

Diplomarbeit

**Computer Vision als neuer Kommunikationsweg der HCI
(Human-Computer Interaction)**

Ausgeführt zum Zweck der Erlangung des Akademischen Grades eines
Dipl.-Ing. (FH) für Telekommunikation und Medien
am Fachhochschul-Diplomstudiengang Telekommunikation und Medien St. Pölten

Unter der Leitung von
LB Dipl.-Ing. (FH) Thomas Zöchbauer

von
Lukas Litzinger
tm0110038062

St. Pölten, am

Unterschrift

Erklärung zum selbständigen Arbeiten

Ich versichere, dass

- ich diese Diplomarbeit selbständig verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt und mich auch sonst keiner unerlaubten Hilfe bedient habe.
- ich dieses Diplomarbeitsthema bisher weder im Inland noch im Ausland einem Begutachter/ einer Begutachterin zur Beurteilung oder in irgendeiner Form als Prüfungsarbeit vorgelegt habe.

Diese Arbeit stimmt mit der vom Begutachter beurteilten Arbeit überein.

St. Pölten, am

Unterschrift

Danksagungen

Ich möchte mich an dieser Stelle herzlich bei allen bedanken, die mir beim Zustandekommen und Realisieren dieser Arbeit geholfen haben. Mein besonderer Dank geht dabei an folgende Personen:

Ich danke Christopher Lindinger, Keyresearcher für Virtuelle Umgebungen im Futurelab des Ars Electronica Centers, dafür, dass er mich in meiner Begeisterung für Medientechnik und Medienkunst bestärkt, und mir unvergessliche Einblicke in die Arbeit im Futurelab ermöglicht hat.

Ich danke meinen Eltern, Andreas und Elisabeth Litzinger, dass sie mir das Studium an der Fachhochschule ermöglicht und finanziert haben.

Abstract (Deutsch)

Wann immer Menschen Computer bedienen, das heißt Informationen zwischen Mensch und Computer ausgetauscht werden, stellen Schnittstellen einen wichtigen Aspekt dar.

Gängige Schnittstellen beschränken sich jedoch hauptsächlich auf visuelle Darstellung der Informationen in Form eines Graphical User Interface (*GUI*).

Alternativ zu herkömmlichen Userinterfaces kann auch der menschliche Körper als Eingabe- und Steuergerät für Anwendungen gesehen werden.

Human-Computer Interaction, oder auch Mensch-Computer Interaktion, ist die Verschmelzung aus Verhaltensforschung und Computer-/Informationstechnologie. Sie beschäftigt sich mit allen Formen und Ausprägungen der Kommunikation zwischen Mensch und Maschine. Das geht vom Wählen beim Telefonieren über das Einschalten der Mikrowelle bis zum Bedienen eines Computerprogramms mit der Maus.

Im medientechnischen Kontext werden natürlich nur Systeme behandelt, die Computer verwenden oder eingebettet haben. HCI Forscher analysieren und entwerfen neuartige Userinterfaces und Interaktionssysteme. Sie waren es, die vor Jahrzehnten den Weg für unsere heutigen GUIs bereiteten.

Heutzutage beschäftigen sich viele HCI Forscher mit der Entwicklung alternativer Input- und Displaytechnologien und möglichen Navigationstechniken in virtuellen Umgebungen. Der neueste und spannendste Bereich dieser Disziplin ist sicher die so genannte Computer Vision.

Computer Vision ermöglicht Computern auf die ursprünglichste aller menschlicher Kommunikationsformen (Gestik, Bewegung) zu reagieren, indem bestimmte Bewegungen oder Körperposen wahrgenommen und unterschieden werden können.

Das Ziel dieser Diplomarbeit ist es, nach der theoretischen Analyse von Computer Vision Systemen als HCI Methode, kostengünstige Tools zur Gestikererkennung oder Computer Vision zu finden, sich damit zu beschäftigen und damit eine Anwendung zu erstellen, die Menschen die Möglichkeit gibt mit einem Computer zu kommunizieren. Die Anwendung soll also auf menschliche Bewegungen reagieren können.

Mit Hilfe der gewonnenen Erkenntnisse soll eine qualitative Aussage getroffen werden, die folgende Frage beantwortet:

In welchen Anwendungsbereichen besitzt eine auf Computer Vision, kurz CV, basierende HCI genug Potential, um in Zukunft eine tragende Rolle zu spielen?

Abstract (English)

Whenever human beings use computers, interfaces play a major role, because information is in some form exchanged between them.

Common interfaces are normally restricted to displaying information in a graphical user interface (*GUI*). Alternatively to such a common user interface the whole human body can be seen as input device for computers or machines.

“Human-Computer Interaction“, short HCI, lies at the intersection between social and behavioural science on the one hand, and computer and information technology on the other. It is concerned with all different ways humans can communicate with machines, whether they are making a simple telephone call, switching on a microwave oven or using the mouse to operate a program on a computer.

In a media engineering context this work will certainly not deal with simple forms of HCI, but focus on systems using computers. HCI scientists analyse and design new user interfaces and interaction systems as for example they have invented the first GUI decades ago.

In these days lots of HCI scientists are trying to invent alternative input and display techniques, and possible ways to navigate through these virtual environments.

In my opinion the most exciting part of this discipline is the so called Computer Vision.

Computer Vision enables computers to react on the most fundamental communication of human beings, namely gestures and movements, because it allows them to recognize and differentiate between certain movements and poses.

The aim of this work is to give a theoretical overview about Computer Vision as a Human-Computer Interaction method, and to find cheap tools to realize Computer Vision. With one of those tools an application, which uses CV as input device and therefore reacts on human movement should be built.

Based on the grown knowledge about Computer Vision an answer to the following question should be given:

In which fields of usage and applications will CV based Human-Computer Interaction be able to play a major roll?

Inhaltsverzeichnis

1. Einführung.....	8
1.1. Motivation.....	8
1.2. Problemstellung	10
1.3. Forschungsleitende Fragestellung.....	10
1.4. Forschungsmethode	11
1.5. Forschungsstrategie.....	11
2. Grundlegendes zur Kommunikation.....	12
2.1. Einführung	12
2.2. Kommunikationsmodelle.....	13
2.2.1. Grundbegriffe von Kommunikationsmodellen.....	13
2.2.2. Das Sender - Empfänger Modell.....	14
2.3. Ziele und Absichten von Kommunikation.....	15
2.4. Kommunikationsformen	16
2.4.1. Grundsätzliche Unterscheidung.....	16
2.4.2. Hierarchische Unterscheidung.....	17
2.4.3. Richtungs-Unterscheidung.....	18
2.5. Mensch-Mensch Kommunikation.....	19
2.6. Computer-Computer Kommunikation.....	20
2.7. Mensch-Computer Kommunikation	22
2.8. Kommunikation und Interaktion.....	24
3. Human-Computer Interaction (HCI)	26
3.1. Einführung	26
3.2. HCI Modelle	29
3.3. Erklärende Modelle oder Beschreibungsmodelle	31
3.3.1. Wahrnehmungsmodell	31
3.3.2. Information processing model	33
3.4. Generative oder Gestaltungsmodelle	34
3.4.1. Interaction-Design.....	34
3.4.2. Software-Ergonomie	35
3.5. Vorhersagende oder Propheziungs-Modelle.....	39
3.5.1. Motorische Verhaltensmodelle	39
3.5.2. User Control Modell	40
3.5.3. GOMS	41
3.5.4. Menschliches Gedankenmodell	43
3.6. Ziele und Absichten von HCI	44
3.7. HCI Formen	45
3.7.1. Inputsysteme	45
3.7.2. Outputsysteme.....	45
3.7.3. Herkömmliche Benutzerschnittstellen	46
3.8. Ausblick und Visionen.....	47
3.8.1. New I/O devices.....	48
3.8.2. Augmented Reality	50
3.8.3. Virtual Reality.....	51
3.8.4. Mass availability	52
3.8.5. Ubiquitous computing.....	52

4. Computer Vision	53
4.1. Einführung	53
4.2. Grundlegendes zur Computer Vision.....	54
4.3. Computer Vision Methoden.....	57
4.3.1. Bildverarbeitung	58
4.3.2. Bildanalyse.....	59
4.4. Anwendungsbeispiele & Zukunftsperspektiven	62
4.4.1. Perceptual User Interfaces	63
4.4.2. Blickrichtungsanalyse	64
4.4.3. Zutritts- & Zugangskontrolle	65
4.4.4. Industrielle Nutzung.....	65
4.4.5. Alltägliche Nutzung	66
4.4.6. Infotainment	67
5. Entwicklung einer Beispielanwendung.....	69
5.1. Idee und Begründung	69
5.2. Ziele der Anwendung.....	69
5.3. Entwicklungsumgebungen	70
5.3.1. Intel Open CV Library	70
5.3.2. Video By The Pixel.....	70
5.3.3. Isadora.....	71
5.3.4. Jitter & Cyclops	71
5.3.5. EyesWeb	71
5.3.6. Myron.....	71
5.3.7. TrackThemColors	71
5.4. Testen verschiedener Entwicklungsumgebungen	72
5.4.1. EyesWeb	73
5.4.2. TrackThemColors	77
6. Erkenntnisse der Arbeit	80
7. Anhänge	82
7.1. Abbildungsverzeichnis.....	82
7.2. Literatur- und Quellenverzeichnis	83
7.2.1. Bücher	83
7.2.2. Online Quellen	84
7.2.3. Weblinks	87
7.3. Glossar	90

1. Einführung

1.1. Motivation

Im Rahmen der fächerübergreifenden Projektarbeit im fünften und sechsten Semester meines Studiums an der Fachhochschule St. Pölten haben einige Mitstudenten/innen und ich eine Anwendung für das Ars Electronica Center in Linz entwickelt.

Das Ziel dieser Anwendung war es, Projektionen auf allen drei Fassaden des Ars Electronica Centers interaktiv zu gestalten.

Uns wurde bei diesem Projekt die ehrenvolle Aufgabe zu Teil, die Projektionen vom Konzept bis zur Installation umzusetzen.

Unser Konzept versuchte, die drei Projektionsflächen zu einem Ganzen zusammen zu fassen, und wir entwarfen eine virtuelle Welt, die sich über alle drei Fassaden des Ars Electronica Centers zog. Die Interaktion sollte laut Konzept so aussehen, dass jeder Passant, der sich vor dem Center aufhielt, einen virtuellen Gegenspieler oder Schatten zugeordnet bekommt, der ihm auf Schritt und Tritt folgt. Dieser sollte dann Interaktionen mit bestimmten Objekten oder Personen in der virtuellen Welt durchführen können, und somit eine Verschmelzung zwischen realer und virtueller Welt ermöglichen.



Abbildung 1: Konzeptvisualisierung AEC Projekt

Allerdings wurde schnell klar, dass die Umsetzung dieses Konzepts ohne der tatkräftigen Unterstützung der Mitarbeiter des AEC Futurelabs in der vorgegebenen Zeit fast unmöglich gewesen wäre.



Abbildung 2: AEC Projekt Südfassade

Deshalb wurde das Projekt in zwei Aufgabebereiche unterteilt.

Das Team des Futurelabs ermöglichte mittels dreier am Dach des AEC angebrachter Kameras und einer eigens für diese Anwendung entwickelten Computer Vision Software die Interaktion.

Wir entwickelten eine Applikation, die es ermöglichte, Daten aus der Computer Vision Software zu empfangen, und konnten somit den eigentlich abstrakten Positionsdaten das virtuelle Gesicht verleihen, in dem wir jedem realen Menschen einen virtuellen Charakter in unserer dreidimensional scheinenden Welt zuordneten.

In diesem Projekt wurde ich zum ersten Mal auf Computer Vision aufmerksam und war sofort davon fasziniert, wie es möglich war, Personen auf einem Video zu erkennen und zu verfolgen.

Die enge Zusammenarbeit mit dem AEC Futurelab hat mich dafür begeistert, gut ein Jahr später Computer Vision als Thema für meine Diplomarbeit zu wählen, um nun selbst herauszufinden, wie Computer Vision funktioniert, wo sie momentan eingesetzt wird und möglicherweise in Zukunft Verwendung finden könnte.



Abbildung 3: AEC Projekt Projektion Timeshift Festival

1.2. Problemstellung

Die Problematik von Computer Vision als Methode der Mensch-Computer Interaktion liegt in der Erstellung kommerziell nutzbarer Anwendungen. Obwohl schon seit Jahrzehnten an diesem Teilgebiet der künstlichen Intelligenz geforscht wird, gelang es erst vor kurzem, sinnvoll nutzbare Anwendungen zu entwickeln.

Es gilt daher herauszufinden, in wie weit diese neuartige Form der Interaktion zwischen Mensch und Computer eingesetzt wird, und vor allem in welchen Anwendungsgebieten Computer Vision überhaupt verwendet werden kann.

Weiters ist es mir ein Anliegen, Entwicklungsumgebungen und Tools zu finden, die ein breit gefächertes Einsatzgebiet ermöglichen, um so die zukünftige Entwicklung der Computer Vision besser einschätzen und bewerten zu können.

1.3. Forschungsleitende Fragestellung

Das Ziel dieser Arbeit ist es, folgende Frage zu diskutieren:

In welchen Anwendungsbereichen besitzt eine auf Computer Vision basierende HCI genug Potential, um in Zukunft eine tragende Rolle zu spielen?

Hierbei gilt es vorab das Begriffspaar Kommunikation und Interaktion genau zu erläutern und ihre Ausprägungen im Bezug auf Mensch-Computer Interaktion zu verdeutlichen.

Weiters gilt es herauszufinden, wann und wie Computer Vision als Methode der Mensch-Computer Interaktion, im Englischen als *HCI* (Human-Computer Interaction) abgekürzt, genutzt und eingesetzt wird. Es ist zu klären, in welchen Anwendungsbereichen Computer Vision derzeit zum Einsatz kommt, um daraus mögliche Schlussfolgerungen auf zukünftige Anwendungsgebiete zu ermöglichen.

Wobei hier mit „zukünftig“ der Zeitraum der nächsten Jahre gemeint ist und der Begriff „tragende Rolle“ nicht nur den kommerziellen Einsatz von Computer Vision zu messen versucht, sondern auch die vielseitigen Einsatzmöglichkeiten dieser HCI Methode zu bewerten sucht.

1.4. Forschungsmethode

Vom wissenschaftlichen Standpunkt aus betrachtet, ist diese Arbeit einerseits eine Inhaltsanalyse, die sich mit den Themen HCI als Kommunikationsform und Computer Vision als Methode der Mensch-Computer Interaktion beschäftigt. Andererseits kann diese Arbeit auch als Experiment betrachtet werden, da versucht wird, eine Anwendung zu erstellen, die Computer Vision nutzt, um Informationsaustausch zwischen Mensch und Computer zu ermöglichen.

1.5. Forschungsstrategie

Zu allererst wird sich diese Arbeit allgemein mit dem Thema Kommunikation beschäftigen, um zu klären, inwiefern Kommunikation zwischen Mensch und Computer stattfindet.

Danach soll die Beschäftigung mit Mensch-Computer Interaktion einen Überblick über diesen, der Informatik zuordenbaren Forschungsbereich geben und gängige Begriffe erläutern.

Es ist zu klären, mit welchen Mitteln Menschen mit leblosen Gegenständen wie Computern interagieren, beziehungsweise kommunizieren und in welchen Anwendungsbereichen Computer Vision derzeit eines dieser Medien darstellt.

Nach der Auseinandersetzung mit Computer Vision als Disziplin der künstlichen Intelligenz, wird ein Überblick aktueller Anwendungsbeispiele gegeben, um zukünftige Anwendungsgebiete und Potentiale dieser Methodik abschätzen und bewerten zu können.

Um das Potential von Computer Vision besser beurteilen und verstehen zu können, folgt eine persönliche Auseinandersetzung mit Computer Vision Tools.

Anhand einer selbst erstellten Beispielanwendung werden dann die Schwierigkeiten und Probleme der Computer Vision veranschaulicht und greifbar gemacht.

Mit Hilfe des im theoretischen und praktischen Teil dieser Arbeit gesammelten Wissens, findet im letzten Abschnitt die Beantwortung der forschungsleitenden Fragestellung statt.

2. Grundlegendes zur Kommunikation

2.1. Einführung

Um die Kommunikation und die daraus resultierende Interaktion zwischen Mensch und Maschine besser verstehen und begreifbar machen zu können ist es wichtig, sich zuerst allgemein mit dem Thema Kommunikation auseinander zu setzen.

Dabei wird schnell klar, dass es sich als schwierig erweist, eine allgemeine Definition für den Begriff Kommunikation zu geben, die einerseits präzise genug ist, aber andererseits für alle Bereiche Geltung hat.

Daher möchte ich an dieser Stelle darauf hinweisen, dass diese Arbeit nur einen allgemeinen Überblick über Kommunikation geben wird, da sie sich im Weiteren näher mit Mensch-Maschine Interaktion befasst.

Das Wort Kommunikation ist eine Kombination aus den lateinischen Wörtern *cum* (gemeinsam, zusammen) und *munire* (binden, bauen), und bezieht sich auf den wechselseitigen Austausch von Gedanken in Sprache, Schrift oder Bild.

Allerdings bedarf diese Definition einiger Interpretation und weiterer Ausführungen. Beispielsweise ist hier noch unklar, ob die Teilnehmer einer Kommunikation immer *Individuen* sein müssen, oder menschliches Bewusstsein die Voraussetzung für Kommunikation ist.

Weiters steht zur Diskussion, ob es Kommunikation ohne aktives Verständnis oder Erinnerung überhaupt geben kann, und ob intentionales Handeln die Grundlage der Kommunikation darstellt.

In technischen Disziplinen wird Kommunikation ganz anders betrachtet, und zwar als Prozess. Nämlich als der Vorgang des Informationsaustausches zwischen zwei komplexen Systemen, wobei Informationsaustausch als das wechselseitige Übermitteln von Daten oder Signalen mit festgelegtem Bedeutungsinhalt definiert wird.

In beiden Ansätzen ist jedoch unumstritten, dass Kommunikation ein, wenn nicht das wesentlichste, Mittel zum Erlangen von Information und Wissen darstellt.

Erst das Sammeln von Daten erlaubt es einem Individuum (oder auch einem Gerät), sich ein Bild oder Modell der Umwelt zu machen und somit jede neu erlangte Information in einen sinnvollen Kontext zu setzen.

Jedes Verhalten kann als Kommunikation gedeutet werden, da jede Reaktion auf etwas oder jemanden einen gewissen Mitteilungscharakter besitzt. Da Verhalten kein Gegenteil besitzt, man sich also nicht - nicht verhalten – kann, folgt daraus, dass es unmöglich ist, nicht zu kommunizieren: "Man kann nicht nicht kommunizieren" (zit. n. Burkart, 2003, S. 9) (erstes von 5 Axiomen zur Kommunikation von Paul Watzlawick, österreichischer Psychotherapeut).

Geht man soweit zu sagen, dass Kommunikation keines intentionalen Senders bedarf, ist Kommunikation die einzige Möglichkeit zu lernen, da dann jedes Wahrnehmen als kommunikativer Vorgang zu sehen ist.

Beschränkt man den Begriff Kommunikation jedoch auf das Wesentliche, nämlich das wechselseitige Übermitteln von Daten oder Signalen, also rein auf den Informationsaustausch, ohne zu berücksichtigen ob dieser Vorgang zwischen Individuen, nicht-menschlichen Lebewesen, Objekten oder Systemen vollzogen wird, wird klar, dass auch zwischen Menschen und Maschinen, also Computern, kommuniziert werden kann.

2.2. Kommunikationsmodelle

Ein Modell an sich ist eine vereinfachte Darstellung der Realität, und hat das Ziel einen komplizierten Sachverhalt leichter verständlich zu machen. Mit Hilfe eines Modells können auch Rückschlüsse auf den ursprünglich darzustellenden komplexeren Sachverhalt getroffen werden. Allerdings ist ein Modell nie als mathematisches Äquivalent zu sehen, sondern eher als eine Metapher.

Im Bezug auf Kommunikation gibt es eine Vielzahl von Modellen, die alle unterschiedliche Ansätze und Auffassungen von Kommunikation haben.

Die Kommunikationsforschung unterscheidet eine Vielzahl von Kommunikationsformen, je nach Absicht oder Intention der Kommunikation, oder zum Beispiel nach der Anzahl der Teilnehmer.

Um die Unterscheidungsmerkmale dieser Kommunikationsformen im nächsten Abschnitt greifbarer zu machen, wird erst eines der gängigsten Kommunikationsmodelle erläutert.

Das Sender-Empfänger Modell von Shannon und Weaver bildet meiner Meinung nach die Grundlage aller Kommunikationsmodelle, da sich auch andere Modelle auf die von diesem Modell geprägten Grundbegriffe beziehen.

2.2.1. Grundbegriffe von Kommunikationsmodellen

- Sender: Der Initiator einer Kommunikation.
- Empfänger: Der Rezipient wird vom Sender angesprochen.
- Kanal: Das Medium über das die Informationsübertragung geschieht.
- Kontext: Die Situation in der kommuniziert wird.
- Botschaft: Die tatsächlich übermittelte Nachricht.
- Code: Die Codierung der Nachricht (verwendeter Zeichensatz).

2.2.2. Das Sender - Empfänger Modell

Im Jahre 1948 haben Claude E. Shannon und Warren E. Weaver, zwei amerikanische Mathematiker, ein sehr allgemeines Modell der Kommunikation geschaffen, das üblicherweise auch Transmissionsmodell genannt wird. Das Shannon-Weaver Modell sieht Kommunikation als Nachrichtenvermittlung und geht von folgendem Sachverhalt aus: „Eine Nachrichtenquelle produziert eine Information, die von einem Empfänger aufgenommen wird und dort an ihr Ziel gelangt. Zu diesem Zweck muss die Nachricht von einem Sender in ein dem Übertragungskanal angemessenes Signal umgeformt (encodiert) werden“ (Burkart, 2003, S.427). Daraus lassen sich sechs Faktoren ableiten, deren Kommunikation bedarf:

- eines Senders
- eines Encoders
- einer Nachricht
- eines Kanals
- eines Decoders
- eins Empfängers

Anhand dieser sechs Elemente ist schon zu erkennen, dass dieses Modell eine eher technikorientierte Vorstellung von Kommunikation hat, und dass keinesfalls festgelegt ist, dass nur Menschen miteinander kommunizieren können.

Es ist eines der frühen Kommunikationsmodelle und erinnert stark an technische Modelle der Informationsübertragung, wie wir sie aus der Informatik kennen. Deshalb wird es von Fachleuten auch oft als das „Informations-Modell“ bezeichnet.

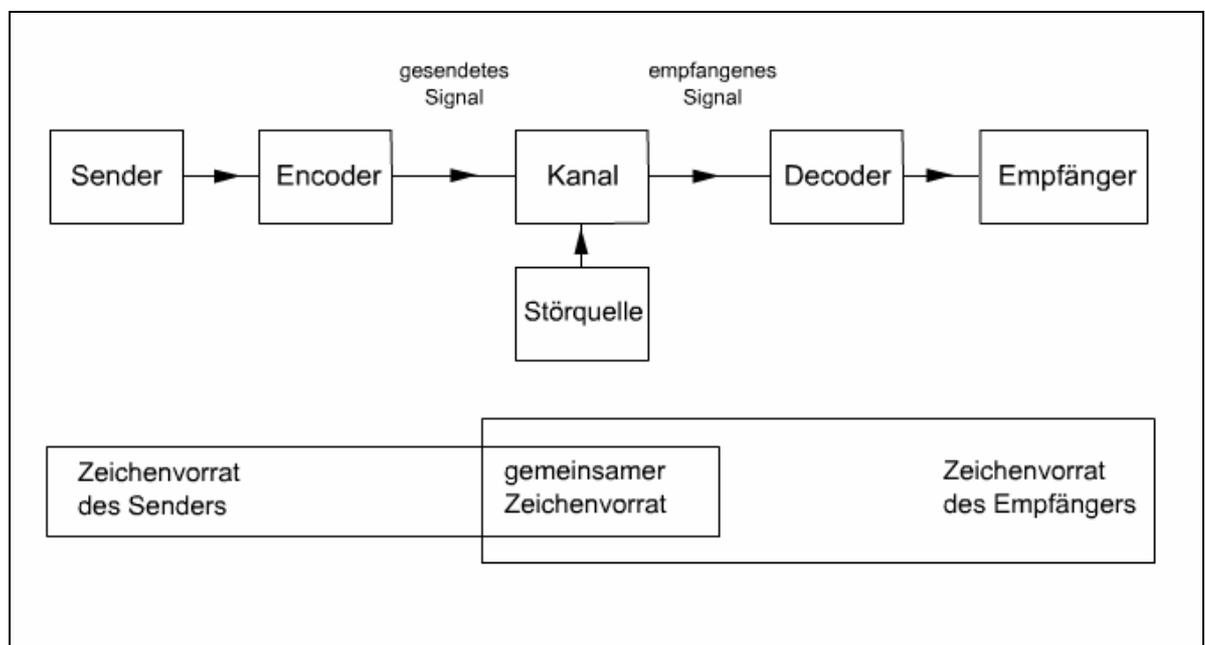


Abbildung 4: Sender-Empfänger Modell (vgl. Burkart, 2003, S.427)

Das Sender-Empfänger-Modell der Kommunikation definiert Kommunikation als Übertragung einer Nachricht von einem Sender zu einem Empfänger. Dazu wird die Nachricht kodiert und als Signal über einen Übertragungskanal übermittelt. Dabei kann

die Nachricht durch Störungen verfälscht werden. Für die erfolgreiche Kommunikation ist es eine Voraussetzung, dass Sender und Empfänger die gleiche Kodierung für die Nachricht verwenden.

Dieses Kommunikationsmodell deckt sich mit dem Konzept der Kommunikation wie es in der Philosophie, der Nachrichtentechnik und der klassischen Informationstheorie verwendet wird und ist deshalb auch sehr weit verbreitet.

Es reicht jedoch in anderen Bereichen kaum aus, Kommunikation allumfassend zu beschreiben. Beispielsweise kann in einem Streitgespräch auch das Ausbleiben einer Nachricht, also das reine Schweigen, eine Information übertragen.

2.3. Ziele und Absichten von Kommunikation

Das augenscheinlichste Ziel jeder Kommunikation ist sicherlich der Informationsaustausch zwischen den Beteiligten.

Stellt man allerdings die Frage: „Warum wird kommuniziert?“, fragt man also nach der eigentlichen Absicht, die hinter Kommunikation steckt, gilt es herauszufinden, was der eigentliche Grund ist, eine Nachricht zu senden.

Hierfür können mehrere Umstände ausschlaggebend sein. Ist die gesendete Nachricht zum Beispiel eine „Antwort“ auf eine zuvor getätigte „Frage“ ist klar, dass das Senden als Reaktion zu betrachten ist und das Ziel der Nachricht die Weitergabe des Informationsgehaltes ist.

In vielen Fällen hat eine gesendete Nachricht jedoch nicht nur die Absicht mit der enthaltenen Information im Empfänger Wissen entstehen zu lassen, sondern oft auch den Empfänger als Resultat zu bestimmten Handlungen zu bewegen, also auf sein Handeln einzuwirken. Ist es tatsächlich die Absicht des Senders den Rezipienten in seinem Handeln und Denken zu beeinflussen, kann von „Überredungskommunikation“ (persuasive communication) (Burkart. 2003, S.465) gesprochen werden.

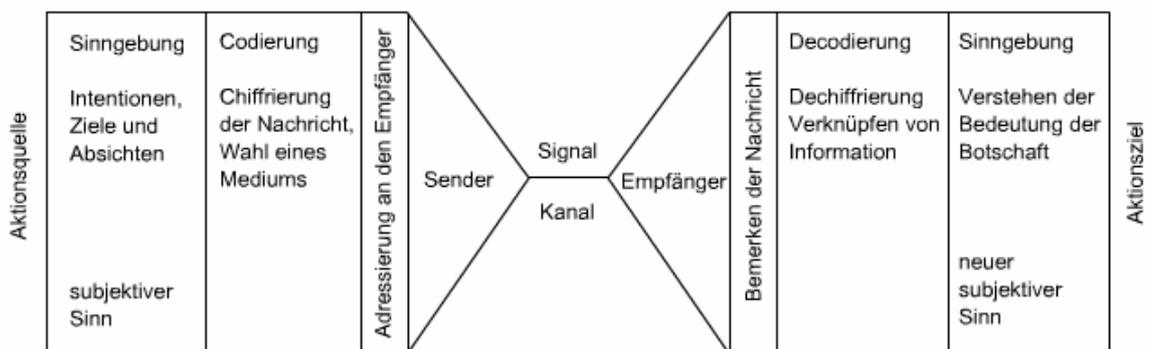


Abbildung 5: Kommunikationsziele (vgl. Burkart, 2003, S.430)

2.4. Kommunikationsformen

Nachdem die kurze Abhandlung des Sender-Empfänger Modells einen Einblick in das wissenschaftliche Verständnis von Kommunikation gegeben hat, sollten die Unterscheidungsmerkmale der einzelnen Kommunikationsformen greifbarer und leichter verständlich sein. Außerdem sollte sie die im Weiteren von mir getroffene Einteilung von Kommunikation in drei Bereiche verständlich machen.

„Kommunikation kann in verschiedenen Arten vor sich gehen: direkt oder indirekt; wechselseitig oder einseitig, privat oder öffentlich, sowie in Anwesenheit oder in Abwesenheit (und damit gegenseitig wahrnehmbar oder nicht wahrnehmbar)“ (Pürer, 2001, S.13). Daher kann Kommunikation auch auf unterschiedlichste Art und Weise kategorisiert und unterteilt werden.

2.4.1. Grundsätzliche Unterscheidung

Hier wird Kommunikation anhand der beteiligten Individuen auf Sender- und Empfänger-Seite unterschieden.

Intimkommunikation

Hierunter versteht der Kommunikationsforscher eine Kommunikationsform, bei der ein Individuum mit sich selbst oder mit einem engen Kreis von Vertrauten kommuniziert (z.B.: Tagebuch).

Individualkommunikation

Individuelle oder auch interaktive Kommunikation bezeichnet eine Form des Kommunizierens, bei der einzelne Individuen wechselseitig Informationen austauschen (z.B.: Telefongespräch).

Gruppenkommunikation

Gruppenkommunikation bezeichnet eine Kommunikation, an der mehr als zwei Individuen beteiligt sind, etwa bei einem Meeting oder einer Sitzung (z.B.: Taxifunk).

Wahlkommunikation

Wahlkommunikation bezeichnet jene Kommunikation, bei der für eine Vielzahl von Empfängern mehrere Nachrichten zur Verfügung gestellt werden, wie etwa in einer Bibliothek (z.B.: Internet).

Massenkommunikation

Massenkommunikation, auch öffentliche Kommunikation genannt, bezeichnet eine Kommunikation zwischen einem Sender und einer unzählbaren Empfängerschaft, wobei die Kommunikation an sich indirekt und nicht wechselseitig ist, und in den meisten Fällen über ein Massenmedium (z.B.: Fernsehen) geschieht.

Interne Kommunikation

Interne Kommunikation wird als verbale und nonverbale Kommunikation zwischen Angehörigen einer bestimmten Gruppe verstanden, mit Sinn und Zweck der Optimierung organisatorischer Abläufe, Informationsverbreitung und Austausch (z.B.: Handzeichen bei Tauchern).

2.4.2. Hierarchische Unterscheidung

Hierarchische Unterscheidung ist ein Begriff, den Klaus Merten (deutscher Kommunikationswissenschaftler) nach der Untersuchung von über 160 Begriffsbestimmungen 1977 prägte.

Demnach unterscheidet er folgende Formen der Kommunikation, bei denen er meiner Meinung nach nicht nur auf Grund der Anzahl der beteiligten Individuen, sondern auch auf Grund der Art der beteiligten Individuen, Kategorien bildet (vgl. Pürer, 2001, S.3).

Subanimalische Kommunikation

Subanimalische Kommunikation ist die organisierende Kommunikation zwischen einzelnen Organismen oder Systemen. Sie beinhaltet sowohl technische als auch naturwissenschaftliche Phänomene, wie etwa die gegenseitige Beeinflussung zweier Magneten.

Animalische Kommunikation

Animalische Kommunikation bezeichnet die Kommunikation zwischen Lebewesen, egal ob zwischen Tieren oder gar zwischen Tieren und Menschen.

Humankommunikation

Dieser Begriff beschreibt ausschließlich die Kommunikation zwischen Menschen.

Massenkommunikation

Massenkommunikation ist eine Art der Humankommunikation, die auf technische Hilfsmittel angewiesen ist, meist einseitig ist und sich an ein *dispersed Publikum* richtet.

Computervermittelte Kommunikation

CVK oder CMC (computer mediated communication) ist eine Kommunikationsform, die aus der Fusion zwischen Telekommunikation, Computerisierung und herkömmlicher elektronischer (Massen-) Kommunikation entstanden ist.

2.4.3. Richtungs-Unterscheidung

Hier wird unterschieden, ob mehrere Empfänger oder auch mehrere Sender an einer Kommunikation beteiligt sind (vgl. Pürer, 2001, S.27).

- One to one Kommunikation (face-to-face)
Ein Sender, ein Empfänger
unidirektional - (z.B.: SMS, Brief, E-Mail)
bidirektional - (z.B.: Gespräch, Telefonat)

- One to many Kommunikation
Ein Sender, viele Empfänger
unidirektional - (z.B.: Befehl, Rede, Fernsehen)
bidirektional - (z.B.: interaktive Webseiten)

- Many to one Kommunikation
Viele Sender, ein Empfänger
unidirektional - (z.B.: TED)

- Many to many Kommunikation
Viele Sender, viele Empfänger
polydirektional – (z.B.: Diskussionsrunde, Forum)

Diese unterschiedlichen Unterscheidungskonzepte zeigen deutlich, dass Kommunikation nicht einfach und vor allem auf unterschiedlichste Art und Weise zu kategorisieren ist (vgl. Pürer, 2001, S.13).

Vom medientechnischen Betrachtungswinkel ist es mir daher wichtig, Kommunikation in drei grundlegende Bereiche zu unterteilen und diese auch getrennt zu behandeln, da die Kommunikation in diesen Bereichen auf höchst unterschiedliche Art und Weise geschieht.

Hierbei erhebe ich allerdings keinerlei Anspruch auf eine vollständige Kategorisierung aller Kommunikationsarten, noch möchte ich bestehende Unterscheidungen in Frage stellen. Mir ist hingegen das Einbinden der Kommunikation zwischen Mensch und Computer, die heutzutage täglich geschieht, in die Kommunikationskategorien wichtig.

- Mensch-Mensch Kommunikation:
Mensch-Mensch Kommunikation bezeichnet meinem Verständnis nach die Kommunikation zwischen Menschen. Sie kann bi- oder unidirektional sein, über ein Medium, durch Bilder, mittels Schrift, durch Sprache oder durch Gesten vermittelt werden.
- Computer-Computer Kommunikation:
Hierunter verstehe ich das Kommunizieren von Computern, Programmen oder Prozessen untereinander.
- Mensch-Computer Kommunikation
Darunter verstehe ich jegliche Arten der Kommunikation zwischen Mensch und Maschine, sei es das Bedienen eines Programms auf einem PC über ein GUI, oder das Einschalten der Mikrowelle.

2.5. Mensch-Mensch Kommunikation

Diese Form der Kommunikation ist wahrscheinlich die komplexeste und am höchsten entwickelte Kommunikationsform, die es überhaupt gibt. Menschen haben unterschiedlichste Möglichkeiten gefunden, miteinander zu kommunizieren, und nutzen die unterschiedlichsten Medien, um Nachrichten zu übermitteln.

Die Grundfesten der menschlichen Kommunikation sind sicherlich Gestik, Sprache und Schrift, wobei die älteste dieser drei sicherlich die nonverbale Kommunikation ist. Menschliches Sozialverhalten ist nicht ohne Berücksichtigung des nonverbalen Systems zu verstehen. Allerdings wird die Bedeutung der „nichtsprachlichen“ Verständigung von den meisten Menschen unterschätzt.

Wenn Menschen nicht verbal kommunizieren, spricht der Körper alleine. Er ist niemals stumm und teilt auch dann, über Signale der Selbstversunkenheit und Abschirmung mit, dass uns zurzeit Kontakte unerwünscht sind.

Der Eindruck, den die Körpersprache macht, ist oft sehr mächtig und Worte haben es schwer, ihn zu dementieren.

Da die Körpersprache auch schwerer bewusst zu beherrschen ist als die verbale Sprache, sind die Botschaften der Körpersprache oft "wahrer", bzw. "echter". Und eben weil sie sich dem Willen weitgehend entzieht, und zwar das Aussenden, wie der Empfang der Signale gleichermaßen, erscheint sie uns selbstverständlicher, irrationaler und sie wird weit weniger bewusst bemerkt als die Sprache der Worte.

Die Sprache ist uns wohl die geläufigste aller menschlichen Kommunikationsformen, und bietet eine bewusste und präzisere Möglichkeit, anderen Menschen Ereignisse zu schildern, Emotionen oder Gefühle auszudrücken.

"Sprache ist eine ausschließlich dem Menschen eigene, nicht im Instinkt wurzelnde Methode zur Übermittlung von Gedanken, Gefühlen und Wünschen mittels eines Systems von frei geschaffenen Symbolen." (John Lyons, 1992, S. 13).

Schrift ist meiner Auffassung nach die Bild gewordene Sprache und bietet Menschen die Möglichkeit, die flüchtige Sprache zeitlich zu bannen.

Informationen konnten dank der Entwicklung der Schrift über lange Zeit festgehalten werden, somit ist die Schrift wohl eine der bedeutsamsten Entwicklungen der menschlichen Kommunikationsgeschichte.

Mensch-Mensch Kommunikation bezeichnet für mich den wechselseitigen Informationsaustausch zwischen Menschen in allen möglichen Ausprägungen.

2.6. Computer-Computer Kommunikation

Betrachtet man diverse Kommunikationsmodelle etwas abstrakter und setzt man nicht voraus, dass menschliches Bewusstsein die Grundlage eines Kommunikationsprozesses bildet, kann man durchaus zu dem Schluss kommen, dass auch Computer oder eigentlich Prozesse des Betriebssystems eines Computers miteinander kommunizieren.

Prozesskommunikation in Betriebssystemen ist heutzutage nicht mehr wegzudenken. Der Abstraktionsgrad und die Effizienz, mit der man diese Art der Prozesskommunikation betreiben kann, differiert je nach Mechanismus erheblich. In der Informatik werden hierbei drei Arten der „Process Interaction“ unterschieden, die sich am Grad der Abhängigkeit der Prozesse zueinander orientieren.

Abhängigkeit	Relation	Gegenseitiger Einfluss	Potentielle Probleme
unabhängig	competition	Keine Beeinflussung der Prozesse untereinander	
indirekt abhängig	cooperation durch sharing	Ergebnis von Prozess A kann vom Ergebnis von Prozess B abhängig sein	Mutual exclusion Deadlock Starvation Datenabhängigkeit
direkt abhängig	cooperation durch Kommunikation	Ergebnis von Prozess A kann vom Ergebnis von Prozess B abhängig sein	Deadlock Starvation

Tabelle 1: Process Interaktion (Schöndorfer, 2001, S. 6)

competition

Competition von Prozessen wird hier der Vollständigkeit halber erwähnt, da diese Art der Prozessinteraktion davon ausgeht, dass Prozesse nicht von einander abhängig sind und daher nicht kommunizieren müssen.

cooperation by sharing

Unter „Sharing“ ist zu verstehen, dass sich zwei oder mehrere Prozesse bestimmte Ressourcen, wie zum Beispiel Speicher oder Drucker etc., teilen. Sie sind daher von einander indirekt abhängig.

Deshalb müssen diese Prozesse synchronisiert werden. Das heißt, dass sie etwa über global zugängliche Flags bekannt geben, ob sie bestimmte Ressourcen in Verwendung haben, andernfalls muss gewartet werden. (Sharing bedarf komplizierter Algorithmen um die Integrität von Daten zu sichern und nicht zu *Deadlocks*, *Starvation*, oder *Mutual Exclusion* zu führen, die hier nicht weiter ausgeführt sind.)

cooperation by communication

Typischerweise werden hier unter den Prozessen Messages ausgetauscht, um Informationen über Ressourcen und deren Verwendung bekannt zu geben.

Die direkte Prozesskommunikation oder Message-Passing bedarf natürlich auch genauer Synchronisation, da Prozesse nicht gleichzeitig senden und empfangen können.

Ich denke, dieser kurze Exkurs in die Informatik zeigt deutlich, dass die Kommunikation zwischen Prozessen ein sehr komplexes Gebiet ist.

Obwohl es zunächst den Anschein haben mag, dass Computer- oder Prozesskommunikation nichts mit menschlicher Kommunikation gemeinsam haben, erkennt man doch Parallelen.

So verbietet unter Menschen etwa „der gute Ton“ das Unterbrechen eines Gesprächspartners und bei Prozessen wird durch strikte Algorithmen sichergestellt, dass immer nur ein Prozess sendet und alle anderen empfangen, um einen reibungslosen Ablauf zu ermöglichen.

Somit ist klar, dass nicht nur Menschen kommunizieren, sondern auch Prozesse untereinander Nachrichten austauschen und eine Form von Kommunikation betreiben.

2.7. Mensch-Computer Kommunikation

Ein Mensch-Computer-Dialog ist in erster Linie symbolisch, das heißt durch Zeichen vermittelte Kommunikation. Neben graphischen Zeichen, z.B. Icons oder Layout-Gestaltung, sind es sprachliche Zeichen, die den Verlauf der Mensch-Computer-Kommunikation beeinflussen.

Da es zu untersuchen gilt, inwieweit es sich bei Mensch-Computer-Dialogen um eine Form von Kommunikation handelt, stellt sich die Frage, ob die Interpretationsmuster, die natürlicher Kommunikation zugrunde liegen, auf Mensch-Computer-Dialoge übertragbar sind. Dabei ist zu berücksichtigen, dass Mensch-Computer-Kommunikation eine sehr neue Form der Kommunikation ist.

Objektiv betrachtet werden von Computern oder Maschinen jedoch keine wirklich intentionalen oder zielgerichteten Handlungen getätigt, da Computer ja „nur“ das tun, worauf sie programmiert wurden und somit keine eigenen Absichten oder Ziele verfolgen.

Trotzdem kann man meiner Ansicht nach von Kommunikation oder zumindest Interaktion zwischen Menschen und Computern oder Maschinen sprechen, da ein Computer ja auf das Verhalten eines Menschen reagiert und umgekehrt. Es werden also Informationen zwischen Mensch und Computer ausgetauscht.

Mensch-Computer Kommunikation genau zu definieren ist schwierig, allerdings kann man meiner Meinung nach einige Voraussetzungen für Interaktion zwischen Menschen und Computern festlegen.

- Das Zusammenwirken von Mensch und Computer ist durch einen **Austausch von Informationen** gekennzeichnet. Der energetische Aspekt des Zusammenwirkens tritt dagegen in den Hintergrund. (Ein Tischler mit einer Handkreissäge bildet im medientechnischen Kontext also noch kein Mensch-Computer-System.)
- Der Informationsaustausch ist **wechselseitig**. Ein Informationsfluss in eine Richtung (wie beim Betrachten einer Anzeigetafel in der Abfertigungshalle eines Flughafens) reicht nicht aus. Der Mensch muss auch eingreifen können.
- Der Informationsaustausch erfolgt wenigstens zum Teil **vermittelt**, das heißt nicht alle Information über die Maschine, den Prozess und die Umgebung sind direkt wahrnehmbar. Typisch für Mensch-Computer-Systeme im medientechnischen Kontext ist, dass Daten über Messgeräte und Sensoren erfasst, transformiert, übertragen und über spezielle Anzeigen (z.B. Zeiger und Skalen, Zählwerke, Bildschirmanzeigen) wahrnehmbar gemacht werden. Auch die Eingriffe des Menschen erfolgen weitgehend vermittelt, das heißt über Bedienelemente (z.B. Schalter, Knöpfe, Schieberegler, Tastaturen).
- In einem medientechnisch relevanten Mensch-Computer-System verfolgt der **Mensch bestimmte Ziele** und bedient sich des Computers zur Umsetzung dieser. Das kann beispielsweise das Erstellen einer Grafik auf einem PC sein. (vgl. Fachgruppe Arbeits- und Organisationspsychologie, 2005)

Genau wie die Kommunikation zwischen zwei Menschen, findet laut der Deutschen Gesellschaft für Psychologie (Fachgruppe für Arbeits- und Organisationspsychologie (Weblink [1]) auch die Mensch-Computer Kommunikation auf mehreren Ebenen statt.

Aufgabenebene

In der Aufgabenebene geht es um die Übereinstimmung zwischen den Zielen des Menschen und dem Funktionsangebot der Maschine.

Dabei ist es vorrangig wichtig, dass der Mensch seine Ziele überhaupt erreichen kann, wobei der Grad der Zielerreichung, also wie genau das vom Menschen angestrebte Ziel umgesetzt wird, als Effektivität bezeichnet wird.

Semantische Ebene

In der Semantischen Ebene geht es um die Übereinstimmung zwischen den Denkprozessen des Menschen und der Daten- oder Algorithmenstruktur des Computers.

Menschen haben eine Vorstellung wie eine Maschine funktioniert. Stimmt dieses mentale Modell mit der tatsächlichen Funktionsweise des Geräts überein, fördert das den Informationsaustausch zwischen Mensch und Maschine.

Syntaktische Ebene

Die Syntaktische Ebene beschäftigt sich mit der Strukturierung von Anzeigen und Ausgaben auf der Maschinenseite und mit Eingabemöglichkeiten auf der Seite des Menschen. Wichtig ist hierbei, dass die Syntax möglichst einheitlich und unkompliziert ist. Informationsanzeigen, die immer an derselben Stelle erscheinen, werden nicht nur weniger oft übersehen, sondern sind auch leichter zu interpretieren.

Lexikalische Ebene

Die lexikalische oder auch Kodierungsebene beschäftigt sich mit der tatsächlichen Repräsentation von Daten. Hierbei stellen sich Fragen, die keine einheitliche Antwort ermöglichen, sondern von Fall zu Fall verschieden zu beantworten sind. Etwa: „Soll die Datenausgabe qualitativ oder quantitativ, analog oder digital, textuell oder in schematischen Bildern erfolgen?“ Wobei diese und andere Fragen meist erst nach eingehender Prüfung durch Psychologen und psychologischen Studien beantwortet werden können.

Alphabetische Ebene

In dieser Ebene geht es um die Feinheiten des Zeichensatzes. Wurde etwa eine analoge Anzeigemethode gewählt, gilt es nun zu klären, ob eine Skala mit Zeiger oder eine Art Thermometerdarstellung verwendet wird.

Hier geht es also um die physischen Merkmale der Aus- und Eingabemedien.

(vgl. Fachgruppe Arbeits- und Organisationspsychologie, 2005)

2.8. Kommunikation und Interaktion

Die Begriffe Kommunikation und Interaktion sind eng miteinander verwandt, und sogar in vielen Bereichen synonym, vor allem, wenn Wechselseitigkeit als Kommunikationsgrundlage gesehen wird.

Es gibt allerdings auch Definitionen, die Kommunikation als eine spezifische Form sozialer Interaktion sehen. Der Begriff der Interaktion ist also meist weiter gefasst als der der Kommunikation. Er zielt oft nicht nur auf manifeste Inhalte des Informationsaustausches ab, sondern umfasst auch stärkere Aspekte der Beziehungsebene. Die Interaktion scheint also der Kommunikation übergeordnet (vgl. Burkart, 2003, S. 30, S.431).

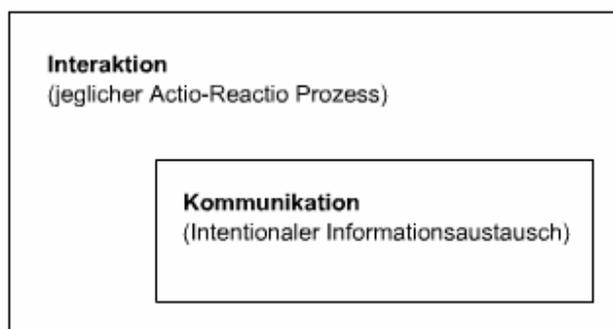


Abbildung 6: Interaktion und Kommunikation

Bei der Interaktion geht es also nicht nur darum was ausgetauscht wird, sondern überhaupt darum, dass und wie etwas vermittelt wird.

Außerdem bezeichnet Interaktion meist einen symmetrischen Prozess, da auf jede Aktion eine Reaktion folgt. Kommunikation hingegen kann auch asymmetrisch ablaufen, da nicht jeder Nachrichtenempfang auch ein erneutes Senden nach sich zieht. Der Begriff „asymmetrisch“ bedeutet in diesem Zusammenhang, dass sich die in beiden Richtungen übertragenen Datenmengen beträchtlich voneinander unterscheiden (Beispiel: video on demand). Sind die von den beiden Beteiligten eingebrachten Informationen und somit die in beiden Richtungen übertragenen Datenmengen etwa gleich groß, spricht man von „symmetrischem“ Informationsaustausch (Beispiel: Videokonferenzen).

Laut Fremdwörterbuch handelt es sich bei dem Begriff Interaktion um einen in Soziologie und Psychologie geläufigen Terminus, mit dem "aufeinander bezogenes Handeln zweier oder mehrerer Personen" oder die "Wechselbeziehung zwischen Handlungspartnern" bezeichnet wird (vgl. Duden, 1982, S. 350f.).

In der Informatik ist der Begriff der Interaktion mit dem Begriff der Kommunikation annähernd identisch: er befasst sich mit Prozessen zwischen Mensch und Maschine respektive zwischen Mensch und Computer. Als interaktiv wird hierbei jener Dialogverkehr bezeichnet, „der es Benutzern von Computern gestattet, ständig

Zwischenergebnisse oder Auswahlmenüs zu sichten und durch neue Tastatur- oder Mauseingaben auf diese zu reagieren“ (zit. n. Pürer, 2001, S.26).

Von Interaktion im technischen Sinne ist die Rede, wenn ein Medium über die Fähigkeit verfügt

- a) mit dem Nutzer in einen wechselseitigen Dialog zu treten – also Kommunikation mit dem Medium gestattet, und wenn
- b) ein Medium synchrone (Chat) oder asynchrone (E-Mail) Kommunikation zwischen zwei räumlich getrennten Kommunikationspartnern ermöglicht – also Kommunikation durch ein Medium.

Eine etwas andere Bedeutung hat der Begriff in Bezug auf die Mensch-Maschine-Interaktion, also der Gestaltung der Benutzerschnittstelle zu Programmen und Hardware. Hierbei handelt es sich ausschließlich um den wechselseitigen Austausch von Informationen zwischen Mensch und Computer. Gemeint ist also der Feed-back-Prozess und damit die Überwindung der gerichteten, einseitigen Kommunikation. „Interaktion ist zunächst nur ein Potential, dessen Realisierung vom jeweiligen Nutzer abhängt“ (Pürer, 2001, S.27).

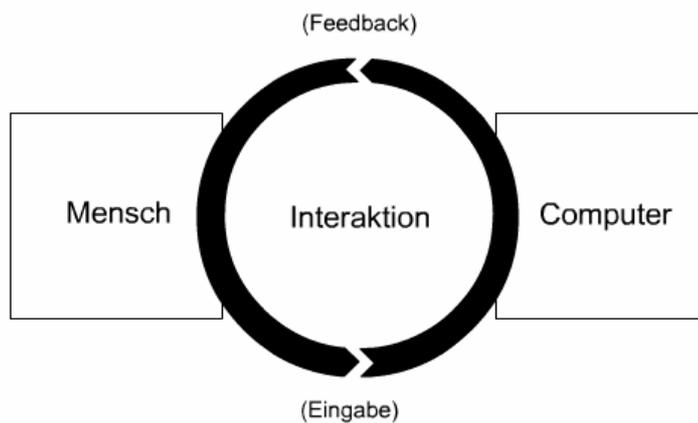


Abbildung 7: Interaktions- oder Feedbackprozess

Interaktivität kann also als die Fähigkeit eines Systems beschrieben werden, passenden Output als Reaktion auf einen bestimmten Input bereitzustellen.

Wobei das Begriffspaar Input und Output (I/O) als Frage und Antwort gesehen werden kann, da es sich, wie schon erwähnt, bei Mensch-Computer Interaction hauptsächlich um den Feedbackprozess handelt.

3. Human-Computer Interaction (HCI)

3.1. Einführung

Wann immer Menschen Computer bedienen, das heißt Informationen zwischen Mensch und Computer ausgetauscht werden, stellen Schnittstellen einen wichtigen Aspekt dar.

Die Mensch-Computer-Interaktion (MCI; englisch Human-Computer Interaction, HCI) als Teilgebiet der Informatik beschäftigt sich mit der benutzergerechten Gestaltung von interaktiven Systemen, also mit allen Formen und Ausprägungen der Kommunikation zwischen Mensch und Maschine, vom Wählen beim Telefonieren, über das Einschalten der Mikrowelle, bis zum Bedienen eines Programms mit der Maus.

Im medientechnischen Kontext werden natürlich nur Systeme behandelt, die Computer verwenden oder eingebettet haben.

Dabei werden neben Erkenntnissen der Informatik auch solche aus der Psychologie, der Arbeitswissenschaft, der Kognitionswissenschaft, der Ergonomie, der Soziologie und dem Design herangezogen. HCI ist also die Schnittstellenforschung zwischen Mensch und Computer und somit eine Verschmelzung aus vielen Disziplinen zweier grundlegender Bereiche, nämlich Verhaltensforschung und Informatik.

Einordnung der Human-Computer Interaction in die wissenschaftlichen Hauptströmungen

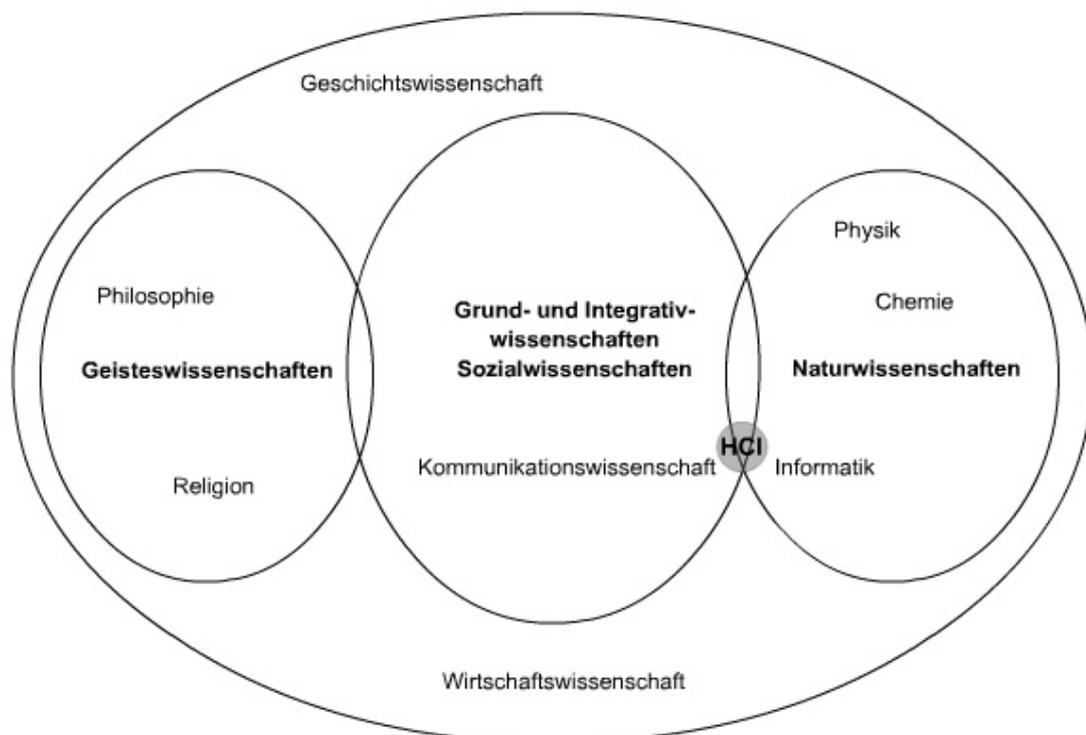


Abbildung 8: Einordnung der HCI in die wissenschaftlichen Hauptströmungen
(vgl. Huber, 2004, S.7)

Dieser Abriss des Begriffes HCI zeigt deutlich, wo die Problem- und Fragestellungen der Mensch-Computer Interaktion liegen. Das Ziel der HCI ist nämlich die Mensch-Computer Schnittstelle für den User so angenehm und benutzerfreundlich wie nur irgendwie möglich zu gestalten. Es gilt also, technische Wege zu finden, um auf unerschwingliche menschliche Empfindungen und Bedürfnisse zu reagieren.

Somit wird also versucht, Regeln und Richtlinien aufzustellen, die bei der Entwicklung von Benutzerschnittstellen (Userinterfaces) beachtet werden sollten, um den Benutzern eine möglichst intuitive und leicht verständliche Möglichkeit der Kommunikation zu bieten.

HCI Forscher analysieren und entwerfen neuartige Userinterfaces und Interaktionssysteme. Sie waren es, die vor Jahrzehnten den Weg für unsere heutigen GUIs (Graphical User Interfaces) bereiteten.

Heutzutage beschäftigen sich viele HCI Forscher mit der Entwicklung alternativer Input- und Displaytechnologien und möglichen Navigationstechniken in virtuellen Umgebungen.

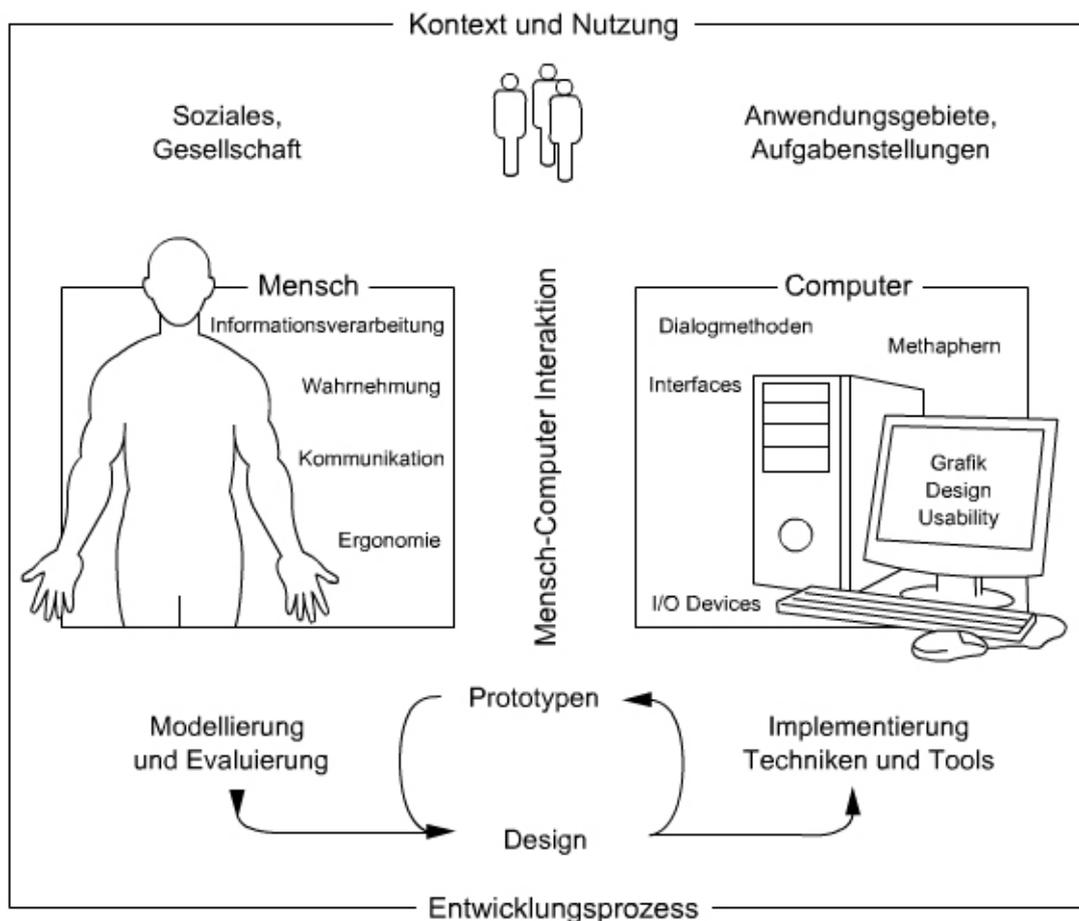


Abbildung 9: HCI Kontext und Prozesse (vgl. Andrews, 2005, S. 2)

Human-Computer Interaction und ihre Forschung versucht sowohl das wissenschaftliche Verständnis für die Wechselwirkung zwischen Mensch und Maschine, als auch den Tools und Programmen, die von Menschen verwendet werden, zu fördern.

Dazu bedient sich die HCI Forschung der Erkenntnisse unterschiedlichster Wissenschaften, wie etwa der Psychologie.

Viele Entwicklungen, die HCI vorangetrieben haben (z.B.: Userinterface Design, Informationsvisualisierung, usw.), beruhen auf den wissenschaftlichen Forschungen und Theorien der Psychologie. Doch auch viele andere Disziplinen wie Mathematik, Architektur, Linguistik, Kunst und Design usw. haben jede für sich relevante Beiträge zur HCI geleistet.

In dieser Arbeit wird jedoch nur auf die aus den Teilgebieten der Psychologie resultierenden Bestandteile von HCI eingegangen. Außerdem wird auf die Haupttheorien und Modelle, die die HCI-Forschung prägen, eingegangen um die Breite der HCI-Forschung zu illustrieren.

Die Theorien und Modelle, die HCI prägen, können in drei Kategorien eingeteilt werden:

- erklärende Theorien (explanatory theories)
- generative Theorien (generative theories)
- vorhersagende Theorien (predictive theories)

Erklärende Theorien bemühen sich, das Verhalten der realen Welt zu erklären und zu beschreiben. Sie stellen meist ein eher begriffliches, konzeptionelles Modell der Welt zur Verfügung und leiten daraus Richtlinien für Interfaces ab.

Vorhersagende oder auch prophetische Theorien bemühen sich, auf die sich ändernden Werte von Teilvariablen zu reagieren und Ergebnisse vorauszusagen. Solche Theorien sind extrem nützlich für die HCI, da sie etwa durch die gezielte Vorhersage von Benutzerverhalten, einen zielorientierten Entwurf von Userinterfaces ermöglichen.

Generative oder auch Gestaltungstheorien versuchen psychologisch nachvollziehbare Richtlinien und Grundsätze aufzustellen, die beim Entwerfen und Designen von Benutzerschnittstellen beachtet werden können. Hierbei kommen meist Usabilitytests oder Umfragen zum Zug.

Indem sie die Theorien und Modelle von den verschiedenen Gebieten der Psychologie vereinigen, sind HCI Forscher im Stande gewesen, sich spezifisch und angewandt mit der Interaktion zwischen Mensch und Computer und den damit verbundenen Problemen zu beschäftigen. HCI-Theorien und Modelle versuchen meist Richtlinien und Grundsätze für das Design und Entwicklung von Schnittstellen zur Verfügung zu stellen.

HCI – Verwandte Begriffe und Synonyme:

MCI: Mensch-Computer Interaktion

MRI: Mensch-Rechner Interaktion

MMI: Mensch-Maschine Interaktion

HCI: Human Centered Interfacedesign

3.2. HCI Modelle

Menschliche Arbeit kann durch Maschinen und Computer sehr erleichtert werden, allerdings spielt beim Zusammenspiel zwischen Mensch und Maschine das Interface eine nicht zu vernachlässigende Rolle.

Um diese Interfaces praktikabel und sinnvoll zu gestalten, werden die unterschiedlichsten Wege verfolgt, die doch eines gemeinsam haben: In jedem Fall wird versucht, zwischen Mensch und Computer zu vermitteln, daher bildet die menschliche Wahrnehmung die Grundlage für fast alle Modelle der Mensch-Computer Interaction.

Wie schon eingangs besprochen, ist ein Modell an sich eine vereinfachte Darstellung der Realität und hat das Ziel, einen komplizierten Sachverhalt auf wenige Parameter zu reduzieren und somit zu vereinfachen.

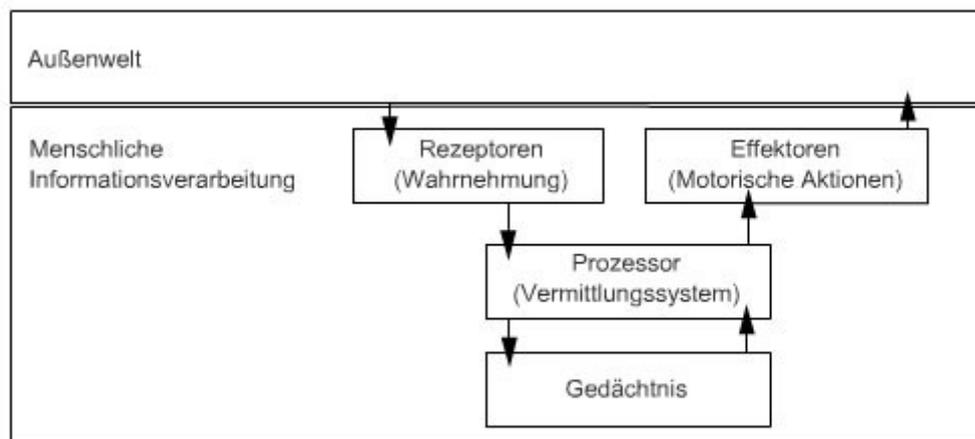


Abbildung 10: Menschliche Informationsverarbeitung (vgl. Carroll, 2003, S.57)

Wie diese vereinfachte Abbildung des menschlichen Wahrnehmungs- oder auch Informationsverarbeitungsapparates zeigt, befindet sich im Zentrum der menschlichen Wahrnehmung ein Vermittlungssystem. Dieses Vermittlungssystem ist in einem Actio-Reactio-Kreislauf zwischen der menschlichen Sinneswahrnehmung und dem Gedächtnis eingebettet und verarbeitet die durch die Rezeptoren gewonnenen Informationen über die Umwelt.

Da der menschliche Wahrnehmungsapparat vielschichtig ist, sind auch die Ansätze der HCI Modelle denkbar unterschiedlich, und versuchen meist einer Komponente des menschlichen Informationsverarbeitungssystems gerecht zu werden.

HCI ist also eine Schnittstellenforschung, daher haben auch alle HCI Modelle eines gemeinsam. Sie versuchen die Schnittstelle zwischen Mensch und Computer greifbarer zu machen und zu vereinfachen. Diese Schnittstelle besteht aus 4 Komponenten: Dem Benutzer oder User, dem System oder Computer, und im Weiteren natürlich aus den Input- und Outputsignalen der Schnittstelle.

Die eigentliche Nutzung der Schnittstelle folgt meist einem Schema:

Der User (Mensch) beginnt die Interaktion und tätigt zur Erreichung seines Zieles eine Eingabe. Dieser Input wird dann vom Interface, also der Schnittstelle, für das System oder den Computer übersetzt. Nach der Abarbeitung der Eingabe wird das Ergebnis als Output ans Interface zurückgegeben, um wieder für den Benutzer abrufbar zu sein.

Der Benutzer kann nun in einer Evaluierungsphase das Ergebnis mit seinem ursprünglichen Ziel vergleichen.

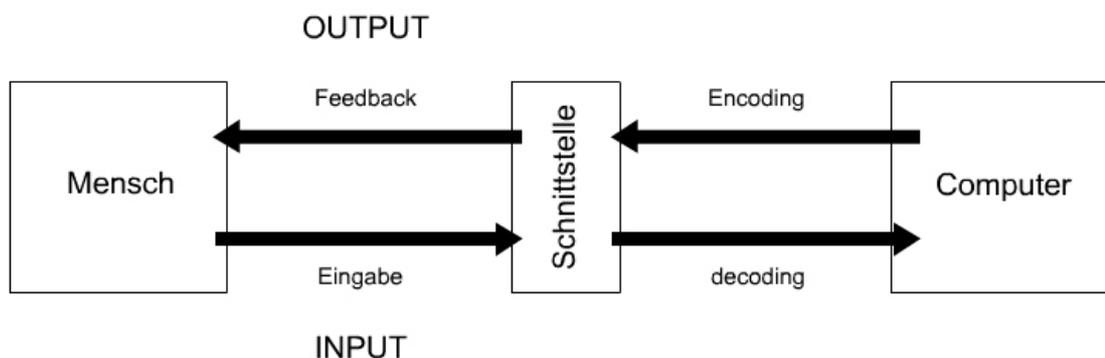


Abbildung 11: Überblick HCI Schnittstelle

Die HCI versucht grundlegende Informationen über das Zusammenspiel zwischen Mensch und Computer zu gewinnen, um diese in die Entwicklung neuer Schnittstellen einfließen zu lassen.

Zusammenfassend kann man sagen, dass HCI sich unterschiedlichster Disziplinen bedient, um die Schnittstelle zwischen Mensch und Computer so komfortabel und intuitiv wie möglich zu gestalten. Dazu bedient sich HCI unter anderem der Erkenntnisse folgender Disziplinen:

- Informatik (Softwareentwicklung und Hardwarekenntnisse)
- Psychologie (Interfacedesign und Benutzerverhalten)
- Kommunikationswissenschaft (Kommunikationsverhalten und Erwartungen)
- Industrial Design (Interaktive Produkte und ergonomisches Design)

Unter Zuhilfenahme der oben genannten wissenschaftlichen Disziplinen versucht HCI folgende Aufgabenstellungen zu bewältigen:

- Interaktion und Kommunikation zwischen Mensch und Computer
- Menschliches Umgangsvermögen mit Computern
- Spezifikation und Entwurf neuer Benutzerschnittstellen
- Design und Gestaltung von Benutzerschnittstellen

3.3. Erklärende Modelle oder Beschreibungsmodelle

3.3.1. Wahrnehmungsmodell

Die menschliche Wahrnehmung wird von visuellen Eindrücken dominiert und geprägt, da mehr als 70 % aller menschlichen Sinnesrezeptoren visuelle Rezeptoren sind und etwa 50 % der menschlichen Hirnrinde von visuellen Eindrücken vereinnahmt werden (vgl. Carroll, 2003, S.11). Daher ist es nahe liegend, sich bei der Gestaltung und dem Design einer Mensch-Computer Schnittstelle an der menschlichen Wahrnehmung zu orientieren.

Da die Wahrnehmung theoretisch bei allen Menschen gleich funktioniert, könnte man eine Mensch-Computer Schnittstelle, die auf den Erkenntnissen visueller Wahrnehmung basiert, weltweit einheitlich umsetzen. Dies ist auch geschehen, und hat uns die wohl am meisten benützten Schnittstellen zwischen Menschen und Computern geliefert, z.B. den Bildschirm, das GUI, Buttons, Icons, und vieles mehr.

Anhand eines vereinfachten 3-Stufen-Modells der menschlichen, visuellen Wahrnehmung können verschiedenste Design- und Gestaltungsrichtlinien für Mensch-Computer Schnittstellen veranschaulicht und abgeleitet werden.

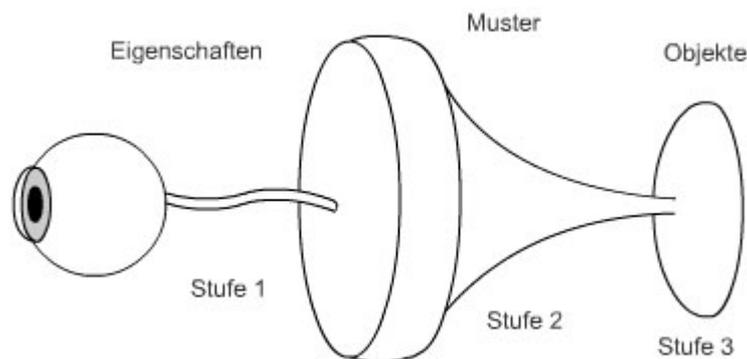


Abbildung 12: Visuelle Wahrnehmung in drei Stufen (Carroll, 2003, S.13)

Die drei Stufen der visuellen Reizverarbeitung zeigen, dass ein visueller Reiz zuerst auf Grund seiner Eigenschaften unterschieden wird, später werden diese Informationen zu Mustern verarbeitet, um zuletzt als komplexe Objekte verstanden und gespeichert zu werden.

Stufe 1:

Stufe eins bezeichnet die erste Phase der visuellen Wahrnehmung und ist von den optischen Eigenschaften und Charakteristika der Photorezeptoren des menschlichen Auges geprägt.

Die aus der Wahrnehmungsforschung stammenden Erkenntnisse bieten seit Jahrzehnten Richtlinien für die Entwicklung von Röhrenmonitoren und Liquid Crystal Displays.

So resultieren daraus etwa die Halbbildfrequenz von 50 Hz, der anzustrebende Kontrastunterschied von 10 zu 1 zwischen Text zu Hintergrund oder die RGB (Rot, Grün, Blau) und HSB (Farbton, Sättigung, Helligkeit) Farbmodelle heutiger Monitore.

Stufe 2:

In der mittleren Stufe der Wahrnehmung werden die in Stufe 1 charakterisierten Impressionen zu Mustern und Konturen verschmolzen. Es werden im wahrgenommenen Bild also logische Strukturen gesucht, Nähe und Ähnlichkeiten werden als Zusammenhang erkannt.

Aus dem Verständnis dieses Wahrnehmungsprozesses lassen sich wiederum weitere Richtlinien für das Layout und die Gestaltung z.B. von Graphischen User Interfaces (*GUIs*) ableiten, die wiederum auf Ausgabemedien dargestellt werden, deren Grundlage in Stufe 1 zu finden ist.

Stufe 3:

Diese Stufe entspricht noch nicht dem menschlichen Gehirn, sondern repräsentiert das so genannte „visual working memory“ (Carroll, 2003, S.18), also das visuelle Kurzzeitgedächtnis.

Hier werden die wahrgenommenen Eindrücke kategorisiert, also als Objekt verstanden und in einen Bezug zu Umwelt und anderen Informationen gesetzt.

Hier wird also die Verknüpfung eines Objekts mit seinen Eigenschaften hergestellt.

Im Vergleich zu den anderen Stufen finden sich hier nicht viele wissenschaftlich belegbare Eigenschaften, die zur Erforschung der Mensch-Computer Schnittstelle hilfreich wären.

Die Erkenntnisse, die sich aus diesem sehr vereinfachten Modell der visuellen Wahrnehmung ableiten lassen, waren die Grundlage für viele Entwicklungen im HCI Bereich. Doch so einleuchtend und nahe liegend dieses Modell auch sein mag, hat es doch auch seine Tücken.

So mag zwar der physische Vorgang des Erkennens von Farben, Formen, Mustern und daraus folgend auch von Objekten bei allen Menschen gleich funktionieren. Doch die dritte Stufe des Modells, die das Verstehen des Gesehenen bewirkt, wird rückbezüglich auch von kulturellen Erfahrungen beeinflusst.

So hat beispielsweise die Farbe Rot im europäischen Raum die Wirkung einer Warnfarbe und deutet somit oft auf Gefahren hin, hingegen wird sie in vielen asiatischen Kulturen als Glücks- und Freudenfarbe gesehen.

3.3.2. Information processing model

Nach dem menschlichen Wahrnehmungsapparat als Möglichkeit Informationen über die Umwelt aufzunehmen, ist auch das menschliche Informationsverarbeitungssystem maßgeblich an unserem Verständnis und Bild der Realität beteiligt.

Der Ansatz, den das Informationsverarbeitungs- oder auch information processing model verfolgt, basiert auf der Annahme, dass menschliches Verhalten, wie Kommunikation, in mehreren Stufen oder Prozessen abläuft.

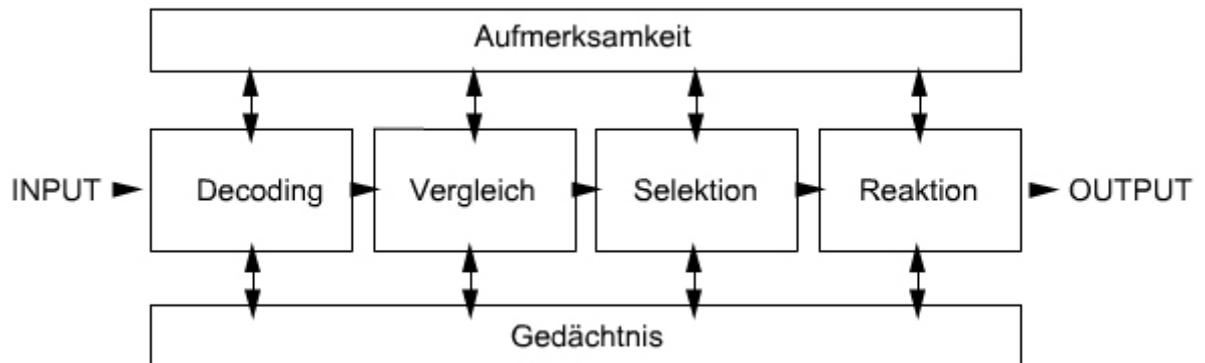


Abbildung 13: information processing model (human error, 2005)

Dieses Modell zeigt ähnlich dem Sender-Empfänger Modell der Kommunikation, dass Information im Menschen vom Input bis zur Reaktion (dem Output) mehrere Verarbeitungsstufen durchläuft. Diese Stufen sind das Dekodieren, der Vergleich, die Selektion und die Reaktion.

Dekodieren bezeichnet in diesem Fall das Filtern von essentiellen Botschaften aus allen vorhandenen Inputs, und Vergleich bedeutet, dass die momentane Situation mit vergangenen Situationen verglichen wird, um mögliche Lösungsansätze zu finden.

Selektion beschreibt den Fakt, dass zwischen mehreren möglichen Verhaltensweisen oder Antworten gewählt werden kann und die passende Reaktion erst selektiert werden muss.

Danach folgt dann das effektive Ausführen einer Reaktion, was wiederum den Output erzeugt.

Alle diese Prozesse finden, wie die Abbildung 14 zeigt, parallel zwischen der menschlichen Aufmerksamkeit und dem Gedächtnis statt.

HCI verwendet die Erkenntnisse aus der Beschreibung der menschlichen Informationsverarbeitung in Interfacedesign, um menschliche Reaktionen auf bestimmte Auslöser zurückführen zu können.

3.4. Generative oder Gestaltungsmodelle

Design und Gestaltungstheorien helfen Interfaces und Schnittstellen, wie etwa das GUI, zu entwickeln. Umfangreiche Tests und Forschungen stellen die Grundlage aller dieser Richtlinien für die Gestaltung von Bestandteilen der Schnittstelle dar.

3.4.1. Interaction-Design

Im Bereich der Gestaltungsdisziplinen setzt sich seit den 90er Jahren immer mehr der Begriff Interaction Design durch. In diesem Feld der Gestaltung geht es um die Planung, Entwicklung und Gestaltung interaktiver Benutzerschnittstellen.

Laut der IxDG (Weblink [2]), der Interaction Design Group, ist Interaction Design eine Disziplin, die definiert, wie interaktive Produkte ihre Funktionalität gegenüber Nutzern kommunizieren, darstellen und wie Nutzer mit ihnen interagieren können.

Gutes Interaction Design macht Produkte, die von Computersoftware bis hin zu Mikrowellen reichen können, nützlicher und bedienbarer, es verbessert die Beziehung zwischen Menschen und interaktiven Geräten oder Produkten.

Das Ziel eines Interaktion Designers ist, durch seine Gestaltung ein Produkt nicht nur nützlicher, sondern vor allem nutzbarer zu machen, wobei der Mensch im Zentrum des Interesses steht. Wichtige Punkte, die gutes Interaktion-Design auszeichnen sind etwa:

- Effektive Kommunikation eines Systems über seine Funktionalität
- Definition von Rückmeldungen auf Benutzerverhalten
- Schaffung klarer System Statusinformationen
- Vermeidung von Benutzerfehlern

Zusammenfassend kann man sagen, dass Interaction Design versucht, die Struktur und das Verhalten von Interaktiven Systemen zu definieren und zu gestalten, sowie deren Benutzbarkeit zu erhöhen.

3.4.2. Software-Ergonomie

Hierbei handelt es sich um den überlegten Entwurf von Software, was bedeutet, dass dem Kommunikations-Design in Bezug auf Orientierung, Navigation, Maskengestaltung, System-Reaktionen, System-Meldungen, Symbolik usw. hohe Aufmerksamkeit gewidmet wird.

In der europäischen Norm EN ISO 9241, wurden Kriterien für das Design ergonomischer Benutzerschnittstellen definiert. Diese Norm besteht insgesamt aus 17 Teilen:

- Teil 1: Allgemeine Einführung
- Teil 2: Anforderungen an die Arbeitsaufgaben - Leitsätze
- Teil 3: Anforderungen an visuelle Anzeigen
- Teil 4: Anforderungen an Tastaturen
- Teil 5: Anforderungen an die Arbeitsplatzgestaltung und Körperhaltung
- Teil 6: Anforderungen an die Arbeitsumgebung
- Teil 7: Anforderungen an visuelle Anzeigen bezüglich Reflexionen
- Teil 8: Anforderungen an Farbdarstellungen
- Teil 9: Anforderungen an Eingabegeräte außer Tastaturen
- Teil 10: Grundsätze der Dialoggestaltung
- Teil 11: Anforderungen an die Gebrauchstauglichkeit - Leitsätze
- Teil 12: Informationsdarstellung
- Teil 13: Benutzerführung
- Teil 14: Dialogführung mittels Menüs
- Teil 15: Dialogführung mittels Kommandosprachen
- Teil 16: Dialogführung mittels direkter Manipulation
- Teil 17: Dialogführung mittels Bildschirmformularen

Die wohl bekanntesten Teile dieser ISO-Norm sind die 1995 als europäische Norm genehmigten „Grundsätze der Dialoggestaltung“ (EN ISO 9241, Teil 10). Sie behandeln die ergonomische Gestaltung von Software und beschreiben sieben allgemeine ergonomische Grundsätze:

- Aufgabenangemessenheit
- Selbstbeschreibungsfähigkeit
- Steuerbarkeit
- Erwartungskonformität
- Fehlertoleranz
- Individualisierbarkeit
- Lernförderlichkeit

Die Grundsätze dieser Norm dürfen jedoch nicht missverstanden werden, sie bilden weder die alleinigen Kriterien nach der sich Usability entwickeln oder beurteilen lässt, noch sind sie ein Raster, durch das ein Userinterface passen muss. Vielmehr sind bei Anwendung der Norm die unterschiedlichen Benutzer- und Aufgabenmerkmale zu berücksichtigen und die Grundsätze in Beziehung zueinander zu setzen.

1. Aufgabenangemessenheit

"Ein Dialog ist aufgabenangemessen, wenn er den Benutzer unterstützt, seine Arbeitsaufgabe effektiv und effizient zu erledigen." (DIN EN ISO 9241, 2005)

Die Effektivität bezeichnet die Genauigkeit und Vollständigkeit, mit der ein Benutzer seine Aufgabe lösen kann. Unterstützen können ihn dabei z.B. Hilfesysteme oder verschiedene Navigationsarten, die mehrere Wege zum Ziel anbieten.

Die Effizienz beschreibt, mit welchem zeitlichen Aufwand der Nutzer seine Aufgabe, z.B. auf einer Website lösen kann. Sie ist sehr stark von der Navigation der angebotenen Schnittstelle abhängig, aber auch von technischen Parametern, wie Ladezeiten und der Performanz des Computers.

Beispiel: Eine Suchmaschine präsentiert auf ihrer Startseite anstelle einer Suchmaske zunächst ein animiertes Logo. Suchmaschinen werden genutzt, um nach bestimmten Informationen zu recherchieren und möglichst schnell relevante Suchergebnisse zu erhalten. Die Startseiten-Animation ist in diesem Fall nicht aufgabenangemessen.

2. Selbstbeschreibungsfähigkeit

"Ein Dialog ist selbstbeschreibungsfähig, wenn jeder einzelne Dialogschritt durch Rückmeldung des Dialogsystems unmittelbar verständlich ist, oder dem Benutzer auf Anfrage erklärt wird." (DIN EN ISO 9241, 2005)

Der User sollte sich intuitiv durch die Benutzeroberfläche bewegen können. Er sollte verstehen, was sie kann, welche Informationen sie vermittelt und wohin er sich begeben muss, um bestimmte Aufgaben zu lösen und seine Ziele zu erreichen, beispielsweise Informationen finden, Kontakt aufnehmen, Einkaufen etc. Besondere Bedeutung haben in diesem Zusammenhang aussagekräftige Link- oder Buttonbezeichnungen, sowie eine aussagekräftige Symbolik.

Beispiel: Links sind so formuliert, dass man sicher vorhersagen kann, wohin sie führen. Auf einer Website sind allgemeine Begriffe wie „Info“ nicht dazu geeignet, dem Benutzer zu beschreiben, was er auf den dahinter stehenden Seiten vorfinden wird: Informationen über Produkte? Über das Unternehmen? Vielleicht Nachrichten?

3. Steuerbarkeit

"Ein Dialog ist steuerbar, wenn der Benutzer in der Lage ist, den Dialogablauf zu starten, sowie seine Richtung und Geschwindigkeit zu beeinflussen, bis das Ziel erreicht ist." (DIN EN ISO 9241, 2005)

Dazu gehört, dem User verschiedene Navigationsformen und/oder Möglichkeiten der Einflussnahme anzubieten, wie etwa die Option, eine Startanimation abbrechen zu können, oder die Steuerung akustischer Informationen zu ermöglichen (Lautstärke-regelung, Ton ein-/ausschalten), aber auch das Anbieten von Thumbnails (Bildvorschau, die bei Bedarf vergrößert werden kann) in Bildergalerien.

Beispiel: Eine Suchmaschine bietet die Möglichkeit, die Zahl der auf einer Seite angezeigten Treffer einzustellen.

4. Erwartungskonformität

"Ein Dialog ist erwartungskonform, wenn er konsistent ist und den Merkmalen des Benutzers entspricht, z.B. seinen Kenntnissen aus dem Arbeitsgebiet, seiner Ausbildung und seiner Erfahrung, sowie den allgemein anerkannten Konventionen." (DIN EN ISO 9241, 2005)

Jede Anwendung sollte sich erwartungskonform verhalten. Das bedeutet, dass das Dialogverhalten der Anwendung konsistent und einheitlich genug sein muss, um dem Benutzer seine mit der Anwendung (oder früher verwendeten, vergleichbaren Anwendungen) gesammelten Erfahrungen im Umgang mit ihr nutzen zu können.

Beispiel: Unterstrichene Wörter sind immer Hypertext-Links.
Ein Benutzer hat auf einer Website die Erfahrung gemacht: Klickt er die unterstrichenen Wörter an, gelangt er zu anderen Seiten. Er schlussfolgert richtig: Die unterstrichenen Wörter sind Links – er kann sie fortan als solche auf Anhieb erkennen. Irgendwann stellt er fest: Nicht alle unterstrichenen Wörter besitzen diese Funktionalität. Manche befördern ihn auf andere Seiten, andere wiederum nicht. Seine diesbezüglichen Erfahrungen mit der Website werden in diesem Moment wertlos. Verwirrung weicht Verärgerung.

5. Fehlertoleranz

"Ein Dialog ist fehlertolerant, wenn das beabsichtigte Arbeitsergebnis trotz erkennbar fehlerhafter Eingaben entweder mit keinem oder mit minimalem Korrekturaufwand seitens des Benutzers erreicht werden kann." (DIN EN ISO 9241, 2005)

Eine Anwendung sollte mögliche Fehlerquellen bereits im Vorfeld eliminieren oder zumindest minimieren. Fehlerhafte Eingaben des Users sollten zu produktiver Hilfestellung führen – nicht zu undefinierten Zuständen. Der Benutzer sollte trotz fehlerhafter Eingaben entweder ohne oder nur mit minimalem Korrekturaufwand sein Ziel erreichen. Dazu gehört auch, dass Fehlermeldungen in einer verständlichen Sprache ausgegeben werden.

Beispiel: Sollten mehrere Eingaben des Users notwendig sein, ist sicher zu stellen, dass die eingegebenen Daten überprüft werden, bevor sie gesendet werden. Somit müssen bei allfälligen Fehlern nur noch die fehlerhaften Felder neu eingegeben werden.

6. Individualisierbarkeit

"Ein Dialog ist individualisierbar, wenn das Dialogsystem Anpassungen an die Erfordernisse der Arbeitsaufgabe, sowie an die individuellen Fähigkeiten und Vorlieben des Benutzers zulässt." (DIN EN ISO 9241, 2005)

Individualisierbare Dialogsysteme erlauben es dem Benutzer, flexible Anpassungen an die Erfordernisse der Arbeitsaufgabe und/oder an seine Vorlieben und seine speziellen Fähigkeiten vorzunehmen.

Individualisierbar bedeutet in diesem Fall, dass das System oder die Anwendung die Fähigkeit besitzt, den User wieder zu erkennen.

So müssen etwa registrierte User einer Website nicht immer alle zur Registrierung notwendigen Userdaten erneut eingeben, sondern können sich einfach mit einem frei gewählten oder vorgegebenem Usernamen und Passwort anmelden.

Beispiel: Wenn der Besucher einer Website zum Lösen seiner Aufgabe individuelle Eingaben oder Einstellungen vornehmen muss, sollte die Website individualisierbar sein.
So kann dem Kunden eines Online-Shops erspart werden, bei jeder Bestellung seine persönlichen Daten erneut eingeben zu müssen.

7. Lernförderlichkeit

"Ein Dialog ist lernförderlich, wenn er den Benutzer beim Erlernen des Dialogsystems unterstützt und anleitet." (DIN EN ISO 9241, 2005)

Die Zeit, die ein Benutzer benötigt, um den Umgang mit einer Anwendung zu erlernen, wird maßgeblich von der sprachlichen und konzeptionellen Ausgestaltung der Oberfläche beeinflusst.

Eine Anwendung sollte so gestaltet sein, dass sie Lernen ermöglicht und den Besucher dazu ermutigt. Dazu gehört beispielsweise eine Online-Tour, die den Besucher durch die Website führt, eine Sitemap, die die logische Struktur der Website erkennen lässt, ein einheitliches Erscheinungsbild oder Ähnliches.

Beispiel: Im Buchungs-System eines Reiseanbieters besteht etwa die Möglichkeit eine Probebuchung vorzunehmen ohne sie abzuschicken, mit dem alleinigen Zweck, dem User den Vorgang des Online-Buchens näher zu bringen.

Abschließend könnte man sagen, dass diese Norm nicht wirklich Revolutionäres enthält. Doch sie gibt grobe Kriterien vor, an denen sich eine Applikation auf ihre Ergonomie und damit auf ihre Effizienz für den Anwender prüfen lässt.

Die häufigste Form von Dialogen zwischen Computern (Software) und Benutzer (Menschen) zielt auf die Auswahl und das Sammeln von Informationen ab. Deshalb bietet diese Norm eine gute Grundlage, um die Effizienz einer Software durch die Integration der sieben Grundsätze zu verbessern.

Diese Grundsätze sind unabhängig von einem bestimmten Dialogsystem zu sehen, also nicht nur auf die Gestaltung und Bewertung von Webseiten, wie in den meisten hier angeführten Beispielen, sondern auf jede Art von HCI anwendbar.

3.5. Vorhersagende oder Prophezeiungs-Modelle

3.5.1. Motorische Verhaltensmodelle

HCI Forschung ist bemüht die Effektivität eines Inputdevices vorherzusagen. Dazu wurden viele verschiedene Devices wie Maus, Joystick, Trackballs usw. auf Eigenschaften wie Genauigkeit, Leichtigkeit oder Geschwindigkeit bei der Benutzung untersucht. Um nun die Performanz dieser Schnittstellen zu ermitteln, bedient man sich der motorischen Modelle.

Fitts Law ist eins der robustesten und wohl meist benutzten Modelle der menschlichen Bewegung (vgl Carroll, 2003, S.35). Fitts' Gesetz ist Teil des ISO 1999 Standards für die Evaluierung von computer pointing devices und versucht Informationsverarbeitung und seine Prozesse messbar und greifbar zu machen, indem die Schwierigkeit von Bewegungsabläufen betrachtet wird. Fitts beschreibt die Bedienzeit für ein Interfacelement in Abhängigkeit von Distanz und Größe des Objekts in einer logarithmischen Funktion. Der nach Fitt errechenbare Index of Difficulty (ID), der in Bits angegeben wird, ist logarithmisch vom Abstand der Ziele zu einander und der Breite (Width) des Ziels abhängig.

Weiters kann mit Hilfe empirisch getesteter Daten die Movement-Time (MT) und Error rate berechnet werden. Diese empirischen Daten beziehen sich auf das Eingabeinterface und müssen ausgetestet werden. So hätte etwa eine Mouse als Bedienungselement andere Parameter als etwa eine Handytastatur. Indem die erhobenen Werte mit dem ID multipliziert werden, erhält man die Movement-Time. Fitts Law sollte bei Tests mindestens mit 240 Observationen belegt werden, also 12 Probanden die 20 Mal eine bestimmte Aufgabe zu bewältigen haben. (vgl. Carroll, 2003, S.37) Fitt ermöglicht es somit, die Datenrate des menschlichen Informationsverarbeitungsprozesses zu berechnen, indem ID durch MT dividiert wird. Diese Datenrate nennt Fitt „Troughput“, also Durchsatz mit der Einheit Bits pro Sekunde.

Es gibt auch die Möglichkeit sich online selbst nach der Methode von Fitt zu testen (Weblink [3]). Hierbei wird schnell klar, wie die Proportionen der Buttons und deren Abstände zueinander die Navigation innerhalb eines Menüs beschleunigen und erleichtern können. Laut Fitt steigt die Pointing- oder Zielzeit mit der Distanz der Objekte. Die Größe der Ziele wirkt sich in so fern aus, dass kleinere Ziele das Zielen verlangsamen.

Fitts Law ist nur auf das Zielen zwischen zwei Menüpunkten ausgerichtet, das heißt, die Zeit zum inhaltlichen Auswählen zwischen mehreren Menüpunkten wird hier vernachlässigt, da ja nur der jeweils andere Menüpunkt als der zuletzt gewählte in Frage kommt.

Meinem Verständnis nach besteht das Navigieren in einem Menü aber aus zwei Prozessen. Erstens muss entschieden werden, welcher der angebotenen Menüpunkte zu wählen ist, um die gewünschte Information zu erreichen und zweitens gibt es den aktiven Wahlprozess, das Klicken.

Fitt beschreibt allerdings nur den zweiten genannten Prozess, daher möchte ich in diesem Zusammenhang auch Hicks Law oder das Hick-Hyman Gesetz erwähnen.

Hicks Law beschreibt die Reaktionszeit bei der Auswahl von Handlungszielen in Abhängigkeit von den verfügbaren Alternativen und ermöglicht es so, vorher zu sagen, wie lange Benutzer von der Navigation „aufgehalten“ werden.

Hicks Law besagt, dass die Auswahl- und Reaktionszeit logarithmisch proportional zur Anzahl der Auswahlmöglichkeiten steht. Daher sind seine Überlegungen auf Auswahlmenüs jeglicher Art anwendbar.

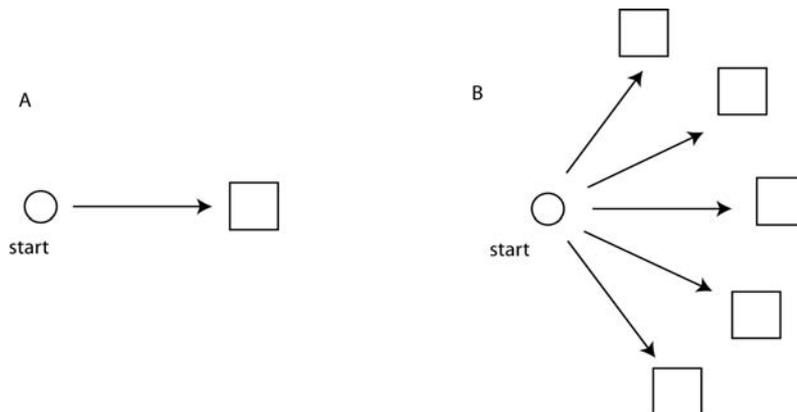


Abbildung 14: Fitts & Hicks Law (Fitts' Law as a Performance Model in Human-Computer Interaction, 2005)

Beispiel A:Fitts Experiment; eine Alternative, also nur eine Möglichkeit zu selektieren

Beispiel B:Hicks Modell mit n gleichwahrscheinlichen Möglichkeiten

3.5.2. User Control Modell

Laut Professor Shneiderman (Weblink [4]) vom Human-Computer Interaction Lab der University of Maryland (Weblink [5]) ist „direct manipulation“ der Schlüssel zu gelungener Mensch-Computer Interaction.

Das wichtigste bei einer HCI Schnittstelle ist seiner Ansicht nach, dass der Benutzer stets die Kontrolle über jegliches Geschehen hat. Daher geht es beim User Control Modell lediglich darum, dass jeglicher Input des Benutzers sofort, also auf direktem Weg, zu sinnvollem Output führt. Dabei sind für Shneiderman folgende Punkte unverzichtbar:

- Sichtbarkeit und Erkennbarkeit aller Interaktionsobjekte
Dies soll dem User ermöglichen, genau vorhersagen zu können, welche Objekte Interaktionen ermöglichen.
- Schnelle, umkehrbare, stufenweise Aktionen
Wo gearbeitet wird, passieren Fehler. Daher ist es wichtig, Fehler in der Bedienung zu berücksichtigen und die Möglichkeit zur Umkehr anzubieten.
- Ersatz von komplexer command-language durch visuelle Interaktionsobjekte
Dieser Punkt ist so simpel wie einleuchtend, da Icons und Symbole viele Information transportieren, die schwer oder umständlich zu formulieren wären.

Dieses sehr uncharakteristische Modell ist meines Erachtens schon gänzlich in heutigen GUIs umgesetzt. Auf Shneiderman geht im Bezug auf direct manipulation

auch der Begriff WYSIWYG ("what you see is what you get") zurück, unter dem man heute leicht abgewandelt eine Anzeigemethode versteht, die Dokumente auf dem Bildschirm so darstellt, wie sie auf dem Ausdruck später erscheinen werden (vgl. direct manipulation, 2005).

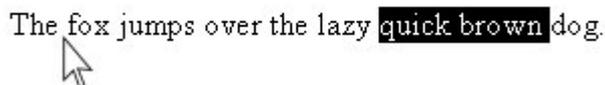
3.5.3. GOMS

GOMS Modelle (Goals, Operators, Methods, und Selection rules) stellen einen wichtigen Aspekt in der Mensch-Computer Interaktion dar, weil sie die Auswirkungen von Design auf den Erfolg, aber auch das Leistungsverhalten einer Anwendung vorhersagen können.

Vereinfacht gesagt, versuchen GOMS Modelle die Aktivitäten des Benutzers in Teilbereiche und Teilaufgaben zu zerlegen, um ihren Informationsgehalt und einzelne Teilziele besser definieren und erklären zu können und aus diesen Subzielen das Gesamtziel ableiten zu können. Dabei geht jedes GOMS Modell vom folgenden Sachverhalt aus: Ein Benutzer hat eine bestimmte Absicht oder ein definiertes Ziel. Er versucht dieses Ziel zu erreichen, indem er Methoden und Verfahren über Bedienungselemente (z.B. Maus, Tastatur) an den Computer weiterleitet. Dabei verwendet er bewusst oder unbewusst Auswahl-Regeln, um die für ihn passenden Methoden zu wählen.

Die Grundsätze von GOMS Modellen können an diesem Beispiel verdeutlicht werden. Der in der unterstehenden Abbildung schwarz hervorgehobene Text „quick brown“ soll in diesem Beispiel vor dem Wort „fox“ eingefügt werden.

GOMS Modelle stellen eine Methode dar, um eine Aufgabe und das Wissen eines Users über das Lösen einer Aufgabe in die Kategorien Goals, Operators, Methods und Selection rules zu unterteilen.



The fox jumps over the lazy quick brown dog.

Abbildung 15: Textbearbeitungsaufgabe (Carrol, 2001, S..59)

Goals sind hierbei die Ziele des Users, also das, was der Benutzer mit einer bestimmten Anwendung erreichen will. In unserem Fallbeispiel ist es das Ziel, den Text zu bearbeiten oder genauer die Worte „quick brown“ vor das Wort „fox“ zu stellen.

Operators sind die Menge aller Aktionen, die dem Benutzer von der Software zur Verfügung gestellt werden, um sein Ziel zu erreichen, also von Kommandoeingaben über Buttons und Menüauswahl bis zu Tastenkombinationen auf der Tastatur.

Der im Englischen als „keystroke level“ bekannte Begriff umfasst hierbei alle Aktivitäten vom Bewegen des Cursors und dem Klicken mit der Maus bis zur Verwendung von Shortcuts wie etwa STRG-X und STRG-V.

Methods sind alle vom Benutzer gelernten aufeinander folgenden Abläufe und Sequenzen von Subzielen, die umgesetzt werden müssen, um das eigentliche Ziel zu erreichen.

Die folgende Auflistung von Subzielen oder zu erledigenden Subtasks ist eine von mehreren Möglichkeiten, um das in Abbildung 16 abgebildete Hauptziel im Programm Microsoft Word zu erreichen:

- Den betreffenden Text markieren
- STRG-X drücken
- Den Cursor an die gewünschte Stelle manövrieren
- Klicken der linken Maustaste
- STRG-V drücken

Diese gereihten Vorgänge müssen vom Benutzer durchgeführt werden, damit er sein Hauptziel erreichen kann. Wobei das Subziel „Text markieren“ wiederum durch mehrere unterschiedliche Methoden erreicht werden kann. (Doppelklick, Klick und Cursorbewegung, SHIFT und Pfeiltasten).

Wenn es also, wie in diesem Fall, mehr als einen Weg zum Ziel gibt, kommen die so genannten Selection rules zum Einsatz.

Selection rules sind die vom User bevorzugten persönlichen Lösungen und Wege zum Ziel. So kann ein Benutzer es bevorzugen, ein falsches Wort zu löschen und es an der richtigen Position neu zu schreiben, während andere User, die etwa Shortcuts nicht beherrschen, über das Menü „Bearbeiten“ ihr Ziel erreichen.

Goals	Hauptziel	Der Text „quick brown“ soll vor dem Wort „fox“ eingefügt werden
	Subziel	Text auswählen (highlighten)
Operators	Aktionen	Maus bewegen Mausbutton klicken Tastatur benutzen
Methods	Abläufe	Für das Hauptziel: 1. Wort löschen und neu eingeben 2. Cut & Paste mit Shortcuts 3. Cut & Paste über das Menü Für das Subziel: 1. Mauszeiger über das Wort ziehen 2. Wort doppelklicken 3. Shift und Pfeiltasten 4. Klicken und Shift + Klicken
Selection rules	Präferenzen	Für das Hauptziel: Wenn das Wort unfertig ist -> Ablauf 1 Wenn das Wort lang ist -> Ablauf 3 Für das Subziel: Wenn es mehr als ein Wort ist -> Ablauf 1

Tabelle 2: GOMS Beispiel für Textbearbeitung (vgl. Correll, 2001, S. 61)

GOMS Modelle eignen sich hervorragend, um Situationen zu analysieren, in denen der User eine Aufgabe bewältigt, die er schon einmal bewältigt hat. Es handelt sich um Problemstellungen, bei denen der Benutzer nicht auf der Suche nach der Lösung ist, sondern bei denen er die Lösung kennt und nur noch handeln muss.

GOMS Modelle können helfen, sowohl qualitative als auch quantitative Aussagen über Programme zu treffen. Quantitativ bieten sie eine gute Abschätzung der absoluten Performance-Zeit, aber auch über die relative Zeit, die man benötigt, einen neuen Lösungsablauf zu lernen, sagen sie viel aus.

Qualitativ werden GOMS Modelle oft verwendet, um Hilfen und Trainings Programme zu verbessern und inhaltlich zu strukturieren. Es hat sich auch herausgestellt, dass diese Modelle effiziente Mittel darstellen um, Tutorials oder Dokumentationen zu erstellen.

Denn ein GOMS Modell ist nichts anderes als ein Modell des Wissens, das benötigt wird, um eine gegebene Aufgabenstellung zu bewältigen. Um einem Benutzer nun zu helfen, muss nur sein Ziel bekannt sein und schon können alle möglichen Lösungswege dargestellt werden.

Bei der Softwareentwicklung selbst können GOMS insofern nützlich sein, da etwa Abläufe für häufige Aufgabenstellungen optimiert werden können und zusätzliche Aktionen zur Umsetzung erstellt werden können. Weiters werden Ziele ohne oder mit uneffizienten Abläufen aufgedeckt, und inkonsistente Methoden können entfernt werden.

GOMS werden seit den frühen Achtzigerjahren von HCI Forschern verwendet, ihre Aussagekraft wurde an vielen Situationen und Programmen getestet (vgl. Correll, 2001, S. 62). (GOMS Weblink [7])

3.5.4. Menschliches Gedankenmodell

Hier ist es leider nicht möglich ein spezifisches Fallbeispiel zu liefern, da diese Modelle ihren Ursprung in der Psychologie finden und ausführlicher Beschreibung bedürfen. Allerdings kann kurz umrissen werden, dass es sich hierbei um Modelle der menschlichen Informationsverarbeitung handelt. Primär gehen menschliche Gedankenmodelle davon aus, dass sich Menschen geistige Modelle ihrer Umwelt bilden, um sich darin zurechtzufinden.

Selbiges gilt auch für Computer und Maschinen, daher sollte es ein ausgesprochenes Ziel von Interfacedesignern sein, Schnittstellen so zu gestalten, dass sie tun, was sich der Benutzer von ihnen erwartet.

Geistige Modelle entstehen nach Ansicht von Psychologen durch Prägung, Erfahrung und Lernprozesse.

Daher ist es wichtig, diesen Modellen zu entsprechen, da eine passende Umsetzung den Usern das Benützen der Schnittstelle erleichtert.

Das bedeutet etwa, dass jemand, der einmal das „X“ am rechten oberen Rand einer Anwendung auf Windows angeklickt und daraus gelernt hat, dass das die aktuelle Anwendung schließt, erwartet, dass dieses Icon bei allen anderen Anwendungen genau denselben Prozess auslöst. (Menschliches Gedanken Modell Weblink [8])

3.6. Ziele und Absichten von HCI

Das Ziel aller HCI Forscher ist es, die bestehenden Schnittstellen zwischen Menschen und Computern zu verbessern und neue Wege der Kommunikation für Mensch und Computer zu ermöglichen. Das Hauptaugenmerk liegt hierbei auf der Erforschung und der Entwicklung alternativer Input- und Displaytechnologien.

Besonderer Wert wird dabei auf funktionelles Interaction Design gelegt, um intuitive Verwendbarkeit und maximale Benutzerfreundlichkeit zu ermöglichen.

Weitere Ziele sind die Entwicklung von neuartigen Navigationstechniken in virtuellen Umgebungen und eine verbesserte Informationsvisualisierung abstrakter, mehrdimensionaler Daten, um sie so auf eine effiziente Art und Weise vermitteln zu können.

Es soll also durch funktionelles und geplantes Design sowohl die Interaktion und Kommunikation zwischen Mensch und Computer, als auch das Umgangs-vermögen mit Computern erleichtert und verbessert werden.

3.7. HCI Formen

Mensch-Computer Interaktion gewinnt immer mehr an Bedeutung, da die Ausbreitung der Informationstechnik in allen Lebensbereichen praktisch alle Menschen auf unterschiedliche Weise zu Benutzerinnen oder Benutzern von Computern macht.

Deshalb wird die Frage der Benutzbarkeit nicht mehr nur für Computer am Arbeitsplatz gestellt, sondern stellt sich viel zentraler, da wir uns immer mehr zu einer Informationsgesellschaft hin entwickeln. Daher gibt es auch unzählige Formen und Ausprägungen der Mensch-Computer Interaktion, auf die hier nicht eingegangen werden kann.

Die durch Mensch-Computer Interaktion hervorgebrachten Benutzerschnittstellen lassen sich in zwei grundlegende Kategorien einteilen, in Input- und in Outputsysteme. Wobei bei dieser Betrachtungsweise meist vom Computer als Bezugsobjekt ausgegangen wird, und die Schnittstellen wiederum in den verschiedensten Ausprägungen auftreten können.

3.7.1. Inputsysteme

Inputsysteme sind alle Methoden und Geräte, die es einem Computer ermöglichen, Informationen aus seiner Umwelt „wahrzunehmen“, also sowohl Systeme über die der Mensch Informationen in einen Computer eingibt, als auch Sensoren, mit denen ein Computer autark Informationen sammelt.

Neben den standardisierten Eingabegeräten wie Keyboard, Mouse, Joystick und Touchscreens gibt es auch noch exotischere Eingabemedien, wie etwa den Datenhandschuh. Außerdem werden Inputsysteme wie optische Sensoren, Näherungsmelder oder Kameras unter dem Synonym Computer Vision verwendet, um menschliche Gesten zu erkennen, Bewegungen zu verfolgen oder andere visuelle Eingaben zu ermöglichen.

3.7.2. Outputsysteme

Outputsysteme bezeichnen jene Methoden und Geräte, die es dem Computer ermöglichen, Informationen an die Umwelt abzugeben. Diese Systeme sind so gestaltet, dass sie die vom Computer bereitgestellten Informationen in einer dem Menschen entsprechenden Art und Weise ausgeben können, also visuell, akustisch, haptisch oder sogar multi-modal.

Akustische und visuelle Feedbackmethoden müssen meiner Meinung nach hier nicht ausgeführt werden, da heutzutage Outputsysteme wie Beamer, Bildschirme, Drucker, Plotter oder Boxen weit verbreitet sind.

Multimodalität hingegen ist kein landläufiger Begriff, deshalb sollte hier kurz darauf eingegangen werden. Multimodal bedeutet, das Bestmögliche aus allen möglichen Systemen zu vereinen, um einen optimal auf die menschlichen Sinne abgestimmten Output zu ermöglichen.

Es handelt sich also um eine Mischung aus akustischem, visuellem und haptischem Feedback, das den Menschen als Ganzes anspricht und einbindet. (Outputsysteme Weblink [9])

3.7.3. Herkömmliche Benutzerschnittstellen

CLI - Command line interfaces

Der große Nachteil von Command line Interfaces ist, dass der Benutzer dem Computer seine Ziele in Form einer bestimmten Syntax mitteilen muss, und diese meist eine längere Lernphase seitens des Users in Anspruch nimmt.

Command line interfaces oder Textinterfaces bilden die Grundlage aller Benutzerschnittstellen und sind trotz fortgeschrittener Entwicklung neuartiger Interfaces noch lange nicht aus der Mode, sondern erfreuen sich gerade bei „heavy users“ großer Beliebtheit, da sie eine schnelle Abwicklung der anstehenden Aufgaben ermöglichen.

GUI – Graphical user interfaces

Das Grafische User Interface ist im Vergleich zum CLI einfacher zu bedienen und ermöglicht auch ungeübten Benutzern das Handhaben eines Computers.

Durch die starke visuelle Prägung des Menschen auf Symbole ist ein gut gestaltetes GUI leicht zu erlernen und intuitiv verwendbar. Mit Symbolen kann innerhalb von Bruchteilen von Sekunden viel Information übermittelt werden, weil sie auf bereits bestehendes Wissen des Benutzers zurückgreifen (Papierkorb Icon usw.).

In der Fachliteratur wird das GUI auch oft als *WIMP* bezeichnet, was abgekürzt für "window, icon, menu, pointing device" steht.

AUI – Auditory user interfaces

Auditory User Interfaces geben dem Benutzer akustische Rückmeldungen auf die von ihm getätigte Eingabe, die ebenfalls akustischer Natur sein kann.

Grundsätzlich teilen sich AUIs in sprachliche und nicht-sprachliche Schnittstellen, wobei sprachliche Schnittstellen Sprachein- und -ausgabe beinhalten, und nicht-sprachliche Systeme jegliche akustische Unterstützung der HCI, wie „Earcons“ oder „Auditory Icons“, (Piepgeräusch bei Fehlermeldungen usw.) beinhalten.

DMI – Direct Manipulation Interfaces

Bei DMIs geht es lediglich darum, dass jeglicher Input des Benutzers sofort, also auf direktem Weg zu sinnvollem Output führt.

Direct manipulation Interfaces stehen daher eng mit taktilen Schnittstellen (für Blinde) und mit GUIs in Verbindung, da sie umgehend auf menschliche Inputs reagieren (Fenster verschieben bei Windows, bitte warten bei Downloads usw.).

Gemischte Interfaces (Mixed interfaces)

Gemischte oder mixed Interfaces sind Schnittstellen, die sich mehrerer Feedbackmethoden bedienen, um ein für den Benutzer optimiertes Feedbacksystem zur Verfügung zu stellen. So ist beispielsweise bei vielen GUIs zusätzlich noch die command line als Schnittstelle verfügbar, und Bestätigungs- und Abbruchsounds unterstützen den Menschen beim Bedienen.

3.8. Ausblick und Visionen

Heutzutage sind beinahe alle menschlichen Möglichkeiten der Kommunikation auch auf Computersysteme anwendbar. So können Mobiltelefone über ein Audiointerface die Nummer des gewünschten Gesprächspartners wählen, während andere Anwendungen mittels Kameras in der Lage sind, auf Bewegungen ihres Benutzers zu reagieren. Allerdings sind noch nicht alle HCI-Schnittstellen robust und anwenderfreundlich genug um sich zu etablieren.

Die wohl bekanntesten und am weitesten verbreiteten Schnittstellen zwischen Menschen und Computern sind visuelle und haptische, also sichtbare und greifbare Input-/Outputdevices (z.B. der Bildschirm, die Maus und die Tastatur). Wobei hier nicht nur die Maus an sich, oder der Bildschirm an sich, als Schnittstelle zu betrachten ist, sondern alle dazugehörigen informationsvermittelnden Medien. Bei der Maus zum Beispiel zählen auch die unterschiedlichst ausgeformten Tasten und Scrollräder als Inputdevice, während bei einem Bildschirm auch die grafische Benutzeroberfläche mit jedem einzelnen Button zur Schnittstelle beiträgt.

Für die Darstellung dreidimensionaler virtueller Umgebungen wurden stereoskopische Brillen entwickelt, die beide Augen mit unterschiedlichen Bildern versorgen und somit eine virtuelle Realität schaffen, in der etwa mit einem Datenhandschuh navigiert und interagiert werden kann.

Aller Wahrscheinlichkeit nach werden in Zukunft Computer im alltäglichen Leben eine sehr viel zentralere Rolle spielen, als sie es heutzutage schon tun. Diese Entwicklung wird durch die Miniaturisierung und die sinkenden Hardwarekosten noch beschleunigt.

Das Verhältnis von Mensch und Computer wird enger und auch die Gestaltung von Schnittstellen und somit HCI werden eine zunehmend wichtigere Rolle spielen.

Die wichtigsten Teilgebiete der HCI in diesem Zusammenhang sind:

- New I/O devices
- Augmented Reality
- Virtual Reality
- Mass availability
- Ubiquitous computing

3.8.1. New I/O devices

Es werden aber auch ständig neuartige Wege und Formen der Mensch-Computer Interaktion erforscht und entwickelt, wie diese Beispiele einiger Spielehersteller zeigen.

Etwa das Spiel „Donkey Kong Jungle Beat“ (Weblink [10]) von Nintendo®, das über Bongo Trommeln zu steuern ist, oder „The Journey to Wild Divine: The Passage“ (Weblink [11]) von The Wild Divine Project, bei dem der Spieler das Spiel über Bio-Feedback, also Puls und Hautleitwert steuern kann.



Abbildung 16: The Passage (The Wild Divine Project, 2005) & Donkey Kong Jungle Beat (donkeykong, 2005)

Ein weiteres Beispiel wäre der „Gametrack“ von Atari. Bei diesem System steuert der Nutzer ein Spiel durch Bewegungen, die von zwei Spezialhandschuhen über Seilzüge mit einer Auswertungseinheit verbunden sind. Anhand des Winkels und der Seillänge kann das System die exakte Position der Hände im Raum feststellen. [Videobeispiel auf beigelegter CD] (Weblink [12])

Mit der „SuperArena“ von Zapgames wurde ebenfalls versucht, der reinen Daumenakrobatik beim Spielen ein Ende zu setzen. Über eine Bodenmatte und vier kabellose Funksensoren an Hand- und Fußgelenken wird die Interaktion mit jeglichen Spielekonsolen ermöglicht (Weblink [13]). [Videobeispiel auf beigelegter CD]

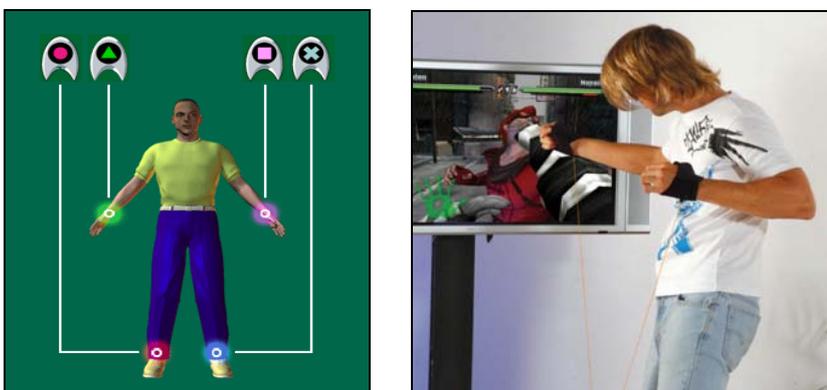


Abbildung 17: Gametrak Super (Arena, 2005) & Superarena (Atari, 2005)

Bei Spielen wie dem Sony Spiel EyeToy® (Weblink [14]) oder dem Spiel „Kick Ass Kung-Fu“ von Animaatiokone Industries [Videobeispiel auf beigelegter CD]



Abbildung 18: Kick Ass Kung-Fu (Animaatiokone Industries, 2005)

(Weblink [15]) nimmt der Spieler über eine Webcam, die oberhalb des Fernsehers platziert wird, aktiv im Videospiel teil. Diese Spiele sind meines Erachtens Vorzeigebispiele von angewandter Computer Vision als HCI Methode.

Ein weiteres Beispiel für neuartige Technologien im Displaybereich wäre etwa das Heliodisplay von IO2Technology, das virtuelle Inhalte direkt in die Luft projizieren kann und sogar Interaktion ermöglicht. (Weblink [16]). [Videobeispiel auf beigelegter CD]

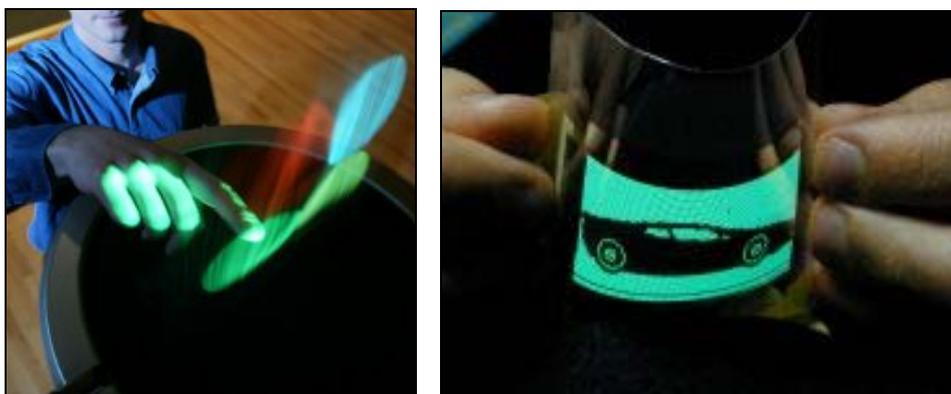


Abbildung 19: Heliodisplay (IO2Technology, 2005) & OLED (Arbeitsgruppe LED, 2005)

Ebenso sind auch die biegsamen, hauchdünnen OLED (organic light-emitting diode) Displays (Weblink [17]) als neuartige Displaytechnologien zu sehen.

Hierbei wird schon deutlich, dass der Zweck der am Computer ausgeführten Anwendung und somit das Ziel des Users eine maßgebende Rolle für die gewünschte HCI Methode bildet.

Es wäre zum Beispiel sinnlos, eine 3D Brille als Outputdevice für Textverarbeitungsprogramme zu wählen, da Buchstaben ja schon in der Realität zweidimensionale Objekte sind. Das zeigt wieder, dass bei der Human-Computer Interaction neben den an der Schnittstelle beteiligten Partnern (Mensch und Computer) auch viele andere Parameter, wie etwa das Ziel des Benutzers auf die Ausformung der Schnittstelle Einfluss nehmen und zu ihrer Gestaltung beitragen.

3.8.2. Augmented Reality

Ein weiterer Begriff, der im Zusammenhang mit der zukünftigen Entwicklung von HCI nicht ungenannt bleiben darf, ist Augmented Reality. Mit diesem Begriff werden technische Entwicklungen bezeichnet, mit denen die menschliche Wahrnehmung der physikalischen Realität um virtuelle, vom Computer erzeugte Inhalte und Informationen erweitert wird.

Beispielsweise gelingt das durch die Verwendung einer AR-Brille, die direkt am Auge die Real-szenen halbtransparent mit digitalen Computerszenen überlagert.

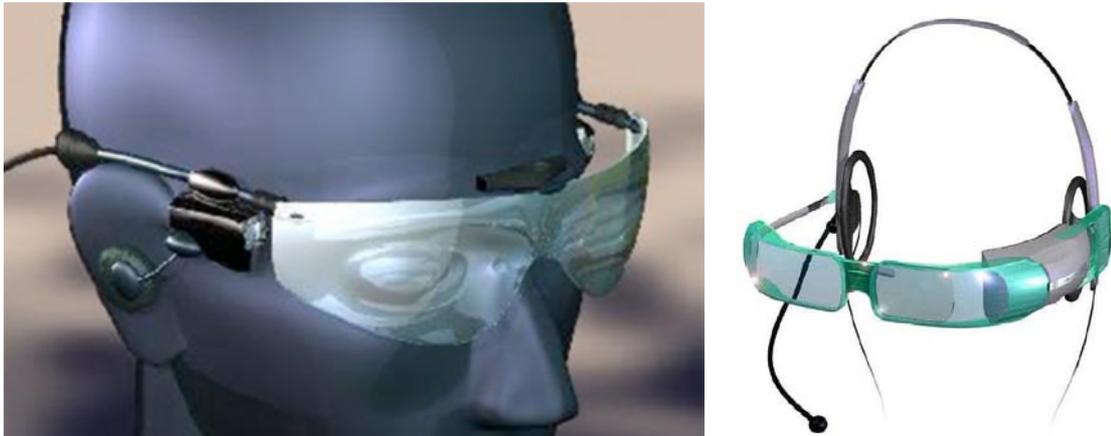


Abbildung 20: Augmented Reality Brille (New HCI Visions and Future PDAs, 2003)

Augmented Reality versucht, die reale mit der virtuellen Welt zu verschmelzen, oder zumindest die reale Welt mit virtuellen Inhalten zu bereichern.

Letztendlich ist es das ausgesprochene Ziel von AR, die Vermischung virtueller und realer Sachverhalte qualitativ so weit voranzutreiben, dass es Menschen unmöglich ist, zwischen realen und virtuellen Objekten zu unterscheiden.

3.8.3. Virtual Reality

Ein wichtiges Schlagwort der Zukunft ist sicher auch Virtual Reality. Hierbei sollen virtuelle Wirklichkeiten erzeugt werden, in denen dem Benutzer das Gefühl des In-der-Welt-Seins vermittelt werden soll. Der Benutzer wird Teil einer virtuellen Welt und somit selbst zum Bestandteil einer ganzheitlichen Benutzerschnittstelle.
(Beispiel: Weblink [18])



Abbildung 21: Beispiele Virtual Reality (New HCI Visions and Future PDAs, 2003)

Eine weitere Schnittstelle, die im Zusammenhang mit der Virtual Reality zukünftig sicher an Bedeutung gewinnen wird, sind so genannte Force-Feedback-Systeme. Sie versuchen, die sensomotorischen Fähigkeiten des Menschen beim Umgang mit virtuellen Objekten zur Geltung zu bringen. Force-Feedback-Systeme beruhen auf den feinsinnigen und kontrollierten Muskelbewegungen der Hand und der Finger, die mit unterschiedlichsten Methoden abgenommen werden können und somit präzise Interaktion ermöglichen.



Abbildung 22: Beispiele Force-Feedback Interfaces (ForceDimensions, 2004)

3.8.4. Mass availability

Mass availability bezeichnet die bereits angedeutete Entwicklung der Hardware. Derzeitige Trends weisen darauf hin, dass Hardware und damit auch Computer bei gleich bleibender Leistung zukünftig immer kleiner werden. Diesen Trend bezeichnet man auch als Miniaturisierung. Die sinkenden Hardwarekosten und die explosionsartige Steigerung von Rechenleistung und Speicherkapazität der Geräte tragen noch weiter zu dieser Entwicklung bei.

3.8.5. Ubiquitous computing

Ubiquitous Computing als Gegenkonzept zur Virtual Reality könnte zukünftig auch an Bedeutung gewinnen. Hierbei soll die räumliche Umgebung durch eine umfassende Integration von computerisierten Artefakten angereichert werden. Bei dieser Vision steht der allgegenwärtige Computer dem Menschen in allen Lebenslagen als Informationsquelle und zur Lösung von Problemen zur Verfügung. Allerdings wirkt dieses Szenario noch auf viele Menschen eher abschreckend (gläserner Mensch).

Zum Abschluss dieses Kapitels bleibt nur noch hinzuzufügen, dass die Einbeziehung zusätzlicher Medien wie Audio, Video, 3D-Visualisierung und Animation durch neuartige Darstellungstechniken und neue Interaktionsmethoden wie Virtuelle Realität, aber auch neue Eingabesysteme, wie Sprache oder Gesten, dem Menschen die Benutzung eines Computers sicher erleichtern werden.

4. Computer Vision

4.1. Einführung

Computer Vision ist eine von vielen Sparten der künstlichen Intelligenz (englisch artificial intelligence). Diese Disziplin der Informatik versucht, vereinfacht gesagt, einem Computer menschliches Sehen beizubringen.

Besser gesagt wird versucht, die optische Wahrnehmung von Menschen zu imitieren, sie auf objektiv messbare Parameter zu reduzieren und im Weiteren auch zu automatisieren. Computer Vision, kurz CV, hat nicht nur das Ziel Bilder zu sehen, sondern auch Informationen daraus abzuleiten und diese zu interpretieren.

CV versucht durch die detaillierte Analyse bestimmter Bildeigenschaften Informationen über die tatsächlichen Inhalte eines Bildes abzuleiten.

Es werden auf Grund der Eigenschaften eines Bildes (Farbwerte, Helligkeit, etc.) Informationen über den im Bild dargestellten Sachverhalt (etwa bewegte Objekte, oder die Dreidimensionalität eines Raumes) ermöglicht.

Computer Vision besteht aus zwei grundlegenden Teilgebieten, dazu zählen die Bildverarbeitung (englisch image processing) und das Extrahieren von Informationen über die Umwelt auf Bildfolgen.

Die eingehenden Bildfolgen können dabei aus digitalen Videokameras, Infrarotkameras, Röntgen oder Ultraschallquellen kommen, auch Laserentfernungsmesser oder andere zwei- bzw. dreidimensionale Inputdaten sind denkbar.

Hinter der Entwicklung von Computer Vision und Gestikererkennung stehen zwei bedeutende Faktoren.

Einerseits ist es die technische Entwicklung von Computern im Allgemeinen, da Rechenleistung und Speicherkapazität eine wichtige Grundlage für die Bildverarbeitung darstellen.

Andererseits ist es die ausgeklügelte und anspruchsvolle menschliche Wahrnehmung, die für uns so selbstverständlich und simpel erscheint und uns glauben lässt, dass einfache optische Aufgaben auch von Maschinen durchgeführt werden können.

Allerdings ist „Sehen“ bei weitem nicht so einfach, wie es auf den ersten Blick scheint. Wir erkennen Flächen, Kanten, Objekte, Tiefe und andere Aspekte auf Grund unserer feinfühligsten Wahrnehmung von Farbnuancen und unterschiedlichster Helligkeitswahrnehmung.

Wollen wir nun einen Computer lehren, wie ein Mensch zu sehen, muss er in der Lage sein, in zahlreichen Subroutinen die unterschiedlichen Eigenschaften und Charakteristika aus dem Vergleich der einzelnen Pixel zu filtern und zu interpretieren.

Der augenscheinlichste Vorteil einer visuellen Mensch-Maschine Kommunikation ist wohl, dass es uns so möglich ist, über große Distanz mit einem Computer zu kommunizieren und das gänzlich ohne physischen Kontakt zu jeglichem Eingabemedium. Im Vergleich zu Audiointerfaces, bei denen der Computer die menschliche Sprache mittels Mikrofon interpretiert, hat Computer Vision den Vorteil, dass sie auch in Bereichen des täglichen Lebens einsetzbar ist, in denen Audiointerfaces auf Grund des akustischen Rauschens versagen würden.

Dies ist hauptsächlich in öffentlichen Räumen auf Grund des hohen Geräuschpegels gegeben.

Die Grundidee der Computer Vision ist es, dem Menschen, so wie er ist, das Bedienen eines Computers zu ermöglichen und das gänzlich ohne Kabel oder Fernbedienungen.

4.2. Grundlegendes zur Computer Vision

Grundsätzlich beschäftigt sich Computer Vision mit der Echtzeitverarbeitung von Bildern und Bildsequenzen, um daraus Informationen über die Umwelt zu extrahieren. Es gibt zahllose Anwendungsgebiete, die von medizinischen Anwendungen über die Nahrungsmittelindustrie bis zur Unterhaltungselektronik reichen.

Kurz gesagt geht es bei Computer Vision darum, Szenen zu verstehen, und Informationen über die Realität zu gewinnen, indem die Eigenschaften der Szene oder besser gesagt die Eigenschaften ihres Abbilds analysiert werden. Der Input, also die Bilder oder Bildfolgen, kommt dabei meist aus digitalen Video Kameras, es können aber auch Infrarotkameras, Röntgenbilder oder auch Kameras mit UV Filtern verwendet werden.

Computer Vision hat sich dabei das höchste aller möglichen Ziele ausgesucht, nämlich die hoch entwickelte und extrem spezialisierte visuelle Wahrnehmung des Menschen zu imitieren, aber auch andere Ziele, wie etwa die Roboter Navigation oder aber auch die Verkehrsüberwachung werden verfolgt.

Zwar wird schon seit Jahrzehnten an diesem Teilbereich der Informatik und der künstlichen Intelligenz geforscht und trotzdem konnten erst die grundlegenden Probleme, wie die Objekt- bzw. Bewegungserkennung oder die stereoskopische Sicht erforscht werden. Dank der Entwicklung der letzten Jahre, die eine enorme Steigerung der Speicherkapazität und der Rechenleistung brachte, wurde auch die Entwicklung der Computer Vision vorangetrieben.

Wie eingangs erwähnt, gibt es viele Toolkits, die es ermöglichen, Daten von einer Kamera zu empfangen und zu manipulieren.

Der Umfang dieser Programme variiert dabei von eher einfach zu bedienenden Anwendungen mit grafischer Benutzeroberfläche bis zu Entwicklungsumgebungen, in denen sogar einzelne Pixel angesprochen und manipuliert werden können.

Egal welches dieser Programme man verwendet, der erste Schritt ist meist dieser:

Das Bild der Kamera wird in eine Matrix (mehrdimensionales Array) umgesetzt, in der jedes Feld zum Beispiel die RGB Farbwerte des betreffenden Pixels enthält.

Um Veränderungen in einer Bildsequenz wahrnehmen zu können, werden aufeinander folgende Matzritzen mit einander verglichen.

Es gibt einige Begriffe, die im Zusammenhang mit Computer Vision näherer Erläuterung bedürfe, da viele Anwendungen diese oder ähnliche Begriffe und die dahinter stehenden Vorgehensweisen verwenden.

Echtzeit Videomanipulation

Bei dieser Art der Bildverarbeitung handelt es sich nicht um Computer Vision im eigentlichen Sinn. Meist wird das empfangene Live-Bild einer Kamera nur leicht verändert und dann wieder für den Benutzer dargestellt. Ähnlich einem Filter in Adobe Photoshop, wird jedes Einzelbild nach einer bestimmten Vorschrift manipuliert und wieder ausgegeben.

Motion Capturing

Der Begriff Motion Capturing steht für Bewegungserfassung. Man versteht darunter eine Technik, die es ermöglicht, menschliche Bewegungen in ein mit Computern lesbares Format umzuwandeln. Dazu werden die menschlichen Akteure meist in einen Anzug gesteckt, der mit verschiedenen weißen Bällen oder Lämpchen beklebt ist. Spezialkameras verfolgen die Bewegungen der Markierungspunkte. Die so gewonnenen Daten werden dann durch eine Motion-Capturing-Software analysiert und können beispielsweise in ein 3D Programm importiert werden.

Motion Tracking

Motion Tracking ist das Echtzeit Equivalent zum Motion Capturing. Tracking versucht, ohne den Benutzer mit zusätzliche optischen Markierungen oder speziellen Anzügen auszustatten, Bewegung in Bildsequenzen wahrzunehmen.

Hierbei wird meist versucht, Blobs (binary large objects) im Bild zu erkennen. Ein Blob ist hierbei ein Bereich von Pixeln, die gewisse Ähnlichkeiten aufweisen, beispielsweise annähernd gleiche Helligkeits- oder die Farbwerte. Die Position dieser Pixelfläche wird nun über mehrere Frames festgestellt und somit wird das Blob getrackt, also die Position ähnlicher Pixel über längere Zeit hinweg lokalisiert.

Objekterkennung

Objekterkennung nützt zusätzlich zum Tracking noch die so genannte Pattern-recognition, bei der nach einem Blob gesucht wird, das einem bestimmten Muster entspricht.

Dies ist die schwierigste Disziplin, da nicht nur die Position des Objekts getrackt werden muss, sondern auch das Objekt als solches erkannt werden sollte.

Nun ist klar, dass ein Computer die Position eines Pixels und seinen Farbwert auslesen kann. Allein aus diesem Wissen lassen sich ein paar Grundregeln für die Interpretation von Videobildern ableiten:

- Das hellste Pixel eines Frames kann bestimmt werden, indem das Pixel mit den höchsten Farbwerten im Bild ermittelt wird.
- Ein Colorblob (Farbbereich) kann bestimmt werden, indem die Zielfarbe definiert wird, ein Abweichungsbereich (Threshold) angegeben wird und dann in einem ebenfalls definierten benachbarten Bereich überprüft wird ob die Pixel innerhalb des Abweichungsbereiches um die Zielfarbe liegen.
- Bewegte Bereiche können erkannt werden, indem die Pixel eines Frames mit denen seines Vorgängers verglichen werden und geprüft wird, ob signifikante Unterschiede der Farbwerte aufgetreten sind.
- Muster können verfolgt werden, indem ein Bereich definiert wird, innerhalb dem das Muster zu suchen ist und nun immer wieder geprüft wird, ob sich darin ein festgelegtes Pixelmuster finden lässt(vgl. Physical Computing, 2005).

Aus diesen Grundregeln folgen wiederum einige Richtlinien für den Aufbau eines Computer Vision Systems:

Um Farbbereiche verfolgen (tracken) zu können, sind konstante Lichtverhältnisse wichtig, da ein Computer die Farbe Rot an sich nicht erkennen kann, sondern nur in der Lage ist festzustellen, ob ein oder mehrere Pixel dem RGB Wert (255,0,0) entsprechen. Ändern sich die Lichtverhältnisse, sind auch die Farbwerte nicht mehr eindeutig zuzuordnen.

Formen oder Muster, deren Positionen verfolgt werden sollen, müssen einigermaßen konstant bleiben, damit sie der Computer in mehreren Frames als gleich bzw. ähnlich erkennt. Eine offene Hand kann zum Beispiel von der Seite gesehen nicht mehr als solche erkannt werden.

Da die Kamera die Informationsquelle für Computer Vision darstellt, legt die Position der Kamera in den meisten Anwendungsfällen fest, welche Informationen man überhaupt erhalten kann. Soll etwa die zweidimensionale Position einer Person in einem Raum festgestellt werden, ist es weitaus schwieriger, diese Information aus dem Bild einer frontal positionierten Kamera aus zu lesen, als aus der Vogelperspektive. Man sollte die Kamera in diesem Fall also möglichst an der Decke des Raumes befestigen.

In vielen Fällen wird die Kamera hinter oder gegenüber der eigentlichen Projektionsfläche statisch platziert, um den Benutzer in die Projektion einzubinden und ihn so als Schatten oder Spiegelbild an der Projektion teilhaben zu lassen.

(Computer Vision Weblink [19])

4.3. Computer Vision Methoden

Die unterschiedlichen Methoden der Computer Vision entspringen den unterschiedlichen Zielen und Applikationen, in denen sie verwendet werden.

Aus diesen angestrebten Zielen lassen sich die folgenden Kategorien für die Einteilung der Computer Vision Methoden bilden. Allerdings sind die Grenzen dieser Kategorien verschwommen und nicht klar definiert, trotzdem möchte ich die Computer Vision in dieser Arbeit unter folgenden Blickwinkeln betrachten:

Bildverarbeitung (Image Processing)

Hierbei handelt es sich um mehr oder weniger einfache Bildmanipulation.

Es wird als Eingangssignal ein Bild verwendet, das nach bestimmten Regeln verändert und manipuliert wird (ähnlich einem Photoshop Filter), um dann wieder als Bild ausgegeben zu werden.

Beispiele für diese primitive Operation wären Rauschunterdrückung (Noisereduction) und Kontrast- oder Schärfverbesserungen.

Bildanalyse (Image Analysis)

In der Bildanalyse wird versucht, die ankommenden Bilder des Eingangssignals gezielt und auf bestimmte Art und Weise zu manipulieren, um so Informationen über den real dargestellten Sachverhalt zu gewinnen und diese Messungen schlussendlich ausgeben zu können.

Beispielsweise wird bei der Bildanalyse das Bild mit Hilfe der Methoden aus der Bildverarbeitung in Regionen oder Objekte unterteilt und Informationen wie Fläche, Kontur, Identität oder Anzahl der Objekte, ausgelesen.

4.3.1. Bildverarbeitung

Der Bildverarbeitung oder auch dem Image Processing (Weblink [19]) liegen zwei unterschiedliche Interessen zu Grunde. Einerseits versucht man das menschliche Verständnis für bildhafte Information zu verbessern und andererseits will man natürlich die Verarbeitung und Vorbereitung von Bildern zum Speichern oder Übertragen optimieren. Grundsätzlich jedoch bezeichnet der Begriff Bildverarbeitung oder auch digitale Bilderverarbeitung das Be- und Verarbeiten von digitalen Bildern mit Hilfe eines Computers.

Ein Bild wird hierbei als zweidimensionale Funktion $f(x,y)$ gesehen, wobei x und y Koordinaten sind und die Amplitude von f der Grauwert oder die Farbzuoordnung des durch die Koordinaten bezeichneten Pixels ist. Wenn nun die Amplitudenwerte dieser Funktion alle endlich und diskret sind, und das Bild aus einer endlichen Anzahl von Elementen mit einer bestimmaren Position besteht, werden diese Bilder digitale Bilder genannt.

Die eben angesprochenen Elemente, die durch die Schnittpunkte der Koordinaten Reihen und Spalten entstehen, unterteilen ein digitales Bild. Die kleinsten Elemente dieser Unterteilungen werden gängigerweise als Pixel bezeichnet.



Abbildung 23: Webcambild normal und Medianfilter (10x10)

4.3.2. Bildanalyse

Wie schon zu Beginn dieses Kapitels erwähnt, wird bei der Bildanalyse eine eingehende Bildersequenz manipuliert und analysiert, um so gezielt Informationen über den dargestellten Sachverhalt zu gewinnen.

Beispielsweise wird bei der Bildanalyse das Bild mit Hilfe der Methoden aus der Bildverarbeitung in Regionen oder Objekte unterteilt und Informationen wie Fläche, Kontur, Identität oder Anzahl der Objekte ausgelesen.

In diesem Abschnitt wird ein kurzer Überblick über die gängigsten Bildanalysemethoden gegeben.

4.3.2.1. Background-Subtraction

Ein Beispiel für eine nützliche und häufig verwendete Manipulation der Bildanalyse ist die „Background-Subtraction“. Diese Methode ermöglicht es, bewegte Objekte oder Bewegung im Allgemeinen in Video- oder Bildsequenzen zu detektieren und auszuwerten.

Wenn Menschen im Allgemeinen von Video oder Bildern reden, geben sie den Termen „Vordergrund“ und „Hintergrund“ eine wichtige Bedeutung.

Der Hintergrund einer solchen Szene kann in groben Zügen als der Teil des Bildes beschrieben werden, dem wenig Aufmerksamkeit geschenkt wird, da wenig Aktivität und somit wenig Bewegung darin stattfindet.

Der Vordergrund hingegen ist der dynamische Teil des Bildes oder Videos. Hier findet Aktivität und somit Bewegung statt. Und genau auf diesen Grundüberlegungen basiert die „Background-Subtraction“. Das Ziel ist es, nur die aktiven Bereiche in Bildsequenzen zu erkennen und somit nur die relevanten Bildbereiche darstellen zu können.

Dazu wird bei der „Background-Subtraction“ ein so genanntes Referenzbild der statischen Szene, also beispielsweise der Bühne ohne Schauspieler, als Definition des Hintergrundes betrachtet. Somit kann jedes bewegte Objekt identifiziert werden, indem das Referenzbild Pixel für Pixel vom aktuellen Bild subtrahiert, also abgezogen, wird.



Abbildung 24: Hintergrundbild | Livebild | Hintergrund vom Livebild subtrahiert

Nach der Subtraktion des aktuellen Bildes vom Hintergrundbild (Referenzbilder sollten erstellt werden, solange keine Bewegung im Bild passiert) erscheint deutlich sichtbar der bewegte Vordergrund. Im Grunde genommen wird nur festgestellt, welche Pixel im Laufe der Zeit ihre Werte geändert haben, da jene Bildteile, in denen keine Änderung geschieht, sich bei der Subtraktion zu Null, also auf Schwarz, reduzieren.

Allerdings sind die theoretisch brillanten Ergebnisse dieses Verfahrens stark von äußeren Einflüssen wie Lichtintensität und vor allem der Qualität der Kamera abhängig.

Um optimale Ergebnisse zu erzielen, müssten hochqualitative Kameras eingesetzt werden. Da das aber in vielen Fällen zu aufwändig oder kostspielig ist, wird meist versucht, einen anderen Weg zu gehen. Hierbei versucht man mit Filtern, das durch die Kamera, das Licht und die Subtraktion erzeugte Rauschen zu minimieren.

4.3.2.2. Blob-Tracking

Das so genannte „Blob-Tracking“ wird verwendet um bewegte Objekte in Videos zu erkennen. Ein Blob (binary large object) ist dabei als eine Folge von zusammenhängenden Pixeln gleichen binären Zustands zu verstehen, die gleichzeitig der ROI (region of interest) entsprechen. Beim „Blob-Tracking“ wird meist versucht, das Zentrum des Blobs, also den Schwerpunkt der Fläche, zu erfassen und dessen Position zu bestimmen. „Blob-Tracking“ in kann in zwei grobe Techniken unterteilt werden. Einerseits gibt es das „Shape-Tracking“, bei dem versucht wird eine bestimmte Form im Bild zu finden und andererseits das „Color-Tracking“, bei dem versucht wird zusammenhängende Bereiche einer bestimmten Farbe zu erkennen.

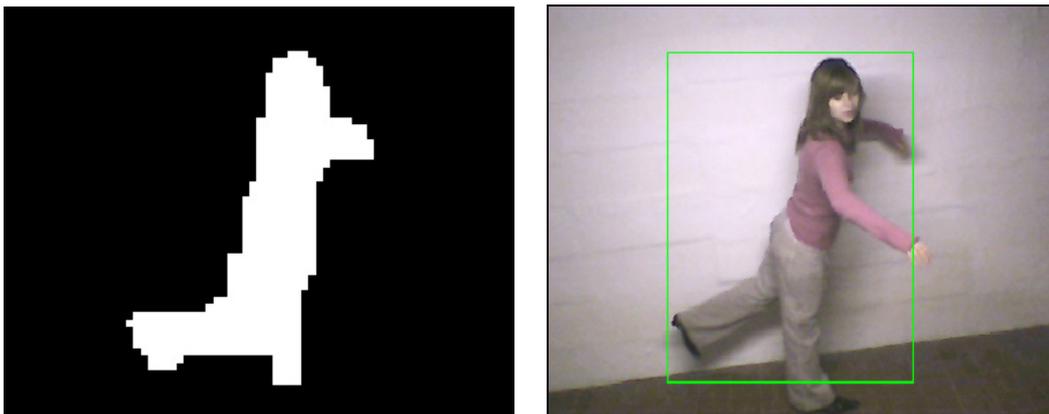


Abbildung 25: „Background-Subtraction“ mit anschließendem „Blob-Tracking“

Im hier gezeigten Beispiel wurde ein durch „Background-Subtraction“ erzeugtes Bild mit Filtern bearbeitet, bis nur noch weiße Pixel für die bewegte Bereiche des Bildes überblieben und der statische Hintergrund vollständig schwarz war.

Somit war es für einen Blob-Tracking-Algorithmus ein Leichtes, die bewegten Bereiche als zusammengehörig zu erkennen.

4.3.2.3. Mustererkennung

Mustererkennung oder „Pattern-Recognition“ ist ebenfalls ein breitgefächertes Forschungsgebiet der künstlichen Intelligenz und selbstlernender Systeme.

Allerdings ist es auch ein weiteres Betätigungsfeld der Computer Vision, das Anwendungsbereiche, wie das Erkennen von Buchstaben und Texten oder aber auch das automatische Detektieren von menschlichen Gesichtern oder andern Körperteilen in Bildern, beinhaltet.



Abbildung 26: Mustererkennung (Automatic Detection of Human Nudes, 1999)

Das oben abgebildete Beispiel zeigt Ergebnisse eines automatisierten Systems, das nackte Menschen in Bildern erkennt. Dazu werden aus dem Bild erst alle Pixel, die keine hautähnlichen Farbwerte haben, entfernt. Danach wird versucht, innerhalb dieser Pixel Muster zu erkennen, die darauf schließen lassen, dass eine menschliche Pose abgebildet ist.

4.3.2.4. Gestikererkennung

Ein primäres Ziel der Gestikererkennung und ihrer Forschung ist es, Systeme zu erschaffen, die spezifische menschliche Gesten und Mimiken erkennen. Diese Systeme müssen dann verstehen, ob diese Gesten Inhalte vermitteln, um sie so zum Beispiel zur Kontrolle von elektronischen Geräten zu nutzen.



Abbildung 27: Mimikererkennung (forsip, 2005)

Diese CV Disziplin erfordert neben detaillierten Kenntnissen aller anderen CV Bereiche, also dem technischen Know How, auch noch eine schwierige psychologische und biologische Komponente. Hier gilt es nämlich auch zu definieren, was Gesten sind, wie diese zu interpretieren und zu verstehen sind und sie auch noch in Videobildern zu erkennen. (Gestikererkennung Weblink [20])

4.4. Anwendungsbeispiele & Zukunftsperspektiven

Einer der Pioniere der Computer Vision als Methode der Mensch-Computer Interaktion ist sicherlich Myron Krueger.

Myron Krueger (1942 in Gary, Indiana USA, geboren) ist einer der ersten Wissenschaftler und Künstler, der die Interaktion von Mensch und Maschine als Kunstform definierte und damit theoretisch wegweisend für die Medienkunst wurde. Bereits Ende der 60er Jahre arbeitete Krueger an interaktiven (Computer) Kunstwerken. Seine interaktiven Environments wie beispielsweise „Videoplace“ wurden weltweit ausgestellt. Unter dem Titel „*Artificial Reality*“, einem von ihm geprägten Begriff als Term um videobasierte und durch Computer interaktive Medien, wurde 1983 seine Dissertation aus dem Jahr 1974 an der Universität von Wisconsin veröffentlicht.

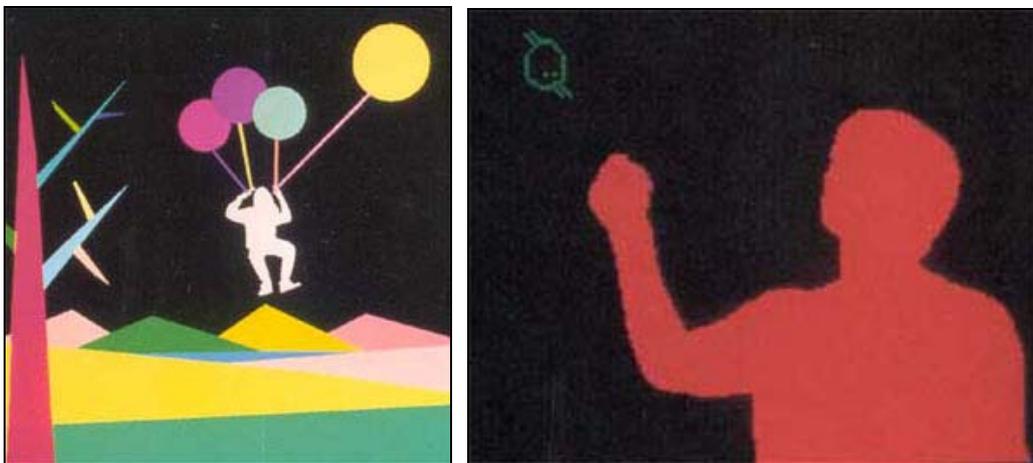


Abbildung 28: Videoplace (Krueger, 1991, S. 110)

Myron Krueger beschäftigte sich schon seit den siebziger Jahren mit berührungslosen Trackingsystemen, die er für seine künstlerischen interaktiven Installationen einsetzte. Er verzichtete auf Ein- und Ausgabegeräte, die einen direkten Körperkontakt voraussetzten, um den Benutzer mehr Freiheit zu gewährleisten. Er konzentrierte sich auf Videowände als Ausgabemedium und Videokameras mit Bilderkennung als Eingabemedien, die den Benutzern ein neuartiges Erlebnis der synthetischen zweidimensionalen Computergrafikwelt ermöglichten.

In der ersten Version von Videoplace, 1975, konnten die Teilnehmer durch die von den Kameras erzeugten Benutzern-Silhouetten miteinander agieren, ohne einen direkten physischen Kontakt zu haben. Später erweiterte Krueger Videoplace um ein Bilderkennungssystem (Computer Vision). Durch die Kombination mit einem Grafiksystem konnte der Benutzer mit computergenerierten Bildern interagieren.

Wie zu Beginn dieses Kapitels erwähnt, gibt es neben Kruegers experimentellen Installationen eine Vielzahl von Anwendungsgebieten, in denen Computer Vision schon heute zum Einsatz kommt.

Um im nächsten und letzten Abschnitt der Arbeit meine eigenen Computer Vision Anwendungen zu erstellen, möchte ich vorab einen kurzen Überblick über aktuelle Forschungsprojekte und Lösungsansätze geben, um argumentieren zu können, in welchen Anwendungsbereichen ich meine Anwendung ansiedeln werde.

Wie einige Beispiele neuartiger Inputtechnologien im Abschnitt 3.8.1 New I/O devices gezeigt haben, sind schon in mehreren spielerischen Anwendungen CV Tools im Einsatz, um das Spielerlebnis zu erhöhen und den Spielspaß auf eine neue Ebene zu heben.

Aber auch rationalere Anwendungsfelder der Computer Vision existieren.

So werden etwa in der Lebensmittelindustrie die Produkte vor dem Verpacken durch Metalldetektoren und Bildanalyser geschickt, um keine Fremdkörper mit zu verpacken. Es gibt aber auch schon komplexere Softwareansätze und Lösungen, wie ich anhand folgender Einzelbeispiele aufzeigen möchte.

Allerdings ist zu erwähnen, dass die hier von mir aufgelisteten Projekte keinerlei Anspruch auf Vollständigkeit erheben, sondern im Gegenteil nur einen Überblick über die unterschiedlichsten Anwendungsfelder, Forschungsansätze und Umsetzungsversuche geben sollen.

4.4.1. Perceptual User Interfaces

Perceptual User Interfaces oder PUIs sind Systeme, bei denen die Bewegung des Users, wie beispielsweise die Bewegung des Gesichts, genutzt wird, um Software oder Programme zu kontrollieren.

Ein Beispiel dafür ist etwa Nouse™, das die Nase des Benutzers sozusagen als Mouse interpretiert (Weblink [21]) und somit die Interaktion zwischen Mensch und Computer ermöglicht, oder aber auch eine vergleichbare Software des Fraunhoferinstituts, das robuste Erkennung von menschlichen Gesichtern in live Videos ermöglicht (Weblink [22]).

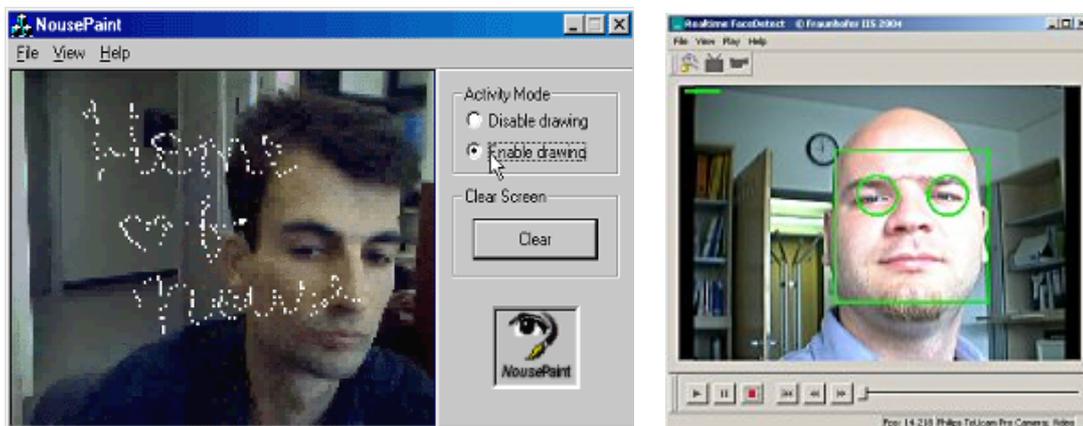


Abbildung 29: Nouse (Perceptual User Interfaces, 2005) & RealTimeFaceDetector (Fraunhofer Institut, 2005)

4.4.2. Blickrichtungsanalyse

Akzeptanz und Attraktivität einer Anzeige werden in der Werbung mittels Blickrichtungsanalyse gemessen. Wie lange ein Zuschauer ein Werbemedium betrachtet, wohin sein Blick schweift, kann beispielshalber durch den Einsatz von Computer Vision zur Verfolgung der Blickrichtung untersucht werden.

Solche oder ähnliche Untersuchungen sind durchaus auch bei Usability-Tests üblich. Das hier beschriebene Beispiel zeigt eine meiner Meinung nach interessantere Anwendungsmöglichkeit dieser Computer Vision Methode, die eher im künstlerischen Bereich der Medientechnik anzusiedeln ist.



Abbildung 30: ART+COM Zerseeher (ART+COM, 2005)

Der „Zerseeher“ von ART+COM (Weblink [23]) ist eine künstlerische Installation, bei der es sich um eine wie ein Bild eingerahmte Rückprojektion auf einer Leinwand handelt, die ein Kunstwerk oder Bild zeigt. Dahinter - für den Betrachter unsichtbar - steht ein Eyetracking-System (Kamera, Computer, Videotracking-Software). Die Kamera nimmt das Auge des Betrachters auf. Der Rechner analysiert dieses Bild in Echtzeit und erkennt das Zentrum der Iris und den Reflektionspunkt eines Infrarotscheinwerfers im Auge. Aus diesen Daten wird genau der Punkt errechnet, auf den der Betrachter blickt. Diese werden dann genutzt, um eine grafische Veränderung des Originals genau an der Stelle herbeizuführen, auf die der Betrachter gerade sieht. Wird das Bild über einen Zeitraum von mehr als 30 Sekunden nicht betrachtet, regeneriert sich das Bild und zeigt den Originalzustand.

So kann nun der Betrachter des Kunstwerks selbst aktiv das Gemälde beeinflussen. Er kann es sozusagen „zersehen“.

4.4.3. Zutritts- & Zugangskontrolle

Der Fingerabdruck, die Netzhautabtastung, aber auch Gesicht, Mimik und Lippenbewegung beim Sprechen eines Passwortes sind Merkmale, die sich hervorragend für die Zutritts-/Zugangskontrolle eignen. Durch biometrische Kontrollsysteme kann die Berechtigung innerhalb eines sicherheitsrelevanten Bereiches für jede einzelne Tür individuell und benutzerfreundlich geregelt werden. Die Ergänzung oder Ersetzung des PINs, Zugangscodes, Schlüssels oder ähnlichen Zugangskontrollen durch biometrische Merkmale, vereitelt den Missbrauch und erhöht die Sicherheit.



Abbildung 31: Iriscan (CNN World, 2005)

Der Mensch wird selbst zum Schlüssel, wie das Beispiel des Schiphol Flughafens in Amsterdam (Weblink [24]) zeigt. Hier werden die gescannten Irisdaten mit denen eines Irispasses verglichen. Ein weiteres Beispiel für CV als Zutrittskontrolle ist die Gesichtserkennung, die es ermöglicht, Personen anhand ihrer Gesichtsstruktur eindeutig in Bildern zu erkennen (Weblink [25]). (Mehr zur Funktionsweise von Iris-scans ist unter Weblink [26] abrufbar)

4.4.4. Industrielle Nutzung

Sogar die heutige hoch technisierte Industrie nutzt Computer Vision, um Abläufe zu automatisieren. So wird beispielsweise eine von Vision Tools - Bildanalyse Systeme GmbH (Weblink [30]) entwickelte Software bei der Fertigung von Metallteilen angewendet, um eine 100%ige Identifikation und somit eine lückenlose Verfolgbarkeit und Historie jedes Produktes zu ermöglichen.

Das System liest und kontrolliert dabei Seriennummern und Codes an verschiedenen Stellen, die in die produzierten Metallteile eingefräst wurden.

Die Metallteile werden dazu an verschiedenen Stellen im Fertigungsprozess von einem Handhabungssystem oder Roboterarm vor eine Kamera gehalten. Bei der Bildaufnahme stehen die Produkte kurz still, um während der Bildauswertung bereits weiter bewegt und verarbeitet werden zu können (Weblink [31]).

4.4.5. Alltägliche Nutzung

Wie schon im Abschnitt 3.8.5 erwähnt, könnte die zukünftige Entwicklung von Hard- und Software dazu führen, dass Computer eine immer wichtigere Rolle in unserem alltäglichen Leben spielen werden.

Dieses Ziel verfolgen auch die Forscher, der Kungl Tekniska Högskolan, in dem sie in einem Forschungsprojekt zum Thema „Gestikbasierende Mensch-Computer Interaktion“ einen Weg fanden, Menschen durch bestimmte Handgesten das Bedienen von herkömmlichen Consumer-Electronics Geräten, wie Fernseher oder Videorekorder zu ermöglichen.



Abbildung 32: homeusage (Computer Vision Based Human-Computer Interaction, 2005)

Die Entwickler sehen den Vorteil der Nutzung von visuellen Inputs darin, dass kein physischer Kontakt zwischen dem Menschen und dem Computer respektive der Maschine entstehen muss, und dass über doch eine beträchtliche Distanz mit dieser Methode kommuniziert werden kann. *[Videobeispiel auf beigelegter CD]*

Außerdem meinen sie, dass Gestik im Vergleich zu Audio oder Sprachinterfaces den klaren Vorteil besitzt, dass sie selbst in lauten Umgebungen eingesetzt werden kann. Beispielshalber könnte man einen Fernseher durchaus durch ein bestimmtes gesprochenes Wort einschalten, allerdings würde das Erkennen des Wortes zum Ausschalten durch das akustische Rauschen des Fernsehers nahezu unmöglich werden (Weblink [27]).

4.4.6. Infotainment

Eine weiteres Einsatzgebiet von Computer Vision stellt möglicherweise der Bereich des Infotainments dar. Infotainment ist ein Kunstwort aus den englischen Worten information und entertainment. Darunter versteht man die unterhaltsame Vermittlung von Bildungsinhalten, aber auch von Scheinwissen, das den Anspruch erhebt, Bildungsbestandteil zu sein. Die technische Bedeutung des Begriffs zeigt die Verknüpfung zwischen dem Vermitteln von Information und Unterhaltung über ein so genanntes Multi-Media-Interface.

Beim Infotainment steht allerdings immer noch der Spaßfaktor im Vordergrund, im Gegensatz dazu wird bei dem Begriff Edutainment Unterhaltung mit Lernen verbunden. Multimediaprogramme werden hierbei zwar zum Spaß und zur Unterhaltung eingesetzt, allerdings wird über diesen Zugang ein beabsichtigter Lernprozess in Gang gesetzt.

Ein meiner Meinung nach sehr gelungenes Beispiel für den Einsatz von Computer Vision im Infotainmentbereich ist das so genannte interaktive Schaufenster der Firma IMS (Weblink [28]).



Abbildung 33: interaktives Schaufenster (IMS Mediensysteme, 2005)

Die genaue Funktionsweise des interaktiven Schaufensters bleibt allerdings ein gut gehütetes Geheimnis. IMC erklärt nur, dass das System innerhalb einer Projektion die Fingerposition des Benutzers erkennen und sofort darauf reagieren kann. Das interaktive Schaufenster lässt somit neue Präsentationsmöglichkeiten für Produkte und Dienstleistungen zu und bleibt auch nach Ladenschluss in Betrieb.

Durch diese interaktive multimediale Projektion in der Auslage eines Geschäfts wird Aufmerksamkeit erzeugt und der sofortige Dialog am POI (point of interest) ermöglicht.

Aber auch Anwendungen mit weitaus verspielteren Hintergründen nutzen Computer Vision, um Unterhaltung am POS (point of sale) zu garantieren.

Ein Beispiel hierfür wäre der Emotional Mirror von studio9 (Weblink [29]). Hierbei wird im Vergleich zum oben genannten interaktiven Schaufenster lediglich versucht, den Kunden am POS zu unterhalten.



Abbildung 34: Emotional Mirror (studio9, 2005)

Der Emotional Mirror zeigt zwar nicht wirklich informative Inhalte, aber er fungiert als Aufmerksamkeitsfänger auf Messen oder in Geschäftsräumen. Laut Hersteller macht er die Kunden selbst neugierig, was wohl das Ziel dieser Anwendung sei.

Beim Aufbau des Emotional Mirrors wird oberhalb eines Bildschirms oder einer Projektionsfläche eine Kamera angebracht, deren Livebilder nach der Verarbeitung durch die von studio9 erstellte Software wieder ausgespielt werden.

Die Interessenten sehen sich also wie in einem Spiegel selbst auf der Videoprojektionsfläche, wobei zusätzlich zu ihrem „Spiegelbild“ ein zeitverzögertes Abbild in mehrfacher Kopie eingespielt wird.

Die Reaktion der Interessenten auf das eigene Abbild wird wieder aufgezeichnet und zeitverzögert ausgespielt. Eine Kettenreaktion setzt ein und die Benutzer gehen verspielt auf ihre zeitverzögerten „Klone“ ein.

Die Interessenten erfinden immer neue Ideen, ihr eigenes (zeitverzögertes) Abbild "auszutricksen", gehen nach und nach aus sich heraus, lassen sich gehen. Darüber vergessen die Besucher die Zeit und haben einfach Spaß [Videobeispiel auf beigelegter CD]. Daher werden von Messe- oder Geschäftspersonal bereitgestellte Broschüren oder Info-Folder bereitwillig angenommen und maximieren so den kundenwirksamen Geschäftsauftritt bei diversen Veranstaltungen.

5. Entwicklung einer Beispielanwendung

5.1. Idee und Begründung

Um das Potential von Computer Vision als HCI Methode besser begreifen und bewerten zu können, habe ich mich in diesem Abschnitt der Arbeit selbst mit Entwicklungsumgebungen und CV Tools beschäftigt und auseinandergesetzt. Dabei beschränke ich mich aus Kostengründen auf möglichst günstige oder sogar kostenlose, also freeware Softwarepakete.

5.2. Ziele der Anwendung

Nachdem ich mir einen Überblick über die in Frage kommenden Tools geschaffen habe, wird zumindest mit einem dieser Programme eine lauffähige Anwendung entwickelt, die den menschlichen Körper als Eingabemedium erfasst und auf seine Bewegungen reagiert.

Das heißt, dass die Anwendung in der Lage sein sollte, die Bewegung des Benutzers zu erfassen, um dann auf unterschiedliche Art und Weise Feedback zu liefern. Es soll also eine klare Reaktion der Software auf die menschliche Bewegung erfolgen.

5.3. Entwicklungsumgebungen

Genauso vielfältig wie die Anwendungsgebiete der Computer Vision, sind auch die Möglichkeiten eine solche Anwendung zu erstellen. Da innerhalb kürzester Zeit die unterschiedlichsten Entwicklungsumgebungen und Tools, mit unterschiedlichster Qualität und Effektivität, entwickelt wurden, ist es meiner Meinung nach unumgänglich, mehrere Entwicklungsumgebungen zu vergleichen und zu testen.

Da es meist mehrere Umsetzungsmöglichkeiten gibt, ist es mir wichtig, einen Kompromiss zwischen Kosten und Nutzen zu erzielen, also die Kosten für Soft- und Hardware so gering wie möglich zu halten und trotzdem dem qualitativen Anspruch an Computer Vision Software standzuhalten.

Wie schon erwähnt, gibt es viele Toolkits, die es ermöglichen, Daten von einer Kamera zu empfangen, zu manipulieren und auch zu interpretieren.

Der Umfang dieser Programme variiert dabei von eher einfach zu bedienenden Anwendungen mit grafischer Benutzeroberfläche, bis zu Entwicklungs- und Programmierumgebungen, in denen sogar einzelne Pixel angesprochen und manipuliert werden können.

Im Rahmen meiner Recherchen zu diesem Thema bin ich auf folgende, nennenswerte Computer Vision Tools gestoßen.

5.3.1. Intel Open CV Library

Diese Funktionsbibliothek ist für den Gebrauch und die Modifizierung durch professionelle Entwickler oder Forscher freigegeben und ermöglicht die Integration in kommerzielle Softwareentwicklungen. Das Hauptaugenmerk der Funktionen dieser Bibliothek liegt auf „realtime“ Computer Vision, also der Echtzeitverarbeitung und Interpretation von Videoinhalten.

Als mögliche Entwicklungsumgebung wird neben C++ unter anderem Matlab genannt und als Beispiele für die Anwendungsfelder werden neben Objektidentifikation und Motion Tracking auch die Human-Computer Interaction erwähnt.

(Open CV Weblink [33])

5.3.2. Video By The Pixel

Bei Video By The Pixel handelt es sich um eine Bibliothek von Java Klassen, die das Auslesen von Videobildern über die Firewire-Schnittstelle und auch rudimentäre Computer Vision und Echtzeit-Videomanipulationsmethoden ermöglichen.

Allerdings ist dieses CV-Tool trotz seiner Komplexität weit nicht so leistungsstark und umfangreich wie, Open CV, dafür aber plattformunabhängig.

(Video By The Pixel Weblink [34])

5.3.3. Isadora

Isadora ist eine grafische Programmierumgebung für Macintosh (derzeit erst als Betaversion für Windows erhältlich), die interaktive Kontrolle über digitale Medien ermöglicht. Das Hauptaugenmerk liegt auf der Echtzeitverarbeitung und Manipulation von digitalen Videosignalen.

(Isadora Weblink [35])

5.3.4. Jitter & Cyclops

Jitter und Cyclops sind Plugins für Max/MSP. Max/MSP wiederum ist eine grafische Entwicklungsumgebung zur Musik-, Audio- und Multimediaverarbeitung. Jitter und Cyclops ermöglichen Max/MSP die Manipulation und Kontrolle von Echtzeit Videosignalen.

(Jitter und Cyclops Weblink [36])

5.3.5. EyesWeb

EyesWeb ist eine kostenlose Softwareplattform zur Entwicklung von Echtzeit-, Musik- und Multimediaanwendungen.

Das intuitive visuelle Programmierinterface erlaubt die einfache und übersichtliche Verwaltung und Verknüpfung von Video und Audio zu einer Gesamtkomposition.

EyesWeb wurde mit dem ausgesprochenen Fokus auf die Analyse von Ausdruck und Bewegung in Video- und Audio-Signalen entwickelt.

(EyesWeb Weblink [37])

5.3.6. Myron

Myron oder auch Webcam-Xtra ist ein Plugin (Xtra) für Macromedia Director, das einfache Computer Vision ermöglicht. Der digitale Videoinput kann hierbei in Echtzeit manipuliert und ausgelesen werden.

Myron ermöglicht Director die Bewegung von Objekten und Menschen wahrzunehmen und zu verarbeiten. Laut den Entwicklern ist es das Ziel von Myron, Computer Vision so einfach und kostengünstig wie möglich zu gestalten.

(Myron Weblink [38])

5.3.7. TrackThemColors

TrackThemColors oder auch TTC ist wohl das bekannteste Echtzeit Video-Xtra für Makromedia Director. Dieses Xtra ermöglicht in zwei unterschiedlichen Ausführungen das Echtzeitmanipulieren von Videosignalen über USB und Firewire. Die Ergebnisse der Manipulationen sind danach direkt über Lingo in Makromedia Director greif- und abrufbar und können somit in jegliche Director-Anwendung einbezogen werden.

(TTC Weblink [39])

5.4. Testen verschiedener Entwicklungsumgebungen

Ich habe mich aus Kostengründen dazu entschlossen, die Multimedia Software EyesWeb und das Director Xtra TrackThemColors näher zu testen.

Die Tests mit EyesWeb wurden durchgeführt, da es eine sehr umfangreiche visuelle Programmierungsumgebung ist, die laut Entwicklern eine Unzahl von Videomanipulationen unterschiedlichster Art und Weise ermöglicht. Außerdem wird EyesWeb von den Entwicklern kostenlos zum Download bereitgestellt.

Die Tests mit TrackThemColors wurden durchgeführt, da ich hier die Möglichkeit hatte in einer, vom Unterricht an der Fachhochschule her vertrauten Entwicklungsumgebung das Potential von Computer Vision besser bewerten zu können.

Beide Programme wurden mit folgendem Versuchsaufbau getestet:

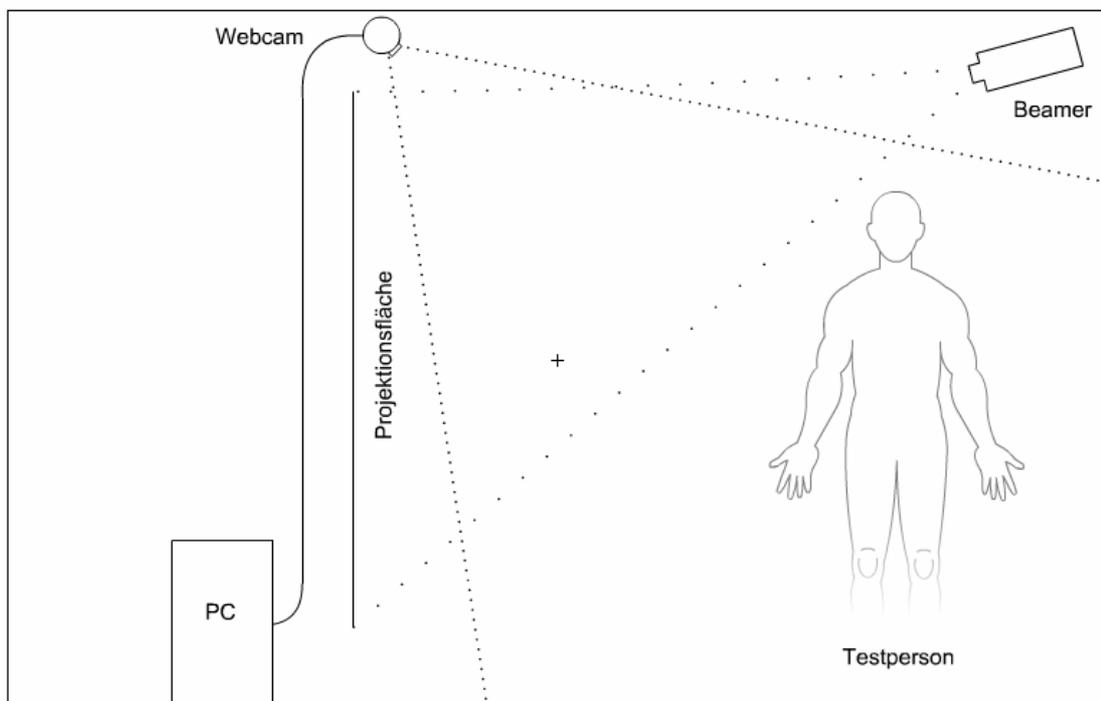


Abbildung 35: Versuchsaufbau

System:

Microsoft Windows XP Professional (Version 2002)

Computer:

AMD Athlon (TM) XP 2800+, 2.08 GHz, 1.00 GB RAM

USB Webcam:

Logitech QuickCam Express

5.4.1. EyesWeb

Die von den Entwicklern gratis zum Download angebotene Software EyesWeb mag auf den ersten Blick unübersichtlich und kompliziert erscheinen.

Allerdings bietet EyesWeb nach eingehender Beschäftigung eine Fülle an Video-, Audio- und Bildmanipulationsmöglichkeiten, sowie die Möglichkeit, unterschiedlichste Ein- und Ausgabearten der Daten zu wählen. Somit ist EyesWeb mehr als geeignet dazu, Echtzeit Multimedia-Anwendungen, welche im Falle von EyesWeb als Patch bezeichnet werden, zu erstellen.

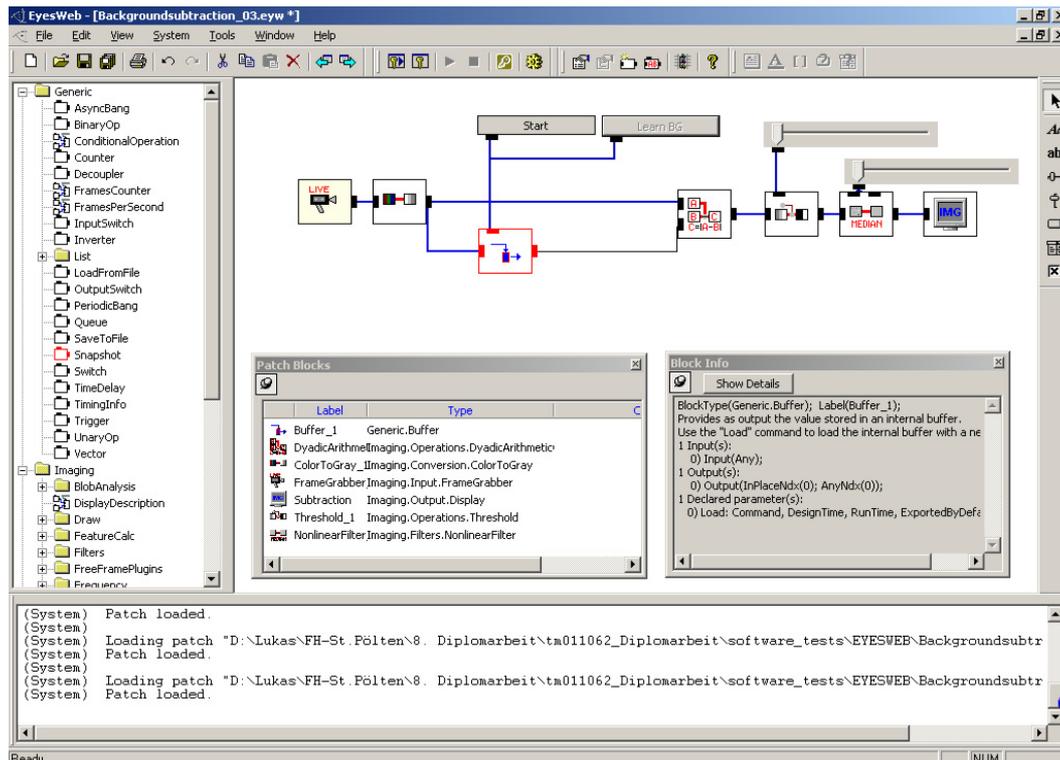


Abbildung 36: Screenshot EyesWeb „Background-Subtraction“ Patch

Der Screenshot in Abbildung 36 zeigt die visuelle Programmierumgebung von EyesWeb. Im linken Bereich befindet sich die Bibliothek, aus der mittels „drag and drop“ die einzelnen Bildverarbeitungs- und Analyseblocks in das Hauptfenster gezogen werden können. Diese Blocks visualisieren die ihnen zugrunde liegenden Funktionen und Klassen, die zumindest einen Ausgang, meist aber mehrere Ein- und Ausgänge zur Verfügung stellen. Optional können auch Manipulationseingänge angelegt werden, um die Funktion des Blocks während der Laufzeit des Patches zu beeinflussen. Beim Erstellen eines Patches werden die Blocks untereinander verbunden. Dabei muss darauf geachtet werden, dass nur bestimmte Datentypen (Ein- und Ausgänge) kompatibel sind.

Ich habe mich dazu entschlossen, in EyesWeb einen vollständigen „Background-Subtraction“ Patch in allen Einzelheiten selbst aufzubauen. Der Prozess der „Background-Subtraction“ lässt sich in folgende Schritte gliedern:

- Graustufenkonvertierung
- Hintergrundsubtraktion
- Schwellwertbestimmung
- Rauschreduktion

Die Entwicklungsumgebung bietet die Möglichkeit, diesen Bildanalyseprozess Schritt für Schritt genau nachzuvollziehen.

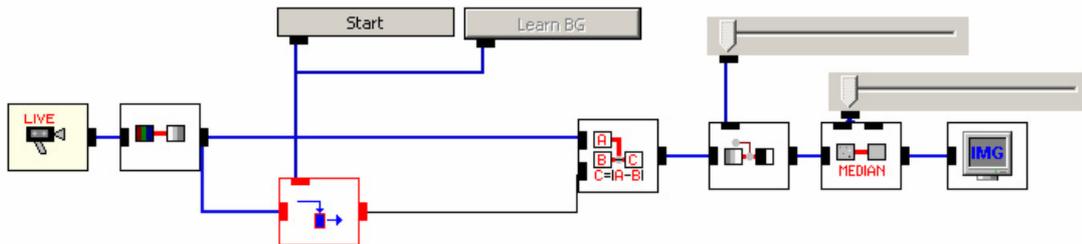


Abbildung 37: „Background-Subtraction“ Patch in EyesWeb

Der oben abgebildete Patch beinhaltet zwei Buttons. Der „Start“ Button löst wie zu erwarten den Start des Patches aus. Ist der Patch gestartet, kann die Funktionsweise der Blocks nicht mehr verändert oder angeglichen werden. Nur Blocks, deren Parameter mit externen Reglern verknüpft sind, können über diese manipuliert werden. Die unterhalb abgebildeten Blocks entsprechen in ihrer Funktionsweise nun genau dem oben angeführten Schema der „Background-Subtraction“.

FrameGrabber



Der „FrameGrabber“ Block empfängt automatisch die Bilder der über USB mit dem PC verbundenen Webcam und kann auch auf deren Einstellungen zugreifen.

ColorToGray



Dieser Block wandelt die vom „FrameGrabber“ empfangenen Bilder in Graustufenbilder um und reduziert dabei die Farbtiefe nach Wunsch auf bis zu ein Bit.

Buffer



Der „Buffer“ kann beliebige Daten auf Befehl zwischenspeichern, daher ist er in meinem Patch mit dem Startbefehl verbunden, um gleich zu Beginn des Patches ein Referenzbild des Hintergrunds zu erstellen. Weiters habe ich diesen Block mit einem weiteren Button verbunden, um während der Laufzeit erneut ein Referenzbild erstellen zu können.

DyadicArithmeticOp



Dieser Block führt die eigentliche Subtraktion durch, indem er die an seinen Inputfeldern angeschlossenen Bilder bitweise subtrahiert. Es wären aber auch andere Operationen wie Addition oder Multiplikation möglich.

Threshold



Der „Threshold“ Block wandelt das Bild in ein Bild mit einem Bit Farbtiefe um, am Ausgang des Blocks wird also tatsächlich ein rein schwarz/weißes Bild geliefert. Dieser Block erzielt über einen Grenzwert nachträglich die Detailliertheit des subtrahierten Bildes.

NonlinearFilter



Der nichtlineare Median Filter ist ein Filter, der zur Rauschreduktion dient, indem er den Mittelwert einer bestimmten Pixelfläche bestimmt und alle diese Pixel auf diesen Wert setzt.

Die folgenden Bilder zeigen den im Buffer-Block gespeicherten Hintergrund, also das Referenzbild, ein repräsentatives Livebild der Webcam, das bitweise subtrahierte Bild nach dem „DyadicArithmeticOp-Block“ und schlussendlich das rauschreduzierte und gefilterte Endergebnis des gesamten „Background-Subtraction“ Patches.

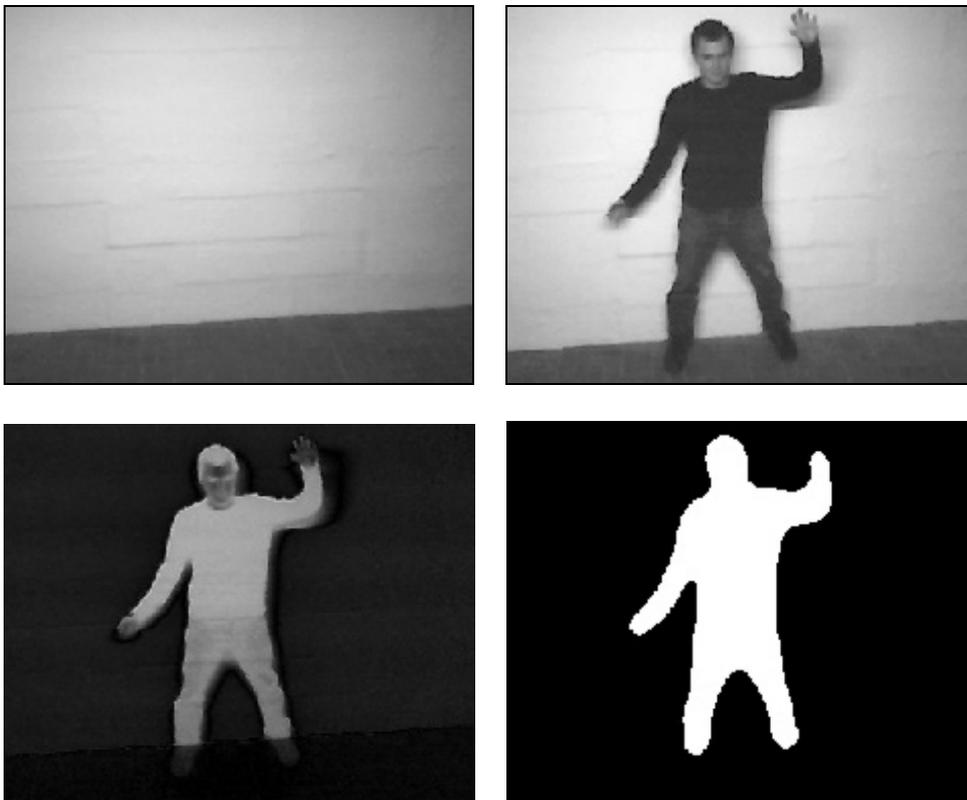


Abbildung 38: „Background-Subtraction“ in EyesWeb

Dieser Einstieg in die Entwicklungsumgebung EyesWeb zeigt schon deutlich, wie vielfältig und umfangreich EyesWeb ist und somit sicher auch andere Computer Vision Tools sind. Dabei lässt der Umfang der in der Software angebotenen Möglichkeiten annähernd die zahllosen Einsatzgebiete und Anwendungsmöglichkeiten erahnen.

Beispielsweise ist es mir gelungen, den „Background-Subtraction“ Patch um ein paar wenige Blocks zu erweitern und somit einen Patch zu erstellen, der ähnliche Effekte erzielt, wie die in Abschnitt 4.4.6 erwähnte Beispielanwendung namens Emotional Mirror.



Abbildung 39: Screenshots des selbst erstellten „Emotional Mirror“

Obige Abbildungen zeigen den spielerischen Umgang eines Benutzers mit seinen zeitverzögerten „Klonen“. Wie zu sehen ist, entstehen immer neue Ideen und Möglichkeiten mit seinem eigenen (zeitverzögerten) Abbild zu spielen.

5.4.2. TrackThemColors

Das Plugin (Xtra) TrackThemColors ermöglicht Macromedia Director unterschiedliche Bildmanipulationen und Bildanalysen auf Livebilder anzuwenden. Da ich mich während der Beschäftigung mit EyesWeb mit dem Thema „Background-Subtraction“ auseinandergesetzt habe, habe ich nun die Möglichkeiten von TTC im Bezug auf „Blob-Tracking“ getestet.

Mit mehr oder weniger komplizierten Funktionsaufrufen aus Lingo, die in der mitgelieferten Hilfe ausführlich beschrieben sind, ist es mir gelungen, die vom Xtra zur Verfügung gestellte Funktion „TrackChangeBlobs“ auf die Livebilder einer Webcam anzuwenden. Die Funktionsweise von „TrackChangeBlobs“ lässt sich überraschenderweise mit oder ohne Referenzbild verwenden und ist über zwei Parameter namens Threshold und Padding genauer einstellbar. Der Parameter Treshold ist für die Umsetzungsgenauigkeit oder auch die Sensibilität auf Bewegung zuständig (je niedriger der Wert, desto sensibler wird auf Bewegung reagiert). Der Parameter Padding ist in erster Line für den Abstand zwischen den Blobs zuständig, er bestimmt aber auch indirekt die Größe der Blobs (sind zwei Blobs weniger weit von einander entfernt als der von Padding bestimmte Wert, werden sie zu einem gemeinsamen Blob zusammen gefasst).



Abbildung 40: Verwendung der Funktion „TrackChangeBlobs“ von TTC

Ich habe eine Directoranwendung erstellt, die die mit der Funktion „TrackChangeBlobs“ ermittelten Daten wie Anzahl, Position und Größe der bewegten Objekte in Arrays verwaltet. Diese sind als globale Variablen (global track_pos, global track_anz) definiert. Der Button „Referenzbild erstellen“ legt ein Hintergrundbild fest, auf das sich „TrackChangeBlobs“ optional beziehen kann. Mit dem Button „Videoeinstellungen“ können die Einstellungen der Webcam geöffnet und geändert werden. Der Button „Projektion öffnen“ ermöglicht es, beliebige Directorfiles der Größe 800x600 px. zu öffnen, an die dann die getrackten Parameter übergeben werden.

Die von mir in diesem Zusammenhang erstellten Beispielanwendungen, sind ähnlich dem Emotional Mirror aus Abschnitt 4.4.6 im Bereich Infotainment oder im Bereich künstlerischer Installationen anzusiedeln.

Die unterstehenden Screenshots zeigen zuerst die von mir als „Pustblumeprojektion“ bezeichnete Installation. Hier werden, sobald sich eine Person der betreffenden Löwenzahnblüte nähert, wie durch einen Windhauch die Samen der jeweiligen Blüte durch die Luft geblasen. Als mögliches Einsatzgebiet für diese Art der Anwendung sehe ich beispielsweise den Warteraum einer Arztpraxis oder die Lobby eines Hotels.



Abbildung 41: Pustblumeprojektion

Eine weitere von mir erstellte Anwendung ist das „interaktive Filmplakat“ (Abbildung 42), auf dem sich der Betrachter nähere Informationen über den Film oder die Schauspieler abrufen kann, indem er einfach vor der entsprechenden Person stehen bleibt. Hier sehe ich als möglichen Einsatzort klarerweise die Foyers von großen Kinos und auch Einkaufsstraßen.



Abbildung 42: Interaktives Filmplakat

Mit der von mir in Director erstellten Anwendung „TrackChangeBlobs“ ist es mir sogar gelungen, die Computer Vision Software des im Abschnitt 1.1 erwähnten AEC Projekts zu imitieren. Das soll heißen, dass es mir gelungen ist, die damals vom Futurelab verwendete Computer Vision Software in meinem Versuchsaufbau durch „TrackChangeBlobs“ zu ersetzen. Es werden die getrackten Positionsdaten nun von „TrackChangeBlobs“ an die damals erstellte interaktive Projektion übergeben. Somit habe ich es geschafft, die Anwendung, die ursprünglich mein Interesse an Computer Vision geweckt hat und somit die Motivation für diese Arbeit darstellt, selbst zu verwirklichen.

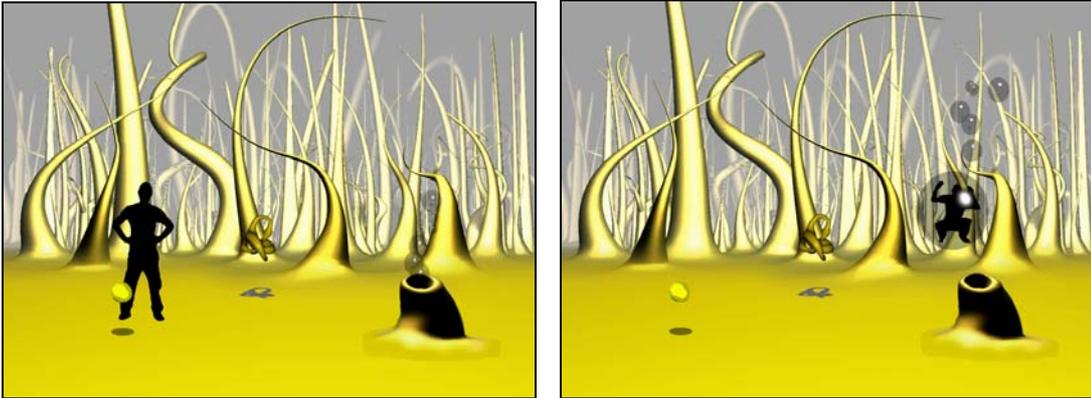


Abbildung 43: AEC Projekt

Abschließend möchte ich noch erwähnen, dass im Vergleich zu TTC und Director EyesWeb zwar intuitiver zu bedienen sein mag, Director bietet aber die Möglichkeit, erstellte Anwendungen als eigenständige Programme ohne zusätzliche Installationen (außer den benötigten Treibern für die Web- bzw. die digitale Videokamera) oder Entwicklungsumgebungen auf anderen Computern zu nutzen. EyesWeb-Patches hingegen benötigen immer die komplette EyesWeb-Umgebung, um lauffähig zu sein. Der eindeutige Vorteil von EyesWeb im Vergleich zu TTC ist, dass EyesWeb dem Benutzer detailliertere und umfangreichere Möglichkeiten bietet Computer Vision zu betreiben.

6. Erkenntnisse der Arbeit

Das Ziel dieser Arbeit war es, folgende Frage zu diskutieren:

In welchen Anwendungsbereichen besitzt eine auf Computer Vision basierende HCI genug Potential, um in Zukunft eine tragende Rolle zu spielen?

Mit Hilfe der durch die intensive Beschäftigung mit den Themen Mensch-Computer Interaktion und Computer Vision gewonnenen Erkenntnisse sollte eine qualitative Aussage getroffen werden.

Nachdem ich derzeitige Anwendungsbereiche aufgezeigt habe, Computer Vision als HCI Methode untersucht und getestet habe, denke ich, dass CV enormes Potential in der Zukunft besitzt. Wobei ich „Potential“ wie folgt verstanden wissen will:

Meiner Meinung nach spiegelt sich das Potential einer Schnittstelle oder Software in ihrem Gebrauch. Wird eine Software von vielen Menschen benutzt, so wird sie sich zukünftig verbessern, da das Feedback der User in weitere Entwicklungen einfließen kann und die Software somit immer besser an die Bedürfnisse der Benutzer angepasst wird. Die Forschung im Bereich der Computer Vision ist sehr vielversprechend und wird sicher Verbesserungen bringen.

Dennoch denke ich, dass eine „tragende Rolle“ der Computer Vision als Medium zur Informationsvermittlung zwischen Mensch und Computer in naher Zukunft unerreichbar bleiben wird. Dies würde sonst bedeuten, dass in alltäglichen Bereichen der Computernutzung eine Verdrängung herkömmlicher HCI Schnittstellen durch CV stattfinden müsste und dazu ist CV noch bei weitem nicht ausgereift genug.

Derzeitige Anwendungen können meiner Meinung nach als sogenannte Insellösungen bezeichnet werden, da sie alle nur in ihrer speziellen Aufbausituationen mit festen Lichtverhältnissen und vordefinierter Umgebung funktionieren. Außerdem sind heutige Computer Vision Systeme, die im Sinne der Mensch-Computer Interaktion genutzt werden, sicherlich noch zu ungenau und fehleranfällig, um sich als gängige Mensch-Computer Schnittstelle durchzusetzen.

Obwohl einige Anwendungsbeispiele im Abschnitt 4.4 gezeigt haben, dass Computer Vision an sich schon sehr robust ist und durchaus Einsatz in der Industrie findet, besteht meiner Ansicht nach im alltäglichen Mensch-Computer Zusammenspiel noch enormer Entwicklungsbedarf für diese Technologie.

Diese Erkenntnis entspricht ganz und gar nicht den Erwartungen, die ich zu Beginn dieser Arbeit hatte. Ich hoffte aufzeigen zu können, dass Computer Vision schon bald in der Lage sein wird derzeitige Schnittstellen zu ersetzen.

Obwohl sich meine Erwartungen nicht erfüllt haben und Computer Vision als HCI Methode trotz jahrzehntelanger Forschung noch nicht an ihrem Ziel angelangt ist, gibt es auch Erfreuliches zu berichten.

Die eingehende Beschäftigung mit Computer Vision Tools und Entwicklungsumgebungen hat mich zu dem Schluss geführt, dass die derzeitigen Softwarelösungen sehr umfangreich sind und qualitativ hochwertige Ergebnisse liefern und das obwohl, oder gerade weil viele der Entwickler im CV Bereich den „open source“ Gedanken vertreten. Denn das bedeutet meiner Meinung nach, dass durch den Austausch von Quellcode und Erfahrungen der Fortschritt beschleunigt wird.

Ich sehe das Potential von Computer Vision eher in künstlerischen und spielerischen Anwendungen, da hier die Spannung und der Reiz des „Neuen“ und „Außergewöhnlichen“ eher zum Tragen kommen, als bei herkömmlichen Anwendungen. Das soll heißen, dass man bei Computerspielen eher bereit ist, auf herkömmliche HCI zu verzichten als etwa in der Arbeitswelt, wie schon einige Spielehersteller mit neuartigen Interfaces bewiesen haben.

Weiters werden in diesem Anwendungsbereich auch Fehler und Ungenauigkeiten der Computer Vision nicht als störend oder behindernd empfunden, was bei Anwendungen in anderen Lebensbereichen sicher der Fall wäre.

Deshalb denke ich, dass sich Computer Vision als Mittel der Mensch-Computer Interaktion in naher Zukunft noch in künstlerischen Installationen, in Anwendungen in öffentlichen Räumen, im Infotainmentsektor oder in spielerischen Anwendungen beweisen wird, bevor sie, wenn überhaupt, zur Vermittlung von Information zwischen Menschen und Computer in herkömmlichen Anwendungen herangezogen wird.

So ernüchternd diese Erkenntnis im Bezug auf meine ursprünglichen Erwartungen auch sein mag, sie bedeutet keineswegs, dass Entwicklungen im Bereich der Computer Vision nicht zielführend und zukunftsweisend wären - oder gar - dass CV nur in künstlerischen Anwendungen Einsatz finden wird.

Im Gegenteil, ein Blick in die Vergangenheit zeigt uns hier, dass zukunftsweisende Technologien schon immer, lange vor ihrer gesellschaftlichen Akzeptanz und Durchsetzung, zuerst in der Kunst Verwendung fanden. Beispielsweise waren es Künstler, die Wege fanden und Regeln erstellten, wie die dreidimensionale Realität auf zweidimensionalem Papier abgebildet werden konnte ohne den räumliche Eindruck zu verlieren. Ein weiteres Beispiel wäre auch die Erfindung des Buchdrucks, den kurz darauf Schriftsteller nutzten, um ihre Werke zu vervielfachen.

Der künstlerische Bereich bediente sich also immer schon der neuesten Technologien seiner Zeit, um möglichst zeitgerecht und ausdruckskräftig zu sein. Für mich ist es also eine positive Erkenntnis, dass Computer Vision schon heute in künstlerischen und spielerischen Anwendungen benutzt und verbreitet wird.

Daher denke ich, dass Computer Vision als zukunftsweisende HCI Methode betrachtet werden sollte, die, obwohl sie derzeit noch in den Kinderschuhen steckt, zukünftig sicher die Kommunikation zwischen Menschen und Computern bereichern wird.

7. Anhänge

7.1. **Abbildungsverzeichnis**

Abbildung 1: Konzeptvisualisierung AEC Projekt	8
Abbildung 2: AEC Projekt Südfassade	9
Abbildung 3: AEC Projekt Projektion Timeshift Festival	9
Abbildung 4: Sender-Empfänger Modell (vgl. Burkart, 2003, S.427).....	14
Abbildung 5: Kommunikationsziele (vgl. Burkart, 2003, S.430)	15
Abbildung 6: Interaktion und Kommunikation	24
Abbildung 7: Interaktions- oder Feedbackprozess	25
Abbildung 8: Einordnung der HCI in die wissenschaftlichen Hauptströmungen	26
Abbildung 9: HCI Kontext und Prozesse (vgl. Andrews, 2005, S. 2)	27
Abbildung 10: Menschliche Informationsverarbeitung (vgl. Carroll, 2003, S.57)	29
Abbildung 11: Überblick HCI Schnittstelle	30
Abbildung 12: Visuelle Wahrnehmung in drei Stufen (Carroll, 2003, S.13).....	31
Abbildung 13: information processing model (human error, 2005)	33
Abbildung 14: Fitts & Hicks Law (Fitts' Law as a Performance Model in Human-Computer Interaction, 2005)	40
Abbildung 15: Textbearbeitungsaufgabe (Carrol, 2001, S..59)	41
Abbildung 16: The Passage (The Wild Divine Project, 2005) & Donkey Kong Jungle Beat (donkeykong, 2005)	48
Abbildung 17: Gametrak Super (Arena, 2005) & Superarena (Atari, 2005)	48
Abbildung 18: Kick Ass Kung-Fu (Animaatiokone Industries, 2005)	49
Abbildung 19: Heliodisplay (IO2Technology, 2005) & OLED (Arbeitsgruppe LED, 2005) ..	49
Abbildung 20: Augmented Reality Brille (New HCI Visions and Future PDAs, 2003)	50
Abbildung 21: Beispiele Virtual Reality (New HCI Visions and Future PDAs, 2003)	51
Abbildung 22: Beispiele Force-Feedback Interfaces (ForceDimensions, 2004)	51
Abbildung 23: Webcambild normal und Medianfilter (10x10)	58
Abbildung 24: Hintergrundbild Livebild Hintergrund vom Livebild subtrahiert.....	59
Abbildung 25: „Background-Subtraction“ mit anschließendem „Blob-Tracking“	60
Abbildung 26: Mustererkennung (Automatic Detection of Human Nudes, 1999)	61
Abbildung 27: Mimikerkennung (forsip, 2005).....	61
Abbildung 28: Videoplace (Krueger, 1991, S. 110)	62
Abbildung 29: Nouse (Perceptual User Interfaces, 2005) & RealTimeFaceDetector (Frauenhofer Institut, 2005)	63
Abbildung 30: ART+COM Zerseher (ART+COM, 2005)	64
Abbildung 31: Irisscan (CNN World, 2005).....	65
Abbildung 32: homeusage (Computer Vision Based Human-Computer Interaction, 2005)	66
Abbildung 33: interaktives Schaufenster (IMS Mediensysteme, 2005)	67
Abbildung 34: Emotional Mirror (studio9, 2005).....	68
Abbildung 35: Versuchsaufbau.....	72
Abbildung 36: Screenshot EyesWeb „Background-Subtraction“ Patch	73
Abbildung 37: „Background-Subtraction“Patch in EyesWeb	74
Abbildung 38: „Background-Subtraction“ in EyesWeb	75
Abbildung 39: Screenshots des selbst erstellten „Emotional Mirror“.....	76
Abbildung 40: Verwendung der Funktion „TrackChangeBlobs“ von TTC.....	77
Abbildung 41: Pustebлумprojektion	78
Abbildung 42: Interaktives Filmplakat	78
Abbildung 43: AEC Projekt	79

7.2. Literatur- und Quellenverzeichnis

7.2.1. Bücher

Keith Andrews „*Human-Computer Interaction / Mensch-Maschine-Kommunikation*“ (Technische Universität, Graz, 2005)

Roland Burkart „*Kommunikationswissenschaft*“ (4.Auflage, Wien, 2003)

J.G. Carbonell, J. Siekmann „*Gesture-Based Kommunikation in Human-Computer Interaction*“ (Genova, 2003)

John M. Carroll „*HCI Models, Theories and Frameworks*“ (San Francisco, 2003)

Sebastian Grange, „*Force Dimensions LLC*“ (Salzburg Research, 2004)

Edith Huber „*Einführung ins wissenschaftliche Arbeiten und Forschungsmethoden*“ (Skriptum der Fachhochschule St. Pölten, 2004)

Pieter Jonker „*An interactive image processing course*“ (Technische Universität, Delft, 2003)

Myron W. Krueger „*Artificial Reality II*“ (1991)

John Lyons „*Die Sprache*“ (München, 1992)

Heinz Pürer „*Grundbegriffe der Kommunikationswissenschaften*“ (Konstanz, 2001)

Cornelia Ruser, „*New HCI Visions and Future PDAs*“ (Universität Zürich, 2003)

Christian Schöndorfer „*Grundlagen von Betriebssystemen*“ (Skriptum der Fachhochschule St. Pölten, 2001)

7.2.2. Online Quellen

Animaatiokone Industries, 2005

Online-Dokument: <http://www.mlab.uiah.fi/animaatiokone/kungfu/en/>

Stand: 10.05.2005

Arbeitsgruppe LED, 2005

Online-Dokument: <http://www.energieland.nrw.de/ag/led/ag-led.htm>

Stand: 10.05.2005

ART+COM, 2005

Online-Dokument:

http://www.artcom.de/index.php?option=com_acprojects&page=6&id=24&Itemid=114&details=0&lang=de

Stand: 10.05.2005

Atari, 2005

Online-Dokument: <http://www.de.atari.com/productpage.php?productID=404>

Stand: 10.05.2005

Automatic Detection of Human Nudes, 1999

Online-Dokumente:

<http://http.cs.berkeley.edu/projects/vision/human/forsyth-ijcv99.pdf>

<http://iris.usc.edu/Vision-Notes/bibliography/applicat825.html>

Stand: 10.05.2005

CNN World, 2005

Online-Dokument:

<http://archives.cnn.com/2002/WORLD/europe/03/27/schiphol.security/>

Stand: 10.05.2005

Computer Vision Based Human-Computer Interaction, 2005

Online-Dokument: <http://www.nada.kth.se/cvap/gvmdi/>

Stand: 10.05.2005

CV Institute, 2005

Online-Dokument:

http://kogs-www.informatik.uni-hamburg.de/~koethe/german_vision.html

Stand: 10.05.2005

Cycling 79, 2005

Online-Dokument: <http://www.cycling74.com/products/jitter.html#graphics>

Stand: 10.05.2005

DIN EN ISO 9241, 2005

Online-Dokumente:

<http://www.kommdesign.de/texte/din.htm>

<http://www2.din.de/>

http://www.kubiss.de/bildung/info/cd_besprechung/iso9241.html

<http://www.lichtung.de/SS2002/Dialoggestaltung.pdf>

Stand: 10.05.2005

direct manipulation, 2005

Online-Dokument:

<http://www.hitl.washington.edu/scivw/EVE/I.D.2.c.DirectManipulation.html>

Stand: 10.05.2005

donkeykong, 2005

Online-Dokument: <http://www.donkeykong.com/final/index.html>

Stand: 10.05.2005

Fachgruppe Arbeits- und Organisationspsychologie, 2005

Online-Dokument: <http://www.aodgps.de/alt/mensch-maschine-systeme.html>

Stand: 10.05.2005

Fitts' Law as a Performance Model in Human-Computer Interaction, 2005

Online-Dokument: <http://www.yorku.ca/mack/phd.html>

Stand: 10.05.2005

Forsip, 2005

Online-Dokument:

http://www.forsip.de/index.php?show=projekte_sipbild&page=2&lang=de#a2

Stand: 10.05.2005

Fraunhofer Institut, 2005

Online-Dokument: http://www.iis.fraunhofer.de/bv/biometrie/download/index_d.html

Stand: 10.05.2005

human error, 2005

Online-Dokument:

<http://www.uscg.mil/hq/g-m/risk/e-guidelines/RBDM/html/vol2/01/v2-01-03.htm>

Stand: 10.05.2005

IMS Mediensysteme, 2005

Online-Dokument: <http://www.ims-mediensysteme.de/pos/schaufenster/index.html>

Stand: 10.05.2005

IO2Technology, 2005

Online-Dokument:

<http://www2.io2technology.com/dojo/204/v.jsp?p=/technology/images>

Stand: 10.05.2005

Online-Enzyklopädie

Online-Dokument: <http://de.wikipedia.org/>

Stand: 10.05.2005

Online-Lexikon

Online-Dokument: <http://www.lexikon-definition.de>

Stand: 10.05.2005

Online-Wörterbuch

Online-Dokument: <http://dict.leo.org/>

Stand: 10.05.2005

Perceptual User Interfaces, 2005

Online-Dokument: http://iit-iti.nrc-cnrc.gc.ca/projects-projets/nouse-img_e.html

Stand: 10.05.2005

Physical Computing, 2005

Online-Dokument: <http://stage.itp.nyu.edu/~tigoe/pcomp/videoTrack.shtml>

Stand: 10.05.2005

studio9, 2005

Online-Dokument:

http://www.studioneun.de/index.htm?starturl=http://www.studioneun.de/produkte/erlebnismaschinen/emotional_mirror/

Stand: 10.05.2005

SuperArena, 2005

Online-Dokument: <http://www.trend-express.com/de/superarena.html>

Stand: 10.05.2005

The Wild Divine Project, 2005

Online-Dokument: <http://www.wilddivine.com/Shop/>

Stand: 10.05.2005

Troika Tronix, 2005

Online-Dokument: <http://www.troikatronix.com/isadora.html>

Stand: 10.05.2005

visiontools, 2005

Online-Dokument:

<http://www.vision-tools.com/ger/Praxis/Anwendungen/Schriftzeichen-Lesesystem/schriftzeichen-lesesystem.html>

Stand: 10.05.2005

7.2.3. Weblinks

Weblink [1]

<http://www.aodgps.de>

Weblink [2]

<http://www.ixdg.org/>

Weblink [3]

<http://www.tele-actor.net/fitts/>

Weblink [4]

<http://www.cs.umd.edu/~ben/>

Weblink [5]

<http://www.cs.umd.edu/hcil/>

Weblink [6]

<http://www.hitl.washington.edu/scivw/EVE/I.D.2.c.DirectManipulation.html>

<http://www.cs.umd.edu/hcil/pubs/books/sparks-of-innovation.shtml>

Weblink [7]

<http://www.acm.org/pubs/citations/journals/tochi/1996-3-4/p320-john/>

Weblink [8]

http://www.tcd.ie/Psychology/Ruth_Byrne/mental_models/

<http://www.si.umich.edu/ICOS/gentleintro.html>

Weblink [9]

<http://www.dcs.gla.ac.uk/~stephen/research.shtml>

Weblink [10]

<http://www.donkeykong.com/final/index.html>

Weblink [11]

<http://www.wilddivine.com/>

Weblink [12]

<http://www.de.atari.com/productpage.php?productID=404>

Weblink [13]

<http://www.trend-express.com/de/superarena.html>

Weblink [14]

<http://www.eyetoy.com/>

Weblink [15]

<http://www.mlab.uiah.fi/animaatiokone/kungfu/en/>

Weblink [16]

<http://www.io2technology.com/>

Weblink [17]

<http://www.oled.at/>

Weblink [18]

<http://www.blui.org/>

Weblink [19]

Computer Vision Links

<http://www.cs.hmc.edu/~fleck/computer-vision-handbook/>

<http://www-cgi.cs.cmu.edu/afs/cs/project/cil/ftp/html/vision.html>

<http://cs-pub.bu.edu/fac/betke/links.html>

<http://www-staff.it.uts.edu.au/~sean/CVCC.dir/about.html#OBJECT%20RECOGNITION>

Computer Vision Testvideo- und Bildsätze:

<http://www.cs.jhu.edu/CIRL/misc/movies.html>

<http://www.cs.bu.edu/groups/ivc/Home.html>

Weblink [20]

<http://www.cee.hw.ac.uk/hipr/html/filtops.html>

<http://www.ph.tn.tudelft.nl/Courses/FIP/noframes/fip.html>

Weblink [21]

http://iit-iti.nrc-cnrc.gc.ca/projects-projets/nouse-img_e.html

Weblink [22]

http://www.iis.fraunhofer.de/bv/biometrie/download/index_d.html

Weblink [23]

http://www.artcom.de/index.php?option=com_acprojects&page=6&id=24&Itemid=114&details=0&lang=de

Weblink [24]

<http://archives.cnn.com/2002/WORLD/europe/03/27/schiphol.security/>

Weblink [25]

<http://vismod.media.mit.edu/vismod/demos/facerec/>

Weblink [26]

<http://www.cybernet.com/~ccohen/gesture.html>

Weblink [27]

<http://www.nada.kth.se/cvap/gvmdi/>

Weblink [28]

<http://www.ims-mediensysteme.de>

Weblink [29]

<http://www.studioneun.de>

Weblink [30]

<http://www.vision-tools.com>

Weblink [31]

<http://www.imn.htwk-leipzig.de/~truebner/mmt/irisscan.html>

Weblink [32]

<http://www.uratek.com/applications.php?1>

http://www.vividgroup.com/products_titles.html

<http://cvlab.epfl.ch/research/augm/augmented.html>

Weblink [33]

<http://www.intel.com/research/mrl/research/opencv/overview.htm>

Weblink [34]

<http://stage.itp.tsoa.nyu.edu/~dano/vbp/>

Weblink [35]

<http://www.troikatronix.com/isadora.html>

Weblink [36]

<http://www.cycling74.com/products/>

Weblink [37]

<http://www.eyesweb.org>

Weblink [38]

<http://webcamxtra.sourceforge.net/>

Weblink [39]

<http://www.smoothware.com/>

7.3. Glossar

<i>cum</i>	lateinisch gemeinsam, zusammen
<i>blob</i>	binary large object
<i>munire</i>	lateinisch binden, bauen
<i>Individuum</i>	lateinisch das Ungeteilte, das Unteilbare. Unter Individuum versteht man ein bewusstes Lebewesen mit Wünschen, Bedürfnissen und Rechten. Das Wort wird allerdings meistens im Zusammenhang mit Menschen gebraucht.
<i>disperses Publikum</i>	Dieser Begriff bezeichnet die Rezipienten von Massenmedien (z.B. Zeitung, Hörfunk, Fernsehen) und geht davon aus, dass diese räumlich voneinander getrennt sind.
<i>TED / Tele-Dialog</i>	(abgekürzt TED) ist ein vom ZDF in Zusammenarbeit mit der Deutschen Bundespost entwickeltes Televoting-Verfahren (unidirektionales Kommunikationsverfahren mit dem Telefon als Rückkanal), welches besonders in Fernsehsendungen für nicht repräsentative Umfragen oder Abstimmungen eingesetzt wurde.
<i>GUI</i>	Grafical User Interface
<i>HCI</i>	Human-Computer Interaction
<i>Deadlock</i>	Angenommen zwei Prozesse benötigen beide Zugriff auf zwei kritische Ressourcen. Wenn nun jeder Prozess eine dieser Ressourcen reserviert hat, warten beide Prozesse bis die jeweils andere Ressource freigegeben wird, um weiter zu laufen. Daher verharren beide Prozesse ohne weiter zu arbeiten. Es kommt also zu einem Stillstand oder Deadlock.
<i>Starvation</i>	Wenn ein Prozess nie Zugriff auf eine von ihm benötigte Ressource erhält und daher untätig verharren muss, wird dieses Stadium als Starvation bezeichnet. Der Prozess „verhungert“.
<i>Mutual Exclusion</i>	Ein Stadium bei dem zwei Prozesse eines Systems eine kritische Ressource (z.B.:Drucker) benötigen. Da immer nur ein Prozess die gewünschte Ressource nutzen kann, besteht die Gefahr von <i>Starvation</i> .