

Der Weg zur humanen Schnittstelle

Matthias Rauterberg*

Kommandosprachen, Maus und die Wahl des Menüs sind die – unzulänglichen – Mittel, über die wir mit unserem Computer kommunizieren. Zur Behebung dieser Unzulänglichkeiten stehen zwei Möglichkeiten zur Verfügung: die virtuelle Realität (virtual reality) und die angereicherte Realität (augmented reality). Der unauffällige Wandel des Computers zum ubiquitären Begleiter hat sich gewaltig auf den Bereich der Mensch-Computer-Interaktion ausgewirkt. Maus und graphisches Display sind allgegenwärtig, der Rechner selber bildet die Grenze zwischen der elektronischen und der realen Welt. Für den Informationstransfer zwischen diesen Welten wenden wir eine Menge Zeit und Energie auf. Eine bessere Integration der virtuellen Welt der Rechner mit der Realwelt ihrer Benutzer könnte hier wesentliche Einsparungen bringen. Der wohl vielversprechendste Weg zu dieser Integration heisst Augmented Reality (AR, mit Computerintelligenz angereicherte Realität). Der Erfolg, den man sich von dieser Methode verspricht, baut auf auf einer grundlegenden menschlichen Fähigkeit auf: die Interaktion mit physisch realen Dingen und Objekten. In diesem Artikel soll eine neue Technik der Interaktion vorgestellt werden: die Natural User Interface NUI (natürliche Benutzungsschnittstelle). Mehrere prototypische Applikationen in verschiedenen Forschungslabors weltweit arbeiten bereits mit NUI, bzw. NUI-Komponenten und führen uns die technische Machbarkeit vor.

Der Nutzen von NUI

Heute sind bereits mehrere Methoden des Mensch-Maschine-Dialogs im Einsatz: Kommandosprache, Funktionstaste, Menüauswahl, Ikone und Fenster. Diese Techniken lassen sich in drei klassische Interaktionsarten bündeln:

• Kommandosprache:

Diese Interaktion (einschliesslich Tastenkombinationen und Softkeys) ist einer der ältesten Arten der Kommunikation mit dem Computer.

Vorteile: Im Kommandomodus hat der Benutzer den grösstmöglichen direkten Zugang zu allen Funktionen und Abläufen.

Nachteile: Der Benutzer hat keine ständige Information über alle vorhandenen Funktionspunkte.

• Menüauswahl:

Sie umfasst starre Menüstrukturen wie Pop-up- und Pull-down-Menüs sowie auszufüllende Formulare. Diese Art der Interaktion wurde erst mit ASCII-Terminals technisch machbar. Mit ihr werden oft zusätzlich noch Funktionstasten für die Dialogführung verwendet.

Vorteile: Alle Funktionen sind durch sichtbare Interaktionspunkte auf dem Bildschirm gekennzeichnet.

Nachteile: Das Aufspüren eines Funktionspunktes in einem tiefer vergrabenen Menü ist äusserst mühsam.

• Direkte Manipulation:

Diese Art der Interaktion konnte erst mit Bitmap-Graphikbildschirmen realisiert werden. Ihre Entwicklung baut auf der Schreibtisch-Metapher auf. Sie setzt voraus, dass eine

möglichst realistische Beschreibung der Arbeitsumgebung (Pult, Ordner, Papierkorb) in der Ein/Ausgabe-Schnittstelle es dem Benutzer erleichtert, sich an die virtuelle Welt elektronischer Objekte zu gewöhnen.

Vorteile: Alle Funktionen werden fortwährend durch sichtbare Interaktionspunkte wiedergegeben (Maus-sensitive Bereiche, etc.) Die Aktivierung gewünschter Funktionen wird durch direkte Manipulation ihrer Abbilder erzielt.

Nachteile: Direkt manipulierbare Schnittstellen haben Mühe mit Variablen oder der Unterscheidung der Darstellung eines individuellen Elementes von der Wiedergabe eines Satzes oder einer Klasse von Elementen.

In all diesen traditionellen Interaktionsarten kann der Benutzer keine realen Objekte in ein- und demselben Schnittstellenraum mit virtuellen Objekten mischen. Sie berücksichtigen auch nicht die zahlreichen Möglichkeiten, mit denen die menschliche Hand mit realen und virtuellen Objekten umgehen kann.

Dieser Gesichtspunkt war einer der Gründe für die Entwicklung von Datenhandschuhen und -anzügen für die Kommunikation mit einem VR-System. Der zweite Beweggrund war die stereoskopische Ausgabefähigkeit, die der Einsatz von Helmdisplays gestattet. In VR-Systemen stecken jedoch einige ernsthafte, auf absehbare Zeit schwer überwindbare Probleme:

- Der Mangel an ertast- und berührbarer Information und infolgedessen die fehlende Abstimmung mit den Feedbacksignalen aus den körpereigenen Tast- und Stellungszereptoren. Zur Bewältigung dieser Schwierigkeiten wurden bereits spezielle, wenn auch noch unzureichende Techniken entwickelt (z.B. simulierte Tastempfindung über Luftdruckpösterchen im Datenhandschuh).
- In der Kontrollschleife Benutzer – Computer herrscht immer eine bemerkbare zeitliche Verzögerung, die eine stabile Wahrnehmung zusammen mit den widersprüchlichen Signalen des Gleichgewichtssinns im Ohr stark erschwert; dies kann dann zu der bekannten Simulatorkrankheit führen.
- Der starke Einfluss fortwährender Kommunikation im Rahmen sozialer Interaktion ist von wesentlicher Bedeutung. Nicht nur das gemeinsame Gespräch, sondern auch die physisch soziale Nähe beeinflusst massgeblich die zwischenmenschliche Interaktion.

Es hat also schwerwiegende Nachteile, wenn sich ein VR-Benutzer vollständig in eine modellierte, virtuelle Welt versenken muss: das Eintauchen in die virtuelle Welt schliesst die gleichzeitige Möglichkeit der Interaktion mit Personen und Objekten in der realen Welt aus. Das Mischen realer und virtueller Objekte ist somit in ein- und derselben Schnittstelle nicht möglich. Personen aber befinden sich nach wie vor die meiste Zeit in der Realität und interagieren mit realen Objekten und anderen realen Personen.

Das menschliche Verhalten in der Realität

Jede Interaktion mit realen Objekten wird von den Gesetzen der Physik dirigiert. Einermassen analog dazu basiert jede Interaktion

mit realen Personen auf sozialen und kulturellen Normen.

In der Kommunikationsforschung zur sozialen Interaktion von Mensch zu Mensch hat man eine grundlegende Unterscheidung herausgefunden: Jede zwischenmenschliche Kommunikation besteht aus einem (Sach-)Inhalts- und einem Beziehungsaspekt. Der grösste Teil aller heutzutage eingesetzten computerisierten Informationssysteme (z.B. Email im Internet) beschränkt die Kommunikation auf den Inhaltsaspekt.

Der Beziehungsteil kann als Metaschicht angesehen werden, die die inhaltliche Schicht begleitet oder kontrolliert. Soziale Kommunikation ist auf dieser Metaschicht multifunktional; das Übermitteln einer Nachricht enthält auf der Metaschicht drei verschiedene Anteile: soziale Beziehungsaspekte, Appellinformation an den Empfänger und Selbstoffenbarungsaspekte des Senders. Alle drei Anteile sind je nach Kommunikation unterschiedlich stark beteiligt. Die Übertragung auf der Metaschicht erfolgt durch Mimik, Gestik, Tonfall und anderes nichtverbales Ausdrucksverhalten. Zusätzlich wurde herausgefunden, dass in der direkten Kommunikation wie beispielsweise beim Geschichtenerzählen auch die Inhaltsschicht oft nichtverbal vermittelt wird und am eindeutigsten an den begleitenden spontanen Gesten der Sprechers abgelesen werden kann. Gebärden fügen dem Sprechen eine weitere Dimension hinzu – gewisse Aspekte von Ereignissen können sogar nur durch Bewegung und nicht durch gesprochene Sprache weitervermittelt werden. Oder jede Art der Vermittlung kann verschiedene Aspekte einer Tatsache weitergeben, womit der Empfänger eine umfassendere Darstellung des Ereignisses bekommt.

Es lassen sich vier Arten von sprechbezogenen Gebärden unterscheiden, die allerdings nichts mit dem Gesichtsausdruck zu tun haben:

- Bildliche Beschreibung eines Teils der Handlung oder Ereignisses durch die Art der Gebärde. Beispielsweise wird der Satz «Er kletterte die Röhre hoch» von einer aufsteigenden Hand begleitet, die den Weg illustriert.
- Metaphorische Gesten, die auch repräsentativ sind, in denen das Beschriebene aber keine physische Form hat. Beispielsweise lässt sich «die Sitzung wollte und wollte nicht aufhören» mit einer rollenden Handbewegung begleiten.
- Hinweisende Bewegungen, die - im physischen Raum vor dem Erzähler - Aspekte der Geschichte lokalisieren, beispielsweise wird «Mario blickte Verena an und Verena blickte zurück» von einer Hand begleitet, die zuerst nach links, dann nach rechts weist.
- Schlagende Gesten, kleine, hiebartige Bewegungen, die sich nicht mit dem Inhalt der begleitenden Rede ändern. Sie haben pragmatische Funktion und treten bei Kommentaren, bei Korrekturen von Aussagen oder bei indirekter Rede auf. Als Beispiel wird «sie redete zuerst, äh nein, als Zweite» von einer sich erst abwärts, dann aufwärts bewegenden Hand begleitet.

Diese vier Gebärdenarten finden sich in fast allen Kulturen wieder. Ihnen allen gemeinsam sind die Typen, nicht die konkreten Ausprägungsformen. Metaphorische Gesten sind

* Matthias Rauterberg ist wissenschaftlicher Mitarbeiter am Insitut für Arbeitspsychologie an der ETH Zürich.

besonders unterschiedlich, wahrscheinlich weil Metaphern selber von einer Sprache oder Kultur zur andern so stark variieren. Wenn Englischsprachige eine Geschichte erzählen, beginnen sie oft mit einer «Übergabe»-Bewegung: ein Überreichen der Information in Form eines Pakets. Chinesen neigen eher dazu, ihre Hände mit der Fläche nach unten zu spreizen, als beschreiben sie eine Landschaft.

Dieses Quartett von gesichtsabhängigen Gesten spielt spezifische Rollen bezüglich des Inhaltsniveaus des Gesprächs, das sie begleiten.

Bildliche Gesten treten in der Inhalts- oder erzählerischen Schicht auf, metaphorische und hinweisende Gebärden in der Inhalts- und Metaschicht und rhythmische Gesten nur auf der Metaebene. Die meisten nichtsprachlichen Gebärden, einschliesslich der Gesichtsausdrücke, sind das Konzentrat des nichtverbalen Kommunikationsteils.

Das inhaltliche Niveau jeder verbalen Kommunikation (gesprochen oder geschrieben) transportiert Tatsachen, Aussagen, Nachrichten und Argumente (der rationale Teil). Die Metaschicht bezieht sich auf die Semantik von Bewegungen, Körpersprache, Gesichtsausdrücke und Gefühlsaspekte (der emotionale Teil). Die emotionalen Aspekte werden oft vernachlässigt und die rationalen Aspekte dafür überbewertet.

Von jedem etwas

Aber nur beide Aspekte gemeinsam bilden die zwei Seiten der Medaille "Kommunikation". In der zwischenmenschlichen Kommunikation können die besten Argumente nicht überzeugen, wenn der passende emotionale Beitrag fehlt. Dies war einer der Gründe, weshalb Internet-Benutzer Smileys erschufen, um ihre textbasierten Kaffeekränzchen mit emotionalen Botschaften anzureichern. Smileys sind eine Art emotionaler Ikonen, sogenannte Emoticons. Eine Studie wies nach, dass Testpersonen, die Emoticons verwendeten, mit der Kommunikation zufriedener waren als jene, die ohne Emoticons arbeiteten. Das Resultat einer weiteren empirischen Studie zeigte den positiven Einfluss von Emoticons auf die Qualität von Aufgabenlösungen.

Ausdrucksweise

Neben all diesen Gesten und Gebärden hat natürlich auch der Gesichtsausdruck einen Einfluss auf die zwischenmenschliche Kommunikation. So haben sich die beiden Forscher Fehr und Exlin mit der konzeptuellen Orientierung beschäftigt und sind zur Ansicht gelangt, dass menschliche Interaktion auf «einem gemeinsamen Code basiert, dessen Signale, Symbole oder Ausdrücke für jeden seiner Benutzer ähnliche Bezugspunkte oder Bedeutungen haben.»

Bei der Untersuchung der visuellen Interaktion, insbesondere in der Erforschung der Augenbewegungen und der zur Orientierung dienenden Kopf- und Körperbewegungen unterscheiden die Autoren zwischen visuellem Verhalten während der Interaktion wie das Starren auf bestimmte Eigenheiten des Gesprächspartners und dem Verhältnis von temporären seelischen Zuständen und kulturellen Unterschieden im Blick-Verhalten.

Die wichtigsten Ergebnisse waren:

- Mitgeteilte Information kann oft nur im Umfeld, in dem sie stattfindet und unter Voraussetzung bestimmter Erwartungen des Empfängers verstanden werden.
- Visuelle Aufmerksamkeit in Form von Blickzuwendungen spielt eine wichtige und manchmal ausschlaggebende kommunikative Rolle in menschlichen Begegnungen. Direkte und indirekte Blicke helfen bei der Strukturierung und Regelung einer Interaktion. Sie organisieren in unserer Kultur zu einem gewissen Grad das abwechselungsweise Sprechen, indem sie Annäherung und Rückzug der Kommunikationspartner signalisieren oder anbieten.

Visuelle Aufmerksamkeit übermittelt nicht nur Information, sondern dient auch der Informationssammlung und erscheint als solche während der Kindheit speziell wichtig.

Das Blickverhalten scheint auch den emotionalen Ton einer Interaktion zu reflektieren, und zwar bei Männern und Frauen unterschiedlich. Schliesslich scheint dieses Verhalten auch gewisse Persönlichkeitseigenschaften der Teilnehmer mitzuteilen.

Wenn der Mensch mit dem Objekt

Aufgabenbezogene Aktivitäten lassen sich in mehrere Arten aufteilen.

- Alltägliche motorische Bewegungen: einzelne Bewegungen mit einem Griff zu einem Zielobjekt, wie der Griff nach einem Hebel oder das Hinweisen auf ein Wort am Bildschirm. Einzelbewegungen erfolgen mit oder ohne Blickkontrolle.
- Wiederholte Bewegungen, die Wiederholung einer Einzelbewegung zu einem oder mehreren ruhenden Zielen, wie das Einschlagen eines Nagels oder das Hämmern auf eine Taste.
- Sequentielle Bewegungen umfassen Einzelbewegungen zu einer Anzahl ruhender Ziele, in regelmässigen oder unregelmässigen Abständen. Als Beispiel gelten das Schreiben auf der Schreibmaschine oder das Angeln nach Teilen in verschiedenen Lagerbehältern.
- Fortwährende Bewegungen verlangen gewisse Änderungen der Muskeln während der Bewegung wie das Bedienen eines Lenkrades oder das Leiten eines Holzstücks durch eine Säge.
- Ruhende Position, das Wahren einer bestimmten Stellung eines Körperteiles während einer bestimmten Zeit. Im Grunde ist dies keine Bewegung, sondern vielmehr das Fehlen jeglicher Bewegung. Zu den Beispielen zählt das Festhalten eines Teiles mit der Hand während des Lötens oder das Einfädeln einer Nadel ohne zu zittern.

Von besonderem Interesse sind hier die zielgerichteten motorischen Aktivitäten, die von einer Person mit der Absicht ausgeführt werden, etwas zu erreichen. Handlungen lassen sich nur durch eine funktionale Sicht, und nicht durch ihre anatomische oder mechanische Form spezifizieren. Das Fangen eines Balls könnte von der linken oder rechten Hand durchgeführt werden. Die Ausgangsposition und die Stellung für das Auffangen des Balles können sich ändern und keine zwei Fangvorgänge würden sich gleichen. Diese Bewegungen aber gelten trotzdem als eine einzige Handlung, da sie dieselbe Funktion haben.

Angewandte Forschung

Diese Ergebnisse aus dem Bereich der Verhaltensforschung lassen sich nun für die Entwicklung einer menschengerechten Schnittstelle zum Computer ausnutzen. Eine die Zeige- und Greifbewegungen der Hand unterstützende Benutzerschnittstelle hat die folgenden Vorteile:

- Sie erlaubt und fördert die beidhändige Interaktion
- Sie gestattet spezialisiertere, kontextsensitive Eingabegeräte
- Sie ermöglicht dem Benutzer mehr gleichzeitige Eingabeaktionen
- Sie unterstützt die hochentwickelten Fähigkeiten des Menschen für die physische Manipulation von realen Objekten
- Sie veräusserlicht bisher nur intern stattfindende Computerabläufe
- Sie erleichtert die Interaktion, indem Schnittstellenelemente direkter und damit auch manipulierbarer werden

- Sie ermöglicht mehreren Benutzern die Zusammenarbeit am selben Arbeitsgegenstand.

Der Weg zum NUI

Die Designstrategie gemäss Augmented Reality (AR) akzeptiert, dass Leute sich die wirkliche Welt gewohnt sind und dass sich diese reale Welt nur unvollständig und unzureichend genau im Computer wiedergeben lässt. Augmented Reality baut daher auf der realen Welt auf und reichert sie mit den Fähigkeiten der Computer- und Informationstechnik an. AR ist die zugrundeliegende Designstrategie für das Konzept der «natürlichen Benutzerschnittstelle» NUI.

Ein System mit einem NUI unterstützt eine Mischung von realen und virtuellen Objekten im Interaktionsraum zwischen Mensch und Computer. Die Computereingabe erkennt (über visuelle, akustische oder andere Sensoren) und versteht die physischen Objekte und Gebärden der Leute, die sich möglichst natürlich bei ihrer täglichen Arbeit verhalten (z.B. Sprach- und/oder handschriftliche Eingabe). Die Computerausgabe beruht auf Musterwiedergabe wie Videoprojektion, Holographie, Sprachsynthese und dreidimensionale Tonmuster. Unabdingbar für unsere Definition eines NUI ist, dass sie inter-referentielle Ein-/Ausgabe erlaubt: dieselbe Modalität kann für Input und Output verwendet werden. So kann z.B. ein projiziertes virtuelles Objekt vom Benutzer direkt für seine nichtsprachlichen Eingaben weiterverwendet werden.

Der Aufbau

Die räumliche Position des Benutzers wird über das Bild von zwei Kameras bestimmt. Somit entsteht gleichzeitig ein stereoskopisches Bild für eventuelle Videokonferenzpartner. Sprache und Ton werden von mehreren Mikrofonen aufgenommen, damit das System ein internes dreidimensionales Modell des Benutzers auch im auditiven Bereich erstellen kann. Eine dritte Kamera zeichnet von oben permanent den Zustand der Arbeit des Benutzers auf, die auf der horizontalen Arbeitsfläche abläuft. Hier sind virtuelle und physische Objekte vollständig überlagert und integriert.

Dieser Aufbau mehrerer paralleler Eingabekanäle erlaubt es, mehrere Ansichten an entfernte Kommunikationspartner weiterzuleiten, wie ein dreidimensionales Gesicht und eine dreidimensionale Ansicht gemeinsamer Arbeitsobjekte. Der multimediale Gesamteindruck entsteht durch die Wiedergabe auf den vertikalen Kommunikationsbereich, durch den Blick von oben auf den horizontalen Arbeitsbereich und durch die vier Lautsprecher, die dem Benutzer einen räumlichen Klangeindruck vermitteln. Der ungenutzte Platz im vertikalen Kommunikationsbereich kann für die Ablage und Präsentation von Arbeitsergebnissen genutzt werden. Natürlich können herkömmliche Ein- und Ausgabegeräte weiterhin zusätzlich genutzt werden. NUIs sind mehrmodusfähig und ermöglichen es dem Benutzer, für jede Handlung den geeigneten und von ihm bevorzugten Interaktionsstil zu wählen.

Da die Menschen Objekte in der physischen Welt am häufigsten und natürlichsten von Hand bewegen, besteht der Wunsch, diese Fähigkeiten an die Mensch-Computer Interaktion zu integrieren. NUIs erlauben es dem Benutzer denn auch, mit echten und virtuellen Objekten im Arbeitsbereich in einer buchstäblich manipulativen Art zu agieren. Der Arbeitsbereich ist notwendigerweise horizontal, damit der Benutzer echte Objekte darauf abstellen kann. Benutzer erhalten Informationen über den Zustand manipulierter Objekte genau an dem Ort, an dem sie die Manipulation vorgenommen haben: Der Raum der Wahrnehmung und der Raum der Handlung stimmen überein.

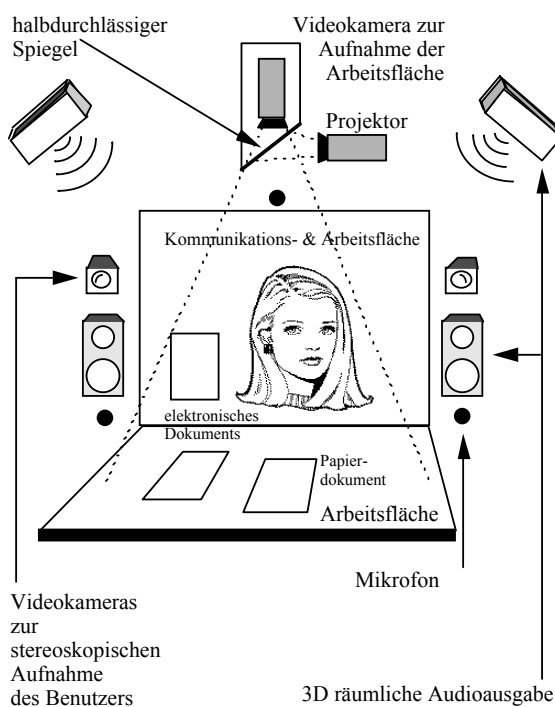
NUI-hafte Applikationen

Bereits bestehen einige Applikationen, die verschiedene NUI-artige Systeme enthalten. Zu diesen Prototypen gehören:

- **BrightBoard:** Das BrightBoard verwendete als erste die Videoeingabe zur Steuerung eines Computersystems. Die Kamera erkennt Befehle auf einem Blatt Papier, einem Whiteboard oder Flip-Chart. Die Bestätigung erfolgt akustisch durch Ton oder einen Sprachsynthesizer. Das BrightBoard beruht zwar auf den Gedanken der Augmented Reality, ist aber nach Definition noch kein NUI, weil es die interreferentielle Ein/Ausgabe nicht unterstützt. Die Eingabe erfolgt visuell und die Ausgabe akustisch. Der Benutzer muss eine Menge Befehle lernen (die er von Hand auf die Tafel schreibt), um das System so zu steuern, wie es auch bei einer Kommandoschnittstelle üblich ist.
- **DigitalDesk:** Pierre Wellner entwickelte den ersten DigitalDesk. Er ist zunächst ein ganz gewöhnlicher Schreibtisch und kann als solcher auch verwendet werden, hat aber zusätzliche Fähigkeiten. Eine auf die Arbeitsfläche gerichtete Videokamera ist über dem Tisch angebracht. Die Aufnahmen der Kamera werden einem Mustererkennungssystem zugeführt, welches erkennt, wohin der Benutzer weist, Dokumente liest, die auf den Schreibtisch gelegt werden, und Rechnungen mittels projiziertem Taschenrechner durchführt. Dazu ist ein computergesteuerter Projektor als Ausgabegerät über dem Pult angebracht, mit dem das System elektronische Objekte auf die Arbeitsfläche und auf echte Papierdokumente projizieren kann. (Dies ist mit herkömmlichen Flachbildschirmen nicht möglich.)
- **DoubleDigitalDesk:** Der DoubleDigitalDesk ist eine erweiterte Version des DigitalDesk. Er gestattet den gemeinsamen Gebrauch von Papierdokumenten. Er erlaubt es Benutzern in zwei weltweit verteilten Büros, ihre physischen Pulte zu teilen, damit jeder des anderen Papierdokumente betrachten, korrigieren und darin herumkritzeln kann.
- **Mosaic:** Diese Applikation baut auf dem DigitalDesk auf. Mosaic befasst sich mit dem Problem, Papier entlang der zeitlichen Dimension auszudehnen. Es erlaubt damit eine Schnittstelle, welche die Vorzüge von einzelnen Papierbildern mit computerkontrolliertem Video kombiniert.
- **Ariel:** Im Rahmen eines europäischen Esprit-Projekts werden neue Möglichkeiten erforscht, um Ingenieure mit einem NUI-basierten System zu unterstützen. Die englische Forschergruppe analysiert die Aufgabe, in Dänemark die weltweit längste Hängebrücke zu konstruieren. In der Praxis arbeiten Ingenieure zunächst oft mit physischen Papierkopien von CAD-Zeichnungen. Um von Hand angelegte Bemerkungen der Online-Version zuzuführen und die Kommunikation zwischen Ingenieuren zu fördern, werden physische Zeichnungen mit handschriftlichen und mündlichen Anmerkungen in elektronischer Form ergänzt. Die somit personalisierte Papierzeichnung eines Ingenieurs wird zur Schnittstelle für das Computersystem.
- **InteractiveDesk:** Ein Computerbild wird auf dem Schreibtisch projiziert mit der zusätzlichen Möglichkeit der Stifteingabe und einem normalen aufrechten Monitor mit Tastatur, womit die gewohnten Bürogeräte einem AR-System einverleibt werden. Der Prototyp unterstützt die Benutzer darin, Eingabemethoden zu wechseln (je nach Aufenthaltsort der Tastatur) und Files mit Hilfe realer Objekte abzurufen (z.B. ein mit zusätzlichen Sensoren ausgestatteter Karteikasten).

- **PlayingDesk:** Eine Forschergruppe unter Leitung des Autors implementierte das Brettspiel Gobang, das mit echten, hölzernen Spielsteinen auf einem projizierten virtuellen Spielfeld gespielt werden kann. Als erstes legt der Spieler einen schwarzen Stein auf ein virtuelles Feld. Das System erkennt die Position, berechnet den nächsten Zug und markiert dieses Feld mit einem projizierten roten Fleck: dem virtuellen Spielstein des Computers. Im Feldversuch – während einer fünftägigen Messe – waren vier Stationen öffentlich zugänglich, jede davon bot dasselbe Spiel mit einer unterschiedlichen Benutzerschnittstelle an (Kommandoschnittstelle, mausbasierte Desktopoberfläche, multimediale Oberfläche mit Touchscreen und das NUI des PlayingDesk). Über 8000 Benutzer versuchten ihr Glück gegen einen der Computer. Die Gewinnchancen für den Menschen sind dann signifikant am grössten gewesen, wenn sie mit dem bekannten und "gewohnten" Spielbrett mit realen Spielsteinen (PlayingDesk mit NUI) spielten, während sie am kleinsten waren, wenn sie an einer der anderen Workstations mit den üblichen Schnittstellen spielten.

In allen vorgestellten NUI-basierten Applikationen erkennt der Computer die Benutzerbefehle durch Video-basierte Mustererkennung und führt die Handlungen durch Projektion seiner Ausgabe auf denselben Arbeitsbereich aus.



Figur 1: Bestandteile einer natürlichen Benutzerschnittstelle.

Technische NUI-Forschung

Um eine natürliche Benutzerschnittstelle zu realisieren, müssen Ein- und Ausgabe durch Mustererkennungstechniken erkannt und entsprechend projiziert werden. Der Benutzer handelt in seiner natürlichen Umgebung mit realen und virtuellen Objekten. Das System hält das Szenario fest und verarbeitet das Bild, um im Zusammenhang bedeutungsvolle Objekte zu erkennen. Dabei könnte es sich um Information auf einem Blatt Papier oder einem Whiteboard handeln, es könnte eine menschliche Handlung sein, es könnte jedes beliebige flache oder dreidimensionale Objekt oder ein projiziertes digitales Objekt sein. Um hier erfolgreich zu sein, sind aus den folgenden Forschungsbereichen effiziente und robuste Algorithmen unerlässlich.

- 2- und 3-dimensionale Objekterkennung
- Scanning und optische Zeichenerkennung
- Gestenerkennung
- Stimmen und Ton-Erkennung
- Mustererkennung allgemein
- Bildverständnis.

Es muss bestimmt werden, welche Bilder wichtig sind und wann das System etwas unternehmen muss. Wenn beispielsweise ein Benutzer mit dem Finger über ein Dokument fährt, um auf ein spezielles Wort zu deuten, und das System mehrere Bilder aufnimmt, wie weiss es, auf welchem Bild der Benutzer in der Endposition ankommt? Auf der Ausgabeseite sind die folgenden Bereiche betroffen:

- zwei- und dreidimensionale hochauflösende Projektion
- dreidimensionale Audiowiedergabe
- Sprach- und Tonsynthese

Die zweidimensionale Projektion ist wohlerforscht und wird im Alltag fleissig eingesetzt – bei der Diaprojektion, Fernsehprojektion etc. Die dreidimensionale Projektion dagegen steckt noch im Forschungsstadium. Vielversprechende Projekte sind digitale Holografie und volumetrische Wiedergabesysteme. Die Echtzeitintegration eines dreidimensionalen Objektes in eine gefilmte (echt) Szene ist ebenfalls schon erfolgt. Anforderungen an solche Projektionen sind flimmerfreie, hochauflösende Bilder.

Im Überblick

Zusammenfassend lässt sich sagen, dass natürliche Benutzerschnittstellen gegenüber herkömmlichen Interaktionsarten und der virtuellen Realität einige Vorzüge haben. Besonders Aspekte der zwischenmenschlichen nichtverbalen Kommunikation und das Festhalten aufgabenspezifischer Aktivitäten (Bewegungen, Stimme und Ton) werden adäquat unterstützt. Die Technologie ist bereits soweit entwickelt, dass viele Forschungsprototypen schon heutzutage die sinnvolle Einsetzbarkeit einiger NUI-Komponenten beweisen. Im Bereich der multidisziplinären Forschung wurden mehrere Bereiche identifiziert, die von dem Autor als Schlüsseltechnologien für die Bewältigung der verbleibenden technischen Probleme identifiziert wurden, insbesondere die Mustererkennung und die hochauflösende Projektion. NUIs unterstützen Menschen in ihren realen Aufgaben und öffnen gleichzeitig einen grossen neuen Raum für die Entwicklung der kommenden Generation von HCI-Technologie.

Literatur

- Arai T, Machii K, Kuzunuki S, Shojima H: InteractiveDESK: A Computer-augmented Desk Which Responds to Operations on Real Objects. In *ACM Proc. of the CHI '95 Companion*, 1995, pp. 141-142.
- Fehr B, Exline R: Social visual interaction: a conceptual and literature review. In *Nonverbal behavior and communication* (A. Siegman & S. Feldstein, eds.), Lawrence Erlbaum, 1987, pp. 225-326.
- Rauterberg M, Stebler R, Mauch T: What is a promising candidate for the next generation of interface technology. In: *Proc. '5th International Conference INTERFACE to Real & Virtual Worlds'*. Montpellier: EC2 & C^{ie}, pp. 95-103.
- Stafford-Fraser Q, Robinson P: BrightBoard: A Video-Augmented Environment. In *ACM Proc. of the CHI '96*, 1996, pp. 134-141.
- Wellner P, Mackay W, Gold R: Computer-Augmented Environments: Back to the Real World. *Communications of the ACM*, 36(7), 1993, pp. 24-26.
- Wellner P: Interacting with Paper on the Digital Desk. *Communications of the ACM*, 36(7), 1993, pp. 87-96.
- Wellner P, Freeman S: The DoubleDigitalDesk - Shared Editing of Paper Documents. *Technical Report EPC-93-108*, Xerox Research Centre Cambridge Laboratory, Cambridge UK, 1993.