

Interaktion Hands-Free. Eine Analyse alternativer Interaktionsmöglichkeiten

Hanna Wagner



BACHELORARBEIT

Nr. S1410456033

eingereicht am
Fachhochschul-Bachelorstudiengang
Kommunikation, Wissen, Medien
in Hagenberg

im Mai 2017

Diese Arbeit entstand im Rahmen des Gegenstands

Course not specified!

im

Sommersemester 2017

Betreuung:

FH-Prof. DI (FH) Dr. Mirjam Augstein

Erklärung

Ich erkläre eidesstattlich, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen nicht benutzt und die den benutzten Quellen entnommenen Stellen als solche gekennzeichnet habe. Die Arbeit wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungsbehörde vorgelegt.

Hagenberg, am 1. Mai 2017

Hanna Wagner

Inhaltsverzeichnis

Erklärung	iii
Kurzfassung	vi
Abstract	vii
1 Einleitung	1
1.1 Motivation	1
1.2 Zielsetzung	1
2 Eingabemethoden	2
2.1 Sprachsteuerung	2
2.2 Augensteuerung	4
2.3 Gestensteuerung	7
2.3.1 Kinnsteuerung	7
2.3.2 Mundsteuerung	8
2.3.3 Kopfsteuerung	9
2.3.4 Fußsteuerung	10
2.4 Muskelsteuerung (Myoelektrische Signale)	11
2.5 Steuerung durch Gehirnaktivität	13
3 Ausgabemethoden	16
3.1 Auditiver Output	16
3.2 Haptischer Output	17
3.3 Visueller Output	17
4 Gegenüberstellung der Methoden	18
4.1 Fördernde und hemmende Faktoren	18
4.2 Einsatzgebiete	18
5 Fazit und Ausblick	19
Quellenverzeichnis	20
Literatur	20
Online-Quellen	21

Abbildungsverzeichnis

2.1	Hierarchie von Eye-Tracking-Systemen [2].	5
2.2	Übersicht über die Eye-Tracking-Systeme von Tobii und SMI Tobii Pro Glasses 2 [16] (a), Tobii Pro Spectrum [17] (b), SMI Red250mobile [18] (c).	6
2.3	Verschiedene Kinnsteuerungen Moso® Kinnsteuerung [14] (a), Smile Smart Kinnsteuerung [20] (b), Sen- sory Kinnsteuerung [21] (c).	7
2.4	Verschiedene Mundsteuerungen: IntegraMouse Plus [23] (a) und QuadStick Gamecontroller [22] (b) . . .	9
2.5	Verschiedene Fußsteuerungen Früwald Joystick [24] (a), Früwald Fußsteuerung [24] (b) und Hidrex Fußsteuerung [25] (c).	10
2.6	BCI Signalverarbeitung [1]	12
2.7	Trainingsoberflächen: 1D-Steuerung [13] (a) und 2D-Steuerung [13] (b)	12
2.8	BCI Signalverarbeitung [5].	14
3.1	Unterteilung der Sprachausgabe [3].	16

Kurzfassung

Abstract

Kapitel 1

Einleitung

1.1 Motivation

1.2 Zielsetzung

Kapitel 2

Eingabemethoden

Die meisten Interaktionen mit elektronischen Geräten jeder Art finden durch den Einsatz unserer Hände statt. Das reicht von der einfachen Bedienung von Geräten (z.B. Einschalten eines Computers) bis hin zur Steuerung von Abläufen mittels Gestik (z.B. Liederwechsel durch eine Wischgestik mit der Hand, die dann von einer Software erkannt und in einen Befehl umgewandelt werden kann). Doch nicht immer gibt es die Möglichkeit mit Hilfe der Hände zu interagieren. Beispielsweise beim Autofahren sollen beide Hände am Lenkrad bleiben, bei anstrengenden und präzisen Arbeiten an Maschinen müssen oft beide Hände benutzt werden oder aber auch für Menschen mit Tetraplegie bzw. Tetraparese, die ihre Hände und Arme nicht oder oft nur sehr eingeschränkt benutzen können, sind alternative Interaktionsmöglichkeiten äußerst interessant. Aus diesem Grund beschäftigt sich dieses Kapitel mit den einzelnen alternativen Eingabemethoden, die es ermöglichen auch ohne den Einsatz von den Händen mit einem System interagieren zu können. Welche Vor- und Nachteile, gesonderte Rahmenbedingungen und welche Einsatzgebiete es im Detail zu den einzelnen Methoden gibt, werden in Kapitel 4 näher erläutert.

2.1 Sprachsteuerung

Eine Alternative zur Interaktion mit den Händen stellt die Steuerung mit Hilfe der menschlichen Stimme dar. Ein Spracherkennungssystem soll im optimalen Fall das Gesprochene genau so gut erkennen und verstehen können wie ein Mensch. Da jeder anders spricht bzw. es auch syntaktische Unterschiede gibt, kann ein System sich an das menschliche Verstehen allerdings nur annähern. Oft werden daher in der Praxis die Systeme speziell für einzelne Szenarien entwickelt bzw. angepasst.

Wird der Inhalt des Gesprochen analysiert und zur Interaktion verwendet, wird von der Spracherkennung gesprochen. Hier gibt es zwei Ansätze, damit das System das Gesprochene versteht:

- Mustervergleich
- Statistische Spracherkennung

Bei einem Mustervergleich werden die einzelnen Wörter verwendet, um diese mit bereits abgespeicherten Mustern und Begriffen zu vergleichen. Jenes Wort, das dem Eingabewort am ähnlichsten ist, wird verwendet. Hingegen bei der statistischen Spra-

erkennung werden Beschreibungen von Lauten und Wörtern verwendet. Darüber hinaus werden Statistiken verwendet, die Wahrscheinlichkeiten von Wortfolgen beinhalten [6].

Bevor das Signal im gesamten Prozess bei dem Analyseschritt angelangt, laufen vorher noch einige Schritte ab. Wenn eine Person etwas spricht werden die analogen Wellen durch elektroakustische Wandler (zum Beispiel ein Mikrophon) in ein elektrisches Signal (Sprachsignal) umgewandelt, damit der Computer diese interpretieren und weiterverarbeiten kann. Auf den gespeicherten Daten werden anschließend verschiedenste Filter angewendet, um störende Signale wie zum Beispiel Hintergrundgeräusche oder Pausen während des Sprechens zu entfernen. Erst dann kann im nächsten Schritt durch eine der bereits beschriebenen Methoden das System das Gesprochene analysieren und so den Sinn erfassen. Ist der Inhalt des Gesprochenen bekannt, kann dieser in einen Text und somit einen Befehl umgewandelt und dadurch der Computer bzw. das Programm gesteuert werden [6].

Es gibt viele Faktoren, die einen Einfluss auf das Sprachsignal haben. So hat jeder Mensch neben seiner einzigartigen Stimme einen unterschiedlichen Dialekt, Sprachgewohnheiten, Emotionen in der Sprach, Pausengewohnheiten und eine verschiedene Sprechgeschwindigkeit [6]. Diese Komponenten sind nicht nur wichtig für die Erkennung was gesprochen wird, sondern viel mehr wer spricht (Sprechereerkennung).

Neben der Spracherkennung gibt es noch die Sprechererkennung, die für die Interaktion mit einem System relevant sein kann. Um erkennen zu können welche Person gerade spricht, muss es Referenzdaten im System geben, die mit dem aktuellen Signal verglichen werden können. Dies geschieht entweder wenn die Person ein bestimmtes Singalwort spricht, dass mit dem abgespeicherten Wort verglichen wird, oder wenn das Sprachsignal lange genug ist, kann auf Grund von statistischen Merkmalen eine Übereinstimmung gefunden werden [6].

Kommunikation durch Sprache ist eines der leichtesten, wenn nicht sogar intuitivsten Mittel zur Verständigung. Daher sollte auch die Interaktion mit einem Computer diesen hohen Standards entsprechen. Allerdings ist im Gegensatz zur gewohnten Face-to-Face-Kommunikation das Gegenüber im Bezug auf Spracherkennung kein intelligenter Gesprächspartner. Ein Computer kann nur schwer lange und komplizierte Sätze und Gedankengänge mitverfolgen und daraus schlecht die Kernaussage zu ziehen. Daher muss dem System verständlich gemacht werden, wann es zuhören soll [12]. Zwei gängige Methoden für den Start zur Interaktion sind:

- das Drücken einer Starttaste
- das Aussprechen eines Aktivierungswortes

Da das drücken eines Startknopfes die Hilfe der Hände verwendet, steht hier das Aktivierungswort im Vordergrund. Durch das Wort 'Alexa' wird beispielsweise der Amazon Echo ¹ aktiviert. Bei Amazon Echo handelt es sich um ein Sprachinteraktionssystem, dass beispielsweise Musik abspielen, den Einkauf bestellen oder auch Nachrichten und

¹<https://www.amazon.com/Amazon-Echo-Bluetooth-Speaker-with-WiFi-Alexa/dp/B00X4WHP5E>

Wetterbeiträge zur Verfügung stellen kann. Im Gegensatz zur Benutzung von Amazon Echo, gibt muss bei der Sprachsteuerung von Apple (Siri)² das System erst auf die eigene Stimme eingestellt, also kalibriert werden. Das Aktivierungswort 'Hey Siri' muss drei Mal einzeln und zwei Mal in einem Satz ausgesprochen werden, damit diese Funktion beispielsweise am iPhone freigeschaltet wird. Anschließend können ebenfalls Wetter oder Nachrichtendaten abgefragt werden, aber auch Anrufe oder Nachrichten können mit Hilfe der Stimme gesteuert werden.

Sobald das System bereit ist den Input aufzunehmen lässt sich zusammenfassend folgendes festhalten, damit eine erfolgreiche Interaktion mit dem System geschehen kann:

- keine zu große Distanz zum Mikrophon beachten,
- laut und deutlich zu sprechen und
- nicht zu schnell zu sprechen.

Zusammenfassend lässt sich sagen, dass die Sprachinteraktion mit dem Computer auf Benutzersicht noch einige Verbesserung aufweist, damit die Bedienung natürlicher und intuitiver wird und der einer Face-to-Face-Kommunikation näher kommt.

2.2 Augensteuerung

Eine weitere alternative Interaktionsmöglichkeit stellt die Steuerung mit Hilfe der Augen dar. Eye-Tracking-Systeme werden verwendet um die Augenbewegungen zu verfolgen. Gleichzeitig wandelt eine Software die Augenbewegungen in einen Mauszeiger um, um so mit dem System interagieren zu können.

Um die Augenbewegungen nachvollziehen zu können, werden sowohl ein Infrarotlichtstrahl, als auch eine Videokamera auf das Auge fokussiert. Eine Software analysiert die Bewegungen im Hintergrund und kann erkennen wo der Fokus auf dem Bildschirm liegt. Für die Umrechnung ist überdies wichtig, dass die Position des Kopfes vom Interaktionssystem erkannt wird [9].

Der Teilbereich des Auges bzw. des menschlichen Sehens, der für die Augensteuerung relevant ist, ist die Unterteilung in peripheres und foveales Sehen. Das foveale Sehen beinhaltet jenen Teil, den das menschliche Auge als scharfen Bildausschnitt wahrnimmt, also auf welchen Ausschnitt es fokussiert. Hingegen enthält das periphere Sehen den Großteil unser Seh Wahrnehmung, das beinhaltet jene Bereiche, die nicht fokussiert sind [9]. Daher ist das foveale Sehen für die Augensteuerung relevant.

Nielsen und Pernice [9] unterscheiden beim menschlichen Sehen zusätzlich zwischen Fixation und Sakkaden. Das menschliche Auge bzw. der Blick bewegt sich nicht kontinuierlich, sondern es handelt sich um schnelle Bewegungen mit Pausen dazwischen. Dies geschieht so schnell, dass es im Alltag nicht bewusst wahrgenommen wird. Wenn das Auge auf einem bestimmten Punkt ruht spricht man von Fixation. Die schnellen Zwischenbewegungen von einer Fixation zur nächsten werden Sakkaden genannt. Die Schwierigkeit der Eye-Tracking-Systeme und der Augensteuerungssoftware

²<https://www.apple.com/ios/siri/>

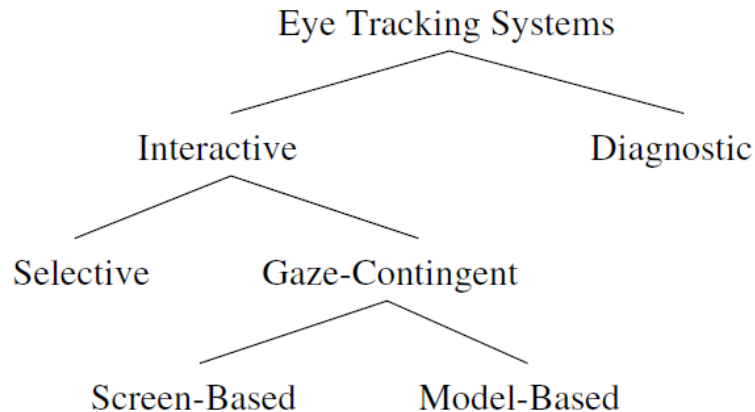


Abbildung 2.1: Hierarchie von Eye-Tracking-Systemen [2].

besteht darin, diese Fixationen zu erkennen, da diese meist zwischen einer hundertstel und einer zehntel Sekunde liegen [9].

Wie in Abb. 2.1 dargestellt, unterscheidet Duchowski [2] zwischen interaktiven und diagnostischen Systemen. Da bei diagnostischen Systemen keine direkte Interaktion stattfindet, sondern die Augenbewegungen aufgezeichnet und im Nachhinein ausgewertet werden, sind für die Analyse von Interaktionsmöglichkeiten ausschließlich interaktive Systeme relevant.

Interaktive Systeme lassen sich wiederum in zwei Bereiche teilen. Selektive Systeme benutzen den Blickpunkt als analoges Eingabegerät, während Blickkontingentsysteme für die Darstellung von komplexen Displays verwendet werden.

Abgesehen davon, wie die Daten aufgezeichnet werden bzw. was aufgezeichnet wird, ist die Unterscheidung zwischen den Eye-Tracking-Systemen relevant. Diese können in intrusive und nicht-intrusive Systeme unterteilt werden. Intrusive Eye-Tracker haben einen direkten Kontakt mit den Anwendern (z.B. durch Kontaktlinsen). Nicht-intrusive Eye-Tracking-Systeme hingegen messen die Blicke mit Hilfe von einer oder mehreren Kameras.

In der Praxis werden eine Vielzahl von verschiedensten Eye-Tracking-Systemen als Basis für die Augensteuerung verwendet. Die Unternehmen Tobii ³ und SensoMotoric Instruments (SMI) ⁴ sind Spitzenreiter auf diesem Gebiet und haben eine umfangreiche Produktpalette bzw. Anwendungsgebiete.

In Abb. 2.2(a) ist ein tragbares Eye-Tracking-System zu sehen. Dieses ist so gestaltet worden, um ein möglichst natürliches Nutzungsverhalten zu erzielen. Es müssen keine Voreinstellungen vorgenommen werden, die Brille kann sofort benutzt werden. Dieses Produkt gibt einen Einblick über das Nutzungsverhalten und weniger über die Interaktion mit einem zusätzlichen System selbst und ist daher als Beispiel für ein in-

³<http://www.tobii.com/group/>

⁴<https://www.smivision.com/>

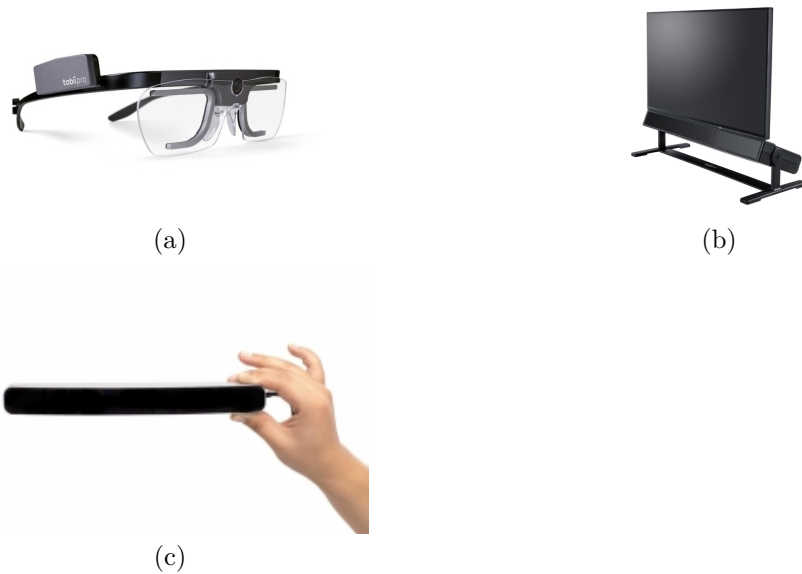


Abbildung 2.2: Übersicht über die Eye-Tracking-Systeme von Tobii und SMI
Tobii Pro Glasses 2 [16] (a), Tobii Pro Spectrum [17] (b), SMI Red250mobile [18] (c).

strusives System angeführt. Bei Tobii Pro Spectrum (Abb. 2.2(b)) handelt es sich um ein nicht-intrusives System, dass die Blickdaten mit einer Geschwindigkeit von bis zu 600Hz aufnimmt. Das Unternehmen SMI hat, wie in Abb. 2.2(c) zu sehen ist, ein mobiles Eye-Tracking-System entwickelt, das die Bewegungen mit bis zu 250Hz aufzeichnen kann. Alle diese Systeme verwenden das Prinzip von Fixation und Sakkaden des Auges, um die Augenbewegungen nachvollziehen zu können.

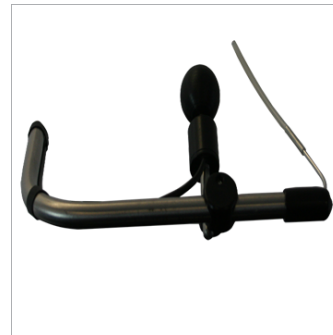
Bevor ein Benutzer mit Hilfe der Augen beispielsweise eine Bildschirmtastatur bedienen kann, muss das System bzw. die Software erst auf den Anwender eingestellt werden. Bei der Kalibrierung tauchen am Bildschirm nacheinander Symbole auf, die fixiert werden müssen. Dadurch kann der Blickwinkel und der Augenabstand vermessen werden um so eine optimale Interaktion zu ermöglichen. Bei der eigentlichen Anwendung erkennt das System dann, wenn der Blick auf einem Punkt länger verharret, also scharf stellt. Verharret der Blick also länger auf einem Buchstaben der Tastatur wird dies als Mausklick gewertet und der Buchstabe oder das Symbol ausgewählt. Am Beispiel von Seetech von Humanelektronik ⁵ beträgt die Fixationszeit etwa 0.5 bis 1.5 Sekunden, je nach Vertrautheit des Benutzers mit dem System [19].

Zusammenfassend kann gesagt werden, dass für die Interaktion mit einer Anwendung bzw. einem System interaktive Eye-Tracking-Systeme für die Nachvollziehbarkeit der Augenbewegungen und eine Software für die Umrechnung in Mausbewegungen relevant sind. Nach einer kurzen Kalibrierung kann mit der Computer durch Fixation einzelne Elemente auf dem Bildschirm auswählen, die dann auf Grund der Fixationszeit in einen Mausklick umgewandelt werden können.

⁵<http://humanelektronik.de/>



(a)



(b)



(c)

Abbildung 2.3: Verschiedene Kinnsteuerungen

Moso® Kinnsteuerung [14] (a), Smile Smart Kinnsteuerung [20] (b), Sensory Kinnsteuerung [21] (c).

2.3 Gestensteuerung

Eine weitere alternative zur Interaktion mit den Händen stellt die Gestensteuerung dar. Es gibt viele verschiedene Arten wie Gesten bzw. Gestik definiert werden kann. In diesem Zusammenhang versteht man unter einer Geste [11]:

die Bewegung von Fingern, Händen und Armen - oder auch weitere Körperteile, wie Kopf, Augen und Lippen - aufgrund einer kommunikativen Absicht. Damit enthält die Bewegung als solche signifikante Informationen, die an den Computer übermittelt werden sollen.

In diesem Abschnitt werden die Kinn-, Mund-, Kopf- sowie Fußsteuerung näher erläutert.

2.3.1 Kinnsteuerung

Bei der Kinnsteuerung werden mit Hilfe des Kinns Bewegungen ausgeführt, um mit einem System interagieren zu können. Die unterschiedlichen Ausführungen bzw. Bedienungen werden anhand der nachfolgenden Beispiele erklärt.

Wie in Abb. 2.3(a) zu sehen ist verwendet das Unternehmen moso®⁶ einen Joystick und zwei Tasten für die Interaktion. Der Joystick besteht aus einem kleinen weichen Ball. Die zwei Tasten sind links und rechts vom Joystick angebracht. Der Joystick kann sich in alle Himmelsrichtungen bewegen und kann so entweder als Mauszeiger, oder zur Steuerung eines Rollstuhls genutzt werden. Die beiden Tasten können als linke und rechte Maustasten verwendet werden [14].

Abb. 2.3(b) zeigt die Version der Firma Smile Smart Technology⁷. Diese besteht aus einem Mini Rollstuhl Joystick, der auf einer schwenkbaren Halterung montiert ist [20]. Die Kinnsteuerung des Unternehmens Sensory Guru⁸ weist im Vergleich zu den anderen beiden keinen Ball auf. Wie in Abb. 2.3(c) dargestellt, gibt es einen Joystick mit einer Mulde für das Kinn. Mit einer Bewegung nach links oder rechts können die beiden äußeren Sensoren aktiviert werden [21].

Durch die Bedienung des Joystickelementes kann der Benutzer beispielsweise die Bewegungen einer Computermouse nachahmen. Darüber hinaus gibt es unterschiedliche Möglichkeiten wie durch ein Menü navigiert werden und Symbole oder andere gewünschte Elemente ausgewählt werden können. Dies geschieht entweder über zusätzlich angebrachte Tasten an der Steuerung oder über eine stärkere Kinnbewegung nach links bzw. rechts [14] [21].

2.3.2 Mundsteuerung

Das Prinzip der Mundsteuerung ist ähnlich dem der Kinnsteuerung. Durch die verschiedenen Interaktionsmöglichkeiten, die durch den Einsatz des Mundes möglich sind, wird versucht eine Interaktion mit beispielsweise einem Computer möglich zu machen.

Bei der in Abb. 2.4(a) gezeigten Mundsteuerung handelt es sich um die IntegraMouse Plus, die mit Hilfe des Unternehmens LIFEtool⁹ entwickelt wurde. Die IntegraMouse Plus wird durch nippen, pusten und bewegen des Mundes gesteuert. Wird die Maus mit einem Computer verbunden so kann durch ein Nippen ein Linksklick erzeugt werden, durch Pusten ein Rechtsklick und durch das Bewegen des Mundstückes kann die Richtung bzw. der Cursor verändert werden. [15].

Ähnlich wie bei der IntegraMouse Plus wird bei dem mundgesteuerten Gamecontroller der Firma QuadStick¹⁰, der in Abb. 2.4(b) zu sehen ist, ebenfalls nippen, pusten und die Bewegungen des Mundes zur Interaktion verwendet. Dieses Modell hat allerdings anstatt einer vier Nipp- bzw. Pustsensoren, um einen regulären Gamecontroller besser imitieren zu können [22].

Bei beiden Steuerungen ist keine Kalibrierung vor der Benutzung notwendig. Das System muss lediglich mit dem gewünschten System per USB-Stick verbunden werden.

⁶<http://www.moso-gmbh.de/>

⁷<https://smilemart-tech.com/>

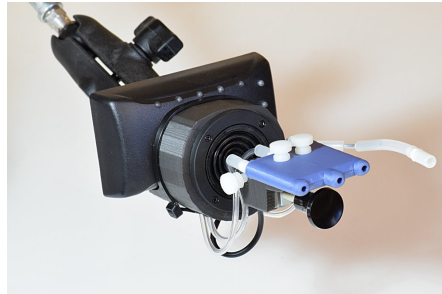
⁸<http://www.sensoryguru.com/>

⁹<http://www.lifetool.at/startseite/>

¹⁰<http://www.quadstick.com/>



(a)



(b)

Abbildung 2.4: Verschiedene Mundsteuerungen:
IntegraMouse Plus [23] (a) und QuadStick Gamecontroller [22] (b)

2.3.3 Kopfsteuerung

Bei der Kopfsteuerung verfolgt eine Kamera entweder die Bewegungen des gesamten Gesichtes, oder einen bestimmten Teil davon (z.B. den Mund). Die Bewegungen werden anschließend von einem System in die gewünschten Befehle umgewandelt.

Es werden grundlegend zwei Arten der Kopfsteuerung bzw. der Nachvollziehbarkeit der Bewegungen unterschieden:

- Gesichtserkennung
- Punkteverfolgung

Bei der Gesichtserkennung wird versucht das Gesicht als Gesamtheit zu erfassen und dieses als Berechnungsgrundlage für die Bewegungen herangezogen. Es gibt verschiedene Arten um die Position des Kopfes zu erkennen. So kann beispielsweise die Hautfarbe dazu verwendet werden. Allerdings ist diese von Mensch zu Mensch verschieden und durchläuft bei der Berechnung bzw. Filterung viele Schritte und ist damit mit großem Rechenaufwand verbunden. Daher ist eine Mischung aus der Haarähnlichen-Gesichtserkennung und dem Camshift-Gesichtstrackingalgorithmus effizienter. In verschiedenen Schritten wird der Unterschied vom Haaransatz zum Gesicht verwendet, ein Farbgesichtsdarstellungshistogramm erzeugt, eine Gesichtswahrscheinlichkeitsberechnung für jedes Pixel und schließlich die Größe und die Winkelberechnung durchgeführt [10].

Bei der Punkteverfolgung wird ein bestimmter Teilbereich des Gesichtes fokussiert, der die Bewegungsgrundlage des gesamten Kopfes darstellt. Naizhong und Jing [8] haben einen Algorithmus entwickelt, der die Mundbewegungen als Grundlage für die Kopfbewegungen verwendet. Dafür wird eine Kamera als Videoinput verwendet und in der Analyse können die verschiedenen Eingabebefehle unterschieden werden. So ergibt beispielsweise ein Kopfschütteln nach Links einen Linksklick oder ein zweimaliges Kopfschütteln nach Rechts einen rechten Doppelklick [8].

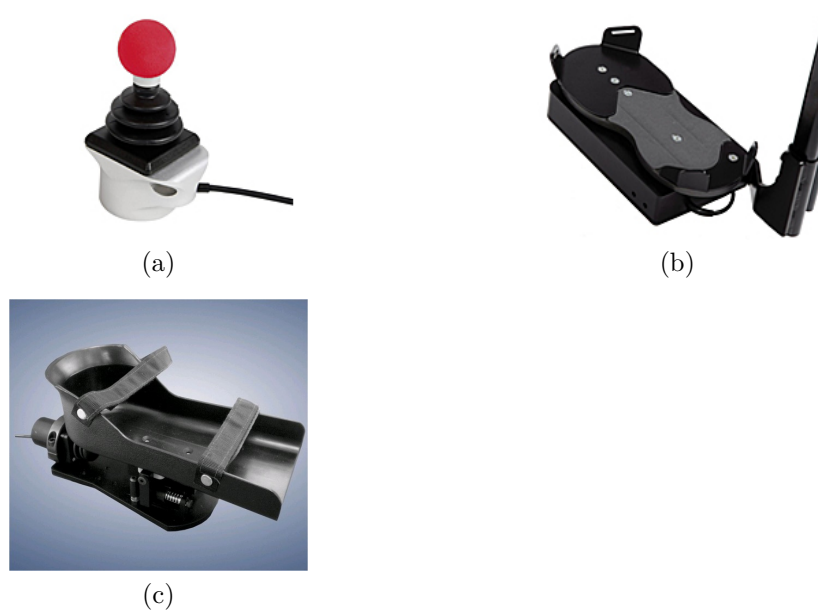


Abbildung 2.5: Verschiedene Fußsteuerungen

Frühlwald Joystick [24] (a), Frühlwald Fußsteuerung [24] (b) und Hidrex Fußsteuerung [25] (c).

Für den Benutzer gibt es einige Punkte, die vor und während der Interaktion ausschlaggebend für eine optimale Verwendung sind. Ähnlich wie bei der Augensteuerung in Kapitel 2.2 muss auch hier das System vorab kalibriert werden, damit eine reibungslose Interaktion möglich ist. Des Weiteren gilt es zu beachten, dass keine störenden Lichtquellen auf das Gesicht fallen und ein möglichst frontaler Blick in die Kamera gewährleistet ist [8].

2.3.4 Fußsteuerung

Es gibt eine Vielzahl an verschiedenen Fußsteuerungen. Manche ähneln sehr der einer Kinnsteuerung und verwenden einen Joystick, um beispielsweise Mausbewegungen zu erzeugen. Darüber hinaus werden verschiedene Tasten für die Auswahl bestimmter Elemente genutzt. Andere wiederum verwenden einen zusätzlichen Schuh bzw. eine Halterung um den Fuß, um die Fußbewegungen zu messen.

Wie in Abb. 2.5(a) zu sehen ist, kann ein gewöhnlicher Joystick, der auch mit dem Kinn bedienbar ist, ebenfalls mit dem Fuß bedient werden. Der Benutzer bewegt den Joystick in die gewünschte Richtung und kann damit die Fahrtrichtung oder die Mausbewegung am Computer steuern [24]. Bei der in Abb. 2.5(c) dargestellten Fußsteuerung der Firma Hidrex ¹¹, handelt es sich um einen Schuhalter, der als Joystick funktioniert. Die Bedienung ist ähnlich der des herkömmlichen Joysticks. Der Fuß muss in die gewünschte Richtung gezogen bzw. bewegt werden, um einen Bewegungsablauf nachvollziehen zu können. Die Fußsteuerung bzw. das Fußpedal des Unternehmens Frühlwald ¹² kann

¹¹<http://www.hidrex.de/>

¹²<https://www.fruehwald.net/>

durch unterschiedlich starke Druckausübung zur Steuerung verwendet werden [24].

Wie die einzelnen Systeme die verschiedenen Gestiken in eine Bewegung umwandeln, ist von System zu System verschieden. Aus Anwendersicht wird allerdings meist ein joystickähnliches Element für die Richtung bzw. Bewegungsrichtung bedient. Zusätzliche Tasten helfen bei der Auswahl von gewünschten Symbolen oder Elementen.

2.4 Muskelsteuerung (Myoelektrische Signale)

Bei der Muskelsteuerung bzw. Verwendung von myoelektrischen Signalen wird das Zusammenziehen eines Muskels oder bestimmter Muskelpartien verwendet, um so mit einem Computer interagieren zu können.

Der Großteil der Muskelsteuerungssysteme verwendet zur Messung der Signale das Prinzip der Elektromyografie (EMG). Bei einer EMG-Schnittstelle handelt es sich um die Messung der Muskelkontraktion und verknüpft diese mit der gewünschten Interaktion. Die Voraussetzung an den Benutzer besteht lediglich, dass die ausgewählten Muskelpartien selbstständig angespannt bzw. aktiviert werden können. Im Zusammenhang mit der Muskelsteuerung kommt die Oberflächenelektromyografie (sEMG) zum Einsatz. Hierbei werden die Elektroden, die für die Messung der Signale ausschlaggebend sind, an der Hautoberfläche angebracht [4].

Abb. 2.6 zeigt die einzelnen Schritte die das Signal im System durchläuft um am Ende die Interaktion mit einem Computer zu ermöglichen. In der Studie von Barszap, Skavhaug und Joshi [1] werden zwei Frequenzbänder verwendet, um eine X- und Y-Richtung des Mauszeigers zu ermöglichen. Zieht sich der Muskel zusammen, so werden alle messbaren Signale aufgefasst und die relevanten herausgefiltert. Die Signale werden anschließend in zwei Partien aufgeteilt und durch zwei separate Bandpassfilter geschickt. Dadurch werden die gewünschten Signale bzw. Frequenzen erneut gefiltert und angepasst. Anschließend können beide Signale durch verschiedenste Algorithmen in einen Wert zwischen 0 und 100 umgerechnet werden, der die Ausprägungsstärke der jeweiligen Koordinate repräsentiert [1].

Damit ein Benutzer durch den Einsatz von myoelektrischen Signalen mit einem Computer interagieren kann, müssen zu Beginn die richtigen Positionen für die Anbringungen der Elektroden gefunden werden. Es gibt verschiedene Ansätze wie die Kalibrierung im Detail aussehen kann. Folgendes Beispiel bezieht sich auf eine Studie von Vernon und Joshi [13].

Dazu werden die Elektroden an der vermuteten Stelle angebracht und mit dem Computer verbunden. Auf dem Bildschirm ist das aktuelle sEMG-Signal zu sehen. Ist das Signal mit vielen Störfaktoren, dem sog. Rauschen verbunden, müssen die Elektroden gereinigt oder in der Regel neu positioniert werden. Ist die optimale Position gefunden worden, muss das System auf die individuellen Muskelkontraktionen des Anwenders angepasst werden. Bei diesem Schritt muss der Benutzer die ausgewählten Muskeln über einen Zeitraum von ein paar Sekunden so fest wie möglich anspannen, um die maximale

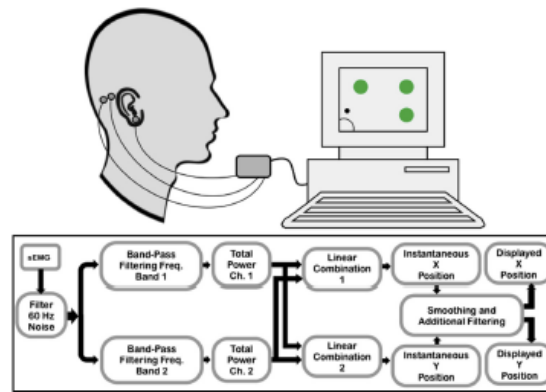
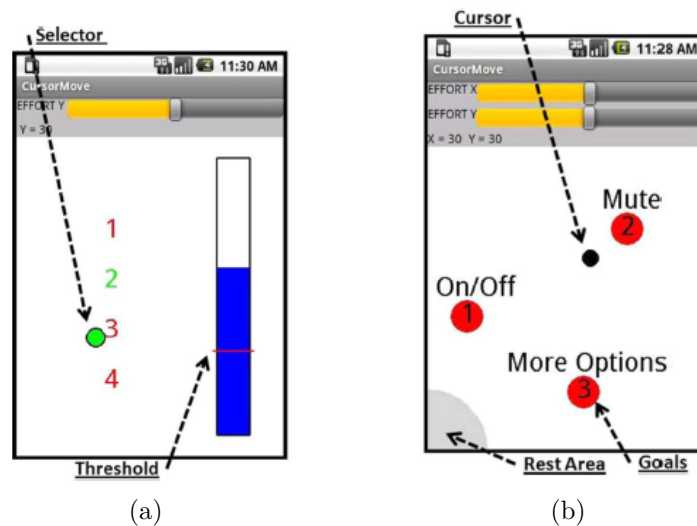


Abbildung 2.6: BCI Signalverarbeitung [1]

Abbildung 2.7: Trainingsoberflächen:
1D-Steuerung [13] (a) und 2D-Steuerung [13] (b)

Frequenz bzw. Leistung zu erfassen. Als nächstes werden zwei verschiedene Testphasen durchgeführt. Der erste Test (Abb. 2.7(a)) dient dazu ein Bewusstsein zu entwickeln, wann und wie stark der gewünschte Muskel angespannt wird. Der grüne Punkt bewegt sich je nach Anspannung durch die verschiedenen Stufen und soll auch durch Entspannen wieder zur Ausgangslage gebracht werden. Bei dem zweiten Test, der in Abb. 2.7(b) dargestellt ist, handelt es sich um eine zweidimensionale Überprüfung. Hier soll der Cursor in X- und Y-Richtung bewegt werden. Dies geschieht ebenfalls durch An- und Entspannen der einzelnen Muskeln. Beide Versuche bauen auf dem Trial und Error Prinzip aus. Der Benutzer muss also durch mehrere Versuche zur richtigen bzw. besten Lösung kommen. Ist der Anwender mit dem 2D-Training vertraut, kann mit der gewünschten Interaktion mit dem Computer begonnen werden. Diese ist auf dem selben Prinzip aufgebaut, allerdings gibt es unterschiedliche Oberflächen, Anordnungen der Elemente und

Unterscheidungen in der Aufgabenstellung selbst [13].

Damit ein Computer mit Hilfe von Muskelkontraktionen gesteuert werden kann, ist ein aufwendiger und gut überlegter Algorithmus notwendig, um die erfassten Signale in das gewünschte Ergebnis zu transferieren. Der Benutzer muss vor der Interaktion einige Schritte durchlaufen, um das System an das Individuum anzupassen und um ein Bewusstsein für die Muskelsteuerung zu erlangen.

2.5 Steuerung durch Gehirnaktivität

Dieses Kapitel beschäftigt sich mit der Schnittstelle zwischen dem Gehirn und dem Computer. In der Literatur wird in diesem Zusammenhang von Brain-Computer Interface (BCI) gesprochen. Ziel ist es eine direkte Verbindung der beiden Komponenten für eine Interaktion herzustellen. Zusammengefasst liest das BCI die Wellen, die vom Gehirn an den verschiedensten Stellen produziert werden ein und übersetzt diese Signale in Befehle, damit mit einem Computer interagiert werden kann [5].

Es gibt verschiedene Aufnahmetechniken von BCI's:

- Invasiv
- Teilweise Invasiv
- Nicht invasiv

Invasive Gehirn-Computer-Schnittstellen werden direkt im menschlichen Gehirn durch eine Operation eingesetzt. Es können hier einzelne oder mehrere Einheiten in verschiedenen Bereichen des Gehirns platziert und anschließend angesprochen werden. Diese Systeme weisen die höchste Qualität von den dreien auf, bringen aber durch die Operation gewisse Risiken mit sich. Bei teilweise invasiven BCI-Systemen werden die Einheiten nicht direkt in den Gehirnzellen sondern im Schädel eingesetzt und können daher die Signale etwas schlechter empfangen. Bei nicht invasiven Systemen handelt es sich um die sicherste und kostengünstigste Variante auf Kosten von schwächerer Gehirnsignalempfangung. Die Signale werden über Elektroden empfangen, die am Kopf befestigt werden. Für diese Technik werden meist mit Hilfe eines Elektroenzephalografen (EEG) die neuronalen Wellen des Gehirns gemessen [5].

Die neuronale Aktivität des Gehirns ist ausschlaggebend für die Messung. Daher ist es wichtig die verschiedenen Signale, die ein EEG misst kurz darzustellen [5]:

- Delta Signal: Hierbei handelt es sich um die langsamsten Wellen mit einer Frequenz zwischen 0.5 und 3.5 Hz, die während des Schlafens oder im komatösen Zustand auftreten.
- Theta Signal: Kommen beispielsweise während des Träumens vor und werden zwischen 3.5 und 7.5 Hz. gemessen.
- Alpha Signal: Dies sind langsamere Wellen, die beim Problemlösen oder Entspannen auftreten und reichen von einer Frequenz zwischen 7.5 und 12 Hz.
- Beta Signal: Diese Wellen sind oft mit Anstrengung wie beispielsweise des Lösen einer mathematischen Aufgabe verbunden (mit einer Frequenz von 12 bis 30 Hz.).

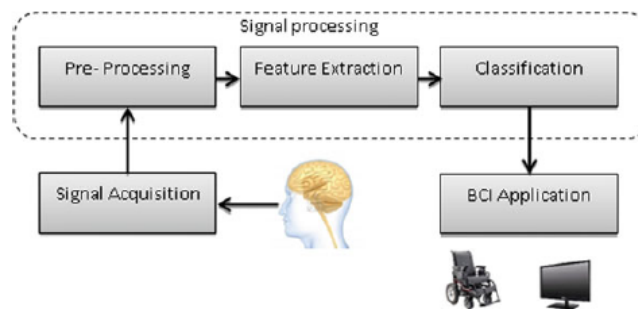


Abbildung 2.8: BCI Signalverarbeitung [5].

- Gamma Signal: Je größer hier die Amplitude ist, desto mehr Stress, Angst oder Panik erlebt die Personen (alle Frequenzen ab 31 Hz.).

Je nach Anwendungsbereich können verschiedene Signale verwendet werden. Für eine Interaktion mit einer Person, die sich nicht in einem komaähnlichen Zustand befindet sind alle Signale ab einer Stärke von 3.5 Hz. relevant.

Für den gesamte Ablauf, der nötig ist um von der technischen Komponente eine Gehirn-Computer-Schnittstelle zu erstellen sind, wie in Abb. 2.8 veranschaulicht, grob drei Schritte notwendig [5]:

- Im ersten Schritt werden die elektrischen Signale von der Kopfhaut oder der Oberfläche des Gehirns erfasst (Signalerfassung).
- Die erfassten Signale werden als nächstes von Störvariablen gereinigt und verstärkt, da diese in der Regel sehr schwach sind. Danach werden die Signale einen Übersetzungsalgorithmus durchlaufen, damit die Absichten des Benutzers erkannt werden können (Signalverarbeitung).
- Im letzten Schritt müssen die Daten von einem Programm bzw. einer Software so manipuliert werden, damit diese für eine Computerinteraktion verwendet werden können.

Um als Benutzer einen Computer mit Hilfe von Gehirnaktivität steuern zu können, muss das System ausreichend eingestellt werden. Als mögliche Kalibrierungsmethode kann ein Mauscursor gedanklich mitverfolgt werden, während dieser sich bewegt. Der Benutzer soll sich hierbei vorstellen er würde die Bewegung mit der Hand oder dem Arm ausführen. Je nach Person zeigt sich eine Anstrengung in dem motorischen Teil des Gehirns unterschiedlich stark und wird dahingehend angepasst. Als nächsten Schritt wird erneut ein Mauscursor gedanklich mitverfolgt. Allerdings erscheint ein zweiter Cursor, der auf Grundlage der bereits erhobenen Daten den momentanen Stand aufzeigt. Diese zwei Schritte werden mehrmals durchgeführt, um ein optimales Ergebnis zu erzielen. Neben der Mausbewegung kann auch ein Mausklick simuliert werden. Die Auswahl eines Menüpunktes kann beispielsweise dann so funktionieren, dass ein Menüpunkt hervorgehoben wird. Will der Anwender diesen auswählen, muss es sich besonders darauf konzentrieren bzw. die Gehirnbereiche müssen eine erhöhte Aktivität vorweisen [7].

Die Interaktion mit Hilfe der Messung der Gehirnaktivität ist zusammengefasst im Gegensatz zu den vorher Beschrieben Methoden eine komplexere und aufwendigere Inputmethode. Auch die Interaktion ist mit viel Aufwand hinsichtlich der Kalibrierung verbunden.

Kapitel 3

Ausgabemethoden

Im Kapitel 2 wurden alternative Eingabemethoden zur Interaktion Hands-Free näher erläutert. Dieses Kapitel befasst sich mit Ausgabemethoden, die aber nicht zwingend an ihre Eingabemethoden gebunden sind. So gibt es nicht nur bei Systemen, die über die Sprache gesteuert werden, ein akustisches Feedback, sondern auch bei Systemen, die beispielsweise über die Augen oder den Mund gesteuert werden.

3.1 Auditiver Output

Eine mögliche Form, wie ein Computer Feedback geben kann, funktioniert über die Sprache. Während bei der Spracheingabe die gesprochene Sprache von einem System in einen Befehl umgewandelt wird, gibt bei der Sprachausgabe der Computer die Informationen in Form von Sprache aus. Da die Sprachausgabe von Menschen gehört und automatisch bewertet wird, ist es wichtig, dass die Sprache von hoher Qualität ist, also möglichst real erscheint.

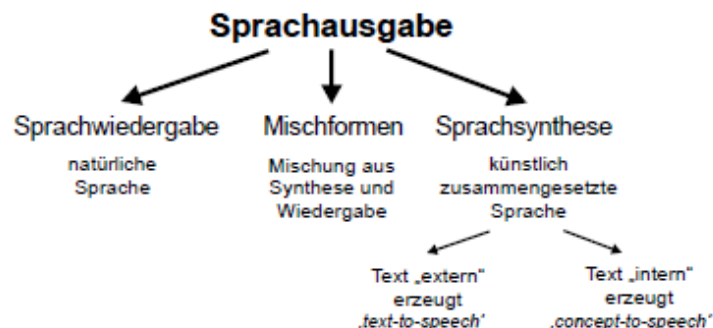


Abbildung 3.1: Unterteilung der Sprachausgabe [3].

Wie in Abb. 3.1 dargestellt gibt es zwei unterschiedliche Ansätze, wie der Computer den Text in Sprache umwandeln kann:

- Sprachwiedergabe
- Sprachsynthese

Bei der Sprachwiedergabe werden Sprachsignale ausgegeben, die zuvor aufgenommen und abgespeichert worden sind. Der Umfang an möglichen akustischen Meldungen also an Wörtern ist aus diesem Grund an die Anzahl der Aufnahmen gebunden und daher begrenzt. Da die Signale bzw. Wörter von einer realen Person eingesprochen wurden, weißt diese Methode eine sehr hohen Sprachqualität auf [6].

Es gibt unterschiedliche Methoden bzw. Verfahren, wie die Sprachwiedergabe im Detail funktioniert. Zusammenfassend werden die Wörter zu Beginn von einer Person eingesprochen, bearbeitet und abgespeichert. Die einzelnen Sprachelemente werden anschließend adressiert, sodass sie vom System später leicht gefunden werden können. Wenn der Computer nun einen Text vorlesen soll, wird im System anhand der gespeicherten Adressen die einzelnen Wörter bzw. Phrasen herausgesucht und verwendet [3].

Im Gegensatz zur Sprachwiedergabe wird bei der Sprachsynthese das Ziel verfolgt alle möglichen Wortkombinationen in ein Sprachsignal umwandeln zu können. Da die Grundlage für die Wörter bzw. Ausgabe nicht von einer Person eingesprochen werden, sondern die einzelnen Wörter mit ihrer Betonung künstlich erzeugt werden müssen, sinkt dadurch die Sprachqualität und das Endergebnis gleicht oft mehr einer Computerstimme als der einer menschlichen.

Bei der Sprachsynthese werden Silben aus verschiedenen Wörtern einzeln abgespeichert und so alle beliebigen Kombinationen zusammengesetzt. Im Anschluss werden Betonungen, Lautstärke und Geschwindigkeit hinzugefügt bzw. angepasst. Bei der sogenannten textgesteuerte Sprachsynthese (‘text-to-speech’) ist dem System der eigentliche Inhalt des Textes nicht bekannt. Es können aber nicht nur einzelne Sätze bzw. ganze Texte verarbeitet werden, sondern auch Konzepte können vom System in Signale umgewandelt werden (‘concept-to-speech’). Hier kann das System einen Text eigenständig kreieren, wenn das Konzept bzw. der Inhalt bekannt ist [3].

Laut Fellbaum [3] besteht die größte Herausforderung der Sprachsynthese darin, die Stimme noch natürlicher und nicht roboterähnlich wirken zu lassen. Daher sind Mischformen zwischen Sprachwiedergabe und Sprachsynthese in der Praxis üblich.

3.2 Haptischer Output

Es gibt verschiedene Arten von haptischem Feedback:

- Thermales Feedback (Wärme)
- Vibration
- Feedback durch ein haptisches Steuerungselement

3.3 Visueller Output

Arten von visuellem Output:

- in Form von Text oder Mediaelementen am Bildschirm
- Grafiken/Symbole (Häckchen oder Kreuze)

Kapitel 4

Gegenüberstellung der Methoden

4.1 Fördernde und hemmende Faktoren

4.2 Einsatzgebiete

Kapitel 5

Fazit und Ausblick

Quellenverzeichnis

Literatur

- [1] Alexander G. Barszap u.a. „Effects of muscle fatigue on the usability of a myoelectric human-computer interface“. *Human Movement Science* 49.5 (Okt. 2016), S. 225–238 (siehe S. 11, 12).
- [2] Andrew Duchowski. *Eye Tracking Methodology. Theory and Practice*. 2. Aufl. London: Springer, 2010 (siehe S. 5).
- [3] Klaus Fellbaum. *Sprachverarbeitung und Sprachübertragung*. 2. Aufl. Heidelberg: Springer, 2012 (siehe S. 16, 17).
- [4] Maria Hakonen u.a. „Current state of digital signal processing in myoelectric interfaces and related applications“. *Biomedical Signal Processing and Control* 18.4 (Apr. 2015), S. 334–259 (siehe S. 11).
- [5] Ahmad Taher Hassanien Aboul Ella und Azar. *Brain-Computer Interfaces. Current Trends and Applications*. Schweiz: Springer, 2015 (siehe S. 13, 14).
- [6] Tobias Kaufmann und Beat Pfister. *Sprachverarbeitung. Grundlagen und Methoden der Sprachsynthese und Spracherkennung*. Heidelberg: Springer, 2008 (siehe S. 3, 17).
- [7] Sung-Phil Kim u.a. „Point-and-Click Cursor Control With an Intracortical Neural Interface System by Humans With Tetraplegia“. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 19.2 (Apr. 2011), S. 193–203 (siehe S. 14).
- [8] Zhang Naizhong und Wen Jing. „Hand-Free Head Mouse Control Based on Mouth Tracking“. In: IEEE, 2015 (siehe S. 9, 10).
- [9] Kara Nielsen Jakob und Pernice. *Eyetracking Web Usability*. Berkeley, CA: New Riders, 2010 (siehe S. 4, 5).
- [10] Christina Pahl, Tan Yong Oon und Eko Supriyanto. „Fuzzy System Based Face Tracking for Head Movement Control in Progressive Health Care“. In: IEEE, 2015 (siehe S. 9).
- [11] Raimund Preim Bernhard und Dachsel. *Interaktive Systeme: Band 2. User Interface Engineering, 3D-Interaktion, Natural User Interfaces*. 2. Aufl. Heidelberg: Springer, 2015 (siehe S. 7).
- [12] Ryszard Tadeusiewicz. „Speech in Human System Interaction“. In: IEEE, 2010 (siehe S. 3).

- [13] Sanjay S. Vernon Scott und Joshi. „Brain–Muscle–Computer Interface: Mobile-Phone Prototype Development and Testing“. *IEEE Transactions on Information Technology in Biomedicine* 15.4 (Juni 2011), S. 531–538 (siehe S. 11–13).

Online-Quellen

- [14] Sep. 2012. URL: <http://www.moso-gmbh.de/de/motion-concepts/sondersteuerungen/kinnsteuerung-moso.html> (besucht am 15.04.2017) (siehe S. 7, 8).
- [15] Sep. 2013. URL: <https://www.youtube.com/watch?v=rpA2y5IPY0A> (besucht am 15.04.2017) (siehe S. 8).
- [16] Apr. 2017. URL: <http://www.tobiipro.com/product-listing/tobii-pro-glasses-2/> (besucht am 03.04.2017) (siehe S. 6).
- [17] Apr. 2017. URL: <http://www.tobiipro.com/product-listing/tobii-pro-spectrum/> (besucht am 03.04.2017) (siehe S. 6).
- [18] Apr. 2017. URL: <https://www.smivision.com/eye-tracking/product/eye-trackingproducteye-trackingproductred250mobile-eye-tracker/> (besucht am 05.04.2017) (siehe S. 6).
- [19] Apr. 2017. URL: <http://humanelektronik.de/produkte/seetech/> (besucht am 29.04.2017) (siehe S. 6).
- [20] Apr. 2017. URL: <https://smilemart-tech.com/assistive-technology-products/wheelchair-controls/wheelchair-chin-control/> (besucht am 15.04.2017) (siehe S. 7, 8).
- [21] Apr. 2017. URL: <http://www.sensoryguru.com/product/chin-mouse-controller/> (besucht am 15.04.2017) (siehe S. 7, 8).
- [22] Apr. 2017. URL: <http://www.quadstick.com/> (besucht am 16.04.2017) (siehe S. 8, 9).
- [23] Apr. 2017. URL: http://www.integramouse.com/en/home/highlights-details/?tx_ttnews%5Btt_news%5D=470&cHash=6457ff0d57ee2de8b5c41ac23b53d9b1 (besucht am 15.04.2017) (siehe S. 9).
- [24] Apr. 2017. URL: <https://www.fruehwald.net/mobilitaet/rollstuehle-u-elektrollstuehle/sondersteuerungen> (besucht am 29.04.2017) (siehe S. 10, 11).
- [25] Apr. 2017. URL: <http://www.hidrex.de/reha/rollstuhlsondersteuerung/r-net-system-penny-u-gilles/1-fuss-steuerung-mit-schuhhalter-inkl-joystick> (besucht am 29.04.2017) (siehe S. 10).