

DISPLAYING INTERPLAY

Entwicklungstrends der Mensch-Maschine-Interaktion

VON ANDREAS KOLB, CHRISTOF REZK-SALAMA,
JOCHEN VENUS

Das gegenwärtige Mediensystem stellt sich den Leuten vor allem als undurchschaubares Gewebe vernetzter Displays dar. Ein großer Teil der wirtschaftlichen, rechtlichen und politischen Kommunikationsroutinen, die vormals von Printprodukten und ihrer Distributionslogik abhingen, werden heute an Displays vollzogen und von Datennetzen vermittelt. Displays konstituieren, so könnte man sagen, die Kontaktzonen zwischen dem Einzelnen und den gesellschaftlichen Strukturen jenseits seiner mikrosozialen Horizonte.

Dieser displayvermittelte Datenaustausch ist mit Blick auf seinen zeitlichen, räumlichen und materiellen Ressourcenbedarf von einer nachgerade revolutionären Effizienz, und dies bei einer außerordentlichen Plastizität seines medienästhetischen Substrats: Displays vermitteln heute Bilder, Klänge und Texte in nahezu beliebigen Kombinationsformen und Darstellungsqualitäten; allein die materielle Oberfläche des Projektionsschirms bzw. des Projektionsgrundes setzt der texturalen Anmutung der angezeigten Daten noch gewisse Grenzen. Die gesellschaftlichen Systeme, deren Kommunikationsstrukturen von diesem Medienumbruch betroffen sind, haben relativ abrupt eine schnellere Taktung und ein neues ästhetisches Gepräge angenommen.

Aus der Omnipräsenz und der gesellschaftlichen Relevanz der Displaytechnologie ergibt sich ein neuer und eminentere Bereich gesellschaftlicher Semantik, ein unscharf begrenztes semiotisches Feld, das sich von der Displaytechnologie selbst und den entsprechenden ingenieurwissenschaftlichen Diskursen über die Images und Einstellungen, die sich aus der Displaynutzung ergeben, bis zu den kulturkritischen Debatten über die ‚telematische Gesellschaft‘ (Flusser) erstreckt.

In der Traditionslinie medienreflexiver Apokalyptik artikuliert sich in diesem Feld die Sorge, dass sich hinter dem Oberflächendesign der Displays die relevanten Vergesellschaftungsstrukturen dem Zugriff des Einzelnen mehr und mehr entziehen. Die Datennetze und die in ihnen ablaufenden Programme zeigen nur noch Eingabemasken vor, die man nach vorgeschriebenen Formen zu bedienen hat, deren Funktion und technologische Basis man aber nicht versteht, geschweige denn zu gestalten in der Lage wäre. Die ungeheure Komplexität gegenwärtiger Informationstechnologie, die selbst von versierten Experten nicht mehr im Ganzen durchschaut werden kann, bildet in der Perspektive dieser Apokalyptik ein blindes Herrschaftssystem aus, dem der Einzelne hilflos ausgeliefert ist.

Genährt werden solche Überzeugungen vor allem von den frustrierenden Erlebnissen im alltäglichen Umgang mit Computersystemen. Unerwartete Ergeb-

nisse, Fehlermeldungen und Systemabstürze brechen über die Nutzer herein wie kleine Naturkatastrophen, und die apokalyptische Vision verschafft vor diesem Hintergrund den kurzfristigen Triumph des negativen Durch- und Überblicks. Vilém Flusser etwa halluzinierte schon in den 1980er Jahren den Advent einer ‚fascistischen Gesellschaft‘ im Zeichen technischer Bilder und traf damit durchaus den Nerv der Zeit:

Die Medien bilden von den Zentren, den Sendern ausgestrahlte Bündel. ‚Bündel‘ heißen lateinisch ‚fascēs‘. Die Struktur der von technischen Bildern beherrschten Gesellschaft ist demnach fascistisch, und zwar ist sie fascistisch nicht aus irgendwelchen ideologischen, sondern aus ‚technischen‘ Gründen. So wie die technischen Bilder gegenwärtig geschaltet sind, führen sie ‚von selbst‘ zu einer fascistischen Gesellschaft.¹

Komplementär zu diesen apokalyptischen Denkfiguren stehen pragmatisch orientierte Gebrauchswertideale, die den frustrierenden Alltagserfahrungen mit positiven Normen begegnen. Sie sollen dafür sorgen, dass sich die Computersysteme und ihre Benutzerschnittstellen in ihrer technischen Evolution mehr und mehr menschlichen Motiven und Fähigkeiten anpassen. Einen institutionellen Rang haben diese pragmatisch orientierten Gebrauchswertideale in der Leitnorm DIN EN ISO 9241 erhalten, die mit Blick auf die Dialoggestaltung der Mensch-Maschine-Interaktion folgende Qualitätskriterien postuliert:

- Aufgabenangemessenheit: Das System soll eine geeignete Funktionalität aufweisen und dabei die überflüssigen Interaktionen minimieren.
- Selbstbeschreibungsfähigkeit: Das System soll durch Hilfen und Rückmeldungen ein Höchstmaß an Verständlichkeit realisieren.
- Steuerbarkeit: Der Dialog soll durch den Benutzer gesteuert werden.
- Erwartungskonformität: Das System soll nach Maßgabe eines realistischen Benutzermodells funktionieren.
- Fehlertoleranz: Benutzungsfehler sollen leicht korrigierbar sein und nicht zum Systemabsturz führen.
- Individualisierbarkeit: Das System soll an besondere Benutzermerkmale und Arbeitskontexte angepasst werden können.
- Lernförderlichkeit: Das System soll leicht zu erlernen sein und den Lernprozess optimal unterstützen.

Sowohl die medienreflexive Apokalyptik als auch die Ideale der Software-Ergonomie dürfen aber nicht mit der Wirklichkeit der technologischen Entwicklung verwechselt werden. Als polemisch bzw. normativ überschießende Reaktionsbildungen weisen sie lediglich auf Problemlagen und abstrakte Zielvorstellungen hin.

1 Flusser, Vilém: *Ins Universum der technischen Bilder*, Göttingen 1985, S. 68.

Diese sind zwar wichtige Determinanten des gesellschaftlichen Status und der Entwicklungstrends der Displaytechnologie. Ein realistischeres Bild gewinnt man aber, wenn man darüber hinaus historische und systematische Erwägungen mit einbezieht, die den ideologischen Motiven der Apokalyptik und der ergonomischen Normierung einen konkreten Gehalt geben. Entscheidend ist dabei eine Weiterung des sachlichen Fokus. Einerseits ist es aufschlussreich, die gegenwärtige Displaytechnologie als Konfiguration im Rahmen einer Geschichte technischer Anzeigen zu perpektivieren, die von primitiven Maßstäben, Sonnenuhren und frühgeschichtlichen Sternwarten über die feinmechanischen Messgeräte der Neuzeit bis zu der modernen elektronischen Messtechnologie reicht. Andererseits ist es sinnvoll, die Displaytechnologie funktional im Rahmen der Mensch-Maschine-Interaktion zu betrachten, denn nach dem Stand der technikgeschichtlichen Entwicklung werden Displays heute insbesondere dazu benötigt, das Zusammenspiel von menschlichem und maschinellem Verhalten zu projizieren und für den menschlichen Anwender kontrollierbar zu machen. Diese technische Funktion erzwingt, dass moderne technische Anzeigen, die das Zusammenspiel von technischen Systemen und ihren Anwendern beobachtbar machen sollen, diese darstellen, *als ob* sie wirkliche Interaktionspartner wären. Die, wenn man so will, *funktionale Fiktion einer Mensch-Maschine-Interaktion* wird somit zum Gegenstand technischer Rationalisierung. Es gilt, die technikhistorisch mit einer gewissen Zwangsläufigkeit emergierende Metapher der Mensch-Maschine-Interaktion technisch möglichst zwingend zu implementieren. Der nächste technik- und medienhistorische Schritt scheint vor diesem Hintergrund der Schritt aus dem *Bildraum des Displays* in den *Aktionsraum der virtuellen Realität* zu sein. Die technologischen Trends jedenfalls sprechen für eine Entrahmung des klassischen Displays und einen Sprung in die dritte Raumdimension.

Im Folgenden sollen kurz einige wichtige Stationen in der Geschichte der Mensch-Maschine-Interaktion rekapituliert werden, um vor diesem Hintergrund die jüngsten Entwicklungen auf diesem Gebiet zu skizzieren und versuchsweise zu evaluieren.

I GESCHICHTE DER MENSCH-MASCHINE-INTERAKTION

Allgemein gesprochen besteht die Leistung technischer Anzeigen vor allem in der distinkten und verlässlichen Darstellung faktischer Zustände und Ereignisse, die sonst gar nicht oder nur schwer und ungenau aufzufassen wären. Die Distinktheit der Darstellung resultiert dabei aus dem systemischen Charakter der Darstellungsoberfläche. Jede besondere Darstellung, die von einer technischen Anzeige vermittelt wird, wird als Ereignis in einem wohldefinierten Ereignisraum präsentiert. Die Verlässlichkeit der Darstellung beruht dagegen auf der naturgesetzlichen Determiniertheit des Darstellungsprozesses. Das Dargestellte steht mit der Darstellung in einer geprüften Kausalbeziehung, so dass die Darstellung als empirischer Nachweis des Dargestellten aufgefasst werden kann. In den Termini der

Peirceschen Semiotik formuliert sind die Darstellungen, die durch technische Anzeigen vermittelt werden, – im Unterschied zur Darstellung mittels *arbiträrer Symbole* oder *abbildender Ikone* – *indexikalisch*, d. h. sie konstituieren einen Zeichenzusammenhang, der zu dem bezeichneten Objekt in einer naturgesetzlichen Beziehung steht.² Die Leistung technischer Anzeigen besteht also in der *automatisierten Produktion skalierter Indexe*. Displays stehen sozusagen am logischen Ende messtechnischer Systeme.

Die Geschichte technischer Anzeigen weist vor dem Hintergrund gegenwärtiger Displaytechnologie zwei entscheidende Wendepunkte auf. Als konstitutive Bestandteile von Messgeräten dienten sie zunächst allein zur differenzierten Beobachtung unabhängiger Naturvorgänge. Technische Anzeigen vermittelten zwischen Mensch und Natur, nicht zwischen Mensch und Technik. Das traditionelle Handwerk kannte keinen expliziten Zeichentausch zwischen Gerät und Nutzer. Die Interaktion zwischen Mensch und Werkzeug war sozusagen implizit und stellte sich als trainierte und geschickte Umgangsweise dar. Die ‚Ergonomie‘ des Werkzeugs bezog sich noch nicht auf allgemeine anthropologische Merkmale, sondern auf einen hinsichtlich Kraft und motorischer Routinen hochgradig spezifizierten Körper. Die ergonomische Raffinesse der über Jahrhunderte fortentwickelten Handwerkszeuge teilte sich nicht dem Lehrling mit, sondern dem durch dauernde Übung am Gerät ausgebildeten Handwerker.

Die Fortschritte im Automaten- und Werkzeugmaschinenbau seit dem ausgehenden 17. Jahrhundert veränderten diese Lage. Sie machten die Geschicklichkeitsvoraussetzungen des Handwerks zwar nicht obsolet, der gesellschaftliche Rang handwerklichen Geschicks ist aber seither beständig gesunken. Bedeutsam war vor allem, dass der technologische Fortschritt den Unterschied zwischen zwei Aspekten deutlich machte, die im Handwerk ungeschieden zusammenfielen: den Unterschied zwischen Kraftentwicklung und performativer Umsetzung der Kraft. Eine moderne Werkzeugmaschine vollzieht ihr technisches Potenzial, ihren ausgeklügelten performativen Mechanismus nur, wenn die entsprechenden Kräfte auf sie einwirken. Zur technischen Herausforderung wurde vor diesem Hintergrund eine standardisierte Produktion und Beherrschung von Kraft. Mit den Kraftmaschinen des Industriezeitalters, zunächst mit James Watts Dampfmaschine und ihrer weiterführenden Rationalisierung in Gestalt der Hochdruckdampfmaschine, fand diese Herausforderung ihre ingeniose Lösung. Die Kraftmaschinen verlangten vom menschlichen Nutzer nicht mehr körperliche Kraft oder feinmotorisches Geschick, sondern, angesichts der Präzisionsanforderung in ihrer Fertigung und entsprechender Aus- und Unfallsrisiken, eine möglichst verlässliche Überwachung des Herstellungsprozesses und des Betriebs. Damit wurde die Beobachtung technischer Prozesse seinerseits zur technologischen Herausforderung, und der Messtechnologie eröffnete sich ein neues Anwendungsfeld. Techni-

2 Vgl. Peirce, Charles S.: „Nomenclature and Divisions of Triadic Relations, as Far as They Are Determined“, in: ders.: *The Essential Peirce. Selected Philosophical Writings*. Jg. 2 (1893-1913), hrsg. v. The Peirce Edition Project, Bloomington 1998, S. 289-299.

sche Anzeigen kamen nunmehr nicht mehr allein bei der technischen Naturbeobachtung zum Einsatz, sondern auch bei der Kontrolle *ereigniskritischer technischer Prozesse*. Die bestimmende technologische Form bestand fortan in der Steuerung von Maschinen anhand von technischen Anzeigen, die skalierte Indexe der respektiven Maschinenzustände übermittelten.

Erst auf dieser paradigmatischen Grundlage, der Maschinensteuerung auf Basis einer quantifizierbaren Datenlage, ließ sich ein technisches Problem denken, dessen Lösung den zweiten Wendepunkt in der Geschichte der Displaytechnologie markiert – die Entwicklung von *Steuerungstechnologie auf der Basis von Datenverarbeitung*. Displays sind also keine späte Folge der Datenverarbeitung, auch wenn die ersten Computer noch nicht über Displays im heute geläufigen Sinne verfügten. Technikhistorisch sind sie ein bedeutsamer Faktor ihres Ermöglichungszusammenhangs. Allerdings schafft die Computertechnologie die Bedingungen für einen logischen Sprung im semiotischen Funktionszusammenhang des Displays, der zur schrittweisen Aufhebung des etablierten Mediendispositivs des Displays führt, zur medialen Überformung und Auflösung der klassischen technischen Anzeige hin zur multimedialen Benutzerschnittstelle und schließlich zur interaktiv anmutenden virtuellen Realität.

Bis zu Beginn der 1960er Jahre war der Begriff Benutzerschnittstelle sehr unscharf. Die Nutzung des Computers beschränkte sich auf die Ausführung langer Berechnungen, weniger auf die visuelle Präsentation der Ergebnisse. Schnittstellen für eine kontinuierliche Intervention des Menschen waren kaum realisiert. Als Ausnahme war der TX-2 am Massachusetts Institute of Technology (MIT) einer der wenigen Computer, die über eine Konsole mit einem Monitor und Tastatur direkt bedient wurden.

Anfang der 1960er Jahre entwickelte Ivan Sutherland auf dem TX-2 sein revolutionäres Computersystem Sketchpad. Dieses System beinhaltete das erste Computerprogramm, das eine graphische Benutzerschnittstelle nutzte. Sutherland bediente das System mit einem Lichtgriffel, indem er direkt auf den Kathodenstrahlmonitor zeichnete, und demonstrierte, dass Computergraphik sowohl für künstlerische als auch für technische Zwecke eingesetzt werden kann.

In den 1960er Jahren griff Douglas Engelbart erneut eine Idee auf, die bereits 1945 von Vannevar Bush in seinem MEMEX-System³ entwickelt wurde: Die Verbindung von inhaltlich verwandten Informationen und Dokumenten über Verknüpfungen (Links). Die Begriffe Hypertext und Hypermedia wurden 1965 von Ted Nelson geprägt.⁴ Engelbart entwickelte in den Stanford Forschungslabo-

3 MEMEX = Memory Expander. Vgl. Bush, Vennvar: „As We May Think“, in: *Atlantic Monthly*, Jg. 176, Nr. 1, 1945, S. 101-108.

4 Nelson, Ted: „A File Structure for the Complex, the Changing, and the Indeterminate“, in: *Proceedings of the 20th National Conference of the Association for Computing Machinery*, 1965, S. 84-100.

ren mit dem NLS⁵ das erste System zur direkten Erstellung von Hypertext-Dokumenten. Dieses System ermöglichte es bereits, zwei Benutzern über Netzwerkverbindungen simultan an einem gemeinsamen Dokument zu arbeiten. Der wenig ergonomische Lichtgriffel wurde ersetzt durch eine Maus, die über den Tisch bewegt wurde. Engelbart führte erste Studien zur Ergonomie von Interaktionsgeräten durch, unter anderem mit einem obskuren, unter dem Tisch montierten Gerät zur Textauswahl mit dem Knie (*knee-device*). Viele der damals erarbeiteten Konzepte finden sich im heutigen Personal Computer wieder, beispielsweise die Unterteilung des Bildschirms in Textfenster. Überlappende Textfenster (Windows) wurden 1969 von Alan Kay entwickelt.⁶

Das erste graphische Computerspiel *SpaceWar* wurde 1962 von Slug Russel am MIT entwickelt und verwendete bereits einen Joystick zur Steuerung. Das erste populäre, kommerzielle Computerspiel war *Pong*, ein einfaches Spiel bei dem ein Lichtpunkt zwischen zwei Benutzern hin und her gespielt wird. Die Steuerung erfolgte über einfache Gamepads. Ende der 60er Jahre wurde die 3D-Rastergraphik populär. Die militärischen Entwicklungen von Flugsimulatoren führten zu den ersten kommerziellen Systemen mit 3D-Echtzeit-Graphikanwendungen. Das erste 3D-Interaktionssystem war der *Lincoln Wand*, ein 3D-Positionserfassungssystem basierend auf Ultraschall, 1966 entwickelt von Larry Roberts.⁷ Eine frühe Anwendung für dieses System war die Modellierung molekularer Strukturen.

II VON DER GRAFISCHEN BENUTZERSCHNITTSTELLE ZUR 3D-INTERAKTION

Parallel zu diesen historischen Entwicklungslinien, in denen sich das Profil einer technischen Implementierung der Metapher von der Mensch-Maschine-Interaktion immer deutlicher herausbildete, entstand auch ein Bewusstsein der besonderen medialen Konfigurationen, die sich im Zusammenspiel technischer Systeme und menschlichen Nutzen ergeben. Die metaphorische Bildquelle der Mensch-Maschine-Interaktion, die Interaktion menschlicher Individuen, und das Zielfeld der Metapher, das produktive Zusammenspiel maschinellen und menschlichen Verhaltens durch dessen mediale Veranschaulichung für den menschlichen Nutzer, können paradigmatisch integriert werden, wenn die semiotischen Kapazitäten von Mensch und Maschine in ihrer Spezifik analytisch begriffen und technisch beherrscht werden.

5 NLS = On Line System. Vgl. English, William K. u.a.: „Display Selection Techniques for Text-Manipulation“, in: *IEEE Transactions on Human Factors in Electronics*, Jg. 8, Nr. 1, 1967, S. 5-15.

6 Vgl. Kay, Alan: *The Reactive Engine*, University of Utah 1969 (Diss.).

7 Vgl. Roberts, Lawrence G.: *Machine Perception of Three-Dimensional Solids*, (MIT Lincoln Laboratory Technical Report, Nr. 315), Cambridge, Mass. 1963 (Diss.).

Im Vergleich zur Kommunikation zwischen menschlichen Individuen ist die Interaktion zwischen Mensch und Computer in ihrer Komplexität stark eingeschränkt. Die Anzahl unterschiedlicher Kommunikationsmedien (z.B. Schrift, Bilder, Sprache, Gestik), die im Zusammenspiel von Mensch und Maschine Verwendung finden, sind deutlich geringer und asymmetrisch auf die ‚Kommunikationspartner‘ verteilt. Dem Mensch stehen zum Senden erheblich weniger Ausdrucksformen zur Verfügung, oft nur Schrift (Tastatur) und deiktische Operationen (Maus). Der Computer kann dagegen mehr und andere Medien für die Ausgabe von Informationen an den Menschen nutzen. Dies ist unter anderem bedingt durch die eingeschränkten Fähigkeiten zur Interpretation und Reaktion des Computers.⁸

Der menschliche Handlungsprozess als Teil der Mensch-Maschine-Interaktion lässt sich für die Zwecke einer technischen Modellierung in drei Teilaspekte untergliedern: die Absicht (Intention), die Planung einer Handlung, sowie die Steuerung (Regulation) der Durchführung. Donald Norman hat in seiner Untersuchung zur Psychologie alltäglicher Dinge ein detailliertes Modell des Handlungsprozesses entwickelt.⁹ Unter der Handlungsregulation versteht man einen kontinuierlichen Prozess, bestehend aus Rückkopplung, Vergleich mit dem Ziel der Handlung und der Modifikation der Handlung. In der Mensch-Maschine-Interaktion erfolgt die Handlungsregulation in der Regel nur einseitig. Der Computer hat nur wenige Möglichkeiten zu prüfen, ob sein Kommunikationsziel vom menschlichen Benutzer verstanden wurde. Er ist daher in seiner Möglichkeit zur Handlungsregulation erheblich eingeschränkt. Die Formen der Handlungsregulation des Menschen auf kognitiver und motorischer Ebene hingegen können kategorisiert werden in

- bewusste, intellektuelle Handlungen (intellektuelle Leistungen des Gehirns basierend auf einer Wissensgrundlage),
- Routinehandlungen (Flexible Handlungsmuster) und
- vollständig automatisierte Handlungen (sensomotorische Bewegungsabläufe).¹⁰

Diese Kategorisierung basiert auf dem Grad der Aufmerksamkeit, die für eine bestimmte Handlung notwendig ist. Bewusste Handlungen erfordern die höchste Aufmerksamkeit. Der Mensch kann daher im Wesentlichen nur eine intellektuelle Handlung gleichzeitig ausführen. Sensomotorische Handlungen, wie beispielsweise das blinde Schreiben auf einer Tastatur, erfordern hingegen nur eine sehr geringe Aufmerksamkeit. Mehrere solcher sensomotorischen Handlungen können auch parallel ablaufen. Das grundlegende Designparadigma für die Mensch-Ma-

8 Vgl. Dahm, Markus: *Grundlagen der Mensch-Computer-Interaktion*, München 2006, Kap. I.

9 Vgl. Norman, Donald A.: *The Psychology of Everyday Things*, New York 2002.

10 Vgl. Hacker, Winfried: *Arbeitspsychologie. Psychische Regulation von Arbeitstätigkeiten*, Bern u.a. 1986.

schine-Schnittstelle fordert daher das Ziel, den Bedarf einer intellektuellen Regulation zu vermeiden. Der Nutzer muss sich auf die übergeordnete Aufgabe, die er mit Hilfe der Maschine zu lösen versucht, konzentrieren. Seine Aufmerksamkeit sollte nicht durch die Benutzerschnittstelle beeinträchtigt werden.

Das Forschungsgebiet der Software-Ergonomie untersucht die Nutzbarkeit und Gebrauchstauglichkeit von Softwaresystemen. Die Effizienz wird hier oft quantifiziert als die durchschnittliche Zeit, die für einen bestimmten Handlungsprozess benötigt wird. Das GOMS-Modell¹¹ von Card et al. zerlegt die Mensch-Maschine-Interaktion in elementare Aktionen und basiert auf einem eher mechanistischen Modell des Menschen.¹² Es erlaubt die Analyse und Quantifizierung der Effizienz anhand heuristischer Regeln und durchschnittlicher Zeiten für elementare Operationen.

Die Mensch-Maschine-Schnittstelle umfasst sowohl reale Interaktionsgeräte (z.B. Tastatur, Maus) als auch virtuelle Interaktionselemente (Elemente einer Graphischen Benutzeroberfläche, wie Knöpfe, Auswahlelemente und Schieberegler). Sowohl die realen als auch die virtuellen Interaktionselemente stellen auf unterschiedlichen Abstraktionsebenen Metaphern für mögliche Operationen dar. Die Analogie von realem und virtuellem Objekt sollte im Idealfall konsistent sein. Eine grundlegende Richtlinie für die Auswahl und Entwicklung von Interaktionsgeräten ist eine intuitive, problemspezifische Nutzung. Als nahe liegendes Beispiel wäre die ideale Steuerung eines Flugsimulators durch Interaktionselemente gegeben, die einer realen Flugzeugkanzel möglichst originalgetreu nachempfunden sind.

Eine mögliche Kategorisierung moderner Interaktionsgeräte kann anhand der mit ihnen assoziierten räumlichen Dimension erfolgen. Sutherlands Sketchpad wird als der Urvater vieler Interaktionsgeräte mit zweidimensionalem Interaktionsraum gesehen. Kamerabasierte Interaktionsgeräte erlauben es, neben den oben beschriebenen Formen der bewussten Regulation, auch die Verarbeitung unbewusster oder unterbewusster Gesten. Eye-Tracking Systeme beispielsweise verfolgen die Augenbewegung eines Betrachters.¹³ Einsatzgebiete findet man neben den kognitiven Wissenschaften, vor allem in der autostereoskopischen Displaytechnik und der Werbung. Die meisten Eye-Tracking Systeme lokalisieren das Zentrum der Pupille in einem Kamerabild und erzeugen eine corneale Reflexion im Infrarotbereich. Auf diese Weise ermitteln sie den (zweidimensionalen) Punkt auf der Bildebene, den der Betrachter anvisiert. Ein kamerabasiertes System zur Spielsteuerung wurde mit dem *Eyeto*y von Sony vorgestellt. Hierbei handelt es sich um ein System zur Erfassung lateraler Bewegungen im Bild. Aufgrund der

11 GOMS = Goals, Operations, Methods, and Selection Rules.

12 Vgl. Card, Stuart K. u.a.: *The Psychology of Human Computer Interaction*, Hillsdale, N.J. 1983.

13 Vgl. Rayner, Keith: „Eye Movements in Reading and Information Processing: 20 Years of Research“, in: *Psychological Bulletin*, Jg. 124, Nr. 3, 1998, S. 372-422.

Einschränkung auf Bewegungen parallel zur Bildebene ist dieses System ebenfalls zu den Geräten mit zweidimensionalem Interaktionsraum zu zählen.

Zu den Interaktionsgeräten mit drei assoziierten Raumdimensionen zählen Spacemouse (bzw. Spaceball), Flystick, Datenhandschuh (Cyberglove) oder Exoskelett (Haptic Workstation).¹⁴ Optional sind einige dieser Geräte auch mit Vibrationselementen zur Umsetzung eines taktilen Feedbacks versehen. Diese tendenziell teuren Geräte werden vorwiegend in virtuellen Umgebungen für Produktdesign und -planung eingesetzt. Ihre Akzeptanz bei den Benutzern ist im Allgemeinen relativ gering, da sie sehr ungewohnt in der Nutzung und wenig komfortabel zu tragen sind. Darüber hinaus muss trotz der vielen Anwendungen von 3D-Interaktionstechniken im Bereich *Virtual Environments* festgehalten werden, dass es nur eine sehr geringe Anzahl übergreifender Interaktionsmetaphern gibt.¹⁵



Abb. 1-3: D-Interaktionsgeräte: Space-Mäuse und Flying-Joystick.

Höhere Akzeptanz bieten kamerabasierte Systeme. Existierende Lösungen basieren vorwiegend auf der Verfolgung (Tracking) natürlicher oder künstlicher Marker. Die Marker werden im Bild erkannt und ihre räumliche Position und Orientierung

14 Vgl. Vince, John A.: *Introduction to Virtual Reality*, London u.a. 2004.

15 Vgl. Mehnert-Spahn, J./Steck, S./Kolb, A.: „A Cross-Platform Approach for User-Interaction in Virtual Environments“, in: *3. Fachtagung zu Virtual Reality und Augmented Reality*, Tagungsband, Fraunhofer IFF, Magdeburg 2006, S. 321-328.

tierung relativ zur Kamera errechnet.¹⁶ Auf optischem Tracking einzelner Punkte basieren auch Motion-Capturing-Systeme, die Gelenkbewegungen des menschlichen Körpers anhand künstlicher Marker erfassen. Diese Systeme verwenden ein kinematisches Modell des Menschen, das den möglichen Bewegungsspielraum vordefiniert und im Vergleich zum natürlichen Bewegungsumfang häufig auch einschränkt.

III BILDBASIERTE 3D-BEWEGUNGSERFASSUNG ALS GRUNDLAGE VON 3D-INTERAKTION

Laser-Triangulierung¹⁷ und Stereo-Technik¹⁸ sind Standard-Techniken zur Erfassung von Entfernungen. Während die Laser-Triangulierung auf die Erfassung statischer Szenen beschränkt ist, kann die Stereo-Technik nur bei ausreichender Objektstruktur eingesetzt werden und ist nur eingeschränkt echtzeitfähig. Alternative Techniken wie strukturiertes Licht¹⁹ oder *Parallel Range Sensing*²⁰ eignen sich zur Erfassung dynamischer Objekte und Szenen. Der Ansatz mit strukturiertem Licht erfordert eine aufwändige Offline-Berechnung der Entfernungswerte. Beim *Parallel Range Sensing* kommt eine zeitlich hochgenaue, hochdynamische Verschlusstechnik (*gating*) zum Einsatz, die eine kompakte und kostengünstige Bauweise verhindert.

Die *Photonic Mixer Detector* (PMD)-Technologie arbeitet mit Amplituden-Modulation im Nah-Infrarot (NIR) Bereich. Die Lichtlaufzeit wird über das direkte Sampling der Korrelationsfunktion zwischen dem Referenzsignal und dem optischen Signal aus der Szene auf dem Pixel realisiert.²¹ Dies erlaubt eine sehr kom-

16 Vgl. Koch, Reinhard u.a.: „Markerless Image-based 3D Tracking for Real-time Augmented Reality Applications“, Vortrag, *Workshop on Image Analysis for Multimedia Interactive Services*, Montreux 2005, <http://www.ist-matris.org/publications/WIAMIS-AR.pdf>, 08.08.2006.

17 Vgl. Winter, Gundolf/Bogen, Manfred: „Teilprojekt B7: Virtualisierung von Skulptur“, in: Finanzierungsantrag, *Kulturwissenschaftliches Forschungskolleg SFB/FK615 Medienumbrüche*, Universität Siegen 2005, S. 489-522; Levoy, Marc u.a.: „The Digital Michelangelo Project: 3D Scanning of Large Statues“, in: *ACM Proceedings SIGGRAPH*, Juli 2000, S. 131-144.

18 Vgl. Ma, Yi u.a.: *An Invitation to 3D Vision. From Images to Geometric Models*, New York u.a. 2004.

19 Vgl. ABW GmbH (Hrsg.): „Optische 3D-Formerfassung mit Codiertem Lichtansatz plus Phasenmessung“, http://www.abw-3d.de/messverfahren/messverfahren_de.php, 08.08.2006.

20 Vgl. 3DV Systems (Hrsg.): „ZCam“, <http://www.3dvsystems.com/products/zcam.html>, 08.08.2006.

21 Vgl. Lange, Robert: *3D Time-of-flight Distance Measurement with Custom Solid-state Image Sensors in CMOS/CCD-Technology*, Universität Siegen, 2000 (Diss.).

pakte Bauweise auf Basis einer Standard-CMOS-Technologie.²² Die herausragenden Eigenschaften der PMD-Technologie, konkret

- die zeitgleiche Erfassung von nahezu 20.000 Entfernungswerten,
- die hohe zeitliche Auflösung der Daten-Akquisition von ca. 20 FPS und
- und kompakte und kostengünstige Bauweise des Sensors

stellen ein herausragendes Potential zu Umsetzung neuer, innovativer Interaktionstechniken dar. Der PMD-Sensor wurde bereits für eine prototypische Gestenerkennung für eine Hand eingesetzt.²³ Hierbei wurde ein modellbasierter Ansatz verfolgt, bei dem die Tiefendaten der PMD-Kamera in Echtzeit auf ein, bislang starres Handmodell abgebildet werden. Hierbei wird von der so genannten Hauptachsen-Analyse (englisch: *principle component analysis*, PCA) Gebrauch gemacht, mit deren Hilfe die wesentliche Ausrichtung der virtuellen Hand an die segmentierten PMD-Daten angepasst wird.

IV ANWENDUNGSPOTENTIALE BILDBASIERTER 3D-INTERAKTION

Für die Untersuchung möglicher Anwendungsszenarien für die Mensch-Maschine Interaktion durch Gestiken kann von einer initialen Unterteilung mit einer groben Charakterisierung der Anwendungsgebiete ausgegangen werden. Bezüglich des Bedarfs und des Nutzens gestischer Interaktion können Softwareapplikationen in unterschiedliche Klassen eingeteilt werden:

- *Abstrakte Anwendungsprogramme ohne direkten Raumbezug für Interaktion und Darstellung*
In diese Kategorie fallen alle Anwendungen, die eine abstrakte Steuerung basierend auf Metaphern (Buttons, Slider etc.) verwenden. Beispiele sind beliebige Datenbank-, Textverarbeitungs- oder Präsentationsanwendungen. Ein Szenario aus dieser Kategorie wäre sehr experimentell und würde dem Entwurf eines neuen, intuitiven Interaktionsparadigmas ein sehr großes Gewicht verleihen. Fiktive Steuerungsmöglichkeiten, wie beispielsweise die Datenbanknavigation im Film *MINORITY REPORT* (Twentieth Century Fox und Dreamworks, 2002), zeigen jedoch durchaus, dass gestische Interaktion für derartige Anwendungen denkbar ist.
- *Unterhaltungsmedien, wie Computerspiele oder digitale Kunst*
In diesem Bereich wurde 3D-Interaktion bereits eingesetzt, ist aber kein zwingender Bestandteil der bekannten Interaktions-Paradigmen. Insbeson-

22 Vgl. Kraft, Holger u.a.: „3D-Camera of High 3D-Frame Rate, Depth-Resolution and Background Light Elimination Based on Improved PMD (Photonic Mixer Device)-Technologies“, Vortrag, *OPTO*, Nürnberg 2004; Schwarte, Rudolf u.a.: „Neue 3D-Bildsensoren für das Technische 3D-Sehen“, Vortrag, *VDE Kongress ‚Ambient Intelligence‘*, Berlin 2004.

23 Vgl. Breuer, Pia: *Entwicklung einer prototypischen Gestenerkennung unter Verwendung einer IR-Tiefenkamera*, Universität Koblenz-Landau 2006 (Diplomarbeit).

dere im Bereich der Computerspiele werden intuitivere und ergonomischere Alternativen zu Maus und Tastatur, etwa Gamepad oder Lenkrad, verwendet. 3D-Interaktionsparadigmen, wie sie in diesem Projekt erforscht werden, könnten die Steuerung eines Spiels mit direktem Raumbezug (z.B. ‚Weg-ducken‘ eines virtuellen Charakters) aber auch mit Indirektion (Beispiel Hand- oder Kopfgestik) umsetzen.

- *Virtual Environments*

In diesem Kontext ist die 3D-Interaktion aktuell am weitesten verbreitet. Anwendungen wie virtuelle Produktpräsentationen oder Produktintegrationsaufgaben finden sich in vielen Investitionsgüterindustrien. Die Umsetzung eines Szenarios in diesem Bereich würde die Möglichkeit eines unmittelbaren Vergleichs mit bekannten, nicht kamerabasierten Interaktionsgeräten eröffnen.

In einer Reihe von Anwendungen mit 3D-Interaktion spielt die Haptik zur Exploration von Gegenständen, beispielsweise von der Oberflächenbeschaffenheit, eine wesentliche Rolle. Naturgemäß ist es bei einem bildbasierten Ansatz, der ohne zusätzliche Geräte auskommen soll, nicht möglich, diese Form der direkten Immersion zu integrieren. Diese Einschränkung kann durch Abbildung der zu vermittelnden Eigenschaften auf andere, im Wesentlichen visuelle Merkmale zu einem gewissen Teil kompensiert werden kann.

Andererseits eröffnet dieser scheinbare Mangel an Immersion bei der taktilen Komponente, Möglichkeiten für neue und innovative Interaktionsformen insbesondere bei Anwendungen mit direktem Raumbezug. Der ‚schwerelose‘ Umgang mit virtuellen Objekten mit unmittelbarem Raumbezug, beispielsweise Automobile oder Häuser, hat das Potential neue Interaktionsformen hervor zu bringen.

V RAUMDISPLAYS UND DIE EXPERIMENTELLE MORPHOLOGIE GESTISCHER KOMMUNIKATION

Die Hoffnungen und Ängste, dass eine Ausweitung des Displays in den Raum zu einer unterschiedslosen Verschmelzung ‚natürlicher‘ und virtueller Interaktionsformen führen könnte, erscheinen vor diesem Hintergrund unbegründet. Die Verräumlichung der Mensch-Maschine-Interaktion eröffnet vielmehr das Feld einer experimentellen Morphologie gestischer Kommunikation, deren sozialpsychologische Implikationen vorerst unabsehbar sind. Denkbar wäre einerseits, dass es im Zuge der massenattraktiven Implementierung einer gestisch basierten Mensch-Maschine-Interaktion zu einer kulturellen Aufwertung und Raffinierung der Gestik kommt; denn für den Prozess der elektronischen Datenverarbeitung müssten die Gebärden in distinkte Einheiten zerlegt werden, die z. T. bedeutungstragend, z. T. aber auch nur als bedeutungsdiskriminierend anzusehen sind (und daher ähnlich wie die Lautsprache auf der Basis einer begrenzten Ausdrucks-

substanz einen unendlichen Formenreichtum zu codieren in der Lage sind²⁴). Schon heute funktionieren die Controller von Computerspielen als semiotisch höchst komplexe Kommunikationsgeräte, die teils gestisch analog, teils deiktisch, teils symbolisch funktionieren und diese Formen darüber hinaus in vielfältigen z. T. kontextabhängigen Kombinationen einsetzen.

Auf den ersten Blick ist allerdings nicht zu sehen, welcher praktische Vorteil in einer gestisch basierten Mensch-Maschine-Interaktion stecken soll. Diese Frage mag jedes Forschungs- und Entwicklungsprojekt auf seine Weise beantworten; eine allgemeine Antwort ist nicht vonnöten. Die computervermittelte Steuerungstechnologie wirft grundsätzlich das technische Problem einer Steuerung 2. Ordnung auf. Das Display einer solchen Steuerungstechnologie muss den Steuerungsprozess und die Möglichkeiten seiner Steuerung durch den Nutzer darstellen. Die technische Herausforderung der Darstellung der Mensch-Maschine-Interaktion ist mit der computervermittelten Steuerungstechnologie gesetzt. Dass auch die gestischen Vermögen des Menschen Gegenstand der Implementierung werden, braucht dann keine weitere Rechtfertigung, als dass dies im Horizont der Problemstellung liegt. Alles Weitere ist durch die moderne Logik technischer Entwicklung, die schon in den 1950er Jahren von Arnold Gehlen beschrieben worden ist, hinreichend motiviert:

[...] [E]s geht um die Ausschöpfbarkeit bestimmter Methoden, und man gewahrt immer deutlicher eine Art Achsendrehung der Fragestellung. Es handelt sich immer weniger darum, für schon definierbare Zwecke die technischen Mittel der Herstellung, für vorgegebene Gegenstandsgebiete die besten Erkenntnismethoden zu finden [...]: die Darstellungsmittel, Denkmittel, Verfahrensarten selbst zu variieren, durchzuprobieren, bis zur Erschöpfung aller Möglichkeiten ins Spiel zu bringen und zu sehen, was dabei herauskommt. Auch in der Technik geht man natürlich, wie früher, oft noch von Zwecken aus und sucht Mittel dafür. [...] Daneben aber und anscheinend zunehmend wichtig gibt es die umgekehrte Art der Problemstellung, nämlich die Frage, was sich Unvorhergesehenes aus einer gegebenen Verfahrensart herausholen lässt.²⁵

Die gesellschaftliche Semantik, von der eingangs die Rede war und die sich um die Entwicklungen der Displaytechnologie rankt, scheint sich unter dem Eindruck dieser Entwicklungen schon umzustrukturieren. Zwar hat man heute immer noch vor allem die an den Bildschirm gebundene, im doppelten Sinne flachen Displays

24 Vgl. Lyons, John: *Einführung in die moderne Linguistik*, München ⁶1984, insbes. S. 56: 2.1.3 Die ‚Doppelgliederung‘ der Sprache.

25 Gehlen, Arnold: *Die Seele im technischen Zeitalter. Sozialpsychologische Probleme in der industriellen Gesellschaft*, Reinbek 1957, S. 28f.

of Interplay im Blick, und dies prägt ganz wesentlich den Vorstellungsgehalt von der Mensch-Maschinen-Interaktion:

Im Englischen wird schon das Betätigen eines Videorecorders als interaction bezeichnet, dabei sollte man in solchen Fällen lieber vom Beginn mehrstufiger Online-Transaktionen sprechen. Diese wachsen mittlerweile explosionsartig an und entwickeln im Ensemble alltäglicher Austauschprozesse eine ‚kritische Masse‘. Das Shopping im Netz, die Teilnahme an Auktionen und das Home-Banking, auch virtueller Behördenverkehr sind heute für viele Routine geworden. [...] [I]n der Netzkommunikation [tendiert, d. Verf.] die Hierarchie zwischen Sender und Empfänger (respektive Produzenten und Rezipienten von Information) [...] zu einer symmetrischen Wechselwirkung. [...] Pseudo-interaktive Anwendungen spiegeln solcherlei nur vor: Das meiste, was als ‚interaktives Fernsehen‘ angepriesen wird oder unter dem Banner ‚interaktiver Kunst‘ läuft, reduziert sich auf die begrenzte Auswahl aus einem starren Menü, erst in zweiter Linie sind minimale Eingriffe in das Medienangebot möglich [...] ‚Echte‘ Interaktivität verlangt mehr: Sie erfordert Einfluss auf Inhalt und Form, [...] die aktive De- und Reprogrammierung des ‚Programms‘ sowie die offene und autonome Mitgestaltung der Netzwerkarchitektur.²⁶

„Echte“ Interaktivität ist nach den Vorstellungen Leggewies und Biebers zunächst einmal allein die Interaktivität aufgeklärter Subjekte, die durch „De- und Reprogrammierung des ‚Programms‘“ (was immer man sich darunter vorzustellen hat) ihre Autonomie verwirklichen. Alles andere sei Pseudo-Interaktivität (ohne Anführungsstriche). Dass Leggewie und Bieber aber nicht ohne Anführungsstriche von ‚echter Interaktivität‘ sprechen wollen und einige Zeilen weiter mit Jaron Lanier unterstellen, „dass das ‚Gegenüber‘ auch die Maschine sein kann“ (ebd.), belegt, dass ihre Vorstellung medialer Interaktivität, die an die Brechtsche Utopie des Kommunikationsapparats anschließt, von ihnen selbst nicht mehr so ganz geglaubt wird und in der gesellschaftlichen Semantik offenbar nicht mehr mit der alten Selbstverständlichkeit verankert ist.

Vergleicht man mit diesen Einschätzungen die Reaktionen auf die neue Computerspielwelt, die mit dem *Eyeto*y möglich geworden ist, einem kamerabasierten Interface, das gestisch vermittelte Informationen verarbeiten kann, zeigt sich deutlich die Bereitschaft, den zur Fiktion der Mensch-Maschine-Interaktion notwendige *suspension of disbelief* zu vollziehen, sobald die eigene Performanz stärker ins Spiel kommt als in den Prozessen gängiger Benutzerschnittstellen. Wenn der gestische Umgang mit Computersystemen alltäglich und breitenwirksam

26 Leggewie, Claus/Bieber, Christoph: „Interaktivität – Soziale Emergenzen im Cyberspace?“, in: dies. (Hrsg.): *Interaktivität. Ein transdisziplinärer Schlüsselbegriff*, Frankfurt a.M./New York 2004, S. 7-15, hier S. 8f.

etabliert ist, so steht zu vermuten, wird die gesellschaftliche Semantik des Displays nicht mehr von Vorstellungen der ‚Ebene‘, der ‚Schnittstelle‘, des ‚Interface‘ reguliert sein, die begriffslogisch mit Vorstellungen der ‚Vermittlung‘, ‚Kommunikation‘ und ‚Rationalisierung‘ harmonieren, sondern sich mehr und mehr den Evidenzen der ‚Szene‘, des ‚Portals‘ und des ‚Environments‘ anpassen, die eher mit Kategorien wie ‚Erleben‘, ‚Möglichkeitsraum‘ und ‚bereichsspezifischer Nutzen‘ zusammen passen.

Dann wird auch die Kritik am Konzept medialer Interaktivität und interaktiver Benutzerschnittstellen, die heute noch überzeugen kann, ihre epistemologische Problematik nach außen kehren. Zwar ist die Kritik, die so genannten interaktiven Medien und Benutzerschnittstellen seien gar nicht interaktiv, insofern richtig, als weder in den Medien noch durch irgendwelche Mensch-Maschine-Schnittstellen ein freier Austausch von Motiven stattfindet, es wird lediglich das durch Hard- und Software definierte Ereignisspektrum verwirklicht; andererseits läuft die Kritik ins Leere, da das Paradigma des freien Austauschs von Motiven, das Miteinander vernunftbegabter Menschen, längst seiner sachlogisch unzutreffenden Prämissen überführt worden ist: sei es soziologisch durch den Nachweis sozioökonomischer Determinanten bürgerlicher Subjektivität, sei es psychologisch durch die Aufdeckung latenter Motive, die das bewusst zielgerichtete Handeln überlagern und funktionalisieren, sei es biologisch durch den empirischen Beleg, dass es für den freien Willen keine neurologische Grundlage gibt.

Jenseits der Displays von Interaktion gibt es also streng genommen gar keine ‚echte‘ Interaktion, es gibt nur technisch mehr oder weniger gerahmte Inszenierungsoberflächen der funktionalen Hypothese, es mit einem Hin und Her bewusst motivierten Verhaltens zu tun zu haben.

Vertritt man eine in diesem Sinne pragmatische Sicht der Dinge und geht davon aus, dass Handlungsmotive und ihre Träger nicht als ontologische Primärgegebenheiten, sondern als hypothetische, rein ‚rechnerische‘ Größen anzusehen sind, die in Wirklichkeitsbeschreibungen angesichts einer sonst nicht mehr zu bewältigenden Komplexität von Verhaltensdeterminanten verwendet werden, gelangt man zu einer gleichsam Turingschen Sicht: Freiheit, Intelligenz, Subjektivität, ja selbst die sublimsten Regungen des Gefühlslebens werden zu Zuschreibungsformen, die grundsätzlich auch Maschinen attribuiert werden können, wenn es sich pragmatisch aufdrängt. Ob und in welchen Einschränkungen Mensch-Maschine-Interaktionen solche Zuschreibungsformen provozieren, ist nicht mit Sachargumenten zu entscheiden, sondern kann getrost dem technischen Fortschritt der Displaytechnologie überlassen bleiben.