zu ertasten. Es geht auch um die Wahrnehmung mit dem ganzen Körper, das Fühlen von Kälte oder vom Widerstand eines Windstoßes und weiterhin sogar um Gefühle als ganzkörperliche Wahrnehmung. Lange wurde der Tastsinn in Philosophie und Psychologie als *Leitsinn* angesehen [18].

Heute setzt sich in der Literatur oftmals eine Abgrenzung durch zwischen der *haptischen Wahrnehmung* und der *taktilen Wahrnehmung*. Diese zwei Ansätze unterscheiden sich durch die Relation des Empfindenden zum Reiz wobei ersteres ein aktives Ertasten voraussetzt, zweiteres eine passive Stimulusaufnahme bedeutet [19].

Auf die biophysiologicalen Grundlagen wollen wir hier nicht eingehen, doch zusammenfassend gesagt kann - ob aktiv oder passiv erfahren - die Haptik Informationen über die physische Umwelt durch mechanische Reize auf unseren Körper vermitteln [18, 44]. Dazu gehören zum einen Attribute wie das Gewicht oder die Härte eines Objekts, die Temperatur eines Körpers oder die Glattheit seiner Oberfläche [8]. Zur aktiven haptischen Exploration eines Gegenstandes mit der Hand identifizierten Ledermann et. al. [25] folgende Prozeduren, die hier mit ihrem englischen Originalbegriff aufgeführt und auf deutsch erklärt werden:

- *lateral motion*: Streichen über einen relativ homogenen Teil einer Oberfläche
- *pressure*: Drücken gegen einen Teil eines Gegenstands, der ansonsten durch eine andere Kraft stabilisiert wird
- *static contact*: Einen Gegenstand bewegungslos berühren, der ansonsten durch eine andere Kraft stabilisiert wird
- *unsupported holding*: Einen Gegenstand von einem Stabilität bietendem Zustand fort heben und ohne Unterstützung in der Hand halten
- *enclosure*: Umfassen eines Objekts
- *contour following*: Möglichst flüssiges dynamisches Entlangfahren an den Kanten eines Gegenstands
- *function test*: Alle ausführende Bewegungen, die die tatsächliche Funktionalität eines Gegenstands betreffen
- *part motion test*: In Bewegung bringen eines Teils eines Objekts durch Krafteinwirkung, während Gegenkraft auf den Rest des Objekts ausgeübt wird

## 2.3 Haptische Rückmeldung in erweiterter und virtueller Umgebung

Generell ist es ein ganz natürliches Bedürfnis des Menschen, nach Gegenständen zu greifen und sie zu berühren [53]. Révész definierte sogar das *stereoplastische Prinzip*, welches er als „Grundprinzip der haptischen Wahrnehmung“ ansieht und besagt, dass im Mensch stets der Drang herrscht, Objekte allumfassend zu betasten um sie haptisch zu erkunden und sie in ihrer Gesamtheit und Plastizität zu begreifen. Diese Exploration steht gegenüber der visuellen Wahrnehmung, die von Natur aus in ihrem Gesamteindruck eingeschränkter ist ([39], S. 165 f)<sup>4</sup>. Zwar haben Versuche ergeben, dass in der Regel visuelle über haptische Reize dominieren, doch kann die Gesamtwahrnehmung durchaus von der Haptik beeinflusst werden [55, 40, 37, 1]. Gleichzeitig ist die Theorie des kognitiven Overloads von Sweller nach Mayer auf multimediale Umgebungen übertragbar und somit höchst aktuell. Die Kapazität des menschlichen Arbeitsgedächtnisses sei demnach begrenzt und sollte kognitiv nicht überladen werden, sodass die Bedienung mehrerer Wahrnehmungskanäle im allgemeinen vorzuziehen ist [32, 51, 33].

Als Vorreiter der AR wird die *Virtual fixture* in den 1990er Jahren von Louis Rosenberg angesehen - eine Entwicklung, bei der ein Nutzer über ein Exoskelett zwei Roboterarme fernsteuerte. Diese und die Umgebung in der sie sich befanden wurden dem Nutzer angezeigt sodass er sie als seine eigenen Arme in dieser Umgebung wahrnehmen konnte. Zusätzlich war das System bereits in der Lage Barrieren und Anweisungen zu generieren, die Umgebung wurde also augmentiert [41]. Im Gegensatz dazu wurde die VR als fester Begriff erstmals 1982 erwähnt, die Entwicklung dort hin liegt aber noch länger zurück. Somit ist es nicht verwunderlich, dass viel Forschung bezüglich haptischer Rückmeldung sich erst einmal auf die VR und weniger auf die MR bezieht [5]. Deshalb werden wir in diesem Artikel nicht darum herum kommen uns oft auf Arbeiten bezüglich der VR zu stützen, doch ist die Exploration von Objekten in der VR und der MR sehr ähnlich.

„The whole idea of VR is to allow to navigate and interact with a VE.“<sup>5</sup>  
(J. Vince [53])

Kein Wunder also, dass die Haptik eines virtuellen Gegenstandes in der VR und somit auch MR zum Verständnis seiner Umgebung von essenzieller Bedeutung

<sup>4</sup> Volle Buchversion unter <https://archive.org/details/formenweltdestas01grev>, letzter Zugriff: 30. Mai 2018, 12:37 MESZ

<sup>5</sup> Zu Deutsch etwa: Die ganze Idee hinter der VR ist die Möglichkeit, durch eine VE zu navigieren und mit ihr zu interagieren.

ist - kann man sie doch nach unserer Beschreibung unter dem Kapitel Haptik als jene Eigenschaften begreifen, die das Objekt für den Tastsinn ausmachen und einem Befühlenden umfassende Informationen über dessen Beschaffenheit geben. Deshalb ist es wichtig, dass auch generierte Objekte fähig sind, eine haptische Rückmeldung zu geben, also eben jene Eigenschaften zu simulieren und zu Verfügung zu stellen. Um dies zu erreichen benötigt es sogenannte *haptische Schnittstellen* (im Folgenden *HI* für engl. *haptic interfaces* nach Burdea et. al. [8]), welche die Rückmeldung an den Menschen übermitteln. Die Bereitstellung ergibt eine Reihe von Vorzügen für den Benutzer, dazu gehört die Erhöhung des Verständnisses vom virtuellen Zustand von Objekten [29], des Realitätserlebnisses (sei es der virtuellen Umgebung oder eines virtuellen Objekts) [44, 52, 29], der Immersion [8] und somit der Präsenz [21, 43], eine bessere und effizientere Leistung der Nutzer innerhalb einer virtuellen Umgebung [36, 35, 41] durch Ermöglichung eines intuitiveren Umgangs mit der Umgebung [48] und Vereinfachung von Aufgaben [15] sowie Zuwachs von Spaß an der auszuführenden Aufgabe [29].

## 2.4 Herausforderungen im Bereich gemischte Realität

Es gibt mittlerweile eine Vielzahl an verschiedenen HIs mit diversen Simulationstechnologien um Rückmeldung über die haptischen Eigenschaften eines virtuellen Gegenstandes zu generieren [47]. Als naheliegende und zum Teil durch kommerziellen Vertrieb bekannte Beispiele sind hier Eingabemethoden wie der taktile *Joystick* (Ende der 1980er bis späte 1990er Jahre), sowie diverse tragbare Systeme (im Folgenden engl. *Wearables* genannt) wie Exoskelette und vibrotaktile Westen oder Handschuhe wie der *CyberTouch Glove*<sup>6</sup> (1998) oder die *PHANTOM*-Schnittstelle (1994) zu nennen, die über einen Aufsatz Kraft auf die Fingerspitze eines Nutzers ausübt [8, 30]. Die Forschung und Entwicklung zur VR hat dabei einge, auch manchmal sperrig anmutende, HIs hervorgebracht.

Lopes et. al. sprechen mehrere Schwierigkeiten bei der Entwicklung von Methoden für eine haptische Rückmeldung in der MR an [29]. Diese liegen vor allem in ihrer Natur selbst, denn Objekte mit denen der Nutzer interagiert müssen im Gegensatz zur reinen VR nicht virtuell sein sondern es kann sich auch um reale, physische Gegenstände handeln. Ebenso sind in der MR auch Teile der Umgebung real. Der Nutzer kann also Eindrücke um ihn herum wahrnehmen und ist nicht wie in der VR vollkommen abgeschirmt vom Sehen

<sup>6</sup> <http://www.cyberglovesystems.com/cybertouch>, letzter Zugriff: 13. Juni 2018, 10:50 WESZ

der ihn beeinflussenden HIs. Sind diese ungünstig gestaltet, könnte das zu einer Minderung des Realitätsgefühls führen, da der Nutzer abgelenkt wird. In jedem Fall ist eine Diskrepanz zwischen dem, was ein Nutzer sieht und dem, was ein Nutzer fühlt zu vermeiden [5]. Außerdem könnte ein HI zu sperrig sein. Abgesehen davon, dass dies den Nutzer wieder von seiner MR-Erfahrung ablenken könnte, weil es für die Umgebung unpassend aussieht, könnte es ihn auch in seiner Bewegungsfreiheit zu stark einschränken, wo er sich doch prinzipiell in einer uneingeschränkten realen Umgebung bewegen könnte. Nach Révész Forschung ist es schon beim realen Raumbewusstsein für das Gefühl, sich bei geschlossenen Augen im Raum zu fühlen von essenzieller Bedeutung, jederzeit seine Kraft- und Bewegungsfreiheit im Raum ausüben zu können. „Die Beziehung zwischen Eigenkörper und Bewegung wirkt also raumbildend“ [39]. Möglich ist auch, dass die Hände belegt sind, diese aber benötigt werden um mit realen Gegenständen aus der gemischten Realität interagieren zu können [29]. Eine Einschränkung des Nutzers findet sich unter vielen Autoren zum Thema als einer der wichtigsten Punkte, der zu vermeiden und gleichzeitig als herausfordernd gilt [29, 48, 44, 10].

Ein Effekt der MR ist auch, dass die visuelle Augmentierung das Sichtfeld des Nutzers einschränken kann [29], dies ist zwar nicht primär ein Problem für die haptische Rückmeldung, könnte aber eines werden, sobald Geräte für die visuelle Erweiterung und die Bereitstellung von haptischer Rückmeldung kombiniert werden. Bis dahin kann die haptische Rückmeldung das Fehlen von visuellen Informationen wegen einer unvollständigen Augmentation durch ein eingeschränktes Sichtfeld der visuellen Schnittstelle ausgleichen. So zum Beispiel bei Lopes et. al. - sie zeigen ein virtuelles Sofa, welches durch den Nutzer an eine reale Wand geschoben werden soll. Zwar können die Nutzer das Ende des Sofas an der Wand nicht immer sehen, durch die gebotene haptische Rückmeldung wissen sie aber, wann das Ziel erreicht ist [29].

Insgesamt kann gesagt werden, dass die Schwierigkeiten von der Gestaltung der HIs und von der Gestaltung und Ausprägung der simulierten Umgebung und der zugelassenen Realität abhängt. Durch die Sichtbarkeit der HIs und durch die Tatsache, dass auch reale Gegenstände zur Gemischten Realität gehören ist die Simulation von haptischer Rückmeldung in der MR schwieriger zu bewältigen als in der VR.

## 3 Generelle Ansätze und Beispiele

Welche Möglichkeiten gibt aber nun genau, ein eigentlich gar nicht physisch existierendes Objekt fühlbar werden zu lassen? In Anlehnung an die im Unterkapitel *Haptik* besprochenen Möglichkeiten der haptischen Wahrnehmung

identifizierten Burdea et. al. als drei grundsätzliche Modalitäten der haptischen Rückmeldung die Simulation von Kraft, taktilen und propriozeptiven Informationen. Letzteres bezieht sich auf die sogenannte Tiefensensibilität und soll dem Menschen Informationen zu seiner Eigenwahrnehmung vermitteln wie beispielsweise die Lage seines eigenen Körpers im Raum. Abweichend zu diesem Schema werden wir in diesem Abschnitte einer etwas anderen Aufteilung folgen. Da bei der MR im Gegensatz zur reinen VR für den Betrachter reale Objekte oder Teile seiner realen Umgebung wahrnehmbar sind zu denen er Bezug aufbauen kann und deren visuelle Wahrnehmung nach Burdea et. al. über deren propriozeptive Wahrnehmung dominiert ([7], S. 271), werden wir auf eine Rückmeldung für letzteres hier nicht weiter eingehen. Wir werden stattdessen im Folgenden die generellen Ansätze der Bereitstellung von taktiler Rückmeldung und Krafrückmeldung sowie die Verwendung von physikalischen Gegenständen besprechen und diese versuchen durch eine knappe Vorstellung exemplarischer HIs für die Verwendung in der MR zu veranschaulichen. Dabei ist zu beachten, dass manche Beispiele sich nicht immer eindeutig einer einzelnen Modalität zuordnen lassen, was aber dem Zwecke der Verdeutlichung keinen Abbruch tun soll: Vielmehr ist der eigentliche Sinn eines HI die Vereinigung der Modalitäten [8].

### 3.1 Taktile Rückmeldung

Bei der taktilen Rückmeldung geht es darum fühlbare Attribute eines Gegenstandes wie die Form und die Beschaffenheit seiner Oberfläche nachzuahmen [29] sowie das Fühlen von Druck und Vibration, Temperatur und Schmerz über die Haut zu vermitteln [38]. Es gibt einige Möglichkeiten, eine taktile Rückmeldung zu simulieren. Dazu gehören zum Beispiel vibrotaktile HIs, die also über Vibrationen Informationen an die Haut geben. Die Benutzung von Vibration ist dabei eine der traditionellsten in der taktilen Rückmeldung, was seine Ursprünge in verschiedenen Projekten des MIT findet, bei dem Vibration auf die Fingerspitzen und Hände ausgeübt wurde [8] aber auch die Übertragung von Vibration auf den Körper und das Gesicht ist möglich [50]. Diese Übertragung passiert zumeist mittels Handhühen, Wearables oder Exoskeletten [20] die zum Beispiel über Elektrostimulation oder Luftpolster Druck auf die Haut ausüben oder sie dehnen [29].

Vor Kurzem wurde auf der *Haptics Symposium*-Konferenz ein Gerät vorgestellt, welches die Form von verschiedenen Gegenständen simulieren kann. Dazu fährt ein kleiner Roboter vor dem Nutzer herum, der mit 288 Bolzen ausgestattet ist. Je nachdem, welche Form simuliert werden soll, kann er diese Bolzen nach oben hin unabhängig voneinander und auf verschiedene Höhe ausfahren. Diese Möglichkeit ist zwar sehr flexibel in der Darstellung verschiedener Gegenstände, ist aber wegen der Sichtbarkeit des Roboters eher für die VR geeignet [45].

Eine andere Möglichkeit zum Bereitstellen von taktiler Rückmeldung ist die Verwendung von Ultraschallwellen, wobei deren Ton als Vibration wahrgenommen wird. Dieser Ansatz eignet sich besonders für die Verwendung in der MR, da der Nutzer zum einen keine einschränkenden Gerätschaften an sich tragen muss und diese Wellen zum anderen unsichtbar übertragen werden können [16] [17].

Ein weiteres Beispiel für eine Verwendung von taktiler Rückmeldung in der MR findet sich in verschiedenen Zeichensystemen. Der Vorteil hier ist, dass die HIs in ähnlicher Form von echten Zeichengeräten gestaltet werden können, was das Realitätserlebnis fördert. Im ersten Fall kreierten Poupyrev et. al. einen haptischen Bildschirm, auf dem mittels eines zugehörigen Stiftes gezeichnet und Text manipuliert werden kann. Dabei gibt eine taktile Rückmeldung durch Ruckeln im Stift Informationen über die Oberfläche des Papiers. Zum einen wird die Papierstruktur simuliert, zum anderen spürt der Nutzer kleine Impulse oder Klicks, wenn er mit dem Stift über eine Zeilenlinie fährt, einen Knopf drückt, Text auswählt oder einen Regler verschiebt. Zusätzlich wurde verschiedenen Zeichenfiguren eine taktile Rückmeldung zugeordnet, indem der Nutzer durch pulsieren im Stift seine Bewegung spüren kann [36]. Auch beim haptischen Zeichen-System *DAB* wird über den Stift eine taktile Rückmeldung gegeben, um das Malen mit verschiedenen Pinseln und Maleffekten spürbarer werden zu lassen [3]. Bei Sugihara et. al. hat der Nutzer die Möglichkeit mittels eines speziellen Stiftes auf dreidimensionalen und zweidimensionalen virtuellen Objekten zu zeichnen die in der realen Umgebung für ihn sichtbar platziert sind. Die taktile Rückmeldung übermittelt dabei Informationen zur Form und Oberflächentextur des Objekts indem der Stift je nach Zeichenrichtung einen Widerstand auf die umgreifenden Finger des Nutzers ausübt und somit auch Reibung simuliert. Als nächsten Schritt nennen Sugihara et. al. das virtuelle Zeichnen auf realen Objekten [49].

### 3.2 Krafrückmeldung

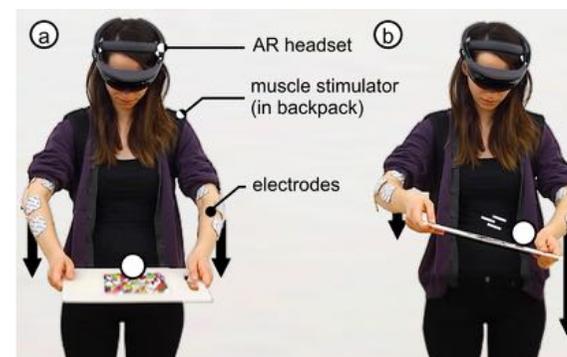
Die Simulation von Kraft und Widerstand ist ein wesentlicher Aspekt im Bereich der haptischen Rückmeldung. Nach Burdeau et. al. gehört zur Krafrückmeldung eines virtuellen Objekts die Simulation von physikalischen Eigenschaften wie Härte, Gewicht und Trägheit [8, 15]. Zusätzlich zeichnet sich die haptische Krafrückmeldung dadurch aus, dass ihr Widerstand aktiv eine Bewegung aufhalten kann [9, 15]. Man stelle sich also beispielsweise einen Nutzer vor, der einen schweren virtuellen Stein vom Boden aufzuheben versucht. Durch dessen Krafrückmeldung spürt er sein Gewicht und nimmt wahr, wie die Hände entgegen der intentionellen Bewegung nach oben gleichzeitig nach unten gezogen werden. Ebenso wird beim Versuch durch eine Wand zu laufen der Körper entgegen dem Versuch der Vorwärtsbewegung aufgehalten oder die

Rückwärtsbewegung am Zug eines virtuellen Tauziehseils verlangsamt. Da die Krafterückmeldung in einfacher Form mechanisch relativ leicht zu bewerkstelligen ist, gehört diese zu den ersten Ansätzen einer haptischen Rückmeldung, der auch schon außerhalb der VR und MR nach unserer Definitionen aus dem Kapitel Virtualität und Realität eingesetzt wurde. Burdea et. al. nennen hier ein Beispiel aus der Telerobotik, bei der Goertz und Thompson in den 1950er Jahren eine Krafterückmeldung in die Fernsteuerung von Roboteroperationen in nuklearen Umgebungen integrierten, der sogenannte *Argonne Remote Manipulator* [8]. Ein auf Goertz' basierendes Projekt namens *GROPE*, bei dem das sogenannte Molekulare Docking<sup>7</sup> simuliert werden kann, brachte die Krafterückmeldung schließlich in den 1960er Jahren in die virtuelle Welt [6]. Später folgen Projekte wie SPIDAR-8, bei dem über Schnüre an den Fingern gezogen wird, während der Nutzer auf einem Bildschirm mit einer Repräsentation seiner Hand eine Art virtuellen Rubik's Cube manipuliert [54] und der bereits erwähnte PHANTOM-Arm [30]. Wie man auch sieht gibt es eine Reihe von HIs, die direkt auf die Hand oder Finger ausgerichtet sind um Kraft auf sie auszuüben. Zwar ist die Verwendung von Krafterückmeldung hauptsächlich in der reinen VR zu finden [29], doch gilt selbiges auch für die MR. Ein Beispiel für eine Krafterückmeldung, die nicht direkt an die Hand gekoppelt ist bildet das Projekt von Knierim et. al., bei dem ein Quadcopter den Benutzer an verschiedenen Stellen des Körpers anstupsst um eine Kollision mit Gegenständen wie Hummeln und Pfeilen zu simulieren [23]. Ein Vorteil dieses Projekts ist, dass das dieses HI sehr flexibel ist und dadurch verschiedene Objekte simulieren kann. Auch Microsoft stellte erst kürzlich *Haptic Links* vor, ein flexibler zweihändiger Controller, der durch Versteifung die Form verschiedener Gegenstände wie eine Waffe oder ein Lenkrad simulieren kann und beim Aufheben eines Gegenstandes so durch Widerstand seine Härte nachahmt [20].

Nun aber zu einem praktischen Beispiel aus der MR. Lopes et. al. arbeiten zum Simulieren von Kraft mit der sogenannten *EMS*, der elektrischen Muskelstimulation. Dabei werden über auf der Haut aufgeklebte Elektroden gezielt kleine Stromimpulse auf die Muskeln eines Menschen abgegeben. Diese Stimulation führt dann zu Kontraktionen der betroffenen Muskeln. Lopes verwendet diese Stimulation in seinen Projekten zumeist auf den Armen des Benutzers [28, 29] aber je nach gewollten Effekt ist auch eine Platzierung zum Beispiel auf den Beinen denkbar [27].

Durch die Kontraktion der Muskeln ist Lopes im Stande Feedback über das

Gewicht und die Trägheit von virtuellen Gegenständen zu geben und Einschränkungen zu simulieren, indem der Nutzer die Kontraktion als ziehende Kraft und/oder Widerstand wahrnimmt. Der Aufbau des Systems ist dabei wie folgt: An den Händen und Armen des Nutzers sind die Elektroden befestigt, während die Steuereinheit dafür in einem Rucksack mitgetragen wird. Zusätzlich trägt der Nutzer ein AR Headset, welches die Umgebung des Nutzers visuell erweitert (siehe Abbildung 2). In verschiedenen Versuchsaufbauten wird mal das Gewicht einer fallenden Murmel simuliert indem die auffangenden Arme mit Hilfe der Elektrostimulation künstlich nach unten gedrückt werden, mal wird beim Verschieben eines virtuellen Sofas ein Zug auf dem Handrücken als Widerstand des Sofas wahrgenommen. In einem weiteren Versuchsaufbau hält der Benutzer eine Tasse in der Hand, die als Dimmschalter für eine Lampe an einem Gemälde benutzt werden kann. Das Headset augmentiert dabei eine Skala zur Dimmung und gibt eine visuelle Informationen über die Begrenzung des Drehschalters. Dreht der Nutzer nun die Tasse bis zur Grenze wird über einen elektrischen Impuls Widerstand im Arm erzeugt, sodass der Nutzer seine Hand mit der Tasse nicht weiter drehen kann. Die Verwendung von EMS in



■ **Abbildung 2** Aufbau einer MR mit EMS nach Lopes unter Verwendung eines AR Headsets, eines Muskelstimulations-Simulators in einem Rucksack und auf bestimmten Muskeln aufgeklebten Elektroden. Die Probandin hält ein reales Tablett in der Hand während das Headset visuell eine Murmel simuliert. Wird das Tablett, also die Unterlage der Murmel, geneigt stellt das EMS-System eine Krafterückmeldung zur daraus folgenden Bewegung der Murmel dar.

Quelle: Lopes et. al. [29], 2018

der MR hält eine Reihe von Vorteilen inne. Zum einen ist dabei ein schnelles und einfaches Prototyping möglich, denn das HI ist schnell auf verschiedene Stellen des Körpers übertragbar [27]. Wie von der haptischen Rückmeldung generell erwünscht, wird bei der Verwendung von EMS das Realitätserlebnis

<sup>7</sup> Unter Molekularem Docking wird ein Verfahren aus der Chemie zur Prognose von Bindungseigenschaften zweier Biomoleküle verstanden.

erheblich erhöht [29]. Zu guter Letzt ist der große Vorteil für die MR, dass diese Herangehensweise in der Regel keine sperrigen Gerätschaften erfordert, die zum einen den Nutzer in seiner Bewegungsfreiheit einschränken würden und zum anderen während des MR-Erlebnisses durch starke Sichtbarkeit die Erfahrung trüben könnten. Fabrizz et. al. haben bereits ein System mit kabelloser Steuerung entwickelt, sodass der Nutzer nur noch die leichten Elektroden an sich trägt [14]. Die Leichtigkeit dieser Anwendung erlaubt sogar ein zusätzliches haptisches Feedback durch die Verwendung eines echten Gegenstandes, wie im Beispiel die Tasse als Repräsentation für den Dimmschalter. Dazu im Folgenden Teil mehr.

### 3.3 Physikalische Gegenstände

Eine Möglichkeit um sowohl Kraft- als auch taktile Informationen über einen Gegenstand zu geben ist die tatsächliche Verwendung physikalisch existierender Gegenstände zur haptischen Repräsentation eines virtuellen Objekts. Das bedeutet im Prinzip, dass ein Benutzer während er einen augmentierten oder virtuellen Gegenstand sieht einen echten Gegenstand berührt, der ähnliche Eigenschaften besitzt. Dabei kann der verwendete Gegenstand zwar „der gleiche“ Gegenstand sein wie später in der virtuellen Umgebung zu sehen ist, muss er aber nicht. Vielmehr wird das Ziel verfolgt, physikalische Gegenstände mit geringer Getreue zu Verfügung zu stellen, die nur relevante Eigenschaften mit dem virtuellen oder augmentierten Objekt teilen [26, 21]. Lindemann et. al. zählen diese Art der Rückmeldung zur *passiven* haptischen Rückmeldung, die also die haptischen Informationen rein durch die dem repräsentierenden Objekt inhärenten Eigenschaften übermittelt, im Gegensatz zum *aktiven* haptischen Feedback, bei dem die entsprechenden Informationen generiert werden [26]. Dieser Ansatz ist nicht komplett getrennt von den vorangehenden Ansätzen zu betrachten, stattdessen geht es hier um das vermitteln von entweder einer oder sogar beider Wahrnehmungsmodalitäten.

Die Verwendung physischer Gegenstände wirkt auf den ersten Blick wie ein naheliegender und einfacher Ansatz, doch birgt er einige zu beachtende Besonderheiten: Zum einen müssen die verwendeten Gegenstände ähnliche Eigenschaften besitzen wie die virtuellen Gegenstände, die sie repräsentieren sollen. Soll beispielsweise die Oberfläche einer virtuellen Betonmauer fühlbar werden, so muss auch das reale Pendant gewisse spürbare Eigenschaften wie eine an Beton erinnernde Rauheit und Kälte besitzen. Soll diese Mauer auch noch mit Moos bewachsen sein, erfordert es ein weiteres Objekt oder einen weiteren Werkstoff zusammen mit dem ersten realen Objekt, welches wieder andere Eigenschaften innehaben muss. Darüber hinaus ist eine solche Mauer

ein nur schwer zu durchdringendes Hindernis. Sollte ein Betrachter also versuchen mit den Händen gegen sie zu drücken, sollte auch der repräsentierende Gegenstand ein ausreichendes Gewicht haben, sodass er es nicht einfach fort schieben kann, oder eine ausreichende Dichte, sodass er nicht einfach hindurch greifen kann.

Zum anderen bedürfte es unter Umständen einem sehr hohen Aufwand, für jedes virtuelle Objekt einen realen Gegenstand zu Verfügung zu stellen. Man stelle sich ein Schwimmbecken voller virtuell hell leuchtenden Glasmurmeln vor, wobei jede Murmel einzeln berührbar sein soll. Nun sind Murmeln nicht das teuerste Produkt in der Anschaffung, doch stünde eine Aufbewahrung so vieler Murmeln in keiner Relation. Genauso wenig wären sie beispielsweise einfach an einen anderen Ort transportierbar, wenn sich die reale Umgebung in ein anderes Schwimmbad verlagern sollte. Um dem entgegen zu wirken gibt es die verbreitete Technik des *User Redirection* (zu Deutsch soviel wie *Benutzer Umleitung*), bei der durch eine Verzerrung der virtuellen visuellen Informationen dem Nutzer ein zur Realität abweichender Bewegungsablauf vorgegaukelt wird um seine Bewegungen und Griffe gezielt zu lenken [11]. Das kann sogar soweit gehen, dass der Nutzer denkt, er laufe einen Korridor entlang, obwohl er sich nur in einem kleinen Bereich im Kreis bewegt [24]. Dies kann dahingehend genutzt werden, dass ein Nutzer denkt, er berühre nacheinander verschiedene Objekte während es sich in Wahrheit immer wieder um das selbe handelt. Der Effekt kann auch durch manuelle Repositionierung des realen Objekts geschehen, sei es durch einen Roboter oder einen Menschen [12]. Als Beispiele sei hier zum einen das Projekt von Azmandian et. al. genannt, bei dem der Nutzer drei virtuelle Würfel sieht, seine Armbewegung durch eine Verzerrung der virtuellen Umgebung aber so umgelenkt wird, dass er immer nach dem selben physikalischen Gegenstück greift ohne es zu bemerken [2]. Zum anderen möchten wir das Projekt *Magic table* von Matsumoto et. al. erwähnen bei dem ein Nutzer immer um den selben vierkantigen Tisch herumläuft, obwohl er in verschiedenen Versuchsaufbauten Tische mit verschiedener Anzahl an Kanten sieht [31].

Die Crux dabei ist, dass diese Projekte in der Regel in der VR realisiert sind, denn es benötigt eine entweder eine programmatische Verzerrung der Umgebung oder eine Umpositionierung der betreffenden physischen Gegenstände. In der MR ist das User Redirecting schwierig umsetzbar ohne das Erlebnis des Nutzers zu beeinträchtigen. In der MR würde der Nutzer sehen, dass es sich um nur ein tatsächliches bzw. immer das selbe Objekt handelt. Außerdem sind die Gerätschaften, die Objekte umplatzieren oft groß und sperrig und somit in der MR für den Benutzer sichtbar.

Ohne ein User Redirecting, wenn also ein physikalisches Objekt nur jeweils ein virtuelles Objekt repräsentieren muss, ist eine Verwendung von physis-

kalischen Gegenständen in der MR allerdings gut nutzbar, wenn etwa der reale Gegenstand für den Nutzer sichtbar bleiben kann, weil er beispielsweise nur augmentiert werden soll oder die Augmentierung sogar den Gegenstand verdecken kann. Oder auch wenn die Sichtbarkeit des realen Gegenstandes der MR-Erfahrung nicht schadet, da er sehr hohe Ähnlichkeit mit dem virtuellen Konterpart hat.

#### 4 Zusammenfassung

Dieser Artikel stellt drei Möglichkeiten vor, fühlbare Eigenschaften wie die Härte, das Gewicht, die Oberflächenstruktur und die Form eines virtuellen Objektes real erlebbar werden zu lassen. Dazu gehören die Krafterückmeldung, die Simulation von physikalischen Kräften und Widerständen, die taktile Rückmeldung, die Simulation von Oberflächeneigenschaften und Formen sowie die Verwendung von physikalischen Objekten mit ähnlichen Attributen ihrer zu repräsentierenden virtuellen Gegenstände. Während für die VR schon viele Schnittstellen existieren, die dem Nutzer haptische Informationen übermitteln, wird die Entwicklung für die MR durch einige Gegebenheiten erschwert wie die Sichtbarkeit der haptischen Schnittstellen oder eine gewünschte Bewegungsfreiheit in der sichtbaren realen Umgebung. Dennoch gibt es bereits eine Reihe an Beispielen für die Verwendung von haptischer Rückmeldung in der Gemischten Realität, bei der sich die HIs annähernd nahtlos in die reale Umgebung einfügen können. Dieser Artikel zeigt dennoch nur einen kleinen Ausschnitt des aktuellen Stands zur haptischen Rückmeldung in der MR auf und die künftige Entwicklung birgt noch viel Potential sodass auf weitere HIs und vielleicht sogar Modalitäten gespannt sein darf.

#### Literatur

- 1 Akshita, Harini Alagarai Sampath, Bipin Indurkha, Eunhwa Lee, and Yudong Bae. Towards multimodal affective feedback: Interaction between visual and haptic modalities. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 2043–2052, New York, NY, USA, 2015. ACM. URL: <http://doi.acm.org/10.1145/2702123.2702288>, doi:10.1145/2702123.2702288.
- 2 Mahdi Azmandian, Mark Hancock, Hrvoje Benko, Eyal Ofek, and Andrew D. Wilson. Haptic retargeting: Dynamic repurposing of passive haptics for enhanced virtual reality experiences. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 1968–1979, New York, NY, USA, 2016. ACM. URL: <http://doi.acm.org/10.1145/2858036.2858226>, doi:10.1145/2858036.2858226.

- 3 Bill Baxter, Vincent Scheib, Ming C. Lin, and Dinesh Manocha. Dab: Interactive haptic painting with 3d virtual brushes. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '01, pages 461–468, New York, NY, USA, 2001. ACM. URL: <http://doi.acm.org/10.1145/383259.383313>, doi:10.1145/383259.383313.
- 4 Gaurav Bhorkar. A Survey of Augmented Reality Navigation. *CoRR*, 4(August):355–385, 2017. URL: <http://arxiv.org/abs/1708.05006>, arXiv:1708.05006, doi:10.1561/1100000049.
- 5 Christoph W. Borst and Richard A. Volz. Evaluation of a haptic mixed reality system for interactions with a virtual control panel. *Presence*, 14:677–696, 2005.
- 6 Frederick P. Brooks, Jr., Ming Ouh-Young, James J. Batter, and P. Jerome Kilpatrick. Project gropohaptic displays for scientific visualization. *SIGGRAPH Comput. Graph.*, 24(4):177–185, September 1990. URL: <http://doi.acm.org/10.1145/97880.97899>, doi:10.1145/97880.97899.
- 7 G.C. Burdea and P. Coiffet. *Virtual Reality Technology*. Academic Search Complete. Wiley, 2003. URL: <https://books.google.de/books?id=0xwgPZbcz4AC>.
- 8 Grigore C. Burdea. Haptic feedback for virtual reality. In *Haptic Feedback for Virtual Reality*, 1999.
- 9 Grigore C. Burdea and Noshir A. Langrana. Virtual force feedback lessons, challenges, future applications. *JRM*, 5:178–182, 1993.
- 10 Cedric Caremel, Gemma Liu, George Chernyshov, and Kai Kunze. Muscle-wire glove: Pressure-based haptic interface. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion*, IUI '18 Companion, pages 52:1–52:2, New York, NY, USA, 2018. ACM. URL: <http://doi.acm.org/10.1145/3180308.3180361>, doi:10.1145/3180308.3180361.
- 11 Cristiano Carvalheiro, Rui Nóbrega, Hugo da Silva, and Rui Rodrigues. User Redirection and Direct Haptics in Virtual Environments. *Proceedings of the 2016 ACM on Multimedia Conference - MM '16*, pages 1146–1155, 2016. URL: <http://dl.acm.org/citation.cfm?doid=2964284.2964293>, doi:10.1145/2964284.2964293.
- 12 Lung-Pan Cheng, Thijs Roumen, Hannes Rantzsch, Sven Köhler, Patrick Schmidt, Robert Kovacs, Johannes Jasper, Jonas Kemper, and Patrick Baudisch. TurkDeck: Physical Virtual Reality Based on People. *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology - UIST '15*, pages 417–426, 2015. URL: <http://dl.acm.org/citation.cfm?id=2807442.2807463>, doi:10.1145/2807442.2807463.
- 13 M. Dessoir. *Ueber den Hautsinn*. Fb&c Limited, 1892. URL: <https://books.google.de/books?id=IzMaAAAAAAAJ>.
- 14 Farzam Farbiz, Zhou Hao Yu, Corey Manders, and Waqas Ahmad. An electrical muscle stimulation haptic feedback for mixed reality tennis game. In *ACM SIGGRAPH 2007 Posters*, SIGGRAPH '07, New York, NY, USA,

2007. ACM. URL: <http://doi.acm.org/10.1145/1280720.1280873>, doi:10.1145/1280720.1280873.
- 15 A. Fischer and J. M. Vance. Phantom haptic device implemented in a projection screen virtual environment. In *Proceedings of the Workshop on Virtual Environments 2003*, EGVE '03, pages 225–229, New York, NY, USA, 2003. ACM. URL: <http://doi.acm.org/10.1145/769953.769979>, doi:10.1145/769953.769979.
  - 16 Euan Freeman, Ross Anderson, Julie Williamson, Graham Wilson, and Stephen A. Brewster. Textured surfaces for ultrasound haptic displays. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, ICMI 2017, pages 491–492, New York, NY, USA, 2017. ACM. URL: <http://doi.acm.org/10.1145/3136755.3143020>, doi:10.1145/3136755.3143020.
  - 17 Euan Freeman, Stephen Brewster, and Vuokko Lantz. Tactile feedback for above-device gesture interfaces: Adding touch to touchless interactions. In *Proceedings of the 16th International Conference on Multimodal Interaction*, ICMI '14, pages 419–426, New York, NY, USA, 2014. ACM. URL: <http://doi.acm.org/10.1145/2663204.2663280>, doi:10.1145/2663204.2663280.
  - 18 M. Grunwald and L. Beyer. *Der Bewegte Sinn: Grundlagen Und Anwendungen Zur Haptischen Wahrnehmung*. Birkhäuser Basel, 2001. URL: [https://books.google.de/books?id=LpDzBQAAQBAJ&printsec=frontcover&hl=de&source=gbs\\_atb#v=onepage&q&f=false](https://books.google.de/books?id=LpDzBQAAQBAJ&printsec=frontcover&hl=de&source=gbs_atb#v=onepage&q&f=false).
  - 19 Martin Grunwald. Der Tastsinn im Griff der Technikwissenschaften? Herausforderungen und Grenzen aktueller Haptikforschung. *Journal of Chemical Information and Modeling*, 53(9):1689–1699, 2009. doi:10.1017/CB09781107415324.004.
  - 20 Christian Holz, Eyal Ofek, Mike Sinclair, Evan Strasnick, and Hrvoje Benko. Haptic links: Bimanual haptics for virtual reality using variable stiffness actuation. In *Haptic Links: Bimanual Haptics for Virtual Reality Using Variable Stiffness Actuation*. ACM, April 2018. URL: <https://www.microsoft.com/en-us/research/publication/haptic-links-bimanual-haptics-for-virtual-reality-using-variable-stiffness-actuation-2>.
  - 21 B.E. Insko and B.E. Insko. Passive haptics significantly enhances virtual environments. *Computer*, page 100, 2001.
  - 22 J. Isdale. Introduction to VR technology. *IEEE Virtual Reality, 2003. Proceedings.*, page 302, 2003. URL: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1191178>, doi:10.1109/VR.2003.1191178.
  - 23 Pascal Knierim, Thomas Kosch, Valentin Schwind, Markus Funk, Francisco Kiss, Stefan Schneegass, and Niels Henze. Tactile Drones - Providing Immersive Tactile Feedback in Virtual Reality through Quadcopters. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '17*, pages 433–436, 2017. URL: <http://dl.acm.org/citation.cfm?doid=3027063.3050426>, doi:10.1145/3027063.3050426.
  - 24 Luv Kohli, Eric Burns, Dorian Miller, and Henry Fuchs. Combining passive haptics with redirected walking. In *Proceedings of the 2005 International Conference on Augmented Tele-existence, ICAT '05*, pages 253–254, New York, NY, USA, 2005. ACM. URL: <http://doi.acm.org/10.1145/1152399.1152451>, doi:10.1145/1152399.1152451.
  - 25 Susan J Lederman and Roberta L Klatzky. Hand movements: A window into haptic object recognition. *Cognitive Psychology*, 19(3):342 – 368, 1987. URL: <http://www.sciencedirect.com/science/article/pii/0010028587900089>, doi:[https://doi.org/10.1016/0010-0285\(87\)90008-9](https://doi.org/10.1016/0010-0285(87)90008-9).
  - 26 Robert W. Lindeman, John L. Sibert, and James K. Hahn. Hand-held windows: Towards effective 2d interaction in immersive virtual environments. In *VR*, 1999.
  - 27 Pedro Lopes, Max Pfeiffer, Michael Rohs, and Patrick Baudisch. Hands-on introduction to interactive electrical muscle stimulation. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '16, pages 944–947, New York, NY, USA, 2016. ACM. URL: <http://doi.acm.org/10.1145/2851581.2856672>, doi:10.1145/2851581.2856672.
  - 28 Pedro Lopes, Sijing You, Lung-Pan Cheng, Sebastian Marwecki, and Patrick Baudisch. Providing Haptics to Walls & Heavy Objects in Virtual Reality by Means of Electrical Muscle Stimulation. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, pages 1471–1482, 2017. URL: <http://dl.acm.org/citation.cfm?doid=3025453.3025600>, doi:10.1145/3025453.3025600.
  - 29 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/10.1145/3173574.3174020>, doi:10.1145/3173574.3174020.
  - 30 Thomas H Massie and J K Salisbury. The PHANTOM Haptic Interface : A Device for Probing Virtual Objects. *ASME Winter Annual Meeting, Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, 55:1–6, 1994. doi:<http://doi.acm.org/10.1145/1029632.1029682>.
  - 31 Keigo Matsumoto, Takeru Hashimoto, Junya Mizutani, Hibiki Yonahara, Ryohei Nagao, Takuji Narumi, Tomohiro Tanikawa, and Michitaka Hirose. Magic table: Deformable props using visuo haptic redirection. In *SIGGRAPH Asia 2017 Emerging Technologies*, SA '17, pages 9:1–9:2, New York, NY, USA, 2017. ACM. URL: <http://doi.acm.org/10.1145/3132818.3132821>, doi:10.1145/3132818.3132821.

- 32 Mark T. Maybury and Wolfgang Wahlster, editors. *Readings in Intelligent User Interfaces*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1998.
- 33 Richard Mayer. *Cognitive theory of multimedia learning*. 01 2005.
- 34 Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. Augmented reality: a class of displays on the reality-virtuality continuum. *Telem manipulator and Telepresence Technologies*, 2351:282–292, 1994. URL: <http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=981543>, arXiv:NIHMS150003, doi:10.1117/12.197321.
- 35 Jacek Polechoński and Dorota Olex-Zarychta. The influence of tactile feedback on hand movement accuracy. *Human Movement*, 13(3):236–241, 2012. doi:10.2478/v10038-012-0027-0.
- 36 Ivan Poupyrev, Makoto Okabe, and Shigeaki Maruyama. Haptic feedback for pen computing: Directions and strategies. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '04, pages 1309–1312, New York, NY, USA, 2004. ACM. URL: <http://doi.acm.org/10.1145/985921.986051>, doi:10.1145/985921.986051.
- 37 Roderick P Power. The dominance of touch by vision: Sometimes incomplete. *Perception*, 9(4):457–466, 1980. PMID: 7422462. URL: <https://doi.org/10.1068/p090457>, arXiv:<https://doi.org/10.1068/p090457>, doi:10.1068/p090457.
- 38 Ralf Pramberger. Taktiles feedback mittels luft zur multimodalen interaktion. *OPUS - Online Publikationen der Universität Stuttgart*, 2014. URL: <http://dx.doi.org/10.18419/opus-3433>, doi:10.18419.
- 39 G. Révész. *Die Formenvelt des Tastsinnes: Erster Band Grundlegung der Haptik und der Blindenpsychologie*. Springer Netherlands, 2013. URL: <https://books.google.de/books?id=L3KEBwAAQBAJ>.
- 40 Irvin Rock and Jack Victor. Vision and touch: An experimentally created conflict between the two senses. *Science*, 143(3606):594–596, 1964. URL: <http://science.sciencemag.org/content/143/3606/594>, arXiv:<http://science.sciencemag.org/content/143/3606/594.full.pdf>, doi:10.1126/science.143.3606.594.
- 41 L. B. Rosenberg. The use of virtual fixtures as perceptual overlays to enhance operator performance in remote environments. *Technical Report AL-TR-0089, Wright-Patterson AFB OH*, 5:178–182, 1992.
- 42 Abdulmotaleb El Saddik, Emil Petriu, Mauricio Orozco, and Juan Silva. The role of haptics in games. In Abdulmotaleb El Saddik, editor, *Haptics Rendering and Applications*, chapter 11. InTech, Rijeka, 2012. URL: <https://doi.org/10.5772/32809>, doi:10.5772/32809.
- 43 Eva-Lotta Sallnäs, Kirsten Rasmus-Gröhn, and Calle Sjöström. Supporting presence in collaborative environments by haptic force feedback. *ACM Trans. Comput.-Hum. Interact.*, 7:461–476, 2000.
- 44 Timothy M. Simon, Ross T. Smith, and Bruce H. Thomas. Wearable jamming mitten for virtual environment haptics. In *Proceedings of the 2014 ACM International Symposium on Wearable Computers, ISWC '14*, pages 67–70, New York, NY, USA, 2014. ACM. URL: <http://doi.acm.org/10.1145/2634317.2634342>, doi:10.1145/2634317.2634342.
- 45 Alexa F. Siu, Eric J. Gonzalez, Shenli Yuan, Jason Ginsberg, Allen Zhao, and Sean Follmer. shapeshift: A mobile tabletop shape display for tangible and haptic interaction. In *Adjunct Publication of the 30th Annual ACM Symposium on User Interface Software and Technology, UIST '17*, pages 77–79, New York, NY, USA, 2017. ACM. URL: <http://doi.acm.org/10.1145/3131785.3131792>, doi:10.1145/3131785.3131792.
- 46 Mel Slater. A note on presence terminology. 3, 01 2003.
- 47 Robert J. Stone. Haptic feedback: a brief history from telepresence to virtual reality. In Stephen Brewster and Roderick Murray-Smith, editors, *Haptic Human-Computer Interaction*, pages 1–16, Berlin, Heidelberg, 2001. Springer Berlin Heidelberg.
- 48 D. J. Sturman, D. Zeltzer, and S. Pieper. Hands-on interaction with virtual environments. In *Proceedings of the 2Nd Annual ACM SIGGRAPH Symposium on User Interface Software and Technology, UIST '89*, pages 19–24, New York, NY, USA, 1989. ACM. URL: <http://doi.acm.org/10.1145/73660.73663>, doi:10.1145/73660.73663.
- 49 Kenji Sugihara, Mai Otsuki, Asako Kimura, Fumihisa Shibata, and Hideyuki Tamura. Mai painting brush++: Augmenting the feeling of painting with new visual and tactile feedback mechanisms. In *Proceedings of the 24th Annual ACM Symposium Adjunct on User Interface Software and Technology, UIST '11 Adjunct*, pages 13–14, New York, NY, USA, 2011. ACM. URL: <http://doi.acm.org/10.1145/2046396.2046404>, doi:10.1145/2046396.2046404.
- 50 Yasuhiro Suzuki, Rieko Suzuki, Junji Watanabe, Ayano Yoshida, and Sakurazawa Shigeru. Haptic vibrations for hands and bodies. In *SIGGRAPH Asia 2015 Haptic Media And Contents Design*, SA '15, pages 9:1–9:3, New York, NY, USA, 2015. ACM. URL: <http://doi.acm.org/10.1145/2818384.2818389>, doi:10.1145/2818384.2818389.
- 51 John Sweller. *Evolution of human cognitive architecture*, volume 43. 12 2003.
- 52 Cecilie Våpenstad, Erlend Fagertun Hofstad, Thomas Langø, Ronald Mårvik, and Magdalena K. Chmarra. Perceiving haptic feedback in virtual reality simulators. *Surgical Endoscopy*, 27:2391–2397, 2012.
- 53 J. Vince. *Introduction to Virtual Reality*. Springer London, 2004. URL: <https://books.google.de/books?id=hfJZyo1LOpGc>.
- 54 Somsak WALAIRACHT, Masahiro ISHII, Yasuharu KOIKE, and Makoto SATO. Two-handed multi-fingers string-based haptic interface device. *IEE-*

*CE Trans. On Information and Systems, D*, 84(3):365–373, mar 2001. URL: <https://ci.nii.ac.jp/naid/110003210484/en/>.

- 55 Lisa Wastiels, Hendrik N.J. Schifferstein, Ine Wouters, and Ann Heylighen. Touching materials visually: About the dominance of vision in building material assessment. *International Journal of Design*, 7(2):31–41, 2013. URL: <https://lirias.kuleuven.be/bitstream/123456789/402721/2/Wastiels2013IJD.pdf>.

# Cockpit Design für autonome Fahrzeuge - bisherige Vorgehensweisen und ihre Veränderungen

Daniel F. Steidl

Ludwig-Maximilians-Universität, Hans-Scholl-Platz 1, 89899  
München, Deutschland [d.steidl@campus.lmu.de](mailto:d.steidl@campus.lmu.de)

---

## Zusammenfassung

Automobile Technologien haben die Gestaltung im Innenraum des Autos schon immer beeinflusst. Derzeitige Entwicklungen und Forschungen autonomer Fahrzeuge versprechen einen weiteren großen Umbruch im Cockpit Design. Während die Interaktionsmöglichkeiten steigen, übernehmen Fahrassistenzsysteme primäre Fahraufgaben. Der Fokus auf Sicherheit verschiebt sich mehr und mehr auf die Unterhaltung des Passagiers während der Fahrt. Das stellt die derzeitigen Richtlinien des Cockpit Designs in Frage. In dieser Arbeit wird ein Design Space im automobilen Kontext vorgestellt und nach Einführung aktueller Technologien diskutiert, ob sich sowohl das Design im Auto verändert als auch der Design Space noch sinnvoll anwendbar ist. Das Resultat zeigt, dass der Fokus beim Gestalten eines Cockpits sich langsam aber sicher verschieben wird. Die Benutzung des Design Spaces kann bei teilautonomen Fahrzeugen noch erfolgen und wird obsolet, wenn das Fahrzeug vollständig autonom fahren kann.

**2012 ACM Computing Classification** H.1.2 [User/Machine Systems]: Human Factors, Human Information Processing.

**Keywords and phrases** HMI, Autonomous Driving, Cockpit, Design Space

## 1 Autonomes Fahren und dessen Auswirkungen auf das Cockpit Design

Die Mobilität ist im Begriff, sich stark zu verändern. An Abbildung 1 sieht man sehr gut, wie die Etablierung von Technologien das Aussehen eines Fahrzeuges in der Vergangenheit geprägt haben. Stand im 20. Jahrhundert eine sichere Fahrt ohne Ablenkung im Vordergrund (siehe Cockpit des Chevrolet links), verschiebt sich der Fokus des Cockpit Designs mehr und mehr in Richtung Unterhaltung und der digitalen Darstellung von Informationen [19] [16] [30]. Neueste industrielle Umsetzungen von Fahrassistenzsystemen haben bereits



© Daniel F. Steidl;

licensed under Creative Commons License CC-BY

Cite as: Daniel F. Steidl. Cockpit Design für autonome Fahrzeuge - bisherige Vorgehensweisen und ihre Veränderungen. In *3rd Seminar on Ubiquitous Interaction (UBIACTION 2019-1)*. Editors: Matthias Hoppe, Jakob Karolus, Thomas Kosch, Pascal Knierim, Albrecht Schmidt. January 31, 2019. Munich, Germany. pp. 16:1–16:24.



■ **Abbildung 1** Darstellung verschiedener Cockpits im Auto. Links der Innenraum eines alten Chevrolet, in der Mitte der Innenraum eines aktuellen Mercedes Benz und rechts der Innenraum eines Tesla Fahrzeugs mit Autopilot. Die Bilder wurden auf <https://pixabay.com/> und <https://unsplash.com/> heruntergeladen und bearbeitet.

jetzt die Nutzung des Autos beeinflusst (siehe Abbildung 1, Bild in der Mitte). Forschung an autonomen Fahrzeugen werden ein weitaus grundlegenden Wandel bei der Bedienung bewirken (Abbildung 1, Bild rechts). Meschtscherjakov et. al. sprechen sogar von bedeutenden Veränderungen in der Kommunikation zwischen Mensch und Fahrzeug [19]. Die primäre Aufgabe eines Autofahrers wird aufgrund automatisierender Technologien zur Nebensache.

Dies wirft die Frage auf, ob der Fokus beim Gestalten eines automobilen Cockpits und bisherige Design Richtlinien sich verändern müssen. Der Schwerpunkt dieser Arbeit liegt daher hauptsächlich auf die Einführung und Diskussion einer automobilen Design Norm und den Einfluss aktueller Technologien darauf. Wann von einem autonomen Fahren die Rede ist, soll im Folgenden geklärt werden. Zusätzlich werden bisher definierte Fahraufgaben sowie Eingabe- und Ausgabegeräte im Fahrzeug erläutert. Ein Interview mit Herrn Dr. Alexander Meschtscherjakov von der Universität Salzburg unterstützen die Aussagen der folgenden Kapitel (siehe Abbildung 9 und 10 im Anhang).

### 1.1 Normen und Level des autonomen Fahrens

Es existieren derzeit verschiedene Normierungen zur Definition des autonomen Fahrens. Diese werden von Instituten wie die der *Society of Automotive Engineers International* (SAE) festgelegt [12]. Die SAE agiert global und ihre Norm ist allgemein anerkannt. Das Ministerium für Verkehr in den USA (NHTSA) hat beispielsweise die Level der SAE vollständig übernommen [22]. In Deutschland ist die *Bundesagentur für Straßenwesen* (BASt) speziell für die Normierung des deutschen Verkehrssystems zuständig [2]. Da die Normierungen der SAE global anerkannt und weit verbreitet sind, werden diese nun vorgestellt und in dieser Arbeit genutzt.

Die Institute normieren die Automatisierung des Autos in aufsteigende

Level, wobei ein Fahrzeug bei Level 0 keine Automatisierung besitzt und das höchste Level ein voll autonomes Fahrzeug beschreibt (siehe Abbildung 2). Das SAE Institut unterteilt die Level noch einmal in zwei Bereiche [12]. Zum einen die Level 0-2, in denen der Mensch überwiegend selbst fährt, sprich seine Umgebung während der Fahrt beobachten muss und bewusst Funktionalitäten des Systems aktiviert. Zum anderen die Level 3-5, in denen das Fahrassistenzsystem die Umgebung überwacht und das Steuern des Fahrzeuges übernimmt. Die genaue Höhe der Level wird anhand spezifischer Kriterien festgelegt: Die Steuerung und Lenkung des Fahrzeuges, die Überwachung der Umgebung während der Fahrt, die manuelle Steuerung des Menschen im Notfall und den Anteil autonomer Systeme in Relation zu den Fahraufgaben des Fahrers (siehe hierzu Kapitel 1.2).

SAE level	Name	Narrative Definition	Execution of Steering and Acceleration/Deceleration	Monitoring of Driving Environment	Fallback Performance of Dynamic Driving Task	System Capability (Driving Modes)
<b>Human driver monitors the driving environment</b>						
0	No Automation	The full-time performance by the human driver of all aspects of the dynamic driving task, even when enhanced by warning or intervention systems	Human driver	Human driver	Human driver	n/a
1	Driver Assistance	The driving mode-specific execution by a driver assistance system of either steering or acceleration/deceleration using information about the driving environment and with the expectation that the human driver perform all remaining aspects of the dynamic driving task	Human driver and system	Human driver	Human driver	Some driving modes
2	Partial Automation	The driving mode-specific execution by one or more driver assistance systems of both steering and acceleration/deceleration using information about the driving environment and with the expectation that the human driver perform all remaining aspects of the dynamic driving task	System	Human driver	Human driver	Some driving modes
<b>Automated driving system ("system") monitors the driving environment</b>						
3	Conditional Automation	The driving mode-specific performance by an automated driving system of all aspects of the dynamic driving task with the expectation that the human driver will respond appropriately to a request to intervene	System	System	Human driver	Some driving modes
4	High Automation	The driving mode-specific performance by an automated driving system of all aspects of the dynamic driving task, even if a human driver does not respond appropriately to a request to intervene	System	System	System	Some driving modes
5	Full Automation	The full-time performance by an automated driving system of all aspects of the dynamic driving task under all roadway and environmental conditions that can be managed by a human driver	System	System	System	All driving modes

■ **Abbildung 2** Level des autonomen Fahrens der SAE International aus [11]. Die dargestellte Tabelle zeigt die Kriterien zur Abstufung der Level. Unterschieden wird in den Level von 0-5. In den Level 0-2 handelt der Fahrer, in den Level 3-5 das System. Neben der Beschreibung der Level sieht man die Kriterien, wann ein Fahrzeug in welchem Level liegt: Steuerung und Lenkung des Fahrzeuges, Überwachung der Umgebung während der Fahrt, manuelle Steuerung des Menschen im Notfall und die Systemfähigkeit.

Level 0 und 1 erfüllen keine dieser Kriterien. Während Level 0 vollständig und ohne Hilfe eines Systems vom Menschen gesteuert wird, erhält der Fahrer in Level 1 bereits Unterstützung von einem Fahrassistenzsystem, welches die

Steuerung des Fahrzeuges unterstützt. Beispielsweise hält eine Geschwindigkeitsregelanlage die aktuelle Geschwindigkeit ohne Betätigen des Gaspedals oder das System lenkt ein, wenn der Fahrer von der Spur abkommt.

Level 2 ist bereits teilweise automatisiert. Fahrassistenzsysteme können minimal die Umgebung interpretieren und in speziellen Situation das Auto steuern. Das System übernimmt beispielsweise das Einparken. In einem Stau oder in der Stadt ist es in der Lage, bis zu einer bestimmten Geschwindigkeit das Auto zu steuern. In einer gefährlichen Situation bremst das Auto automatisch ab. Wie bereits erwähnt, hat der Fahrer dabei die volle Kontrolle über das System und muss die Umgebung im Auge behalten.

Mit Level 3 gibt der Nutzer das erste Mal die Kontrolle an ein Fahrassistenzsystem ab, welches durchgehend die Umgebung beobachtet und darauf reagieren kann. Dies ermöglicht beispielsweise das Überholen langsamerer Fahrzeuge ohne das Eingreifen des Fahrers. Jedoch ist das System nicht auf jede Situation vorbereitet. Ist eine Entscheidung nicht möglich, übergibt es die Steuerung an den Fahrer. Dies kann vor allem in extremen oder gefährlichen Situationen wie schlechtem Wetter oder einem Unfall passieren.

Das Fahrassistenzsystem kann ab Level 4 auch in außergewöhnlichen Situationen reagieren. Es ist theoretisch eine komplette Fahrt ohne Eingreifen des Fahrers möglich. Manuellen Steuerung ist trotzdem noch gegeben. Die hohe Automatisierung kann beispielsweise nur Gebietsweise verfügbar sein und zwingt den Benutzer außerhalb dieser Region das Auto selbst zu steuern.

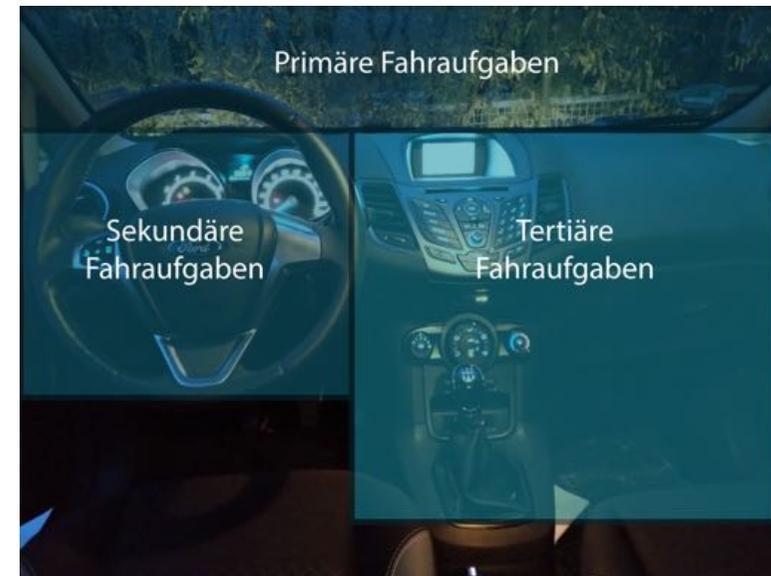
Ein voll autonomes Fahrzeug wird mit Level 5 definiert. Alle Fahraufgaben werden von einem Fahrassistenzsystem übernommen und der Fahrer muss nicht mehr eingreifen. In diesen Fahrzeugen werden kein Lenkrad, Gaspedal oder Bremse mehr verbaut.

Die am meisten gefahrenen Autos befinden sich derzeit in den Level SAE 0 und 1. In Level SAE 2 fahren bereits erste höherpreisige Modelle bekannter Automarken wie Audi [4]. Für Level SAE 3 gibt es erste Bestrebungen. Ein für den Endverbraucher verfügbares Fahrzeug ist jedoch noch nicht auf dem Markt. Dies gilt auch für die Autos der Firma Google und dem Tochterunternehmen Waymo [34]. Diese Fahrzeuge sind in Level SAE 5 einzuordnen und eher noch Prototypen der Forschung.

Gleichwohl haben bereits einige Hersteller teilweise autonome Fahrzeuge zum Verkauf angekündigt [4]. Die Möglichkeit besteht, dass der Fahrer damit langfristig das Fahren an ein Fahrassistenzsystem überträgt. Damit ändern sich auch die Nutzung des Fahrzeuges während der Fahrt. Die primären Aufgaben des Fahrers verschieben sich und damit auch die Nutzung des Innenraums.

## 1.2 Fahraufgaben in einem Fahrzeug

Fahraufgaben des Benutzers (engl. *driving tasks*) werden bisher in drei Bereiche aufgeteilt: die primären, die sekundären und die tertiären Fahraufgaben [26][32][33].



■ **Abbildung 3** Primäre, sekundäre und tertiäre Fahraufgaben aus [33]. Primäre Aufgaben stellen im Wesentlichen das Fahren des Autos dar. Sekundäre Aufgaben bündeln die Interaktion mit der Umgebung. Tertiäre Aufgaben fassen die Informationsbeschaffung und Unterhaltung im Auto zusammen.

Mit den primären Aufgaben ist sowohl Gas geben, Bremsen als auch das Steuern des Fahrzeuges gemeint. Die Teilnahme am Verkehr ist ebenso darunter zu verstehen wie die Vermeidung von Unfällen. Etwas genauer betrachtet lassen sich die primären Fahraufgaben in die Ebenen *Stabilisierung*, *Führung* und *Navigation* untergliedern [32]. Unter *Stabilisierung* bündeln sich die Längs- und Querverführung sowie die nötigen Funktionen, um einen passenden Abstand zu einem anderen Fahrzeug oder Objekt zu halten (Bremsen, Gas geben, Lenken). Auch das Halten der Fahrspur zählt zur *Stabilisierung*. Die Ebene der *Führung* beinhaltet die Wahl des Fahrweges oder der Geschwindigkeit. Unter der Ebene der *Navigation* gliedern sich die "[...] Wahl des Weges vom Start zum Ziel abhängig von örtlichen, zeitlichen oder auch meteorologischen Bedingungen [...]" [26][32].

Die sekundären Fahraufgaben umfassen alle weiteren Interaktionen mit Verkehr und Umwelt, die nicht mit den Aufgaben der Stabilisierung einhergehen. Darunter fallen die Funktionalitäten, die in der Regel um - oder auf dem Lenkrad angebracht sind wie beispielsweise Hupen, Blinken oder die Nutzung des Abblendlichts.

Die Fahraufgaben, die überwiegend der Unterhaltung, Komfort und der Informationsbeschaffung dienen, sind den tertiären Aufgaben untergeordnet. Sie beinhalten die Interaktion mit dem Fahrassistenzsystem, dem Radio oder dem Navigationsgerät. Die Nutzung einer Sitzheizung fällt ebenso in diese Kategorie wie das Bedienen einer Klimaanlage.

### 1.3 Eingabe- und Ausgabegeräte

Im vorherigen Abschnitt 1.2 sind die Stufen der Fahraufgaben eingeführt worden. Ähnlich teilen die Autoren Kern & Schmidt die verschiedenen Eingabegeräte in einem Fahrzeug ein [14]. In Relation zur Sicherheit während der Fahrt werden sie in primäre, sekundäre und tertiäre Geräte eingeteilt [33] [32].

Primäre Eingabegeräte ermöglichen laut Kern & Schmidt die Steuerung des Fahrzeugs und werden für eine spezielle Aufgabe benutzt. Die Platzierung dieser Geräte erfolgt um den Fahrer herum, um sie leichter zu erreichen. Als primär sind sowohl das Lenkrad als auch die Pedale im Fußraum einzuordnen. Sekundäre Eingabegeräte sind meist um das Lenkrad herum platziert. Sie übernehmen die Aufgabe der Interaktion mit der Umgebung wie Hupe, Scheibenwischer, Abblendlicht oder Blinker. Tertiäre Eingabegeräte ermöglichen die Interaktion mit Informations- oder Entertainmentssystemen. Die Platzierung erfolgt normalerweise auf der Mittelkonsole.

Je wichtiger die Aufgabe, desto wichtiger ist dabei die Rückmeldung der Ausgabegeräte. Die haptische Rückmeldung beim Steuern oder dem Betätigen des Gaspedals sollten dann präzise, schnell und so verständlich wie möglich sein. Dagegen ist die Rückmeldung des Radios weniger wichtig. Ausgabegeräte sind weiter in visuelle, auditive und haptische Rückmeldung aufgeteilt, die auch in Kombination auftreten können.

### 1.4 Bisherige Interaktionsmöglichkeiten in einem Fahrzeug

Zur Interaktion haben Kern & Schmidt Eingabe- und Ausgabemöglichkeiten erarbeitet und vorgestellt [14].

Für die Eingabe wurden acht verschiedene Typen definiert. Meist sind dies analoge Knöpfe. Nicht mehr im Einsatz sind dagegen Mechanische Knöpfe mit haptischen Rückmeldung. Schieberegler bestimmen die Richtung, in der die Klimaanlage an den Lüftungsausgängen die Luft bläst. Drehknöpfe können weiter zur Regelung bestimmter Definitionsbereiche genutzt werden. Während



■ **Abbildung 4** Kern & Schmidt haben 8 verschiedene Eingabemöglichkeiten definiert: Knöpfe (a), mechanische Knöpfe (b), kontinuierliche und diskrete Drehknöpfe (c und d), Lenksäulenhebel (e), Multifunktionsregler (f), Schieberegler (g), Touch Displays (h), Pedale (i) und Daumenräder [14].

kontinuierliche Drehknöpfe beispielsweise zur Einstellung der Lautstärke des Radios zuständig sind, wendet man diskrete Drehknöpfe zur Anpassung der Temperatur an. Lenksäulenhebel sind oft am Lenkrad angebracht und beispielsweise zuständig für die Steuerung der Scheibenwischer oder des Blinker. Das Daumenrad wurde oft an Multifunktionslenkrädern angebracht, um die Lautstärke zu regeln. Als einer der wichtigsten Eingabemöglichkeiten zählen immer noch die Pedale im Fußraum. In vielen Fahrzeugen sind weiter Multifunktionsregler zu finden, welche gedreht und manchmal sogar wie ein Steuerknüppel benutzt werden können.

Ausgabemöglichkeiten im Fahrzeug sind durch menschlichen Sinne beschränkt: Sehen, Hören und Fühlen. Visuell ist das beispielsweise die Anzeige der Geschwindigkeit. Ein weiteres Beispiel ist die Darstellung eines Zustandes. Das kann in Form der Beleuchtung eines Knopfs oder Symbols geschehen. Digitale Displays tragen zur Unterhaltung oder der Informationsbeschaffung bei. Das kann das aktuelle Musikstück eines Radiosenders sein, aber auch die kartographische Beschreibung einer Strecke. Für die auditive Rückmeldung sind Lautsprecher verbaut und wurden bisher zum größten Teil benutzt, um Musik abzuspielen. Vermehrtes Auftreten von Spracheingabesystemen (vgl. hierzu Kapitel 2.2) sorgen dafür, dass die auditive Ausgabe die Steuerung anderer Geräte unterstützt. Haptische Ausgabemöglichkeiten können in Form von Vibrationen am Lenkrad geschehen. Dies soll als warnende Reaktion fun-

gieren, falls der Fahrer von der Fahrspur kommen sollte. Die Autoren erwähnen zusätzlich die Ausgabemöglichkeit über das Riechen. Hierfür wurde jedoch bisher kein Gerät etabliert.



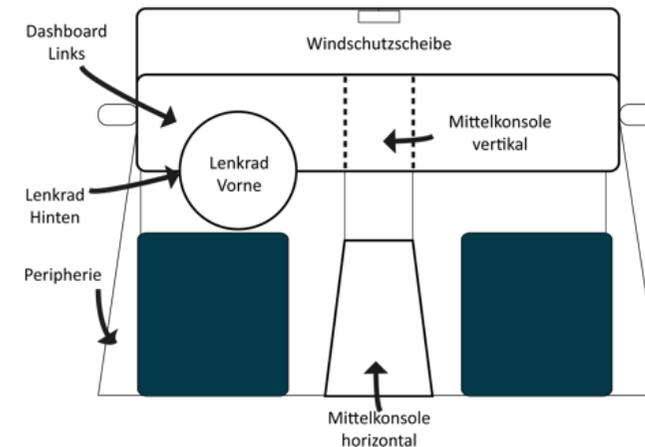
■ **Abbildung 5** Ausgabemöglichkeiten aus [14]: Analoge Geschwindigkeitsanzeige (a), digitale Geschwindigkeitsanzeige (b und c), Anzeigelampe (d), geformte Kontrollleuchte (e), multifunktionaler Display (f) und digitaler Display (g).

Heutige Technologien mit unterstützender Funktion beim Fahren sind in Bezug der eben aufgeführten Geräte und Interaktionsmöglichkeiten weitaus schwerer einzuordnen (siehe hierzu Kapitel 2.2). Sie dienen nicht nur allein der Sicherheit, der Unterhaltung oder dem Komfort. Einige dieser Technologien greifen bereits jetzt in allen Fahraufgaben gleichzeitig ein. Dabei hat besonders die Übernahme primärer Fahraufgaben eine Verschiebung der Nutzung des Autos während der Fahrt zur Folge. Es beeinflusst die Art und Weise, wie ein Cockpit im Auto gestaltet und aufgebaut ist. Sind dadurch bisherige Design Normen und Vorgaben überhaupt noch sinnvoll einzusetzen? Muss das Auto in einem neuen Kontext gestaltet werden? In den folgenden Kapiteln sollen diese Fragen diskutiert werden. Es wird sowohl die bisherige *Design Space* Richtlinie im automobilen Bereich erklärt als auch Technologien der Level SAE 1 bis 3 vorgestellt, die seitdem immer mehr in Benutzung sind. Letztendlich wird anhand der erarbeiteten Informationen über die Gestaltung im Auto und den Sinn des bisherigen Design Space diskutiert.

## 2 Design Space im Auto und heutige Technologien

In verschiedenen Bereichen der Mensch-Computer-Interaktion (HCI) werden *Design Spaces* definiert, um so das Gestalten darin zu vereinfachen und zu beschleunigen. Kern & Schmidt haben 2009 einen ersten *Design Space* für den automobilen Bereich erstellt [14]. Die Definition des *Design Space* und die daraus resultierenden Grafiken sind im folgenden Kapitel beschrieben. Im Anschluss werden verschiedene Technologien vorgestellt, die in den letzten Jahren immer häufiger in Fahrzeugen zu finden sind.

### 2.1 Automobil Design Space der letzten Jahre

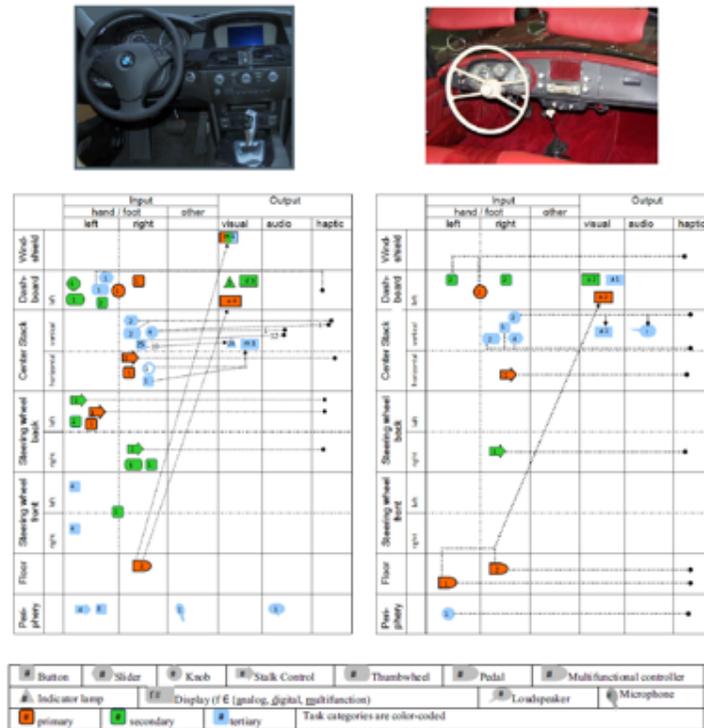


■ **Abbildung 6** Mögliche Interaktionsumgebung zwischen Menschen und Fahrzeug aus [14]. Diese bestehen aus Windschutzscheibe, Instrumententafel, Mittelkonsole und Peripherie.

Die Motivation für einen *Design Space* im automobilen Bereich hatten Kern & Schmidt aufgrund der steigenden Anzahl an Interaktionsmöglichkeiten im Auto [14]. Das steigere die Gefahr, den Benutzer eines Autos während der Fahrt zu sehr abzulenken. Die Autoren wollten durch einen definierten *Design Space* einen besseren Überblick der Platzierung der Geräte im Cockpit erzielen.

Ein *Design Space* für automobiler *User Interfaces* unterscheidet sich grundlegend von anderen. Der Fahrer sei in seiner Mobilität während der Fahrt eingeschränkt, wodurch die Wahl des Ortes eines Gerätes zu berücksichtigen

ist. Es erfordert eine fixe Platzierung der Geräte in der Nähe des Fahrers. Die Interaktion erfolgt dabei meist mit Hilfe der Extremitäten. Im Auto seien weiter sichere, nicht ablenkende Geräte notwendig. Bei einer primären Aufgabe sollten schnelle und klar zu verstehende Rückmeldungen erfolgen.



■ **Abbildung 7** Design Space eines spezifischen Auto Typs aus [14]. Links für einen BMW 3er aus dem Jahr 2007, rechts aus dem Jahr 1957. Er fokussiert sich auf die Platzierung der Eingabe- und Ausgabegeräte am Cockpit und deren Interaktion zwischen Mensch, Fahrzeug und den Geräten. Durch die Einfärbung erkennt man die Einteilung in Fahraufgaben, durch die Linien die Verbindung zwischen Eingabe- und Ausgabegeräte. Die Formen stellen unterschiedliche Geräte dar, die Zahlen entsprechen der Anzahl des Gerätes an dieser Stelle.

Die in Kapitel 1.3 beschriebenen Eingabe- und Ausgabegeräte können anhand einer definierten Interaktionsumgebung eingeteilt werden: Windschutzscheibe, Armaturenbrett, Mittelkonsole, Lenkrad, Fußraum und die Peripherie (vgl. Abbildung 6). Anhand von 706 Fotografien unterschiedlicher Fahrzeu-

ge und Marken wurden zwei Kategorien von *Design Spaces* erstellt. Einer illustriert die Struktur eines einzelnen Modells. Die Autoren vergleichen hier den BMW 5er von 1956 und 2007 (siehe Abbildung 7). Der andere zeigt eine Auswahl an Modellen eines Herstellers (siehe Abbildung 8).

Beim zuerst genannten *Design Space* liegt der Fokus auf die platzierten Geräte und die Interaktion mit einem bestimmten Körperteil (Abbildung 7). In der ersten Spalte sind die möglichen Platzierungen zu erkennen, während in der ersten Zeile zwischen Eingabe- und Ausgabegeräte unterschieden wird. Die Eingabegeräte sind wiederum in die möglichen Körperteile eingeteilt, mit der die Interaktion geschieht. Unter *other* kann beispielsweise die in der Peripherie liegende Spracheingabe liegen. Wie bereits erwähnt unterteilen Kern & Schmidt die Ausgabegeräte in einen visuellen, auditiven und haptischen Bereich (vgl. hierzu Kapitel 1.3). Darunter können passend zur Platzierung Geräte eingefügt werden. Verschiedene Formen stellen jeweils einen Gerätetypen dar und sind dadurch beliebig erweiterbar. Die Farbgebung der Formen illustriert die Stufe der Fahraufgabe, die Zahlen die Anzahl des Gerätes an dieser Stelle. Anhand der gestrichelten Linien erkennt der Betrachter die Verbindungen zwischen Eingabe- und Ausgabegerät. Pfeile bedeuten eine direkte Verbindung, Punkte eine indirekte.

Der zweite *Design Space* bündelt die Struktur vieler Auto Modelle eines Herstellers (vgl. Abbildung 8). Hier wird nicht in Subkategorien unterschieden. Der Fokus liegt auf die generelle Platzierung und dessen Variationen. Anhand der Abbildung sieht man beispielsweise, dass BMW typübergreifend eine einheitliche Platzierung durchführt, während Renault eine größere Variation an Anordnungen besitzt. Die Zahlen in Klammern zeigen primäre, sekundäre und tertiäre Eingabegeräte und deren entsprechend mögliche Anzahl.

Der *Design Space* von Kern & Schmidt beschreibt die Platzierung von Eingabe- sowie Ausgabegeräte sowie deren Interaktionsmöglichkeiten im Auto. Dies kann die Darstellung eines bestimmten Fahrzeugs sein, aber auch die Bündelung verschiedener Fahrzeuge einer Automarke. Wichtig ist hierbei das Ablesen der möglichen Anzahl der Geräte an einer bestimmten Stelle. Bisher konnte hierdurch jede Technologie platziert werden. Jedoch ist dies aufgrund von Technologien mit unterstützenden Funktionalitäten aller Fahraufgaben immer schwieriger darzustellen und könnte den Design Space entscheidend verändern.

## 2.2 Heutige unterstützende Technologien beim Fahren und ihre Wirkung

Seit Einführung der Design Spaces im automobilen Kontext haben sich neue Technologien im Fahrzeug etabliert. Im Folgenden wird eine Auswahl vorgestellt

	Eingabe		Ausgabe
	Rechts	Links	
Windschutzscheibe			(0,1,1)
Instrumententafel	(0,4,2)	(1,0,0) (1,0,0)	(4,2-5,0)
Mittelkonsole		(3,0,35-50)	(0,0,17-20)
Lenkrad (Rückseite)	(0-2,3-5,0)	(0,3,0)	
Lenkrad (Vorderseite)	(0,0,4)	(0,1,0) (0,0,4)	
Fußraum	(0-1,0,0)	(2,0,0)	
Peripherie	(0,0,11)		(0,0,3-4)
	Eingabe		Ausgabe
	Rechts	Links	
Windschutzscheibe			(0,0,2)
Instrumententafel	(0,2-4,2-11)	(1,0,0) (0-2,0-2,0)	(0-4,0-3,0)
Mittelkonsole		(2,0,6-41)	(0-4,0-3,1-18)
Lenkrad (Rückseite)	(0,3,0-4)	(0,5,4)	
Lenkrad (Vorderseite)	(0-2,0,0)	(0,1,0) (0,0,0-2)	
Fußraum	(0-1,0-1,0)	(2,0,0)	
Peripherie	(0,0,8)		(0,0,3-5)

■ **Abbildung 8** Design Space einer Auswahl an Auto Modellen eines Herstellers aus [14]. Die obere Tabelle zeigt den allgemeinen Design Space von BMW, der Untere den allgemeine Design Space von Renault. Es soll den generellen Aufbau der Fahrzeuge eines Autoherstellers im Cockpit Design darstellen. Durch die Nummernbereiche lässt sich ablesen, wie viele Geräte an welcher Interaktionsumgebung zu finden sind.

und ihre Wirkung auf die primäre Aufgabe des Fahrers diskutiert.

#### ■ Head Up Display

Ein Head Up Display ist eine Projektion von Informationen auf der Windschutzscheibe des Fahrzeugs und ein Ausgabegerät. Die Projektion kann von oben nach unten, oder von unten nach oben geschehen. Meist ist die Abbildung auf der Seite des Fahrers und nur dort zu sehen [6]. Es existieren allerdings Studien und Projekte, die die gesamte Windschutzscheibe nutzen, um den Betrachter bei der Fahrt mit wichtigen Informationen zu versorgen [15] [9] [35] [6]. Durch ein Head Up Display ist es dem Fahrer möglich, den Blick nicht von der Straße wenden zu müssen und die Hände am Steuer lassen zu können [35]. Warnzeichen, Geschwindigkeit, die Wegbeschreibung des Navigationsgerätes oder Musik können so peripher vom Fahrer wahrgenommen werden. Das verringert die Wahrscheinlichkeit, durch Ablenkung

einen Unfall zu verursachen. Das Head Up Display kann anhand dieser Beschreibung als Unterstützung primärer Fahraufgaben bezeichnet werden und nimmt dadurch Einfluss auf die Art und Weise, das Fahrzeug zu steuern (vgl. Kapitel 1.2).

#### ■ Touch Display

Große Touch Displays werden in den letzten Jahren immer häufiger in Fahrzeugen verbaut. Dabei nehmen sie beinahe die gesamte Mittelkonsole ein. Sie vereinen einige grundlegende Möglichkeiten der Eingabe im Auto. Die Größe ermöglicht eine übersichtliche Darstellung dieser Bedienelemente. Neue Unternehmen wie Tesla verbauen sie als Grundausstattung und auch günstige Fahrzeuge wie Ford oder Opel haben sie bereits in der Standardausstattung verzeichnet [31] [7] [23]. Ein Touch Display ersetzt die immer größer werdende Anzahl analoger Eingabegeräte. Es dient außerdem der Übersicht tertiärer Fahraufgaben und bietet gleichzeitig eine Kombination aus Eingabe- und Ausgabemöglichkeit an [14].

#### ■ Spracherkennung

Die Spracherkennung ist eine weitere Möglichkeit, während der Fahrt mit den Eingabegeräten des Fahrzeugs zu interagieren. [10] Es soll den Fahrer unterstützen, Navigation, Telefonie oder Musik zu steuern, ohne bei der Fahrt gestört zu werden. Die Nutzung der Spracheingabe könne dabei sowohl über ein Bluetooth fähiges Smartphone als auch ein eingebautes Spracheingabe System erfolgen. Nichtsdestotrotz hat die Technologie im Innenraum des Autos mit einigen Herausforderungen wie Störgeräusche zu kämpfen. Das System interpretiert dadurch einen Befehl des Nutzers tendenziell falsch [21]. Zusätzlich haben Spracheingabesysteme meist explizite Befehle, die der Fahrer wissen muss [25]. Es gibt einige Bemühungen, die Interaktion im Auto mit Sprache durch neue Systeme wie Lippenerkennung und Neuronale Netzwerke zu verbessern [21] [29] [17]. Spracherkennung ist eine Technologie, die möglichst jede Interaktion während der Ausübung der primären Fahraufgabe vereinfachen soll.

#### ■ Blick- und Gesichtserkennung

Weitere Eingabemöglichkeiten im Fahrzeug können über Blick- oder Gesichtserkennung erfolgen. Dies kann sowohl implizite als auch explizite Interaktion sein [3]. Bei Impliziter Interaktion überwacht beispielsweise ein System die Anzahl der Wimpernschläge des Fahrers. Sind diese zu häufig, deutet dies auf einen müden Fahrer hin. Das System schlägt daher Alarm und weist ihn darauf hin, eine Pause zu machen. Für die Gesichtserkennung kann ein Kamerasystem anhand des Gesichtsausdrucks die Emotionen des

Fahrers erkennen und schlägt passende Musik vor [3]. Bei der Möglichkeit einer expliziten Interaktion ist der Benutzer in der Lage, mit Hilfe von Blickkontakt weitere Eingabegeräte im Auto zu steuern [13] [5]. Ähnlich verhält es sich mit expliziter Interaktion über das Gesicht. Durch Lächeln oder Grimassen schneiden besteht die Möglichkeit, Musik zu halten oder weiter zu schalten [3]. Blick- und Gesichtserkennung dient zur Unterstützung aller Fahraufgaben und sorgt vor allem für eine zentralere Steuerung der Geräte im Fahrzeug.

### ■ Fahrassistenzsystem

Fahrassistenzsysteme bündeln die bisher vorgestellten Technologien und sorgen für eine einheitliche Steuerung dieser. In der Regel werden diese *Advanced Driving Assistant System* genannt (ADAS) [24] [16]. Es unterstützt den Fahrer bei der Ausführung der primären Fahraufgabe, welches beispielsweise durch Müdigkeitserkennung oder einem eingebauten Navigationsgerät erfolgen kann. Durch Sensoren in und um das Fahrzeug kann es in der Lage sein, die Steuerung des Autos abzunehmen [8]. Darunter fallen beispielsweise die Parkhilfe, automatische Notbremse oder der Autopilot im Stau. Ein Fahrassistenzsystem ermöglicht die zentrale Steuerung vieler Funktionen über Touch Displays, Spracheingabe, Blickkontakt oder Gesichtserkennung. Es ermöglicht ferner die Synchronisation eines Smartphones mit Hilfe einer Bluetooth Verbindung. Einige Hersteller bieten bereits die Spiegelung eines Android Betriebssystems auf dem Display an [1].

Anhand der soeben genannten Technologien lässt sich erkennen, dass diese bereits die primäre Aufgabe des Fahrers beeinflussen. Zentral von einem ADAS gesteuert, unterstützen sie während der Autofahrt sowohl die Navigation, Stabilität als auch die Führung des Fahrzeugs (siehe hierzu auch Kapitel 1.2). Head Up Displays, Touch Displays, Sprach-, Blick- und Gesichtserkennung bündeln einige Interaktionsmöglichkeiten des Cockpits und tragen dazu bei, dass sich das Aussehen im Auto verändert. Ob aktuelle Technologien überhaupt noch auf den aktuellen Design Space von Kern & Schmidt abzubilden sind, soll nun diskutiert werden.

## 3 Cockpit Design: Neu ausrichten?

Der Innenraum des Fahrzeuges erhält derzeit immer mehr Technologien und Funktionen, die den Fahrer in seiner primären Fahraufgabe unterstützen (siehe Kapitel 2.2). Dies hat zwei Auswirkungen auf die Art und Weise, wie das Aussehen des Cockpit gestaltet sein sollte.

Zum einen erfolgt ein direkter Einfluss auf die Benutzung des Autos. Bisher wurde das Cockpit hauptsächlich so gestaltet, dass es den Fahrer bei der Fahrt

unterstützt und gleichzeitig nicht zu sehr ablenkt. Ab dem SAE Level 2 werden jedoch erste Fahraufgaben zeitweise übernommen (siehe 1.1). Beispielsweise könnte das System während eines Staus in der Stadt die Autofahrt übernehmen. Es entsteht der Widerspruch, dass der Fahrer in diesem Zeitraum keine Aufgabe hat und gleichzeitig konzentriert seine Umgebung beobachten soll, falls eine gefährliche Situation entsteht und der Fahrer eingreifen muss. Ab Level SAE 3 wird ihm diese Entscheidung abgenommen und er muss nur noch eingreifen, wenn das System mit einer Situation nicht umgehen kann. Dabei kann schnell Langeweile aufkommen, weil das bisherige Cockpit Design nur geringfügig zur Unterhaltung beigetragen hat (vgl. hierzu auch Abbildung 9 im Anhang). Wenn etwas passiert, muss rechtzeitig reagiert werden. Laut Kim et. al. hat dies Potential für schwerwiegende Unfälle durch menschliches Versagen [15]. Hier muss sich also der Fokus beim Gestalten des Innenraums verändern.

Zum anderen ist ein immer größer werdende Zahl an Funktionen Ursache, dass Technologien über Fahrassistenzsystem gebündelt werden. Dies ermöglicht eine zentrale und übersichtliche Interaktion [27] [5] [8]. Große Touch Displays sind in Begriff, Standard zu werden und ersetzen viele der in Kapitel 1.3 vorgestellten analogen Knöpfe, Hebel und Tasten. Gleichzeitig kann das Touch Display sowohl als Eingabe- als auch als Ausgabemöglichkeit fungieren und entwickelt sich somit zu einer zentralen Interaktionsumgebung [14] [28]. Anhand der steigenden Relevanz des Head Up Displays sieht man auch die komprimierte Darstellung wichtiger Informationen direkt im peripheren Blickfeld des Fahrers. Bereits Kern & Schmidt haben diesen Bereich hervorgehoben [14]. Die Windschutzscheibe würde sich zu einer neuen wichtigen Stelle für Ausgabegeräte entwickeln. Die Steuerung ist auch nicht mehr nur über die Mittelkonsole möglich, sondern lässt sich zentral über Sprach-, Blick- und Gesichtserkennung bedienen (vgl. Kapitel 2.2). Diese Technologien sind noch nicht ausgereift und können die bisherige analoge Steuerung nicht vollständig ersetzen. Trotzdem haben sie Potential, die Bedienung und das Aussehen des Cockpits in naher Zukunft weitreichender zu verändern, als es derzeit bereits der Fall ist.

Bezogen auf den aktuellen Design Space aus Kapitel 2.1 sind durch die vorgestellten Technologien je nach SAE Level weitreichende Auswirkungen zu erwarten (eine Zusammenfassung ist auf Tabelle 1 zu finden). Die ersten beiden SAE Level entsprechen der Technologie, die bereits bei der Erstellung des Design Space etabliert waren. Interessant werden die Veränderungen im Fahrzeug ab SAE Level 2.

Wie bereits erwähnt, muss im SAE Level 2 im Bezug auf das Design des Cockpits sowohl auf Sicherheit als auch auf die Unterhaltung Wert gelegt werden. Die Steuerung wird erstmals nach einer bewussten Entscheidung des Fahrers übernommen. Ausgabegeräte und unterhaltende Maßnahmen sind

dezent einzusetzen, da der Fahrer weiterhin konzentriert seine Umgebung beobachten muss. Als mögliche Platzierung macht daher das Head Up Display an der Windschutzscheibe Sinn. Es projiziert notwendige Informationen in das periphere Sichtfeld des Fahrers, ohne zu sehr abzulenken. Das System kann über das Head Up Display den Fahrer auf eine gefährliche Situation aufmerksam machen, damit dieser die manuelle Steuerung übernehmen kann. Ein Touch Display auf der Mittelkonsole ermöglicht weiter eine schnell zu erreichende Interaktion mit Fahrassistenzsystem. Insgesamt ist der in Kapitel 2.1 vorgestellte Design Space in Level SAE 2 noch sinnvoll einzusetzen. Platzierung sowie Interaktionsmöglichkeiten konzentrieren sich weiterhin in Reichweite des Fahrers. Dieser sitzt durchgehend am Steuer und hat dadurch einen festen Sitzplatz.

In Level SAE 3 übernimmt das Fahrassistenzsystem überwiegend das Steuer. Gleichzeitig überwacht es die Umgebung und der Fahrer wird erstmals zu einem Passagier [16]. Dadurch rücken unterhaltende Maßnahmen in einen stärkeren Fokus. Die Interaktionsumgebung der Windschutzscheibe kann dadurch vollständig genutzt werden. Beispielsweise wäre es nun möglich, Videos und Interaktionsmöglichkeiten wie Soziale Medien auf Augenhöhe anzuzeigen zu lassen. Notwendig ist die Erweiterung des Design Space auf den gesamten Innenraum, wie Meschtscherjakov et. al. bereits vorgeschlagen haben [20]. Das Lenkrad könnte einklappbar sein und der Sitzplatz drehbar. Dadurch wäre auch der hintere Bereich ein nutzbarer Platz. Beispielsweise ließe sich dort ein Tisch platzieren um darauf an einem Laptop zu arbeiten. Alle weiteren Fenster können zur Platzierung von Head Up Displays benutzt werden. Im Falle eines Systemfehler müssen Ausgabegeräte rechtzeitig jegliche Form von Unterhaltung abschalten, damit der Benutzer möglichst schnell die Steuerung des Fahrzeugs übernehmen kann (vgl. hierzu auch Abbildung 9 und 10 im Anhang). Ähnlich zu Level SAE 3 verhält es sich in SAE Level 4. Durch die meist geographische Einschränkung autonomen Fahrens ist es in diesem Level sehr wichtig, rechtzeitig den Wechsel von manuellen zu autonomen Fahren und umgekehrt zu kommunizieren.

Das SAE Level 5 bedeutet ein vollständig autonomes Fahren, wodurch primäre Geräte wie Lenkrad und Pedale nicht mehr benötigt werden. Der Benutzer eines voll autonomen Fahrzeugs ist zu jeder Zeit Passagier, wodurch sich seine bisherigen tertiären Aufgaben vollständig zu primären Aufgaben gewandelt haben. Der Sitzplatz ist nicht vorgegeben, wodurch eine feste Platzierung von Eingabe- und Ausgabegeräten ebenfalls nicht mehr notwendig ist. Der Design Space von Kern & Schmidt besitzt durch die flexible Interaktionsumgebung im Level SAE 5 keine sinnvolle Anwendung mehr.

Wichtig zu erwähnen ist, dass autonome Fahrzeuge noch sehr kostspielig sind und tendenziell nur in der Luxusklasse verfügbar sein werden. Für viele

■ **Tabelle 1** Auswirkungen der SAE Level auf den Design Space von Kern & Schmidt.

SAE Level	Auswirkung auf den Design Space
0-1	Keine. Level waren etabliert, als der Design Space vorgestellt wurde
2	Bündelung der Interaktionsmöglichkeiten auf Windschutzscheibe und Mittelkonsole. Dezent Platzierung unterhaltender Interaktionsmöglichkeiten.
3-4	Erweiterung auf den gesamten Innenraum notwendig. Interaktionsmöglichkeiten auf allen Fenstern sowie in der Mitte des Fahrzeugs.
5	Keine sinnvolle Anwendung des Design Space mehr möglich. Durch vollständig autonomes Fahren fällt die fixe Platzierung der Eingabe- und Ausgabemöglichkeiten um den Fahrersitz herum weg. Lenkrad und Pedale werden nicht mehr eingebaut.

Menschen erschwingliche Modelle können dadurch erst viel später davon profitieren. Gleichzeitig ist die Übernahme der Steuerung für viele Menschen neu [18] [19]. Der erste Kontakt mit Fahrassistenzsystemen erzeugen beim Fahrer das Gefühl, die Kontrolle zu verlieren. Die Gesellschaft muss erst Erfahrung mit selbst fahrenden Fahrzeugen machen, bis diese auch akzeptiert werden können [16]. Zu klären ist auch die rechtliche Situation selbst fahrender Fahrzeuge der Level SAE 3 bis 5. Bisher ist bei einem Unfall der Fahrer Schuld und muss dafür haften. Wenn der Fahrer selbst zu einem Passagier wird, ist Schuldfrage weitaus schwieriger zu beantworten. Es muss also national wie auch international eine rechtliche Grundlage geschaffen werden, um autonomes Fahren flächendeckend zu ermöglichen (vgl. Abbildung 9 und 10 im Anhang). All diese Herausforderung sorgen dafür, dass der aktuelle Design Space vor allem bei tendenziell günstigeren Fahrzeugen mit SAE Level 0-2 eingesetzt werden kann.

#### 4 Fazit und Ausblick

In dieser Arbeit wurden heutige Institute vorgestellt, die für die Normen der Level des autonomen Fahrens zuständig sind. Es wurden Fahraufgaben, Eingabe- und Ausgabegeräte und Interaktionsmöglichkeiten im Auto definiert, sowie der bisherige *Design Space* im automobilen Kontext vorgestellt. Moderne Technologien wurden eingeführt und ihr Einfluss auf die primäre Fahraufgabe betrachtet. Es wurde argumentiert, dass sich dadurch das Cockpit Design im Auto verändert und damit auch direkten Einfluss auf den derzeitigen Design Space besitzt. Dies variiert jedoch je nachdem, welchem SAE Level ein Fahrzeug einzuordnen ist.

Zusammenfassend lässt sich schließen, dass das Design im Auto sich verändern soll und wird. Bisher mögen Fahrassistenzsysteme den Fahrer noch nicht vollständig in seinen primären Fahraufgaben unterstützen, doch das kann sich bereits in den nächsten zehn Jahren ändern. Ab diesen Zeitpunkt

müssen Fahrzeuge ab dem Level SAE 2 so gestaltet sein, dass sie dem Fahrer rechtzeitig mitteilen, wenn dieser das Fahrzeug manuell steuern soll. Und es muss den Benutzer während der Fahrt unterhalten und vor allem konzentriert halten können.

Anhand der in Kapitel 4 diskutierten Punkte ist es in Zukunft zu erwarten, dass in den nächsten 5-10 Jahren eine stetig wachsende Anzahl teilautonomer Fahrzeuge entstehen wird. Es wird jedoch bis auf Weiteres bei einem gemischten Verkehr aus nicht autonomen - und teilautonomen Fahrzeugen bleiben [16]. Sowohl die Kosten notwendiger Technologien als auch die Akzeptanz in der Bevölkerung sprechen noch gegen eine schnelle Umstellung auf selbst fahrende Fahrzeuge. Außerdem ist es unumgänglich, rechtliche Grundlagen des autonomen Fahrens zu schaffen. Bis dahin ist der Design Space von Kern & Schmidt eine sinnvolle Darstellung, um einen Überblick auf Eingabe- und Ausgabegeräte und deren Interaktionsmöglichkeiten in einem Fahrzeug der SAE Level 0-2 zu erhalten. Ab dem Level SAE 3 muss der Design Space zumindest um den gesamten Innenraum erweitert werden, genauso verhält es sich in Level SAE 4. Im Level SAE 5 ist der Design Space nicht mehr sinnvoll, da die Geräte nicht mehr um den Fahrer herum platziert werden müssen.

Heutige Technologien verändern den Innenraum eines Autos und werden in den nächsten Jahren durch autonomes Fahren einen grundlegenden Wandel im Cockpit Design bewirken. Dafür ist ein neuer Design Space notwendig, der in der Lage ist, sich an den Level des autonomen Fahrens zu orientieren. Mit Hilfe der zusammengetragenen Erkenntnisse soll ein neuer Design Space konzipiert werden, um einen effizienten Gestaltungsprozess zu gewährleisten.

**Danksagung.** Mein besonderer Dank gilt meinen Betreuern Professor Lewis Chuang und Tobias Benz.

---

#### Literatur

- 1 Android. Android auto, 2019. Letzter Aufruf am 08.02.2019. URL: <https://www.android.com/auto/>.
- 2 BAST. Rechtsfolgen zunehmender fahrzeugautomatisierung. Letzter Aufruf am 08.02.2019. URL: [https://www.bast.de/DE/Publikationen/Foko/Downloads/2012-11.pdf?\\_\\_blob=publicationFile](https://www.bast.de/DE/Publikationen/Foko/Downloads/2012-11.pdf?__blob=publicationFile).
- 3 Michael Braun, Sarah Theres Völkel, Gesa Wiegand, Thomas Puls, Daniel Steidl, Yannick Weiß, and Florian Alt. The smile is the new like. In Slim Abdennadher and Florian Alt, editors, *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia - MUM 2018*, pages 383–389, New York, New York, USA, 2018. ACM Press. doi:10.1145/3282894.3289729.
- 4 Audi Dialoge. Automatisiertes fahren, 2019. Letzter Aufruf am 08.02.2019. URL: <https://audi-dialoge.de/de/automatisiertes-fahren>.
- 5 David Dobbstein, Marcel Walch, Andreas Köll, Ömer Şahin, Tamino Hartmann, and Enrico Rukzio. Reducing in-vehicle interaction complexity. In Jonna Häkikila, editor, *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia*, pages 311–313, New York, NY, 2016. ACM. doi:10.1145/3012709.3016064.
- 6 Anup Doshi, Shinko Yuanhsien Cheng, and Mohan Manubhai Trivedi. A novel active heads-up display for driver assistance. *IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics : a publication of the IEEE Systems, Man, and Cybernetics Society*, 39(1):85–93, 2009. doi:10.1109/TSMCB.2008.923527.
- 7 Ford. Ford sync 2 mit touchscreen & sprachsteuerung, 2019. Letzter Aufruf am 08.02.2019. URL: <https://www.ford.de/service/betriebsanleitungen-hilfe/ford-sync-bluetooth/sync2-mit-touchscreen>.
- 8 Benjamin Franz, Jan P Haccius, Dennis Stelzig-Krombholz, Matthias Pfromm, Michaela Kauer, and Bettina Abendroth. Evaluation of the sagat method for highly automated driving. In *Proceedings 19th Triennial Congress of the IEA*, volume 9, page 14, 2015. URL: <https://www.iea.cc/congress/2015/154.pdf>.
- 9 Renate Haeuslschmid, Susanne Forster, Katharina Vierheilg, Daniel Buschek, and Andreas Butz. Recognition of text and shapes on a large-sized head-up display. In Oli Mival, Michael Smyth, and Peter Dalsgaard, editors, *Proceedings of the 2017 Conference on Designing Interactive Systems - DIS '17*, pages 821–831, New York, New York, USA, 2017. ACM Press. doi:10.1145/3064663.3064736.
- 10 Zhang Hua and Wei Lih Ng. Speech recognition interface design for in-vehicle system, 2010. URL: [http://dl.acm.org/ft\\_gateway.cfm?id=1969780&type=pdf](http://dl.acm.org/ft_gateway.cfm?id=1969780&type=pdf), doi:10.1145/1969773.1969780.

- 11 S.A.E. International. Automated driving levels of driving automation are defined in new sae international standard j3016. Letzter Aufruf am 08.02.2019. URL: [https://web.archive.org/web/20170903105244/https://www.sae.org/misc/pdfs/automated\\_driving.pdf](https://web.archive.org/web/20170903105244/https://www.sae.org/misc/pdfs/automated_driving.pdf).
- 12 S.A.E. International. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. Letzter Aufruf am 08.02.2019. URL: [https://www.sae.org/standards/content/j3016\\_201806/](https://www.sae.org/standards/content/j3016_201806/).
- 13 Dagmar Kern, Angela Mahr, Sandro Castronovo, Albrecht Schmidt, and Christian Müller. Making use of drivers' glances onto the screen for explicit gaze-based interaction, 2010. URL: [http://dl.acm.org/ft\\_gateway.cfm?id=1969792&type=pdf](http://dl.acm.org/ft_gateway.cfm?id=1969792&type=pdf), doi:10.1145/1969773.1969792.
- 14 Dagmar Kern and Albrecht Schmidt. Design space for driver-based automotive user interfaces. In Albrecht Schmidt, editor, *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, page 3, New York, NY, 2009. ACM. doi:10.1145/1620509.1620511.
- 15 Naeun Kim, Kwangmin Jeong, Minyoung Yang, Yejeon Oh, and Jinwoo Kim. "are you ready to take-over?". In Gloria Mark, Susan Fussell, Cliff Lampe, m.c schraefel, Juan Pablo Hourcade, Caroline Appert, and Daniel Wigdor, editors, *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '17*, pages 1771–1778, New York, New York, USA, 2017. ACM Press. doi:10.1145/3027063.3053155.
- 16 Andrew L. Kun, Susanne Boll, and Albrecht Schmidt. Shifting gears: User interfaces in the age of autonomous driving. *IEEE Pervasive Computing*, 15(1):32–38, 2016. doi:10.1109/MPRV.2016.14.
- 17 Weifeng Li, Yicong Zhou, Norman Poh, Fei Zhou, and Qingmin Liao. Feature denoising using joint sparse representation for in-car speech recognition. *IEEE Signal Processing Letters*, 20(7):681–684, 2013. doi:10.1109/LSP.2013.2245894.
- 18 Alexander Meschtscherjakov, Wendy Ju, Manfred Tscheligi, Dalila Szostak, Sven Krome, Bastian Pfleging, Rabindra Ratan, Ioannis Politis, Sonia Baltodano, and Dave Miller. Hci and autonomous vehicles. In Jofish Kaye, Allison Druin, Cliff Lampe, Dan Morris, and Juan Pablo Hourcade, editors, *CHI 2016*, pages 3542–3549, New York, New York, 2015. The Association for Computing Machinery. doi:10.1145/2851581.2856489.
- 19 Alexander Meschtscherjakov, Manfred Tscheligi, Dalila Szostak, Sven Krome, Bastian Pfleging, Rabindra A. Ratan, Ioannis Politis, Sonia Baltodano, Dave Miller, and Wendy Ju. Hci and autonomous vehicles: Contextual experience informs design. *Unknown*, 2016. URL: <http://dl.acm.org/citation.cfm?id=2856489>.
- 20 Alexander Meschtscherjakov, David Wilfinger, Nicole Gridling, Katja Neureiter, and Manfred Tscheligi. Capture the car!: qualitative in-situ methods to grasp the automotive context, 2011. URL: [http://dl.acm.org/ft\\_gateway.cfm?id=2381434&type=pdf](http://dl.acm.org/ft_gateway.cfm?id=2381434&type=pdf), doi:10.1145/2381416.2381434.
- 21 Rajitha Navarathna, Patrick Lucey, David Dean, Clinton Fookes, and Sridha Sridharan. Lip detection for audio-visual speech recognition in-car environment. In *10th International Conference on Information Science, Signal Processing and their Applications (ISSPA 2010)*, pages 598–601. IEEE, 2010. doi:10.1109/ISSPA.2010.5605429.
- 22 NHTSE. Automated vehicles for safety. Letzter Aufruf am 08.02.2019. URL: <https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety#issue-road-self-driving>.
- 23 Opel. Infotainment cascada, 2019. Letzter Aufruf am 08.02.2019. URL: <https://www.opel.de/opel-erleben/infotainment/infotainment-cascada.html>.
- 24 A. Pauzié. A method to assess the driver mental workload: The driving activity load index (dali). *IET Intelligent Transport Systems*, 2(4):315, 2008. doi:10.1049/iet-its:20080023.
- 25 Bastian Pfleging, Stefan Schneegass, and Albrecht Schmidt. Multimodal interaction in the car: combining speech and gestures on the steering wheel, 2012. URL: [http://dl.acm.org/ft\\_gateway.cfm?id=2390282&type=pdf](http://dl.acm.org/ft_gateway.cfm?id=2390282&type=pdf), doi:10.1145/2390256.2390282.
- 26 Tony Matthias Poitschke. *Blickbasierte Mensch-Maschine Interaktion im Automobil*. dissertation, Technische Universität München, 2011. URL: <https://mediatum.ub.tum.de/doc/1072203/document.pdf>.
- 27 A. Riener, A. Ferscha, F. Bachmair, P. Hagmüller, A. Lemme, D. Muttenthaler, D. Pühringer, H. Rogner, A. Tappe, and F. Weger. Standardization of the in-car gesture interaction space, 2013. URL: [http://dl.acm.org/ft\\_gateway.cfm?id=2516544&type=pdf](http://dl.acm.org/ft_gateway.cfm?id=2516544&type=pdf), doi:10.1145/2516540.2516544.
- 28 Albrecht Schmidt, Anind K. Dey, Andrew L. Kun, and Wolfgang Spiessl. Automotive user interfaces: human computer interaction in the car, 2010. URL: [http://dl.acm.org/ft\\_gateway.cfm?id=1753949&type=pdf](http://dl.acm.org/ft_gateway.cfm?id=1753949&type=pdf), doi:10.1145/1753846.1753949.
- 29 Michael L. Seltzer, Dong Yu, and Yongqiang Wang. An investigation of deep neural networks for noise robust speech recognition. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 7398–7402. IEEE, 2013. doi:10.1109/ICASSP.2013.6639100.
- 30 Stavros Tasoudis and Mark Perry. Participatory prototyping to inform the development of a remote ux design system in the automotive domain. *Multimodal Technologies and Interaction*, 2(4):74, 2018. URL: <https://www.mdpi.com/2414-4088/2/4/74/pdf>, doi:10.3390/mti2040074.
- 31 Tesla. Tesla models, 2019. Letzter Aufruf am 08.02.2019. URL: <https://www.tesla.com/models>.

- 32 M. Tonnis. *Towards automotive augmented reality*. dissertation, Technische Universität München, 2008. URL: <https://mediatum.ub.tum.de/doc/652549/652549.pdf>.
- 33 M. Tonnis, V. Broy, and G. Klinker. A survey of challenges related to the design of 3d user interfaces for car drivers. In *IEEE Symposium on 3D User Interfaces, 2006*, pages 127–134, Piscataway, N.J, 2006. IEEE Service Center. doi:10.1109/VR.2006.19.
- 34 Waymo. Homepage, 2019. Letzter Aufruf am 08.02.2019. URL: <https://waymo.com/>.
- 35 Wen Wu, Fabian Blaicher, Jie Yang, Thomas Seder, and Dehua Cui. A prototype of landmark-based car navigation using a full-windshield head-up display system. In A. El Saddik, K. S. Candan, I. Cheng, A. Basu, Qing Li, Rynson W.H Lau, Benjamin Wah, Howard Leung, and Cha Zhang, editors, *Proceedings of the 2009 workshop on Ambient media computing - AMC '09*, page 21, New York, New York, USA, 2009. ACM Press. doi: 10.1145/1631005.1631012.

## A Anhang

Interview mit  
Dr. Alexander Meschtscherjakov  
über Cockpit Design (teil-) autonomer Fahrzeuge

Am 17.01.2019

### 1. In welchen SAE Level sehen Sie derzeit heutige Fahrzeuge?

- Level 0-2 normalerweise. Fahrer trägt rechtliche Belangung.
- Level 3-4 rechtlich noch nicht erlaubt, da Auto Schuld (nur mit Fahrer). Grundsätzlich bereits möglich.
- Beispiel *Digibus* in Salzburg, der autonom fahren kann.

### 2. Deutsche Hersteller werden in den nächsten Jahren tendenziell Automobile der SAE Level 3 herstellen, sprich teilweise autonome Systeme verbauen. Wie könnte sich dies auf das Cockpit Design im Auto auswirken?

- Hängt stark von der eigentlichen autonomen Fahrzeit ab.
- Problematik der Schuldfrage: 10 Sekunden Reaktionszeit vom System kalkuliert, danach ist der Fahrer schuld, falls ein Unfall gebaut wird.
- Schlafen nicht erlaubt, da bis Level 3 grundsätzlich noch manuell eingegriffen werden muss.
- Mögliche Änderungen des Designs:
  - Sitz umdrehen/ Lenkrad einklappen (Zeitfaktor muss beachtet werden!)
  - Noch größere Displays (Touch)
  - Head Displays/ Windshield Displays für Entertainment
  - Darstellung, welches Level gerade aktiviert ist (im Verbraucherkontext *automatisiert/ nicht automatisiert*)
  - Bei der Platzierung könnte sich anhand dieser Möglichkeiten neue Stellen entstehen
  - Content Rotation Studien (Tablet auf Rondell legen, um so den Content von allen Seiten betrachten zu können)
  - Evtl. Technologie aus dem VR/ AR Bereich

### 3. Was für eine Auswirkung hat dies auf den Fahrer und dessen Nutzererfahrung im Auto?

- Sehr groß! Bisher Sicherheit im Fokus: primäre Aufgabe war das Fahren.

■ **Abbildung 9** Interview mit Dr. Alexander Meschtscherjakov - Teil 1

- in Zukunft vom Fahrer zum Passagier: primäre Aufgabe wird Entertainment.
- Durch Navi schleicht sich eine schlechtere Orientierung des Benutzers ein.
- Schlechter Autofahren durch Automatismen → *irony of automation*.
- Motion Sickness!
- *Overtrust*: zu viel Vertrauen durch automatisches System.
- UX: Akzeptanz, Joy-of-Use → am besten keine manuelle Steuerung, das gesamte Straßennetz profitiert von ein voll automatisierten Verkehrsumgebung.
- Über AR User Experience einbauen: Information über die Umgebung anzeigen.
- Design Space: Vibrationen, Lichtverhältnisse (äußere Reize).
- Ergonomisch sinnvoll: Platzierung auf Augenhöhe.

**4. Was könnte man als Designer eines Innenraumes tun, den Fahrer sowohl auf die Fahrt zu konzentrieren als auch Langeweile entgegen zu wirken?**

- Problem eher auf Level 2: Immer aufpassen, immer konzentriert sein.
- Level 3: Möglichkeit des Wechsel von Automatik auf Manuell besser (Zeit des autonomen Fahrens länger).
- Beispiel: Wahrnehmung Fahrer/ Beifahrer während der Fahrt. Bremszeit wird als viel enger angesehen, wenn man selbst nicht eingreifen kann.
- Beispiel Fußballer auf einer Ersatzbank. Jeder Spieler ist emotional voll dabei, auch wenn er nicht spielt.
- Gradwanderung Sicherheit/ Entertainment: Möglichkeit Entertainment herunterzufahren und auf Sicherheit hinzuweisen (*judge and dodge law*)
- *Spielerische Kontexte* möglich: Autos zählen, Hörbuch hören.

■ **Abbildung 10** Interview mit Dr. Alexander Meschtscherjakov - Teil 2

# Evaluation der Fähigkeit von haptischen Interfaces für dynamische Materialsimulation in Virtueller Realität

Yannick Weiß

Ludwig-Maximilians-Universität München, München, Deutschland  
Yannick.Weiss@campus.lmu.de

## Zusammenfassung

Haptik gilt als der nächste große Schritt in der Weiterentwicklung virtueller Realitäten. Die Möglichkeit der realistischen Simulation von verschiedenen Materialien spielt dabei eine große Rolle. In dieser Arbeit wurden Ansätze für haptische Interfaces zusammengetragen und anhand ihrer theoretischen Fähigkeit der realistischen und veränderbaren Darstellung von Materialeigenschaften evaluiert. Die untersuchten Interfaces wurden in drei Kategorien geteilt: Mechanische Interfaces, die Druck durch Vibrationen, Exoskelette oder externe Systeme wie Roboter erzeugen. Shape Changing User Interfaces, bei denen reale Objekte dynamisch ihre Eigenschaften verändern können. Und Haptic Rendering Methoden, die das Erzeugen von Kräften frei in der Luft ermöglichen. Es stellte sich heraus, dass jedes Interface individuelle Vor- und Nachteile mit sich bringt und keine der untersuchten Interfacearten eigenständig als Favorit gewertet werden kann. Eine Kombination mehrerer Methoden ist notwendig, um ein möglichst realistisches haptisches Feedback zu gewährleisten. Die Verbindung von Jamming-Methoden und Pneumatik zeigt dabei das größte Potential, um eine hohe Reichweite an verschiedenen realistischen Materialien zu simulieren.

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

**Keywords and phrases** Haptische Interfaces; Haptisches Feedback; Materialsimulation; Shape Changing; Virtuelle Realität.

## 1 Ausgangssituation

Das Eintauchen in virtuelle Umgebungen, in denen frei interagiert werden kann, ermöglicht eine Vielzahl an neuen Möglichkeiten und Erfahrungen. Während visuelle und auditive Simulationen weiter fortschreiten, geraten die restlichen Sinne nun langsam ebenfalls in den Fokus. Unser ausgeprägter Tastsinn stellt



© Yannick Weiß;  
licensed under Creative Commons License CC-BY  
Cite as: Yannick Weiß. Evaluation der Fähigkeit von haptischen Interfaces für dynamische Materialsimulation in Virtueller Realität. In *3rd Seminar on Ubiquitous Interaction (UBIAC-TION 2019-1)*. Editors: Fiona Draxler, Matthias Hoppe, Jakob Karolus, Thomas Kosch, Pascal Knierim, Albrecht Schmidt. January 31, 2019. Munich, Germany. pp. 17:1–17:21.

dabei das größte Potential für virtuelle und erweiterte Realitäten dar. Durch Anfassen fallen uns Interaktionen am leichtesten und das Gefühl der Präsenz hängt stark davon ab. Es existieren bereits eine Vielzahl an verschiedenen Methoden für haptische Interfaces, sowohl in kommerziellen Produkten auf dem Markt, als auch in prototypischen Umsetzungen in der Forschung. Für die Zukunft ist dabei vor allem der Realismus dieser Interfaces und Geräte interessant. Eine wichtige Rolle spielt dabei die realistische Darstellung von Materialeigenschaften. Für Virtuelle und Erweiterte Realität müssen diese Materialien und ihre Eigenschaften jedoch zusätzlich beliebig variierbar sein, um eine möglichst hohe Reichweite an unterschiedlichen Materialien und Objekten darstellen zu können.

Diese Arbeit beschäftigt sich daher mit unterschiedlichen Ansätzen für haptische User Interfaces und ihrem Potential der realistischen Darstellung von Materialeigenschaften, sowie ihre Fähigkeit der dynamischen und programmgesteuerten Anpassung dieser Eigenschaften. Schlussendlich soll die Frage beantwortet werden, welche dieser Methoden — oder Kombination dieser Methoden — für dynamische Materialsimulation in Virtueller Realität am geeignetsten sind.

Da die meisten Interfaces in dieser Hinsicht weder untersucht noch verglichen wurden, basiert diese Evaluation lediglich auf Folgerungen, die aus den theoretischen Fähigkeiten erschlossen wurden. Um die wirkliche Eignung der Interfaces festzustellen benötigt es daher zusätzliche Vergleiche in der Praxis und Nutzerstudien.

## 2 Haptisches Feedback

Der Begriff Haptik umfasst alle Interaktionen zur Manipulation oder Wahrnehmung der Umgebung durch physischen Kontakt. Um diesen Kontakt virtuell zu simulieren, können haptische Interfaces genutzt werden. Damit ein Objekt als physisch wahrgenommen werden kann, müssen bei einer Berührung bestimmte Kräfte auf den Menschen einwirken, die Widerstand suggerieren. Die Wahrnehmung unterteilt sich dabei in zwei Aspekte: Kinästhetisches und Taktiler Feedback [29].

### 2.1 Kinästhetisches Feedback

Kinästhetische Informationen werden innerhalb des Körpers durch Mechanorezeptoren an Gelenken, Sehnen und Muskeln gesammelt. Diese Informationen bestehen aus der Stärke und Richtung der verspürten Kraft, die gegen die Gelenke wirkt. Geht man von einem statischen, nicht verschiebbaren Objekt aus, so ist die Stärke dieser Kraft vom Grad des Eindrucks (engl.: indentation), wie weit das Objekt eingedrückt wird, und der Steifigkeit (engl. stiffness), wie

fest das Material an diesem Punkt ist, abhängig [29]. Drückt man beispielsweise mit der Hand gegen eine Betonwand, deren Material in der Regel eine enorm hohe Steifigkeit besitzt, so entsteht eine sehr hohe Gegenkraft auf den gesamten Arm. Drückt man stattdessen gegen eine weiche Matratze, ist diese Kraft schwächer, nimmt jedoch spürbar zu je weiter man in diese hineindrückt.

### 2.2 Taktiler Feedback

Im Gegensatz zu kinästhetischen Kräften, die von den Eigenschaften des Gesamtobjektes, wie dessen Material und Form, abhängig sind, geben taktile Informationen Auskunft über die Oberflächenbeschaffenheit des Objektes. Die Wahrnehmung findet dabei auf der Haut statt, die eine Vielzahl an Rezeptoren besitzt. Dadurch kann die genaue Verteilung der Kräfte auf der gesamten Kontaktfläche, z.B. der Handfläche, erkannt werden. Diese Kräfte, die wir als Oberflächeneigenschaften wahrnehmen, entstehen durch Reibung und Textur [29]. Reibungskräfte entstehen, wenn mit der Haut bzw. der Kontaktfläche über die Oberfläche des Objektes gefahren wird. Sie sind tangential zur Objekt Oberfläche und gegen die Fahrtrichtung gerichtet. Texturen bestehen aus kleinen Erhöhungen und Vertiefungen auf der Oberfläche, die dementsprechend mehr oder weniger Druck auf die Haut ausüben, wenn diese flach auf dem Objekt liegt. Bewegt sich die Kontaktfläche über die Textur entstehen wiederum Kräfte tangential zur Oberfläche [31]. Eine Textur kann gleichmäßig oder unstrukturiert sein. Zusammen mit der gespürten Reibung nimmt das Gehirn dadurch Oberflächeneigenschaften wahr, beispielsweise Rauheit, Härte und Klebrigkeit [7].

## 3 Mechanische Interfaces

Diese Kategorie beinhaltet Interfaces, die durch mechanische Vorgänge zu haptischen Erfahrungen führen. Dazu gehören die meistverbreiteten Methoden für haptisches Feedback. Das Vibrationsfeedback, welches in den herkömmlichen VR-Kontrollern implementiert ist, und das mechanische Exoskelett, das in Datenhandschuhen meistens verwendet wird, stellen den derzeitigen State-of-the-Art für den Konsumentenbereich dar. In der professionellen Anwendung, gerade im Bereich Medizin und Operationstraining, findet man häufig haptische Systeme, die mithilfe von externen Installationen präzises und dynamisches Feedback generieren können.

### 3.1 Vibrationsfeedback

Haptisches Feedback wird in herkömmlichen VR-Systemen, beispielsweise der HTC Vive oder Oculus Rift, über Vibration der Controller suggeriert. Diese

können dabei in unterschiedlichen Stärken, Längen und Intervallen pulsieren [10]. Vibrationsfeedback ermöglicht es lediglich, die Stärke der auftretenden kinästhetischen Kräfte zu emulieren. Richtung dieser Kräfte, sowie taktile Informationen können darüber nicht vermittelt werden. Durch andauernde Vibrationen in variablen Amplituden lassen sich verschiedene Steifigkeiten von Objekten simulieren. Die dynamische Bandbreite der möglichen Stufen ist durch die Möglichkeit der präzisen Kontrolle von Amplitude und Intervall sehr hoch. Da diese Vibrationen jedoch nicht natürlich bei Kräften auftreten und das Vibrationsfeedback auf die gesamte Hand gleichermaßen ausgewirkt wird, können virtuelle Materialien und deren Eigenschaften nicht ausreichend kommuniziert werden.

### 3.2 Exoskelett

Die häufigste genannte und genutzte Lösung für ein verbessertes haptisches Feedback sind Datenhandschuhe [13] [22] [36]. Sie erlauben kinästhetisches Kraftfeedback in verschiedenen Stärken und Richtungen. Um diese Kräfte auszuwirken, werden in der Regel mechanische oder hydraulische Bremsen oder Kolben verwendet, die der Kraft des Nutzers gegenwirken. Dazu werden Exoskelette (Stützstrukturen) verwendet, die durch das Blockieren der Bremsen oder Gegenkraft der Kolben zu dem gewünschten Druck führen können. Neuere Modelle, z.B. die „HaptX Gloves“ ([16]) ermöglichen Druck aus vielen Richtungen und individuell für jeden Finger. Die kinästhetischen Kräfte haben jedoch ein strenges Limit, das zusätzlich stark mit der Richtung der Kraft variiert. Außerdem erzeugt ein Exoskelett immer einen Widerstand, selbst wenn sich die Hand des Nutzers sowohl in der Realität, als auch in der virtuellen Umgebung, ohne Kontakt in der Luft befindet. Die gegenwirkende Kraft kann also niemals komplett abgeschaltet bzw. auf den normalen Wert des Luftwiderstandes gebracht werden. Auch die Sensorik und Berechnung kann einen Nachteil der Handschuhe darstellen. Da die Bremsen und Kolben aktiv in der richtigen Position und zur richtigen Zeit wirken müssen, kann es bei kleinen Abweichungen durch die Auflösung der Sensoren bereits zu Fehlern kommen. Auch die Berechnung, die, im Gegensatz zu der kontinuierlichen Bewegung des Menschen, nur in Intervallen möglich ist, führt zu leicht verspäteter Aktivierung oder Freigaben. All das führt zu virtuellen Energieverlusten — weniger Kraft muss zum Greifen aufgebracht werden, als beim Lösen freigegeben wird, wodurch die haptische Wahrnehmung von der Realität abweicht [29].

Die Emulation von Materialien mit verschiedenen Graden von Steifigkeit ist durch Datenhandschuhe präzise und dynamisch möglich. Energieverluste und die relativ geringe maximale Steifigkeit bilden hierbei die größten Hürden. Taktiles Feedback, also Vermittlung von Oberflächentextur und Reibung, ist durch Exoskelette allein ebenfalls nicht umsetzbar.

### 3.3 Encounter-Type Interface

Encounter-Type Interfaces üben ebenfalls Druck durch mechanische Gegenkräfte aus. Im Gegensatz zu den Datenhandschuhen werden sie jedoch nicht am Körper getragen, sondern extern installiert. Um virtuelle Objekte trotzdem an richtiger Position darstellen zu können, müssen daher die Hände des Nutzers verfolgt und die Kontaktfläche demensprechend verschoben werden. Da Objekte im virtuellen Raum willkürlich platziert sein können, muss die Verschiebung der Kontaktfläche in der Luft und in möglichst vielen Dimensionen möglich sein. Dafür eignen sich besonders Roboterarme [41] [26], aber es wird auch bereits mit neuartigen Ansätzen, wie z.B. Drohnen [1] [21] [18], experimentiert. Um weiche und harte Materialien zu simulieren, können Sensoren die Position der Hand und/oder die ausgeübte Kraft des Nutzers messen, und die Position der Kontaktfläche, sowie die Gegenkraft dementsprechend anpassen. Durch die Möglichkeit der freien Positionierung und Verschiebung im Raum, können Kräfte in verschiedene Richtungen entwickelt werden [20]. Kinästhetik kann dadurch sehr realistisch dargestellt werden. Die Kraft eines Roboterarms ist zusätzlich hoch genug, um selbst die härtesten Materialien darzustellen. Allerdings erben Encounter-Type Interfaces viele derselben Schwächen von tragbaren Exoskeletten: Das Tracking und die Berechnung sind fehlerbehaftet und nur in Intervallen möglich und die Simulation von verschiedenen Oberflächen ist nicht möglich.

### 3.4 Externe Endeffektoren

Bei dieser Methode wird die Hand des Nutzers durch ein reales physisches Element erweitert. Diese Erweiterung, die nun den Endeffektor — das letzte Glied der kinematischen Kette — des Menschen bildet, erfährt nun statt der Hand des Nutzers das haptische Feedback und leitet dieses weiter. Das Element ist dabei direkt mit einer Vorrichtung, z.B. einem Roboterarm, verbunden und kann dadurch präzise Positions- oder Kraftveränderungen erfahren. Der Endeffektor kann dabei ein beliebiges greifbares Objekt sein: Lövgquist et al. [23] verwenden dafür eine Nadel, Minsky [7] einen Joystick. Das System „PHAN-ToM“ [24] funktioniert stattdessen über ein fingerhutartiges Objekt, in das der Finger direkt hineingelegt werden kann. Kinästhetische Kräfte können durch diese Systeme präzise und realitätsnah dargestellt werden, sowohl in Stärke als auch Richtung. Die Simulation der Steifigkeit verschiedenen Materialien ist einer der Hauptaufgaben dieser Systeme. Lövgquist et al. [23] simulieren beispielsweise Steifigkeiten und Oberflächenspannungen von unterschiedlichen Schichten des Rückengewebes. Wie Minsky [7] zeigt, können diese Systeme ebenfalls Feedback über Oberflächentextur geben. Diese Informationen können allerdings nicht direkt über die Haut des Nutzers rezipiert werden, da

sich diese in keinem direkten Kontakt befindet. Daher werden diese durch minimale kinästhetische Kräfte ausgedrückt, die als Textur oder Reibung des Endeffektors interpretiert werden, ähnlich der haptischen Erfahrung beim Schreiben mit einem Bleistift auf unterschiedlichen Unterlagen. Dadurch wird das haptische Feedback jedoch auf die gesamte Hand gleichmäßig angewandt. Anstatt präzisen wahrnehmbaren Druckverteilungen auf der Kontaktfläche, können also nur haptische Informationen für einen einzigen Punkt gesammelt und weitergegeben werden.

## 4 Shape Changing Interfaces

Shape Changing Interfaces zählen zu den Tangible User Interfaces (TUI). Tangible (deutsch: greifbare) Interfaces ermöglichen Repräsentation und Kontrolle von digitalen Informationen über reale, physische Objekte. Diese können dadurch sowohl zur Eingabe durch den Nutzer, z.B. durch Berührung oder Drücken, als auch zur Ausgabe von Informationen dienen [37]. Shape Changing Interfaces können diese digitalen Informationen durch die Veränderung der eigenen Objekteigenschaften darstellen. Dazu gehören unter anderem Veränderungen in Ausrichtung, Form und Größe. Rasmussen et al. [28] und Coelho et al. [6] geben einen detaillierten Überblick über die Mechanismen und Eigenschaften dieser Interfaces. Für diese Arbeit sind davon vor allem die Möglichkeit der programmierbaren Veränderung von Steifigkeit/Zähigkeit und Oberflächentextur der Objekte interessant.

### 4.1 Pneumatische Aktuation

Pneumatik ist ein Teilgebiet der Fluidtechnik, welches sich mit verdichteter Luft — Druckluft — beschäftigt [14]. Obwohl die in Shape Changing Interfaces verbauten pneumatischen Systeme sehr komplex sein können, so ist das grundlegende Konzept dahinter sehr einfach: In das Objekt wird Luft hinein- oder herausgepumpt, was den inneren Druck erhöhen bzw. verringern kann. Unterschiedliche Membran- und Basismaterialien, sowie elektrische Bauteile, ermöglichen dabei verschiedene Reichweiten von Steifigkeit. Dasselbe Prinzip ermöglicht das Erzeugen von Textur auf der Oberfläche, indem kleine Luftblasen in der Membran separat mit Druckluft gefüllt werden. Das Muster — Dichte, Frequenz und Abfolge der Erhöhungen — kann dabei durch die unabhängige Aktivierung einzelner Punkte oder Reihen individuell angepasst werden. Talhan und Jeon [34] geben einen umfassenden Überblick über verschiedene Systeme, Herangehensweisen, Anwendungsfälle und Tests von pneumatischen haptischen Interfaces in der Medizin. Oft zusammen mit Jamming-Methoden (siehe folgende Abschnitte) werden diese Interfaces in

Operationstrainings für eine Vielzahl von Simulationen, z.B. von Tumoren und Hautkrankheiten, verwendet.

Pneumatik ermöglicht die Veränderung von Steifigkeit und Oberflächentextur und kann somit realistisches kinästhetisches und taktiles Feedback an den Nutzer vermitteln. Die mögliche Reichweite der Steifigkeit ist durch den Charakter der Luft jedoch relativ gering. Um ein relativ hartes Material zu simulieren, müsste man einen enormen inneren Druck erreichen, gegen den die Aktuatoren oder schlussendlich die Membran nicht standhalten kann.

## 4.2 Particle Jamming

Particle Jamming ist eine Technik, die bereits verbreitete Anwendung in der Softrobotik findet [39] [3]. Dort wird sie vor allem dafür verwendet, die Steifigkeit der Greifarme von Robotern dynamisch zu verändern. Im weichen Zustand können Objekte dadurch behutsam gegriffen und getragen werden, während die Veränderung zu steiferem Material mehr Druck auf Objekte ermöglicht, um diese beispielsweise zu verschieben.

### 4.2.1 Granular Jamming

Beim Granular Jamming werden granulare Medien verwendet, wie beispielsweise Sand, Kunststoffgranulat oder verschiedene Arten von Pulver. Diese Materialien haben eine spezielle physikalische Besonderheit. Sie können sowohl einen flüssigen, als auch einen starren Charakter annehmen. Wenn sie nicht unter Druck stehen, können sie sich frei verformen und fließen. Werden die einzelnen Granulate jedoch unter Druck zusammengepresst, erstarren sie und verhalten sich wie ein festes Material, lassen sich dementsprechend schwerer, aber dafür anhaltend verformen — behalten also die Verformung bei. Wird der Druck wieder herabgesetzt, kehrt das Material in seinen ursprünglichen nahezu flüssigen Zustand zurück [11]. Dieses Verhalten unter Druck entsteht durch „interparticle friction“ [5] — also Reibung zwischen den einzelnen Partikeln — und durch die Bildung von sogenannten „force chains“ [4] entlang der Richtung der einwirkenden externen Kraft, wodurch sich diese auf die gesamte Kette verteilt.

Um dieses Verhalten zu ermöglichen, müssen die einzelnen Partikel des Granulats eng zusammengepresst werden. Als Medium wird hierfür meistens Luft verwendet. Im Gegensatz zu der in Kapitel 4.1 beschriebenen Methode, nimmt beim Pneumatic Jamming die Steifigkeit des Materials jedoch durch das Erzeugen eines Unterdrucks zu, d.h. je weniger Luft in dem Material vorhanden ist, desto starrer wird es. Pneumatik ist besonders geeignet, da es sehr günstig, leicht, sicher und kontrollierbar ist [34]. Außerdem ermöglicht es die Aktivierung innerhalb von Millisekunden [5].

Die Reichweite an Veränderung ist bei der Granular Jamming Methode sehr hoch. Wall et al. [39] evaluierten verschiedene Prototypen und erreichten dabei einen Anstieg der Steifigkeit um das 8-fache. Durch die breite Auswahl an Materialien kann diese Reichweite dem Anwendungsfall weiter angepasst werden — was jedoch nicht dynamisch während des Nutzens funktioniert. Während Kaffeepulver eines der größten Veränderungspotentiale besitzt [5], können beispielsweise auch Glasperlen als Jamming-Material verwendet werden, die durch ihre rundlichen Formen und glatten Oberflächen zwar keine sehr hohen Steifigkeitswerte erreichen können, sich im entspannten Zustand dahingegen noch flüssiger verhalten [9].

#### 4.2.2 Layer Jamming

Layer Jamming beschreibt eine Erweiterung der Granular Jamming Methode, die das Jamming innerhalb sehr dünner und leichter Objekten ermöglicht. Layer Jamming wurde von Kim et al. für medizinisches Training von minimalinvasiven Operationen entwickelt [19]. Die beim Granular Jamming freien Elemente des Granulats werden beim Layer Jamming durch raue Schichten ersetzt. Durch Erzeugung eines Vakuums zwischen den Schichten werden diese eng aneinandergespreßt, wodurch sie, durch die Reibung zwischen den Schichten, starrer werden und das Gesamtmaterial steifer machen. Durch viele Variationsmöglichkeiten der Strukturen und der Materialien können durch Layer Jamming sehr individuelle Steifigkeits-Reichweiten erreicht werden. Jifei Ou et al. [25] haben dafür verschiedene Strukturen und Materialien untersucht und evaluiert.

### 4.3 Smart Materials

Smart Materials besitzen die Möglichkeit ihre molekulare Struktur, und damit ihre Steifigkeit, durch Reaktion auf äußere Stimuli — z.B. Licht, elektrischer Strom oder Temperatur — zu verändern [2]. In diesem Abschnitt werden verschiedene Smart Materials betrachtet, die alle die Möglichkeit zur dynamischen Anpassung ihrer Materialeigenschaften besitzen. Da Smart Materials extern aktiviert werden, benötigen sie wenig Platz und können somit auch als dünne Schichten verwendet werden.

#### 4.3.1 Shape Memory Materials

Shape Memory Materials (SMMs) — beispielsweise Shape Memory Polymers (SMP) oder Shape Memory Alloys (SMA) — heben sich besonders durch die Fähigkeit hervor, nach Entzug des Stimulus in ihren vorherigen Zustand bzw. ihre ursprüngliche Form zurückkehren zu können. Die kontrollierte Form-Änderung macht diese Materialien sehr interessant für Shape Changing, die

damit einhergehende Veränderung der Materialeigenschaften wird jedoch oft vernachlässigt. Mit der Ausnahme von speziellen Polymeren, die sich durch Lichteinwirkung oder chemische Reaktionen aktivieren lassen, verändern sich die Meisten SMMs durch externe Temperaturanpassung. Shape Memory Alloys verwenden Legierungen aus Metallen, die schon bei relativ niedrigen Temperaturen zu schmelzen bzw. erweichen beginnen, und so in einen biegsamen Zustand wechseln. SMAs haben damit eine sehr hohe maximale Steifigkeit bei niedrigeren Temperaturen, auf Kosten der möglichen Reichweite (weniger als vierfache Veränderung der Steifigkeit). Shape Memory Polymers haben bei Erhitzung dagegen eine höhere Reichweite an reversibler Veränderung (über das 100-fache), können jedoch nicht dieselbe Härte erreichen [30]. Die Aktivierung durch Temperatur bedeutet jedoch, dass das Erhitzen und vor allem das Abkühlen mehrere Sekunden benötigt und von der Umgebung abhängig ist [39]. Außerdem müssen die Objekte für haptisches Feedback berührt werden, was wiederum Probleme aufwirft: Ist die Aktivierungstemperatur zu hoch, wird sie bei Berührung als Schmerz empfunden oder könnte den Nutzer sogar verletzen. Gleichzeitig beeinflusst der Nutzer wiederum die Temperatur des Materials durch seine Körperwärme.

#### 4.3.2 Smart Fluids

Smart Fluids sind Flüssigkeiten, die ihre Viskosität (Zähflüssigkeit) als Reaktion auf z.B. ein Magnetfeld oder elektrischen Strom verändern. Im Gegensatz zu temperaturaktivierten Materialien findet diese Veränderung deutlich schneller statt und ist weniger von der Umgebung abhängig. Da es sich um Flüssigkeiten handelt, kann jedoch, selbst bei der Erzeugung starker elektrostatischer oder magnetischer Felder, keine hohe maximale Steifigkeit erreicht werden [17].

#### 4.3.3 Piezoelektrische Materialien

Ähnlich wie manche Smart Fluids werden piezoelektrische Materialien durch elektrischen Stromfluss aktiviert. Sie verändern dabei aber lediglich ihre Form um wenige Millimeter [17]. Trotzdem können sie zur Darstellung von Materialeigenschaften beitragen. Sie können als Schichten oder einzelne Fäden verarbeitet werden. Ihr großer Vorteil, z.B. gegenüber SMMs, ist die Aktivierung durch Strom, der die individuelle Ansprache einzelner Stränge ermöglicht, um somit beispielsweise eine Oberfläche im mikroskopischen Bereich zu verändern. Derzeit werden diese Materialien vor allem auf ihre Eignung für haptisches Feedback auf Touchscreens getestet [27], eine Verwendung für Virtuelle Realität ist aber auch sehr gut möglich. Herausforderungen sind dabei noch die sehr hohe erforderliche Spannung und der Preis dieser Materialien [17].

## 5 Mid-Air Haptic Rendering

Mid-Air Haptic Rendering beschreibt bestimmte Interface-Methoden, die fähig sind, physikalische Kräfte und somit haptisches Feedback freistehend — in der Luft — zu erzeugen [29]. Diese Interfaces können in der Regel entweder ausschließlich taktiles oder kinästhetisches Feedback generieren, da die Erzeugung einer stärkeren Kraft in der Luft gleichzeitig zu einer breiteren Verteilung führt, wodurch taktile Informationen nicht mehr dargestellt werden können.

Die Darstellung von taktilen Informationen in der Luft ist bereits ein sehr fortgeschrittenes Feld mit kommerziellen Produkten wie Ultrahaptics' STRATOS ([38]). Die Darstellung wird dabei durch die Generierung von leichten aber präzisen Kräften erreicht, z.B. durch Ultraschall oder Luftwirbel. Kinästhetisches Feedback, welches deutlich stärkere Kräfte erfordert, kann beispielsweise durch Magnetfelder oder Luftdüsen generiert werden.

### 5.0.1 Ultraschall

Die verbreitetste Methode für taktiles Feedback in der Luft ist das Ultraschall Rendering, bei dem durch eine Vielzahl an Lautsprechern akustische Wellen im Ultraschallbereich ausgesandt werden. Diese können so gerichtet werden, dass sie präzise an bestimmten Punkten interferieren, um so einen empfindbaren Druck zu generieren [8]. Um dadurch eine Oberflächentextur zu simulieren, kann auf mehrere Punkte zur selben Zeit Druck ausgeübt werden oder die Druckpunkte zyklisch gewechselt werden — wie Freeman et al. ([12]) beweisen, ist die mögliche Geschwindigkeit des Wechsels dabei hoch genug, dass die sequenziell aktivierten Punkte als simultan existierend empfunden werden. Weiterhin zeigen Freeman et al. die hohe Variationsvielfalt, die mit solch einem System möglich ist, da man sowohl Intensität und Wechselgeschwindigkeit (zwischen den Punkten), als auch die Form und Frequenz der Wellen selbst manipulieren kann, was wiederum Unterschiede in der taktilen Empfindung hervorruft.

### 5.0.2 Luftwirbel

Das Generieren von Luftwirbeln (engl.: air vortex) ermöglicht ebenfalls taktiles Feedback. Ein Luftwirbel oder Luftvortex entsteht, wenn Luft impulsartig aus einem zylinderähnlichen Objekt (z.B. einem Rohr) gepresst wird. Dabei entsteht ein Ring, der seine Form und kinetische Energie über längere Distanzen beibehalten kann. Treffen mehrere dieser Ringe gleichzeitig auf die Hand des Nutzers, entsteht der Eindruck einer Berührung [32]. Die mögliche Auflösung ist jedoch sehr gering und die Latenz von der Generierung bis zum Auftreffen auf der Handfläche recht hoch [15].

## 5.0.3 Magnetismus

Magnetisches Rendering ermöglicht die Darstellung von volumetrischen Objekten in der Luft durch die Erzeugung eines Magnetfelds. Durch eine Reihe von Elektromagneten kann das entstehende Feld beliebig angepasst werden, um bestimmte Formen darzustellen. Durch an die Hand des Nutzers angebrachte magnetische Scheiben kann die magnetische Abstoßungskraft ein kinästhetisches Feedback bewirken. Im Gegensatz zu Ultraschall und Luftwirbel können Magnetfelder relativ hohe Kräfte produzieren, um so den Eindruck eines festen Materials zu generieren [42]. Taktile Informationen können durch diese Methode jedoch nicht erzeugt werden. Die Natur von Magnetfeldern stellt eine weitere Herausforderung für die Materialsimulation dar: Die magnetische Abstoßungskraft wird nicht erst am Mittelpunkt spürbar, sondern nimmt langsam bei Annäherung zu. Laut Zhang et al. [42] nimmt die Kraft zum Quadrat der Distanz zum Mittelpunkt der konzentrierten Kraft zu. Somit wird, unabhängig von Stärke und Position des Magnetfelds, immer dasselbe Material wahrgenommen, da bei Annäherung bzw. Eindringen immer dieselbe Kraftsteigerung erzeugt wird. Dennoch ist die Darstellung verschiedener Materialeigenschaften möglich, indem man sich dem gleichen Prinzip wie z.B. der Encounter-Type Interfaces bedient. Durch Tracking der Hand bzw. der Finger kann die Magnetfeldstärke so variiert werden, dass es für die derzeitige Position die momentan erwartete Kraft ausübt.

### 5.0.4 Luftdüsen

Luftdüsen sind ein weiterer Ansatz für freistehendes, kinästhetisches Feedback. Durch geeignete Ausrichtung und ausreichenden Luftdruck können virtuelle Objekte bis zu einer bestimmten Steifigkeit simuliert werden. Je mehr Düsen installiert werden, desto höher können die Auflösung oder die ausgeübte Kraft werden [34]. Da die Düsen in der Regel feststehend sind, sind die möglichen Richtungen der Kraft sehr beschränkt. Tsalamlal et al. [35] entwarfen daher ein Konzept, bei dem nur eine Düse verwendet wird, die sich aber durch einen Roboterarm beliebig ausrichten lässt. Auch bei Luftdüsen besteht das Problem, dass die Kraft nicht punktuell ist, sondern auch über dem Objekt noch gespürt werden kann. Auch hier wäre also ein aktives Tracking und Anpassen notwendig. Im Vergleich zu magnetischem Rendern benötigen Luftdüsen keine zusätzlichen getragenen Geräte, können dafür Objekte jedoch nur mit geringerer Auflösung und weniger realistisch darstellen.

Name	Feedback Fähigkeiten		
	Kinästhetik Stärke	Kinästhetik Richtung	Taktil
<b>Mechanische Interfaces</b>			
Vibrationsfeedback	✓	✗	✗
Exoskelett	✓	✓	✗
Encounter-Type Interface	✓	✓	✗
Ext. Endeffektor	✓	✓	✓
<b>Shape Changing Interfaces</b>			
Pneumatik	✓	✓	✓
Granular Jamming	✓	✓	✗
Layer Jamming	✓	✓	✗
Shape Memory Materials	✓	✓	✗
Smart Fluids	✓	✓	✗
Piezoelekt. Materialien	✗	✗	✓
<b>Mid-Air Haptic Rendering</b>			
Ultraschall	✗	✗	✓
Luftwirbel	✗	✗	✓
Magnetismus	✓	✓	✗
Luftdüsen	✓	✓	✗

■ **Tabelle 1** Überblick der Fähigkeit der ausgewählten Interfaces zur Erzeugung der verschiedenen Typen von haptischem Feedback.

## 6 Überblick

In den vorherigen Kapiteln wurden verschiedene Methoden zur Darstellung haptischen Feedbacks vorgestellt und individuell evaluiert. Durch die Tabellen 1, 2 und 3, wird zusätzlich ein Gesamtüberblick über diese Methoden und ein direkter Vergleich durch bestimmte Kriterien gegeben. Tabelle 1 zeigt auf, dass nur zwei Interfaces — Pneumatik und Endeffektor Interaktion — beide Feedbackarten simulieren können. Vibrationsfeedback, welches nur die Stärke der kinästhetischen Kräfte suggerieren kann, spaltet sich ebenfalls ab. Die restlichen Interfaces können dabei klar in rein taktile (Piezoelekt. Material, Ultraschall, Luftwirbel) und rein kinästhetische (die restlichen Interfaces) unterteilt werden. In Tabelle 2 werden die Eigenschaften der Interfaces genauer benannt. Ein passiver Zustand, der kein kontinuierliches Tracking des Nutzers und Berechnung sowie Anpassung des Feedbacks benötigt, ist dabei klar der aktiven Alternative vorzuziehen. Shape Changing Interfaces haben hier einen klaren Vorteil. Die Art der Kontaktfläche kann ebenfalls eine sehr wichtige Rolle spielen. Das Vibrationsfeedback und die indirekte Interaktion durch

Name	Interface Eigenschaften		
	Zustand	Kontakt	Freie Hand
<b>Mechanische Interfaces</b>			
Vibrationsfeedback	Aktiv	Punkt	✗
Exoskelett	Aktiv	Volumen	✗
Encounter-Type Interface	Aktiv	Fläche	✓
Ext. Endeffektor	Aktiv	Punkt	✗
<b>Shape Changing Interfaces</b>			
Pneumatik	Passiv	Volumen	✓
Granular Jamming	Passiv	Volumen	✓
Layer Jamming	Passiv	Fläche	✓
Shape Memory Materials	Passiv	Volumen	✓
Smart Fluids	Passiv	Volumen	✓
Piezoelekt. Materialien	Passiv	Fläche	✓
<b>Mid-Air Haptic Rendering</b>			
Ultraschall	Passiv	Fläche	✓
Luftwirbel	Passiv	Fläche	✓
Magnetismus	Aktiv	Volumen	✗
Luftdüsen	Aktiv	Volumen	✓

■ **Tabelle 2** Eigenschaften der Interfaces. *Zustand* = Beschreibt, ob das Interface aktives Tracking der Hände und fortwährende Berechnung und Anpassung der auszubehenden Kräfte benötigt; *Kontakt* = Art der möglichen Kontaktfläche; *Freie Hand* = Keine Geräte an den Händen/Fingern.

Endeffektoren sind in dieser Hinsicht benachteiligt, da sie nur Informationen für einen Punkt zur selben Zeit liefern können. Informationen einer ganzen Fläche ermöglichen dagegen bereits mehrere DOFs (Degrees of Freedom) bei der Darstellung taktiler, aber auch kinästhetischer Informationen. Die Fähigkeit, das haptische Feedback für ein ganzes Volumen zu simulieren, ermöglicht eine sehr realitätsnahe und organische Interaktion mit dem Interface. Vor allem das Greifen und Zusammendrücken von Objekten erzeugt erst einen deutlichen Eindruck der Materialeigenschaften. Zuletzt ist noch wichtig, ob die Hände während der Nutzung des Interfaces frei bleiben können. Geräte, die man tragen oder greifen muss, verringern in der Regel den Komfort, die Immersion und oft die Freiheitsgrade der Interaktion. Tabelle 3 zeigt schließlich die Kriterien der dynamischen Veränderung der Steifigkeiten. Wichtig ist dabei die Geschwindigkeit, die relative Reichweite der Veränderung, sowie die maximal erreichbare Steifigkeit. Die Veränderung der Steifigkeit ist bei den meisten Interfaces schnell genug, dass ein Mensch keine Verzögerung bemerkt. Ausnahme sind hier vor allem die temperaturaktivierten SMMS. Die Relative

Name	Geschw.	Veränderung der Steifigkeit	
		Rel.Reichw.	Max.Steif.
<b>Mechanische Interfaces</b>			
Vibrationsfeedback	schnell	mittel	gering
Exoskelett	schnell	hoch	mittel
Encounter-Type Interface	schnell	hoch	hoch
Ext. Endeffektor	schnell	hoch	hoch
<b>Shape Changing Interfaces</b>			
Pneumatik	mittel	gering	gering
Granular Jamming	schnell	hoch	hoch
Layer Jamming	schnell	hoch	hoch
Shape Memory Materials	langsam	mittel	hoch
Smart Fluids	schnell	hoch	gering
<b>Mid-Air Haptic Rendering</b>			
Magnetismus	schnell	hoch	mittel
Luftdüsen	schnell	hoch	mittel

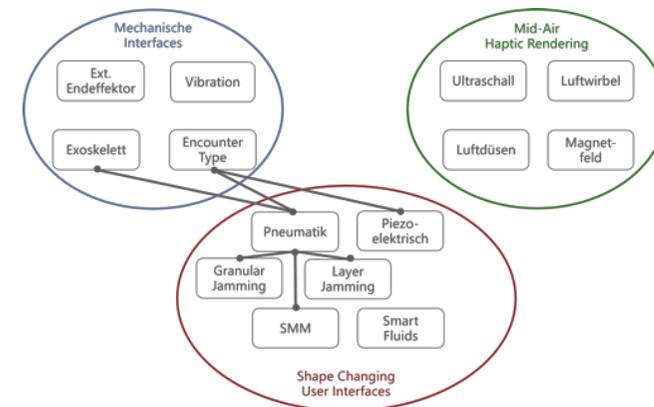
■ **Tabelle 3** Übersicht über die Kriterien des Verhaltens der Interfaces bei der dynamischen Veränderung der Steifigkeiten. *Geschw.* = Geschwindigkeit der Veränderung; *Rel.Reichw.* = Die relative Reichweite der Veränderung der Steifigkeit; *Max.Steif.* = Die maximale erreichbare Steifigkeit

Reichweite und maximale Steifigkeit spielen zusammen eine sehr wichtige Rolle um sowohl feste als auch weiche Materialien darzustellen. Endeffektor Interaktion und Encounter-Type Interfaces bestehen hier dadurch, dass sie selbst härteste Materialien wie Metalle darstellen können. Sie können aber trotzdem durch geschicktes nachgeben auch sehr weiche Objekte darstellen. Jamming Methoden haben ebenfalls eine hohe Reichweite und maximale Steifheit, eignen sich also auch für ein großes Spektrum an Materialien.

## 7 Multimodalität

Wie in Tabelle 1 zu erkennen ist, besitzen nur wenige Methoden die Möglichkeit alle Feedback-Arten darzustellen, die für eine realistische, dynamisch veränderbare Simulation von Materialeigenschaften notwendig sind. Die Methoden, die sowohl kinästhetische als auch taktile Informationen vermitteln können, kämpfen dagegen mit anderen deutlichen Schwächen. Wenn das haptische Feedback nur auf ein Element angewendet wird, welches durch Greifen zum Endeffektor des Menschen wird, können taktiles Feedback und Druckverteilungen nur interpretiert anstatt wirklich rezipiert werden (siehe Kapitel 3.4). Pneumatische Aktuation ermöglicht dagegen zwar variierbares taktiles Feed-

back, die Steifigkeit, und damit die kinästhetischen Kräfte beim Eindringen; kann dagegen aber nur sehr beschränkt angepasst werden (siehe Kapitel 4.1). Um die Nachteile und Unzulänglichkeiten einzelner Methoden aufzuheben oder zumindest zu verringern, könnten verschiedene dieser Methoden miteinander kombiniert werden. Multimodale Ansätze gibt es für Virtuelle und Erweiterte Realität sehr häufig, beispielsweise die Verbindung von Sprach- und Gestensteuerung. Abbildung 1 zeigt die besprochenen Interfacemethoden und ihre Kombinationsmöglichkeiten in einem Graphen auf. Knotenpunkte, die direkt miteinander verbunden sind, können sich gegenseitig ergänzen, um ein möglichst realistisches haptisches Feedback zu erzeugen. Im Folgenden werden die Kombinationsmöglichkeiten aller Interfaces besprochen und in Abschnitt 7.4 das vielversprechendste davon hervorgehoben.



■ **Abbildung 1** Übersicht über mögliche sinnvolle Kombinationen mehrerer haptischer User Interface Methoden. Methoden, die zusammen genutzt werden können und gemeinsam sowohl kinästhetisches als auch taktiles Feedback vermitteln können, sind mit einer Linie verbunden.

### 7.1 Mechanische Interfaces

Die mechanischen Interfaces lassen sich, mit Ausnahme des Encounter-Type Interfaces, nur sehr schwer mit anderen Methoden verbinden. Vibrationsfeedback kann zwar sehr einfach eingebunden werden, es bringt aber im Vergleich zu den meisten anderen hier vorgestellten Methoden keinen zusätzlichen Vorteil. Datenhandschuhe mit Exoskeletten würden diese Vorteile mit sich bringen. Da sie jedoch getragen werden, können viele der sinnvollen Kombinationen

(taktile Interfaces) derzeit nur sehr begrenzt implementiert werden: Es besteht nur wenig Platz zur Ausdehnung oder Einbau von Akteuren (Pneumatik), die Stromversorgung ist begrenzt (Piezoelektrik) und die kinästhetische Kraft wirkt durch direkte Berührung (Mid-Air Interfaces, siehe 7.3). Encounter-Type Interfaces können dagegen mit vielen weiteren Methoden kombiniert werden, da sie externe Systeme sind. Stanley et al. [33] zeigen die Verbindung eines Shape Changing Interface, welches über pneumatisches Granular Jamming funktioniert, mit einem Encounter-Type Interface. Die Endeffektor-Methode kann, durch die indirekte Vermittlung des haptischen Feedbacks, nicht mit anderen Methoden verbunden werden.

## 7.2 Shape Changing Interfaces

Shape Changing Interfaces lassen sich sehr leicht kombinieren. Durch die in sich geschlossene Aktivierung und den danach passiven Zustand könnten alle dieser Methoden miteinander, oder mit z.B. einem Encounter-Type Interface oder einem Exoskelett verbunden werden. Besonders interessant sind dabei jedoch Pneumatik und Piezoelektronik, da diese taktiles Feedback ermöglichen. Layer Jamming ermöglicht zusätzlich Veränderungen der Steifheit auf sehr engen Raum und kann damit auch sehr gut in verschiedenen Kombinationen verwendet werden. In Abschnitt 7.4 wird die Kombination von Layer Jamming mit pneumatischer Aktuation für taktiles Feedback genauer erläutert.

## 7.3 Mid-Air Haptic Rendering

Haptic Rendering Interfaces lassen sich schwer miteinander oder mit anderen Methoden kombinieren. Die Taktile — Ultraschall und Luftwirbel — können nicht innerhalb einer Kontaktfläche bei Berührung erzeugt werden. Sie sind somit nicht mit beispielsweise einem Encounter-Type Interface kombinierbar. Die kinästhetischen Mid-Air Interfaces — über Magnetismus oder Luftdüsen — lassen wiederum keine oberflächenverändernden Methoden wie z.B. Piezoelektrik zu, da sie aus Entfernung wirken. Übrig bliebe noch die Kombination untereinander. Der Druck, der bei Magnetismus und Luftdüsen jedoch generiert wird, trifft ebenfalls auf die Haut und würde so die Wahrnehmung der taktilen Informationen von Ultraschall und Luftwirbeln übertönen.

## 7.4 Favorit: Pneumatik und Layer Jamming

Layer Jamming und Pneumatische Aktuation ergänzen sich sehr gut. Da Layer Jamming grundsätzlich bereits das Erzeugen eines Vakuums benötigt, was durch pneumatische Aktuatoren möglich ist, können beide Methodiken durch denselben Systemaufbau implementiert werden. Yao et al. ([40]) beschreiben

einen möglichen Aufbau, bei dem mehrere verschiedene Schichten übereinandergelegt werden. Eine oder mehrere dieser Schichten sorgen dabei durch Layer Jamming für die variable Anpassung der Steifigkeit, während auf den anderen Schichten die Oberflächenstruktur durch Pneumatik — durch Luftblasen, Luftkanäle, etc. — erzeugt wird. Zusätzliche Schichten, beispielsweise zur Ermittlung von Berührungen durch leitende Metalle, können hinzugefügt werden. Diese Kombination kann somit sowohl kinästhetisches als auch taktiles Feedback vermitteln, hat eine schnelle Änderungsgeschwindigkeit, sowie hohe relative Reichweite und Maximalsteifigkeit, und kann als dünne Schicht für viele verschiedene Anwendungen verwendet werden. Eine weitere Verbindung mit Granular Jamming für Volumen oder Encounter-Type Interfaces für freie Positionierung wäre dadurch zusätzlich denkbar.

## 8 Fazit

Aus den vorangehenden Kapiteln lässt sich entnehmen, dass jede Form von haptischen Interfaces ihre individuellen Vor- und Nachteile in der Darstellung von Materialeigenschaften besitzen und eine Kombination mehrerer Interfacearten für realistisches und veränderbaren haptisches Feedback vorteilhaft ist. Shape Changing User Interfaces zeigen durch das passive und realistische Feedback eine bessere Alternative zu den herkömmlich verwendeten Systemen auf. In Verbindung mit Encounter-Type Interfaces oder Exoskeletten könnten sie weitläufig zur Materialsimulation in virtueller Realität genutzt werden. Aus der Evaluation hervorgehend wäre die Kombination von Pneumatik und Layer Jamming als vielversprechendste Variante zu werten, wenn die Simulation von realistischen und veränderbaren Materialeigenschaften gefordert ist. Mechanische Interfaces werden vermutlich jedoch weiterhin genutzt und erweitert werden, da sie außerhalb der Materialsimulation enorme Vorteile bieten können, wie Positions- und Formveränderung der virtuellen Objekte. Mid-Air Haptic Rendering befindet sich noch in den Kinderschuhen und zeigt nicht nur bei der Simulation von Material sehr große Probleme. Könnten diese Hürden in der Zukunft jedoch überwunden werden, so würde diese Art von Interface das größte Potential für virtuelle Realität darstellen, da sie dem utopischen Gedanken des Manifestierens realer Objekte aus der Luft am nächsten kommen.

### Literatur

- 1 M. Abdullah, M. Kim, W. Hassan, Y. Kuroda, and S. Jeon. Hapticdrome: An encountered-type kinesthetic haptic interface with controllable force feedback: Example of stiffness and weight rendering. In *2018 IEEE Hap-*

- tics Symposium (HAPTICS)*, pages 334–339, March 2018. doi:10.1109/HAPTICS.2018.8357197.
- 2 Musavir Bashir, Chih Fang Lee, and Parvathy Rajendran. Shape memory materials and their applications in aircraft morphing: An introspective study. *Journal of Engineering and Applied Sciences*, 12, 10 2017.
  - 3 Eric Browna, Nicholas Rodenberga, John Amendb, and Heinrich M. Jaegera. Universal robotic gripper based on the jamming of granular material. 2010.
  - 4 M. E. Cates, J. P. Wittmer, J.-P. Bouchaud, and P. Claudin. Jamming, force chains, and fragile matter. *Phys. Rev. Lett.*, 81:1841–1844, Aug 1998. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.81.1841>, doi: 10.1103/PhysRevLett.81.1841.
  - 5 N. G. Cheng, M. B. Lobovsky, S. J. Keating, A. M. Setapen, K. I. Gero, A. E. Hosoi, and K. D. Iagnemma. Design and analysis of a robust, low-cost, highly articulated manipulator enabled by jamming of granular media. In *2012 IEEE International Conference on Robotics and Automation*, pages 4328–4333, May 2012. doi:10.1109/ICRA.2012.6225373.
  - 6 Marcelo Coelho and Jamie Zigelbaum. Shape-changing interfaces. *Personal Ubiquitous Comput.*, 15(2):161–173, February 2011. URL: <http://dx.doi.org/10.1007/s00779-010-0311-y>, doi:10.1007/s00779-010-0311-y.
  - 7 Margaret Diane Rezvan Minsky. *Computational haptics : the Sandpaper system for synthesizing texture for a force-feedback display*. PhD thesis, Massachusetts Institute of Technology, 09 2005.
  - 8 Brygida Dzidek, William Frier, Adam Harwood, and Richard Hayden. Design and evaluation of mid-air haptic interactions in an augmented reality environment. In Domenico Prattichizzo, Hiroyuki Shinoda, Hong Z. Tan, Emanuele Ruffaldi, and Antonio Frisoli, editors, *Haptics: Science, Technology, and Applications*, pages 489–499, Cham, 2018. Springer International Publishing.
  - 9 J. Rembisz N. Corson H. M. Jaeger E. Steltz, A. Mozeika. Jamming as an enabling technology for soft robotics. volume 7642, pages 7642 – 7642 – 9, 2010. URL: <https://doi.org/10.1117/12.853182>, doi:10.1117/12.853182.
  - 10 Facebook Technologies, LLC. Haptic feedback. <https://developer.oculus.com/documentation/pcsdk/latest/concepts/dg-input-touch-haptic/>, 2018. Accessed: 2019-01-07.
  - 11 Sean Follmer, Daniel Leithinger, Alex Olwal, Nadia Cheng, and Hiroshi Ishii. Jamming user interfaces: Programmable particle stiffness and sensing for malleable and shape-changing devices. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*, UIST '12, pages 519–528, New York, NY, USA, 2012. ACM. URL: <http://doi.acm.org/10.1145/2380116.2380181>, doi:10.1145/2380116.2380181.
  - 12 Euan Freeman, Ross Anderson, Julie Williamson, Graham Wilson, and Stephen A. Brewster. Textured surfaces for ultrasound haptic displays. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, ICMI 2017, pages 491–492, New York, NY, USA, 2017. ACM. URL: <http://doi.acm.org/10.1145/3136755.3143020>, doi:10.1145/3136755.3143020.
  - 13 Duce Gomez, Grigore Burdea, and Noshir Langrana. Integration of the rutgers master ii in a virtual reality simulation. pages 198 – 202, 04 1995. doi:10.1109/VRATS.1995.512496.
  - 14 Horst-W. Grollius. *Grundlagen der Pneumatik*. Hanser, 3 edition, 2012.
  - 15 Happy, Pragti, and Niranjan Bhattacharyya. Rendering 3d haptic shapes in mid-air using ultrasound. *Journal of Management Engineering and Information Technology (JMEIT)*, 2, 2015.
  - 16 HaptX Inc. Haptx | haptic gloves for vr training, simulation, and design. <https://haptx.com/>, 2018. Accessed: 2019-01-07.
  - 17 Chris Harrison and Scott E. Hudson. Texture displays: a passive approach to tactile presentation. In *CHI*, 2009.
  - 18 Matthias Hoppe, Pascal Knierim, Thomas Kosch, Markus Funk, Lauren Futami, Stefan Schneegass, Niels Henze, Albrecht Schmidt, and Tonja Machulla. Vrhapticdrones: Providing haptics in virtual reality through quadcopters. In *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*, pages 7–18. ACM, 2018.
  - 19 Y. Kim, S. Cheng, S. Kim, and K. Iagnemma. A novel layer jamming mechanism with tunable stiffness capability for minimally invasive surgery. *IEEE Transactions on Robotics*, 29(4):1031–1042, Aug 2013. doi:10.1109/TRO.2013.2256313.
  - 20 Yaesol Kim, Hyun Jung Kim, and Young J. Kim. Encountered-type haptic display for large vr environment using per-plane reachability maps. *Computer Animation and Virtual Worlds*, 29(3-4):e1814, 2018. e1814 cav.1814. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/cav.1814>, arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1002/cav.1814>, doi:10.1002/cav.1814.
  - 21 Pascal Knierim, Thomas Kosch, Valentin Schwind, Markus Funk, Francisco Kiss, Stefan Schneegass, and Niels Henze. Tactile drones-providing immersive tactile feedback in virtual reality through quadcopters. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 433–436. ACM, 2017.
  - 22 Michael L. Turner, Daniel H. Gomez, Marc R. Tremblay, and Mark Cutkosky. Preliminary tests of an arm-grounded haptic feedback device in telemanipulation. 1998.
  - 23 Erik Lövquist, Zsuzsanna Kulcsár, Mikael Fernström, Annette Aboulafia, and George Shorten. The design of a haptic simulator for teaching and assessing spinal anaesthesia. 2008.
  - 24 Thomas Massie and John Kenneth Salisbury. The phantom haptic interface : A device for probing virtual objects. 1994.

- 25 Jifei Ou, Lining Yao, Daniel Tauber, Jürgen Steimle, Ryuma Niiyama, and Hiroshi Ishii. jamsheets: Thin interfaces with tunable stiffness enabled by layer jamming. In *Proceedings of the 8th International Conference on Tangible, Embedded and Embodied Interaction*, TEI '14, pages 65–72, New York, NY, USA, 2013. ACM. URL: <http://doi.acm.org/10.1145/2540930.2540971>, doi:10.1145/2540930.2540971.
- 26 Javier Posselt, Lionel Dominjon, Alexandre Bouchet, and Andras Kemeny. Toward virtual touch: investigating encounter - type haptics for perceived quality assessment in the automotive industry. In *14th annual EuroVR conference - EuroVR 2017*, Dec 2017.
- 27 Ivan Poupyrev, Shigeaki Maruyama, and Jun Rekimoto. Ambient touch: Designing tactile interfaces for handheld devices. pages 51–60, 01 2002. doi:10.1145/571985.571993.
- 28 Majken Rasmussen, Esben W. Pedersen, Marianne Petersen, and Kasper Hornbæk. Shape-changing interfaces: A review of the design space and open research questions. *Conference on Human Factors in Computing Systems - Proceedings*, 05 2012. doi:10.1145/2207676.2207781.
- 29 K. Salisbury, F. Conti, and F. Barbagli. Haptic rendering: introductory concepts. *IEEE Computer Graphics and Applications*, 24(2):24–32, March 2004. doi:10.1109/MCG.2004.1274058.
- 30 Bryan Schubert and Dario Floreano. Variable stiffness material based on rigid low-melting-point-alloy- microstructures embedded in soft poly(dimethylsiloxane) (pdms). 2013.
- 31 Jeremy Shopf and Marc Olano. Procedural haptic texture. In *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology*, UIST '06, pages 179–186, New York, NY, USA, 2006. ACM. URL: <http://doi.acm.org/10.1145/1166253.1166281>, doi:10.1145/1166253.1166281.
- 32 Rajinder Sodhi, Ivan Poupyrev, Matthew Glisson, and Ali Israr. Aireal: Interactive tactile experiences in free air. *ACM Trans. Graph.*, 32(4):134:1–134:10, July 2013. URL: <http://doi.acm.org/10.1145/2461912.2462007>, doi:10.1145/2461912.2462007.
- 33 A. A. Stanley and A. M. Okamura. Controllable surface haptics via particle jamming and pneumatics. *IEEE Transactions on Haptics*, 8(1):20–30, Jan 2015. doi:10.1109/TOH.2015.2391093.
- 34 Aishwari Talhan and Seokhee Jeon. Pneumatic actuation in haptic-enabled medical simulators: A review. *IEEE Access*, PP:1–1, 12 2017. doi:10.1109/ACCESS.2017.2787601.
- 35 M. Y. Tsalamlal, P. Issartel, N. Ouarti, and M. Ammi. Hair: Haptic feedback with a mobile air jet. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2699–2706, May 2014. doi:10.1109/ICRA.2014.6907246.
- 36 Michael L. Turner, Ryan P. Findley, Weston B. Griffin, Mark R. Cutkosky, and Daniel H. Gomez. Development and testing of a telemanipulation system with arm and hand motion. 2000.
- 37 B. Ullmer and H. Ishii. Emerging frameworks for tangible user interfaces. *IBM Systems Journal*, 39(3.4):915–931, 2000. doi:10.1147/sj.393.0915.
- 38 Ultrahaptics Limited. Touch development kit | first-generation mid-air haptics. <https://www.ultrahaptics.com/products-programs/>, 2018. Accessed: 2019-01-13.
- 39 V. Wall, R. Deimel, and O. Brock. Selective stiffening of soft actuators based on jamming. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 252–257, May 2015. doi:10.1109/ICRA.2015.7139008.
- 40 Lining Yao, Ryuma Niiyama, Jifei Ou, Sean Follmer, Clark Della Silva, and Hiroshi Ishii. Pneu: Pneumatically actuated soft composite materials for shape changing interfaces. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, pages 13–22, New York, NY, USA, 2013. ACM. URL: <http://doi.acm.org/10.1145/2501988.2502037>, doi:10.1145/2501988.2502037.
- 41 Y. Yokokohji, R. L. Hollis, and T. Kanade. Wysiwyf display: A visual/haptic interface to virtual environment. *Presence*, 8(4):412–434, Aug 1999. doi:10.1162/105474699566314.
- 42 Q. Zhang, H. Dong, and A. El Saddik. Magnetic field control for haptic display: System design and simulation. *IEEE Access*, 4:299–311, 2016. doi:10.1109/ACCESS.2016.2514978.