FACHHOCHSCHULE ERFURT UNIVERSITY OF APPLIED SCIENCES

Entwicklung eines Natural User Interface mit Hilfe moderner AR und AI Technologien

Masterarbeit

Angewandte Informatik

an der

Fachhochschule Erfurt

von

Sebastian Rieger

Abgabedatum 15.11.2017

Bearbeitungszeitraum	24 Wochen
Matrikelnummer	10286908
Betreuer der Masterarbeit	Prof. Rolf Kruse
Zweitbetreuer der Masterarbeit	Prof. Steffen Avemarg

Erklärung

Ich, Sebastian Rieger, versichere hiermit, dass ich die vorliegende Masterarbeit mit dem Thema „Entwicklung eines Natural User Interface mit Hilfe moderner AR und AI Technologien“ selbstständig und nur unter Verwendung der angegebenen Quellen und Hilfsmittel angefertigt habe.

Ort Datum

Unterschrift

Inhaltsverzeichnis

1 Abstract

2 Einleitung

In den letzten zwanzig Jahren hat sich die Bedienung von Computern grundlegend geändert. Vor nicht all zu vielen Jahren gab es nur die Möglichkeit mit Hilfe von Maus und Tastatur mit einem Computer zu interagieren.

Mit dem Aufkommen von Touch-Screens jedoch änderte sich auch die Benutzung von Computern. Es war nun möglich, direkt mit dem Bildschirm zu interagieren ohne den Umweg über die Maus.

Als dann wenig später die ersten Sprachsteuerungen auf den Markt kamen, änderte sich die Interaktion mit dem Computer erneut. So ist es nun möglich, Computern mittels Sprache Befehle zu erteilen oder Texte zu sprechen, welche automatisch transkribiert werden.

Heute ist es mit manchen Smartphones schon möglich, Nachrichten wie SMS zu schreiben und zu versenden, ohne das Telefon überhaupt in die Hand zu nehmen. Dies ist nur durch neuste Entwicklungen in der *künstliche Intelligenz* (KI) möglich.

Im selben Zeitraum hat sich parallel auch die *Virtuelle Realität* (VR) entwickelt. Sie versucht Menschen mit Hilfe von verschiedenen Brillen in eine virtuelle Welt zu versetzen. Die Entwicklungen solcher VR Systeme wurde vor allem von der Spieleindustrie getrieben, da sie immer neue Versuche unternimmt, Spieler besser in die Welt des Spiels zu versetzen.

Eine Abstufung der VR ist die *Augumented Reality* (AR), welche versucht die analoge, reale Welt um digitale Inhalte zu erweitern. Hierbei sind die Möglichkeiten für den Einsatz von AR fast unbegrenzt. Es ist also quasi möglich, jede menschliche Tätigkeit durch die AR zu unterstützen.

Genau hier setzt der Schwerpunkt der Arbeit an. Im Verlauf soll versucht werden, ein *Natural User Interface* (NUI) unter Zuhilfenahme moderner AR und KI Technologien zu erstellen.

Hierfür werden aktuelle AR und KI Systeme daraufhin untersucht, wie sie im Zusammenspiel ein NUI bilden können, welches allein durch Sprache und Gesten mit einem Computer interagiert.

Es soll eine übliche Tätigkeit mit Hilfe einer AR Technologie in der realen Welt abgebildet werden, welche dann von einem Menschen nur unter Zuhilfenahme von Sprache und Gesten gesteuert und bearbeitet werden kann.

3 Natural User Interfaces

Als Natural User Interfaces NUI werden Benutzeroberflächen von Computern und Programmen bezeichnet, welche sich durch Gesten, Sprache, Berühren, Tippen und Wischen bedienen lassen. In der heutigen Zeit begegnen wir ständig solchen NUIs. Zum Beispiel beim mehrmals täglichen Griff zum Smartphone oder der Smartwatch. Über einen berührungsempfindlichen Display, werden Apps heute schon über Gesten gesteuert. [DP15]

Diese Steuerung funktioniert zum Teil bei der jüngeren Generation, so genannten „digital natives“, also Menschen die im digitalen Zeitalter geboren sind, schon automatisch und unterbewusst. Jeder, der schon einmal eine App bedient hat kennt die Geste zum löschen aus einer Liste! Die Bewegung des Eintrags nach links oder rechts löscht das entsprechende Element, getreu nach dem Motto „Aus den Augen aus dem Sinn“. [Dig17]

Diese Art von Gesten ist so intuitiv für Menschen, das auch „digital immigrants“, also Menschen, die erst im Erwachsenenalter den Umgang mit digitalen Geräten gelernt haben diese nach einmaliger Erklärung gelernt haben.

Aber das natürliche Verhalten von uns Menschen geht noch viel weiter. Diese einmal auf einem Smartphone gelernte Geste in einer App, wird unterbewusst auch auf andere Applikationen übertragen. Eine App, welche diese einfache und mittlerweile etablierte Geste nicht unterstützt wird umgehend Kritik ernten, da sie mit altbekanntem bricht und so ein „Natürliches“ arbeiten nicht mehr gegeben ist.

3.1 Natürliche Bedienung durch Gesten

Was jedoch genau macht eine Geste „Natürlich“? Die Beantwortung dieser Frage ist im gleichen Maße Trivial, wie Komplex.

Dies bedeutet, sobald einmal eine „Natürliche Geste“ gefunden wurde, erscheint wie gegeben und wird nicht in Frage gestellt, weil sie ja „Logisch“ ist. Das Komplexe ist jedoch solche „logischen Gesten„ erst einmal zu finden.

Hierfür müssen viele Aspekte aus den Verhaltenswissenschaften beachtet werden:

- Arbeitswissenschaft
- Psychologie
- Soziologie
- Pädagogik

Ergänzend zu diesen Aspekten gibt es die EN ISO 9241, welche Richtlinien der Mensch-Computer-Interaktion beschreibt. Ins besondere der Teil 110 “Grundsätze der Dialoggestaltung„ kann Helfen passende Bedienmöglichkeiten für ein NUI zu finden. [ISO11] Die Grundsätze sind:

- Aufgabenangemessenheit - geeignete Funktionalität, Minimierung unnötiger Interaktionen
- Selbstbeschreibungsfähigkeit - Verständlichkeit durch Hilfen / Rückmeldungen
- Steuerbarkeit - Steuerung des Dialogs durch den Benutzer
- Erwartungskonformität - Konsistenz, Anpassung an das Benutzermodell
- Fehlertoleranz - unerkannte Fehler verhindern nicht das Benutzerziel, erkannte Fehler sind leicht zu korrigieren

- Individualisierbarkeit - Anpassbarkeit an Benutzer und Arbeitskontext
- Lernförderlichkeit - Minimierung der Erlernzeit, Metaphern, Anleitung des Benutzers

3.2 Erstellung eines Use Cases

Anhand dieser Grundsätze kann nicht nur ein System, welches mit Tastatur und Maus bedient wird definiert werden, sondern auch ein NUI. Dies wird nun Beispielhaft an der Geste für "löschen„ gezeigt um sicher zu Stellen, dass im späteren Verlauf der Arbeit ein NUI anhand genau dieser Grundsätze definiert werden kann.

Aufgabenangemessenheit Das Wischen zum Löschen eines Eintrags ist schnell und einfach zu erlernen. Ebenso kann die Geste schnell wiederholt werden um möglichst schnell auch mehrere Elemente löschen zu können.

Selbstbeschreibungsfähigkeit Die Geste ist zwar einfach und mittlerweile auch bekannt, aber es muss auch sichergestellt sein, dass neue Nutzer die Geste lernen können. Dies kann zum Beispiel durch ein kurzes wackeln beim ersten öffnen angezeigt werden. Dem Nutzer wird durch diese Bewegung suggeriert, dass diese Elemente bewegbar sind.

Steuerbarkeit Wischt der Nutzer nun bewusst oder durch Neugier ausgelöst vom Wackeln das Element aus dem Bildschirm muss immer die Möglichkeit bestehen das Element wiederherstellen zu können. Dies ist nur konsequent, da eine so einfache Geste auch einmal versehentlich gemacht werden kann.

Erwartungskonformität Ist diese Geste einmal etabliert, muss sie konsequent in allen Listen implementiert sein, da es sonst zum Bruch der Erwartung kommt und der Nutzer unnötiger verwirrt wird.

Fehlertoleranz Zu Fehlertoleranz gehört das wiederherstellen eines Eintrags wie schon unter Steuerbarkeit beschrieben. Zusätzlich kommt jedoch hinzu, das die Geste eventuell vom Gerät falsch interpretiert wurden ist und der Nutzer gar nichts löschen wollte. Auch in solchen fällen muss eine Wiederherstellung möglich sein.

Individualisierbarkeit Das wischen ist eine tolle Geste! Aber in welche Richtung? Nach links oder rechts wischen zum löschen? Genau hier muss die App schlau reagieren, denn die richtige Antwort ist: Beides muss möglich sein. Bekanntlicher weise gibt es Links- und Rechtshänder, wobei das Gerät zumeist mit der dominanten Hand gehalten wird. Für einen Rechtshänder welcher den Daumen zum wischen links hat, ist das Wischen nach rechts leichter als nach links. Genau umgekehrt ist es bei Linkshändern. Diese wischen eher nach Links.

Dieses Verhalten kann auch selbst nachgestellt werden, egal ob der Nutzer Links- oder Rechtshänder ist. Die Geste wird unterbewusst durch beide Hände in unterschiedliche Richtungen ausgeführt.

Lernförderlichkeit Bei der Lernförderlichkeit trifft ganz klar das Sprichwort "Aus den Augen aus dem Sinn„ zu. Der Nutzer möchte etwas "weg„ haben. Warum es also nicht außerhalb des Displays platzieren. Es wird nicht mehr gesehen und ist somit "weg„. Diese einfache Analogie ist für jeden Nutzer zu verstehen und umzusetzen.

3.3 Auswertung des Use Cases

Anhand des vorgestellten Use Cases wurde gezeigt, dass es möglich ist ein NUI zu designen. Hierbei müssen natürlich die Gegebenheiten des jeweiligen Gerätes beachtet werden, wie zum Beispiel das ein Smartphone in beiden Händen gehalten werden kann. Dadurch kann sich gegebenen Falls der Use Case erweitern oder verändern.

1991 beschrieb Mark Weiser schon, wie sich Computer immer mehr in unseren Alltag integrieren und langsam immer unkenntlicher werden. Die Grenze zwischen realer Umwelt und virtueller Welt verschwimmt heute immer mehr. [Wei91]

Dies hat natürlich auch zur Folge, das sich die Schnittstelle zwischen Mensch und Computer ständig ändert und weiterentwickelt. Musste man 1997 zu beginn von Google noch Stichworte mit Maus und Tastatur eingeben um das Internet zu durchsuchen, reicht heute schon ein Sprachbefehl.

Der Computer wertet mit verschiedenen Algorithmen das gesprochene aus, und versucht dem Nutzer eine direkte Antwort zu geben, während man 1997 von Google eine Liste mit Webseiten bekam, auf welchen die Schlagworte zu finden waren.

4 Augmented Reality

In der heutigen Zeit ist fast jedem der Begriff Augmented Reality geläufig. Jedoch gibt es im wissenschaftlichen Umfeld keine einheitliche Definition. Georg Klein definiert die AR als „Anreicherung der realen Welt um computergenerierte Zusatzobjekte“. [Kle]

Viele Abhandlungen zu diesem Thema beziehen sich auf das „Reality-Virtuality Continuum“, welches von Milgram, Takemura, Utsumi und Kishino beschrieben wird (Abbildung 1) Dieses stellt bildlich dar, wie sich eine AR-Anwendung einordnen lässt. Links ist die reale Umgebung (Real Environment) und rechts die Virtuelle Umgebung (Virtual Environment) zu sehen.

Heutige VR-Anwendungen müssen also ganz rechts eingeordnet werden. Zwischen der realen und der virtuellen Umgebung gibt es jedoch noch weitere Abstufungen. So gibt es die „Augmented Reality“, welche die reale Welt um virtuelle Elemente ergänzt. Sie bezieht sich eher auf die reale Umgebung, weshalb sie weiter links angeordnet ist. Die „Augmented Virtuality“ ist das genaue Gegenteil der AR. Hier wird die virtuelle Welt des Computers um reale Gegenstände ergänzt.

Im wissenschaftlichen Umfeld wird als Oberbegriff für „Augmented Reality“ und „Augmented Virtuality“ oft „Mixed Reality“ oder „Enhanced Reality“ verwendet. [MBRS11]

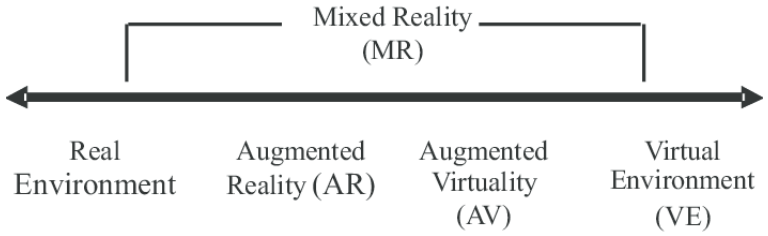

Abbildung 1: Reality-Virtuality Continuum [MTUK]

Der Begriff der *Augmented Virtuality* (AV) ist heute weniger gebräuchlich, da es keine wirklichen Anwendungsfelder gibt, in denen eine virtuelle Welt um reale Gegenstände ergänzt werden muss. Im Umkehrschluss wird heute jedoch die AR immer bedeutender, da bei fast jeder menschlichen Tätigkeit eine Unterstützung durch Computergenerierte Objekte oder Informationen möglich ist.

Im medizinischen Umfeld können Ärzten zum Beispiel Positionen von empfindlichen Gewebe oder zu entfernenden Fremdkörpern bei einer Operation angezeigt werden. Eine andere Einsatzmöglichkeit wäre das unterstützende Anzeigen von Gefahren oder Fluchtwegen für Feuerwehrleute, die sich in einem stark verrauchten Brandobjekt befinden. Computer können durch Infrarot und Wärmebild das Sichtvermögen eines Feuerwehrmannes in einer Gefahrensituation so entscheidend erhöhen, um sich selbst und andere zu retten.

Ein Gerät, welches später für die Erstellung eines NUI dienen kann, muss folgende Kriterien erfüllen:

- Das Gerät muss Freihändig bedienbar sein.
- Das Gerät muss die AR-Inhalte exakt zur realen Welt wiedergeben.
- Das Gerät muss Handgesten und Sprache des Nutzers erkennen können.

4.1 Typen von visuellen Ausgabegeräten

Im folgenden Abschnitt werden mögliche visuelle Ausgabegerätetypen genauer beschrieben und auf ihre Einsatzmöglichkeiten untersucht.

4.1.1 Handheld-Geräte

Als Handheld werden in Geräte bezeichnet, die wie der Name schon sagt von einer Person in der Hand gehalten werden. Im AR-Umfeld sind hier Smartphones oder Tablets gemeint, welche heutzutage durch ihre Dual- oder Quad-Core-Prozessoren und teilweise mehrere Gigabyte RAM in der Lage sind, auch komplexere AR-Anwendungen zur Ausführung zu bringen.

In Smartphones und Tablets nimmt die Kamera das Bild der realen Welt auf und stellt es in Echtzeit auf dem Display dar. Ein Programm (App) ist dann in der Lage, die aufgenommene reale Welt um virtuelle Elemente zu erweitern und diese innerhalb des Kamerabildes auf dem Bildschirm zu platzieren.

Hierbei ergibt sich jedoch ein Problem, da sich der Blickwinkel es Betrachters vom Blickwinkel der Kamera unterscheidet, was zu einer unsauberen Platzierung der Objekte im Bild führt. In Abbildung 2 ist zu sehen, wie sich der Blickwinkel der Kamera und der Blickwinkel des Betrachters unterscheiden.

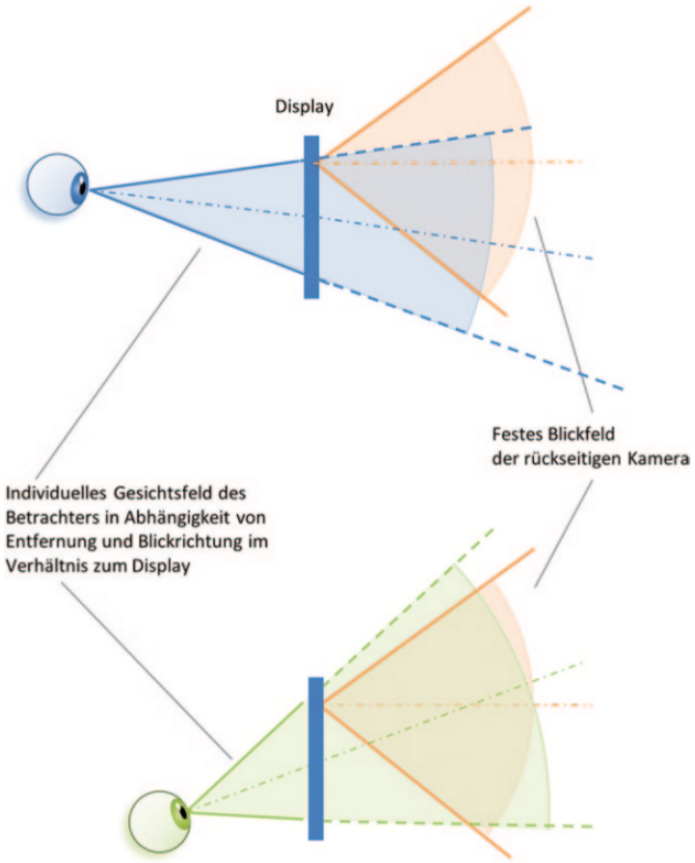

Abbildung 2: Unterschiedliche Blickfelder von Betrachter und Kamera bei Handheld-AR [DBGJ13]

Um den Unterschied zwischen den Blickwinkeln auszugleichen, muss das entsprechende Gerät die beiden Blickwinkel aufeinander abstimmen. Dies kann jedoch kein zur Zeit auf dem Markt befindliches Gerät leisten.

Einen ersten Ansatz hierzu hat die University of California gemacht, indem eine Gruppe Forscher ein Tablet mit einem Kinect Sensor verbunden haben. Mit Hilfe des zusätzlichen Sensors und dem „KinectFusion Algorithmus“ konnten sie eine „Magic Lens“ entwickeln, die beide Blickwinkel in etwa aufeinander abstimmt. [BLT+12]

In Abbildung 3 ist einmal der Unterschied zwischen einer Aufnahme mit Magic Lens und ohne schematisch dargestellt.

Im linken Bildabschnitt ist gut zu sehen, dass Bäume und andere Gegenstände im Bildschirm exakt an der Position in der realen Welt liegen. Im Gegensatz dazu sind Bäume und andere Gegenstände im rechten Bildabschnitt merklich verzerrt.

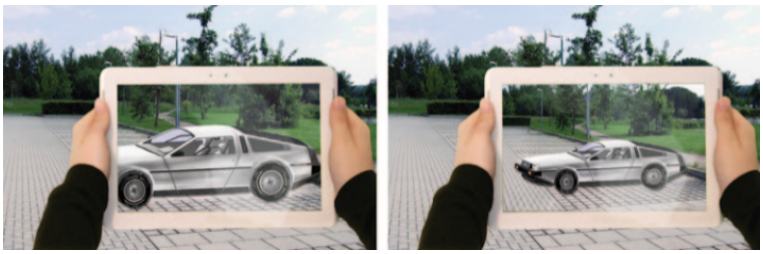

Abbildung 3: Schematische Darstellung des Magic Lens Effekts [DBGJ13]

Ein weiteres Problem ist, das AR-Anwendungen für die optimale Skalierung und Platzierung von virtuellen Objekten im Raum zusätzlich Tiefeninformationen des aktuellen Bildes benötigen, denn nur dann ist auch gewährleistet, dass virtuelle Objekte in der richten Größe im realen Raum positioniert werde.

Dieses Problem versucht gerade das Projekt Tango von Google zu lösen. Mit Hilfe eines Lasers und einer Infrarotkamera können Tiefeninformationen zum aktuellen Bild aufgenommen und verarbeitet werden. [Goo17]

Nur mit der Tango Technologie ist es bisher möglich, AR-Objekte richtig skaliert und im richtigen Blickwinkel präzise im Raum zu positionieren. Weitergehende Informationen sind auf den Developer-Seiten[1] von Google zu finden. Bisher unterstützen jedoch nur das „Lenovo Phab2“ und das „Asus ZenFon AR“ diese Technologie, da hier eine zusätzliche Sensoreinheit im Smartphone verbaut sein muss.

Abschließen kann über AR-Anwendungen ausgesagt werden, dass sie grundsätzlich lauffähig auf Handheld Geräten wie Tablets oder Smartphones sind. Die Bedingungen sind heutzutage jedoch noch nicht optimal, wie eben dargelegt wurde.

4.1.2 Video-See-Through-Displays

Video-See-Through Geräte existieren im Vergleich zu AR-Anwendungen schon ziemlich lange. Schon Anfang der 1940er Jahre wurden erste Head-Up-Displays in Kampfflugzeugen eingesetzt, um die Piloten mit zusätzlichen Informationen zu versorgen, welche immer im Blick behalten werden müssen. [Wik17a]

Von Anfang an, wurden die zusätzlichen Informationen auf eine Glasfläche projiziert, um das Sichtfeld nicht unnötig einzuschränken. Es wurde entweder direkt auf das Cockpit Fenster projiziert oder die Piloten hatten kleine durchsichtige Displays in den Helm eingearbeitet, auf die mit Hilfe von Spiegeln projiziert werden konnte.

Mit der Veröffentlichung des iPhones im Jahre 2007 wurde das Thema AR immer interessanter, da mit ihm die technische Grundlage gelegt wurde. 2008 erschien der AR-Browser Wikitude wodurch das Feld immer mehr Fuß fasste und sich nun rasant weiter entwickelt. [Jöc14]

Die durch die immer größerer Verbreitung von AR-Anwendungen wurde auch die Hardware entsprechend weiterentwickelt und angepasst, was zu den ersten AR-Brillen führte.

Durch die Kombination von Handheld-Geräten und Head-Up-Displays entstanden bis 2014 die ersten Video-See-Through-Brillen oder auch AR-Brillen genannt. Die Entwicklung dieser Brillen

[1] `https://get.google.com/tango/`

steckt noch in den Kinderschuhen, aber dennoch gibt es schon zwei auf dem Markt verfügbare Geräte, die für die spätere Entwicklung eines NUI in Betracht gezogen werden können. Zum einen ist das die Epson Moverio BT-200 und zum anderen die Microsoft HoloLens, welche zur Zeit nur für Entwickler verfügbar ist.

Die HoloLens und die Moverio sind beide Prismen basiert. In Abbildung 4 ist die Funktionsweise von Prismen basierten See-Through-Displays schematisch Dargestellt. Ein normales Display ist an einer Seite eines Prismas angebracht. Das Prisma leitet das Licht des Displays weiter, so dass es in das Auge des Betrachters weitergeleitet wird. Weil das Prisma nach vorn durchsichtig ist, wird auch das Umgebungslicht direkt in das Auge des Betrachters geleitet. Innerhalb des Prismas vermischen sich die beiden Bilder, wobei das Displaylicht das Umgebungslicht überlagert. Für den Betrachter sieht es somit aus, als würden die Objekte in der realen Welt zu finden sein.

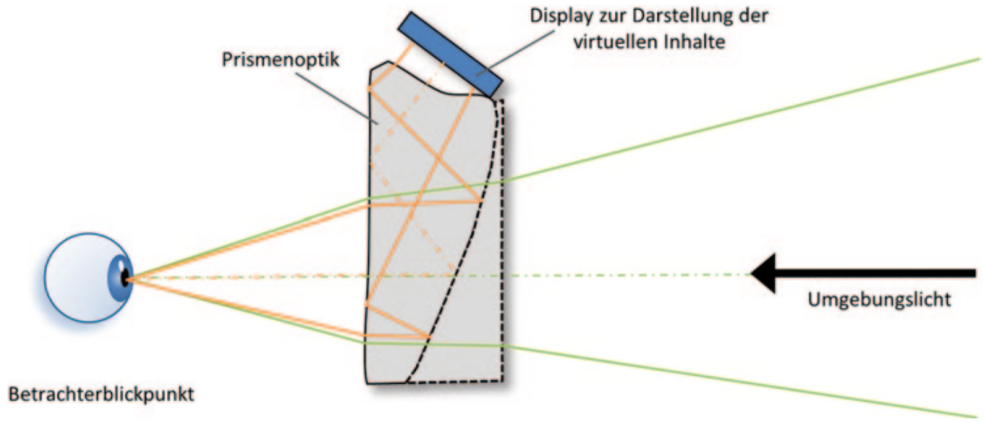

Abbildung 4: Schematische Darstellung des Aufbaus eines See-Through-Displays [DBGJ13]

Die Microsoft HoloLens wurde 2015 von Microsoft vorgestellt und ist momentan die Wahrscheinlich beste AR-Brille auf dem Markt. Sie verfügt über mehrere Kameras und Tiefensensoren, womit die HoloLens in der Lage ist, sich selbstständig im Raum zu orientieren.

Ausgestattet mit einem Intel Atom x5-Z810 Prozessor und 2 Gigabyte RAM, ist die Brille sehr gut aufgestellt und kann Programme und Berechnungen ohne zusätzliche Hardware ausführen. Als Betriebssystem kommt Windows 10 zum Einsatz. Die Akkulaufzeit wird mit zwei Stunden angegeben, wobei dies wie immer abhängig von der geforderten Leistung ist. [ARV17]

Abbildung 5: Aufgesetzte HoloLens [Wik17e]

Für Entwickler hat Microsoft ein *Software Development Kit* (SDK) bereitgestellt mit dessen Hilfe es möglich ist Anwendungen für die HoloLens zu entwickeln. [Hol17]

Über verschiedene Sensoren kann die HoloLens Bewegungen und Gesten des Nutzers wahrnehmen, erkennen und verarbeiten.

Die HoloLens ist standardmäßig in der Lage, grundlegende Gesten zu erkennen, mit denen zum Beispiel ein Tippen, Halten, Rotieren und Zoomen möglich ist. [Hol17] Auf genauere technische Details zur Programmierung der HoloLens wird im Kapitel ????? genauer eingegangen.

Eine Besonderheit der HoloLens ist natürlich das mit Microsoft ein großer Softwarekonzern hinter der Entwicklung steht und somit ein schnelles voranschreiten der Technik quasi sicher ist.

Die Epson Moverio BT-200 eine weitere AR-Brille ist die sogenannte Epson Moverio BT-200, welche auf Android-Basis arbeitet. Im Gegensatz zur HoloLens, kann die Moverio wie eine normale Brille getragen werden und sieht auch so aus.

Intern sorgt ein nicht genauer benannter 1,2 GHz Dual-Core-Prozessor für die nötige Rechenleistung. Die Brille verfügt über 8 Gigabyte Speicher, der sich mittels SD-Karten erweitern lässt.

Ähnlich wie die HoloLens verfügt auch die Moverio über eine Tiefenbildkamera, welche es ermöglicht AR-Inhalte exakt zu platzieren. Ein Beschleunigungssensor registriert zusätzlich die Bewegungen des Kopfes des Trägers.

Abbildung 6: Epson Moverio mit Controller [VRS16]

Gesteuert wird die Brille über ein Controller-Kästchen, welche im Lieferumfang enthalten ist. Dieses verfügt zur Steuerung verschiedene Button wie Home oder Zurück, welche auch von Android bekannt sind. Zusätzlich kann das System über einen eingebauten Touchscreen im Controller Gesten des Nutzers erkennen und interpretieren.

Der Akku soll 6 Stunden durchhalten, was ein längeres Vergnügen verspricht, als es bei der HoloLens gegeben ist.

Der „Google Play"-Store ist auf der Brille nicht verfügbar, jedoch portiert Epson einige Anwendungen damit diese auf der Brille lauffähig sind.

4.2 Auswertung der Möglichkeiten für ein Natural User Interface

Für die Erstellung eines NUI gibt es frontendseitig zwei Möglichkeiten. Zum einen gibt es die Handheld-Geräte, zum anderen gibt es noch die AR-Brillen.

Als Handheld-Gerät kann heute wie dargelegt fast schon jedes Smartphones oder Tablet verwendet werden. Ohne eine eingebaute Magic-Lens 4.1.1 oder die Google Tango Technologie, ist das präzise arbeiten mit solchen Geräten nicht möglich. Hinzu kommt noch, dass der Benutzer des Interfaces immer mit einer Hand das entsprechende Gerät halten muss, wodurch diese dann nicht mehr zum Arbeiten zur Verfügung steht. In der Tabelle 1 wurden alle Handheld Geräte zusammen betrachtet, da es hier praktisch keine Unterscheide gibt.

Anforderung	Priorisierung	Handheld Geräte	HoloLens	Moverio BT-200
Freihändig bedienbar	sehr hoch	X	✓	X
exakte AR-Inhalte	hoch	X	✓	✓
Handgestenerkennung	sehr hoch	✓	✓	✓
Spracherkennung	sehr hoch	✓	✓	✓

Tabelle 1: Tabellarischer Vergleich der betrachteten Geräte

Das ständige halten eines Smartphones oder Tablets ist auch anstrengend für den Nutzer und seine Armmuskulatur.

Handheld-Geräte sind eine gute und auch einfache Lösung um Filme zu schauen oder Spiele zu spielen. Eine Langzeitnutzung wie in einem 8 stündigen Arbeitstag steht jedoch außer Frage.

Hier kommen die AR-Brillen ins Spiel. Sie ermöglichen eine Anreicherung der realen Welt ohne dass ein Nutzer permanent ein Gerät halten muss, da die Brillen durch die natürliche Kopfform halten.

Die Epson Moverio Brille ist ein erster Ansatz um AR-Brillen einer breiten Masse zur Verfügung zu stellen. Jedoch kann sie durch viele Nachteile nicht mit der HoloLens mithalten. Im Gegensatz zur Moverio kann sich die HoloLens selbstständig im Raum orientieren.

Ein weiterer K.O.-Punkt ist, dass die HoloLens Handgesten des Nutzers direkt erkennt, ohne wie bei der Moverio auf ein zusätzlichen Controller zurück greifen zu müssen.

Im Vergleich zu Handheld-Geräten ist die Moverio nicht besser, da auch ständig ein Gerät zur Steuerung in der Hand des Nutzers verbleiben muss und somit nicht frei für andere Aufgaben ist.

Die Microsoft HoloLens ist bisher die einzige Möglichkeit um ein NUI zu entwickeln, wie es die Arbeit vorsieht. Dies geht auch aus Tabelle 1 hervor. Sie kann völlig freihändig gesteuert werden, ohne das der Nutzer zusätzliche Hardware zum Steuern oder Anzeigen in der Hand halten muss.

Im Verlauf der Arbeit wird durch die klare Überlegenheit der HoloLens diese als einzige weiter betrachtet und für die Entwicklung verwendet. Alle anderen Alternativen konnten nicht überzeugen.

5 Artificial Intelligence

Als *Artificial Intelligence* (AI) oder im deutschen auch *artifizielle Intelligenz* oder *künstliche Intelligenz*, bezeichnet, in der Informatik, ein System welches eine Automatisierte Intelligenz besitzt. Eine eindeutige Definition für AI gibt es nicht, da es schon in der Psychologie an einer genauen Definition mangelt. [Wik17b]

In der heutigen Zeit werden Systeme und Algorithmen die auf neuronalen Netzen basieren als annähernd intelligent bezeichnet. Ein Beispiel hierfür ist „Google Translate“ welches zwischen verschiedenen Sprachen mittels neuronalen Netz übersetzt. [Mer16]

Es gibt aber auch intelligente Systeme wie „IBM Watson“, welche nicht auf neuronalen Netzen basieren und annähernd intelligent wirken. Diese Intelligenz ist das Resultat einer sehr großen Wissensbasis und verschiedenen kombinatorischen Algorithmen. Um zu zeigen wie Leistungsfähig Watson ist, trat er 2011 in der Quizshow „Jeopardy“ an und gewann gegen die weltbesten Spieler. [IBM17]

5.1 Sprachassistenten

Im allgemeinen Sprachgebrauch werden als AI die heute üblichen Sprachassistenten wie Siri, Alexa, Cortana oder Google Assistent bezeichnet. Das im Verlauf der Arbeit zu entwickelnde NUI soll auf einen dieser Assistenten aufbauen.

Die Entwicklung einer eigenen Spracherkennung ist im Rahmen dieser Arbeit nicht möglich, da dies sehr viel Zeit und Arbeit in Anspruch nimmt. Selbst Firmen wie Google oder Amazon benötigten mehrere Jahre um ihre Assistent auf das aktuelle Leistungsniveau zu heben.

5.2 Funktionsweise der Sprachassistenten

Auch wenn die genauen Implementierungen der Sprachassistenten verschiedenen ist, so haben sie alle eins gemenisam und zwar ist dies der Weg der Sprachverarbeitung.

Die Funktionsweise von Siri udn Co. ist recht einfach gehalten. Der Nutzer löst mit einem Stichwort, „Hey, Siri...“, „Ok, Google...“ (Abbildung 7 1.)und so weiter, die Benutzung aus. Der darauf folgende Sprachbefehl wird aufgezeichnet und komprimiert an einen Server weitergeleitet (Abbildung 7 2.), welcher den gesprochenen Text auswertet und in schriftliche Befehle umwandelt (Abbildung 7 3.). Diese Befehle werden dann umgehend zurück an das Endgerät des Nutzers Versand (Abbildung 7 4.), welches nun auf die Befehle reagieren kann (Abbildung 7 5.). In Abbildung 7 ist dieses vorgehen noch einmal schematisch Dargestellt. [Ebe17]

?????Wie sieht es besser aus? Verweis auf Abb. oder einzelnen Punkte angeben? oder punkte ganz weglassen? ?????

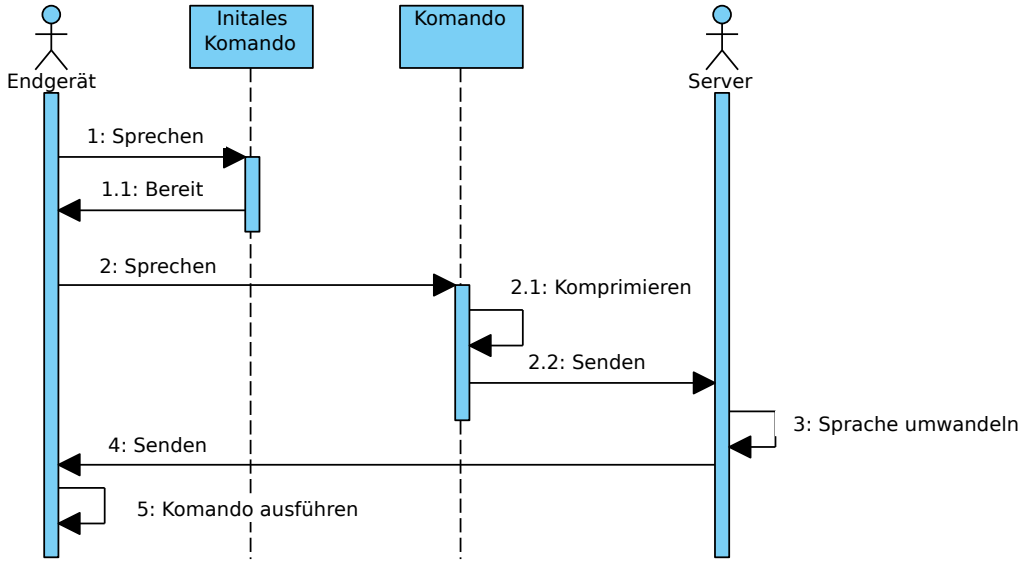

Abbildung 7: Schematische Darstellung der Kommunikation von Sprachassistenten

Hierbei ist zu beachten, dass jegliche Kommunikation zwischen Endgerät und Server bei allen Assistenten über HTTPS gesichert ist.

Innerhalb der Server, welche die Sprache zu Text wandeln, kommen fast immer neuronale Netze zum einsatz, da nur sie in der Lage sind schnell und eindeutig die Spachbefehle zu wandeln. Durch die gehäufte Verwendung mit immer neuen Stimmen und Befehlen lernen die Systeme immer mehr hinzu und können so die Anfragen immer schneller und besser umwandeln. Dies hat zur Folge, das ein System besser wird, je häufiger es genutzt wird.

In den nun folgenden Abschnitten werden die genannten Assistenten auf ihre Möglichkeiten untersucht.

5.3 Untersuchung der Sprachassistenten

Es soll nun untersucht werden, in wie weit die heute gängigen Assistenten genutzt werden können, um eine Sprachsteuerung für ein NUI (siehe Kapitel ?????) umzusetzen.

Hierbei muss der Assistent folgende Kriterien erfüllen:

- Lauffägig auf der Microsoft HoloLens
- Offene Programmierschnittstelle (SDK)
- Möglichkeit eigener Sprachkomandos

5.3.1 Siri

Der wohl älteste und bekannteste Sprachassistent ist „Siri“. Siri ist eine Software zur Spracherkennung und wurde von Apple im Jahre 2010 zusammen mit der Firma Siri Inc. von Apple aufgekauft. Schon ein Jahr später wurde der Assistent im Zuge der Veröffentlichung des iPhone 4s vorgestellt und veröffentlicht.

Siri ist eine proprietäre Software und nur iOS, machOS, watchOS und tvOS einsetzbar, weshalb eine Verwendung auf der HoloLens nicht möglich ist. Zwar verfügt Siri über ein offenes SDK, dieses ist jedoch nur für Apple eigene Betriebssysteme ausgelegt. [Sir17a] [Sir17b]

5.3.2 Alexa

Alexa ist der Name des Sprachassistenten, welcher von Amazon entwickelt wird und im November 2014 veröffentlicht wurde. Seither entwickelt Amazon seinen digitalen Assistenten stetig weiter. [Wik17c]

Grundsätzlich funktioniert auch Alexa wie schon im Abschnitt ????? beschrieben. Alexa arbeitet mit sogenannten Skills. Diese ermöglichen es Entwicklern den Assistent mit neuen Funktionen zu erweitern. Will ein Nutzer diesen Skill verwenden, muss er ihn zunächst innerhalb seines persönlichen Assistenten installieren. Danach beherrscht Alexa die entsprechenden Kommandos des Skills. [Ama17b]

Die eigentlichen Skills, also kleine Programme, welche innerhalb der Amazon Cloud ausgeführt und durch den Assistenten gestartet werden können entweder in Java oder Node.js geschrieben werden. Um diese später zu veröffentlichen muss der Entwickler über ein Amazon Developer Account verfügen.

Amazon verfolgt zur Zeit eine große Marketing-Offensive, welche für verschiedene Produkte der Amazon Echo Serie, welche Alexa integriert haben. Zusätzlich ist es möglich Alexa auch auf eigenen Geräten verfügbar zu machen. Hierfür wird das *Amazon Voice Service* (AVS) benötigt. Dieses SDK ermöglicht es den Amazon Assistent auch auf Geräten verfügbar zu machen, welche nicht von Amazon hergestellt wurden. [Ama17a]

Mit den Skills und dem AVS-SDK, welches C++ basiert ist, ist es mögliche alle Geräte die ein Mikrofon besitzen mit dem Amazon Assistenten auszustatten.

Ein weiterer Vorteil von Alexa sind die sogenannten Speechcons. Diese werden speziell von Amazon zur Verfügung gestellt, um eine möglichst natürliche und vor allem richtig betonte Aussprache zu gewährleisten. Speechcons sind kurze Phrasen, welche in der Regel Umgangssprachlich genutzt werden, wie zum Beispiel der Ausspruch „ich glaub mein Schwein pfeift“ oder „Pustekuchen“. Zusätzlich steht mit „Whispers“ auch ein Modus bereit, welcher Alexa flüstern lässt. [Pak17] [Ale17]

All dies, macht Alexa zumindest für den Zuhörer momentan zu den besten Assistenten auf den Markt.

5.3.3 Cortana

Nutzer von „Windows 10“, einer Xbox oder eines Windows Phone kennen sie bereits, „Cortana“, den digitalen Assistenten, welcher von Microsoft entwickelt wurde. Er wurde im Jahr 2014 veröffentlicht und steht neben den genannten Microsoft Geräten auch für Android und iOS zur Verfügung. Grundsätzlich kann Cortana eine Suche starten, Programme ausführen oder Termine und Notizen verwalten. [Wik17d]

Über Skills können Entwickler ähnlich wie bei Alexa zusätzliche Funktionen erstellen und für Nutzer freigeben. Das von Microsoft bereitgestellte Skills Kit ist momentan noch nicht freigegeben und befindet sich aktuell in einer öffentlichen Beta-Phase.

Skills für Cortana werden ähnlich wie die von Alexa in der Clound ausgeführt und können in .NET oder Node.js geschrieben werden. [Mic17]

5.3.4 Google Assistent

5.3.5 Auswertung der Möglichkeiten

6 Beschreibung eines Natural User Interface

7 Versuch der Entwicklung eines Natural User Interface anhand der Microsoft HoloLens und ...

8 Fazit

9 Zusammenfassung

10 Abkürzungsverzeichnis

KI *künstliche Intelligenz*

VR *Virtuelle Realität*

AR *Augumented Reality*

NUI *Natural User Interface*

AV *Augmented Virtuality*

SDK *Software Development Kit*

AI *Artificial Intelligence*

AVS *Amazon Voice Service*

Abbildungsverzeichnis

Tabellenverzeichnis

Literatur

[Ale17] Speechcon Reference. Technical report, Amazon Inc., Oktober 2017.

[Ama17a] Alexa Voice Service . Technical report, Amazon Inc., Oktober 2017.

[Ama17b] Alexa Skills Kit. Technical report, Amazon Inc., Oktober 2017.

[ARV17] Microsoft HoloLens. Technical report, ARVRZone, Mai 2017.

[BLT+12] Domagoj Baricevic, Cha Lee, Matthew Turk, Tobias Höllerer, and Doug A. Bowman. A Hand-Held AR Magic Lens with User-Perspective Rendering. Technical report, University of California, 2012.

[DBGJ13] Ralf Dörner, Wolfgang Broll, Paul Grimm, and Bernhard Jung. *Virtual und Agmented Reality*. Springer Vieweg, 2013.

[Dig17] Digital Native. `https://de.wikipedia.org/wiki/Digital_Native`, Juli 2017.

[DP15] Raimund Dachselt and Bernhard Preim. *Interaktive Systeme*. Springer-Verlag, 2015.

[Ebe17] Christoph Ebert. Grundlagen der Spracherkennung – so funktionieren Alexa, Cortana, Siri & Co. `https://entwickler.de/online/mobile/grundlagen-spracherkennung-alexa-cortana-siri-579769393.html`, Februar 2017.

[Goo17] Google Tang. Technical report, Google, 2017.

[Hol17] Hololens. Technical report, Microsoft, Mai 2017.

[IBM17] Watson. `https://www.ibm.com/watson/?cm_mc_uid=91647690093115002014743&cm_mc_sid_50200000=1500201474`, Jul 2017.

[ISO11] Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten. Technical report, International Organization for Standardization, Juli 2011.

[Jöc14] Marja-Liisa Jöckel. Augmented Reality: Definition, Anwendung, Apps. *entwickler.de*, August 2014.

[Kle] Georg Klein. *Visual Tracking for Augmented Reality*. PhD thesis.

[MBRS11] Anett Mehler-Bicher, Michael Reiß, and Lothar Steiger. *Augmented Reality Theorie und Praxis*. Oldenbourg Verlag, 2011.

[Mer16] Johannes Merkert. Google: Translate-KI übersetzt dank selbst erlernter Sprache. https://www.heise.de/newsticker/meldung/Google-Translate-KI-uebersetzt-dank-selbst-erlernter-Sprache-3502351.html, November 2016.

[Mic17] Create intelligent, personalized experiences for users with the Cortana Skills Kit . Technical report, Microsoft Inc., Oktober 2017.

[MTUK] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. Augmented Reality: A class of displays on the reality-virtuality continuum.

[Pak17] Ingo Pakalski. Amazon lässt Alexa natürlicher klingen. *Golem*, Main 2017.

[Sir17a] Creating an Intents App Extension. Technical report, Apple Inc., September 2017.

[Sir17b] Creating an Intents App Extension. Technical report, Apple Inc., September 2017.

[VRS16] Epson Moverio BT-200. https://www.virtual-reality-shop.co.uk/epson-moverio-bt-200/, Mai 2016.

[Wei91] Mark Weiser. The Computer for the 21 st Century. *Scientific American*, page 12, September 1991.

[Wik17a] Head-Up-Display. https://de.wikipedia.org/wiki/Head-up-Display, Mai 2017.

[Wik17b] Künstliche Intelligenz. https://de.wikipedia.org/wiki/K%C3%BCnstliche_Intelligenz, Jul 2017.

[Wik17c] Amazon Alexa. https://en.wikipedia.org/wiki/Amazon_Alexa, Oktober 2017.

[Wik17d] Cortana (Software). https://de.wikipedia.org/wiki/Cortana_(Software), September 2017.

[Wik17e] Microsoft HoloLens. https://de.wikipedia.org/wiki/Microsoft_HoloLens, Mai 2017.