

Multimodale Interaktion in der Virtuellen Realität

I. Wachsmuth, I. Voss, T. Sowa, M.E. Latoschik, S. Kopp & B. Jung

Technische Fakultät, Universität Bielefeld

Zusammenfassung

Virtuelle Realität oder Virtual Reality (VR) bezeichnet ein neuartiges Kommunikationsmedium, das die unmittelbare Wechselwirkung des Menschen mit räumlich organisierten rechnergenerierten Darstellungen erlaubt. Verbunden mit körperlich verankerter Interaktion finden insbesondere gestische Eingaben starkes Interesse. Dieser Beitrag gibt einen Überblick über Forschungsarbeiten im Labor für Künstliche Intelligenz und Virtuelle Realität an der Universität Bielefeld, mit denen Grundlagen für den Einsatz gestischer und sprachlicher Interaktionstechniken entwickelt werden; als Erprobungsdomäne dient ein Szenario des virtuellen Konstruierens. Für die schnelle Erfassung komplexer Hand-Armgesten werden derzeit Datenhandschuhe und Körper-Tracker eingesetzt. Die Auswertung erfolgt mit wissensbasierten Ansätzen, die atomare Formelemente der Gestik symbolisch beschreiben und zu größeren Einheiten zusammensetzen. Ein zweites Thema ist die multimodale Interaktion durch sprachlich-gestische Eingaben, z.B. wenn auf einen Gegenstand gezeigt ("dieses Rohr") oder eine Drehrichtung ("so herum") signalisiert wird. Schließlich wird dargestellt, wie die Ansätze zur Formbeschreibung von Gesten für die Synthese natürlich wirkender gestischer Ausgaben mit einer artikulierten, anthropomorphen Figur übertragen werden können, die in laufenden Arbeiten mit Sprachausgaben koordiniert werden.

1 Einleitung und Forschungsüberblick

Virtual Reality beinhaltet wesentlich die Abkehr von üblicher Bildschirmausgabe und damit verbundenen WIMP-Techniken (windows, icons, menus, pointer) und ersetzt sie durch betrachterabhängige dreidimensionale Ein- und Ausgabeverfahren (Burdea & Coiffet 1994; Barfield & Furness 1995; van Dam 1997). Statt vor dem Bildschirm zu sitzen treten Anwender mehr oder weniger stark in eine "greifbare" synthetische Welt ein, in der sie sich bewegen und die Wirkungen ihres Tuns unmittelbar erfahren können. Maus und Tastatur sind in diesem Zusammenhang unnatürlich; die Interaktion basiert zumeist auf direktem Manipulieren mit dem Datenhandschuh, unterstützt durch Trackingsysteme.

Dennoch sind die Interaktionsmöglichkeiten des Benutzers in vielen Fällen zunächst auf die Exploration einer virtuellen Welt und simple Objektbewegungen beschränkt. Ein Hauptziel ist daher der Brückenschlag zwischen Bilderzeugungssystemen, die errechnete Visualisierungen einem überwiegend passiven Benutzer zur Verfügung stellen, und interaktiven Systemen, die Benutzereingriffe in der visualisierten Szene unmittelbar umsetzen können. Zur Unterstützung der Benutzerinteraktion in VR wird gefordert, dass die synthetischen Darstellungen möglichst umfassend Eigenschaften und Manipulationsmöglichkeiten besitzen sollen, die der Mensch in Analogie zu den dargestellten realen Objekten intuitiv voraussetzt. Hierzu sind wesentliche Impulse für die VR-Technik durch Methoden der Künstlichen Intelligenz zu erwarten bzw. im Ansatz erbracht worden (Wachsmuth 1998); mittlerweile hat sich ein Begriff "Intelligenter virtueller Umgebungen" zu etablieren begonnen (Luck & Aylett 2000; Jung & Milde 2000).

Für industrielle Anwendungen sind besonders VR-Systeme interessant, mit denen sich Modelle realer Objekte bereits vor dem Bau physikalischer Produktmodelle realistisch darstellen

und explorieren lassen (Dai & Göbel 1994). Die Erstellung sog. virtueller Prototypen aus CAD-basierten parametrischen Grundbauteilen findet bereits im Fahrzeug- und Flugzeugversuchsbau Einsatz (Dai 1998). Dabei bildet die Unterstützung der VR-Technik durch intelligente Funktionen den Ausgangspunkt für ein "virtuelles Konstruieren". Die Grundidee besteht darin, Benutzern die Möglichkeit zu geben, mit virtuellen Objekten ähnlich zu interagieren wie aus der realen Umgebung gewohnt. Dies erfordert zunächst eine Simulation physikalischer Objekteigenschaften, wobei zwei Richtungen verfolgt werden: (1) physikalische Simulationen (z.B. Metaxas 1996; Gupta et al. 1997) mit noch sehr beschränkten Möglichkeiten, da extrem rechenintensiv, und (2) physikrekonstruierende, algorithmisch effiziente Ansätze wie Kollisionsberechnung und Modellierung von "Schnappmechanismen" zur teilautomatischen Objektpositionierung; siehe z.B. (Milne 1997; Jayaram et al. 1997; Gausemeier et al. 1998; Drews & Weyrich 1998; Jung et al. 2000).

Verbunden mit körperlich verankerter Interaktion finden insbesondere gestische Eingaben starkes Interesse; dabei können kommunikative (distante) und manipulative Gesten unterschieden werden. Frühe Ansätze zur kommunikativen Gestik in VR (z.B. Böhm et al. 1992) betrachten eher eingeschränkte, symbolische Gesten. Daneben wird meist Zeigegestik zur Objekt- und Ortsreferenzierung ausgewertet (Cavazza et al. 1995, Lucente et al. 1998). Bereits kombiniert mit Spracheingabe dienen des weiteren ikonische Gesten der Kommunikation von Formbeschreibungen oder Positionsänderungen (Sparrell & Koons 1994), bei (Lucente et al. 1998) zudem der Skalierung von Objekten. Schließlich werden mimetische ("vormachende") Gesten zur kontinuierlichen Interaktion mit 3D-Objekten untersucht (siehe Abschnitt 2.1). Im Hinblick auf manipulative Gestik wurde die Interaktion mit virtuellen Objekten zunächst oft nur auf Umwegen, über "Space Menus", 3D-Widgets oder sequentielle Spracheingabe erzielt (Weimer & Ganapathy 1989; Hauptmann & McAvinney 1993). Es wurden auch erste VR-Systeme für zweihändige Manipulation entwickelt (Mapes & Moshell 1995; Shaw & Green 1994; Mine et al. 1997; Cutler et al. 1997).

Zum Einsatz kommen heute vielfach großprojizierende, komfortable Mehrbenutzerumgebungen, die an reale Arbeitssituationen anschließbar und für kooperative Bearbeitung multimedialer Daten geeignet sind: Stereoprojektionen auf Großbildwänden ("Walls") und Bildtischen ("Responsive Workbench"; Krüger et al. 1995), mehrseitige Stereoprojektionen ("Caves"; Cruz-Neira et al. 1993), oder "Holobench" und "Holoscreen" als kostengünstigere Zweiseiten-Varianten. In neueren VR-Modellierungssystemen sind dabei Funktionen für eine netzwerkweite Verteilung (Distributed Virtual Reality) bereits berücksichtigt (siehe z.B. Fellner & Hopp 1997). In Design- und Konstruktionsanwendungen lassen sich Entscheidungen örtlich verteilter Arbeitsgruppen so gemeinsam treffen und interaktiv umsetzen ("Concurrent Engineering"). Es lässt sich erwarten, dass solche Plattformen eine Basis für zahlreiche Ingenieuraufgaben im virtuellen Entwurf sein werden, was einen Hintergrund für unsere im Folgenden dargestellten Arbeiten bildet.

2 Ansätze zur multimodalen Interaktion in Virtueller Realität

Dieser Beitrag gibt einen Überblick über Forschungsarbeiten im Labor für Künstliche Intelligenz und Virtuelle Realität an der Universität Bielefeld, mit denen Grundlagen für den Einsatz gestischer und sprachlicher Interaktionstechniken entwickelt werden. Seit mehreren Jahren ist unser Forschungsleitziel der Entwurf intuitiver Mensch-Maschine-Schnittstellen für interaktive 3D-Grafiksysteme, deren grundsätzliche Machbarkeit mit implementierten Systemprototypen demonstriert wird. In frühen Arbeiten ließen sich bereits mit dem VIENA-System (Wachsmuth & Cao 1995) einfache Zeigegesten (über Datenhandschuh) zur Ergänzung verbaler Eingaben auswerten. Dabei wurden in die Grafikwelt gekoppelte

Softwareagenten zur Berechnung der Raumreferenz bei natürlichsprachlichen Benutzer-eingaben eingesetzt. Einbezogen wurden Verkörperungen von Raumreferenzsystemen durch eine anthropomorphe Figur, die sich gestisch äussern kann (Jörding & Wachsmuth 1996) sowie auch Agentenlernverfahren zur Adaptation an Benutzerpräferenzen (Lenzmann & Wachsmuth 1997). Hierbei wurde eine Grundkonzeption für die Realisierung mehrmodaler Schnittstellen entworfen, die in den weiteren Vorhaben verfeinert wurde (Sowa et al. 1999), dies sowohl im Hinblick auf multimodale Eingaben (siehe Abschnitt 2.1) als auch für die Konzeption einer in die virtuelle Umgebung eingebrachten anthropomorphen Figur, die multimodale Äußerungen generieren soll (Abschnitt 2.2).

Ein Schwerpunkt unserer Arbeiten sind jetzt wissensbasierte Ansätze zum virtuellen Konstruieren, durch die realitätsnahes Manipulieren der virtuellen Umgebung erreicht wird. In einem zweiten Fokus werden sprachlich-gestische Eingaben in der Mensch-Maschine-Interaktion bearbeitet. Die beiden Forschungsrichtungen ergänzen sich, z.B. sind bei Ausnutzung von Domänenwissen Vagheiten in Benutzereingaben und Ungenauigkeiten der VR-Eingabegeräte ausgleichbar. Die unten näher beschriebenen Arbeiten bauen auf einem wissensbasierten Simulationssystem auf, dem Virtuellen Konstrukteur, mit dem die interaktive Montage komplexer Aggregate aus CAD-basierten Grundbausteinen auf einer bildschirmpräsentierten virtuellen Werkbank möglich ist (Jung & Wachsmuth 1998). Weitere Arbeiten betreffen die gestische und sprachliche Handhabung virtueller Objekte an einer interaktiven Wand mit 2.5m x 3m Stereoprojektion. Im Gegensatz zu den einleitend beschriebenen Ansätzen werden dabei nicht vordefinierte zeichensprachliche Gesten, sondern natürliche koverbale Gesten untersucht (Fröhlich & Wachsmuth 1998; Latoschik et al. 1998). Dabei werden hochaufgelöste räumliche Darstellungen CAD-basierter Bauteilmodelle in realistischer Größe präsentiert und über VR-Eingabegeräte (Datenhandschuhe, Positionssensoren, Spracherkennungssystem) interaktiv gehandhabt. In stärker auf Interaktionen im Greifraum bezogenen Szenarien kommt ferner eine Responsive Workbench (RWB) zum Einsatz.

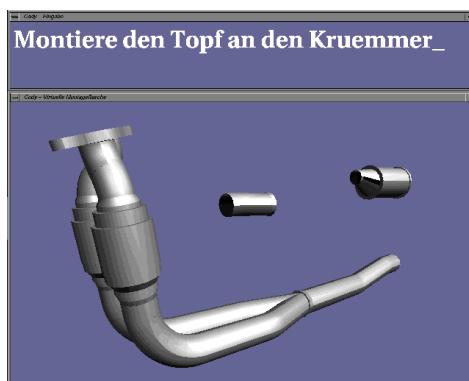


Abb. 1: Montage CAD-basierter Bauteile mit dem Virtuellen Konstrukteur: Links wird die Abgasanlage eines Automobils mittels sprachlicher Instruktionen gefügt (IGES-Datensatz der Volkswagen AG). Rechts werden Rahmenteile eines Kleinfahrzeugs unter sprachlich-gestischen Instruktionen virtuell zusammengesetzt, bei zusätzlicher Auswertung der Blickrichtung.

Mit dem Virtuellen Konstrukteur steht ein wissensbasiertes System zur Verfügung, das interaktive Montagesimulationen mit dreidimensional modellierten Bausteinen ermöglicht. Grundlage dafür ist ein allgemeiner Ansatz zur Modellierung der Bauteilverbindungsmöglichkeiten (Jung et al. 2000); er berücksichtigt verschiedene Arten von Verbindungsstellen virtueller Bauteile, z.B. den Schaft einer Schraube oder das Gewinde einer Mutter, sowie verschiedene Verbindungstypen, z.B. Schrauben oder Stecken. Neben den Standard-

interaktionen herkömmlicher Graphiksysteme wie Navigation und Objekttranslation erlaubt der Virtuelle Konstrukteur die Echtzeitsimulation von montagebezogenen Manipulationen: passgenaues Fügen und Trennen von Bauteilen und Aggregaten und die Modifikation erzeugter Aggregate durch Relativbewegung (Rotation und Translation) von Bestandteilen gemäß verbindungsartspezifischer Freiheitsgrade. Auf Basis verschiedener Baukastensysteme, die von Baufix-Holzbauteilen bis hin zu industriellen, CAD-basierten Grundbausteinen (z.B. die Auspuffanlage eines VW Polo) reichen, wurden Ansätze entwickelt, mit denen das System die beim virtuellen Konstruieren entstehenden Baugruppen automatisch erkennt. Dies ermöglicht u.a. deren funktionsbezogene Benennung in sprachlichen Instruktionen. Benutzereingaben erfolgen in der Desktop-Version des Virtuellen Konstrukteurs durch mausbasierte Manipulation und sprachliche Instruktion. An der interaktiven Wand erfolgen die Benutzereingaben dagegen mit sprachunterstützter Gestik (Abb. 1).

In diesem Szenario wurden verschiedene Problemstellungen der Realisierung multimodaler Interaktion in der virtuellen Realität bearbeitet und Lösungskonzepte entwickelt, die in Tabelle 1 aufgeführt sind und in den folgenden Abschnitten kurz erläutert werden.

PROBLEM	LÖSUNGSKONZEPTE
Korrespondenz zeitlich paralleler Eingaben	a) Symbol-Integratoren b) erweitertes ATN
explizite Beschreibung von Gesten	merkmalsbasiert, kompositionell: a) logische Operatoren auf Attributsequenzen b) HamNoSys (mit Erweiterungen für temporale Merkmale)
Abstraktion von Körperdaten	Körpermodell und Handmodell auf der Basis von Aktuatoren
Einbindung kontinuierlicher Interaktion	Motion-Modifikatoren
Auflösung von Vagheit	a) Komplementierung Sprache und Gestik b) Auswertung externer Wissensquellen (Virtueller Konstrukteur)

Tab. 1: Übersicht über entwickelte Lösungskonzepte für Probleme der multimodalen Interaktion

2.1 Multimodale, insbesondere gestische Eingabe-Interaktion

Multimodale Interaktion betrifft hier die Umsetzung von Benutzereingriffen in visualisierten 3D-Szenen anhand sprachbegleiteter Gesteneingaben. Im 1999 abgeschlossenen SGIM-Projekt (Latoschik et al. 1999) wurden Grundlagentechniken entwickelt, die mit verschiedenen Sensoren Informationen über die Bewegungen der oberen Extremitäten und die Position eines Benutzers bei der Interaktion an der Stereo-Projektionswand erschließen. Dazu gehört die signaltechnische Erfassung und Bedeutungsanalyse von Körpergestik (vor allem Arme und Kopfstellung des Benutzers) und ihre Kopplung in Anwendungssysteme. Die gestische Kommunikation wird durch Spracheingabe unterstützt, mit der unter anderem Objekttypen und -positionen mitgeteilt werden, um Vagheiten im gestischen Eingabekanal aufzulösen. Die signaltechnische Erfassung der Körperbewegung (Abb. 2) erfolgt mit einem Sensorset (Flock of Birds) und Datenhandschuhen (CyberGloves). Die verwendete Spracherkennung, ein an der Universität Bielefeld entwickelter Forschungsprototyp (Fink et al. 1998), arbeitet benutzerunabhängig, inkrementell und kontinuierlich.

Die Gestenerkennung basiert auf der Detektion definitorischer Merkmale, die sowohl die Form als auch den zeitlichen Verlauf einer Geste betreffen und die mit logischen Operatoren auf registrierten Sequenzen solcher Attribute verknüpft werden. Als Formmerkmale werden Fingerstellung, Handorientierung und -position betrachtet. Zeitlich-expressive Elemente, die auf das Vorliegen einer bedeutungstragenden Geste hinweisen, sind Ruhepunkte, hohe Beschleunigungen, Symmetrien und Abweichungen von Ruhestellungen bei Handspannung und der Handposition. Zur Detektion der Form- und Expressionsmerkmale wurden Erkener entwickelt, die auf den Sensordaten der Datenhandschuhe und Positionssensoren aufsetzen. Abbildung 3 zeigt exemplarisch den Bewegungspfad der rechten Hand bei aufeinander folgenden Zeigegesten. Deutlich zu erkennen ist sowohl ein "Overshooting" bei der Armrückstellung in die absolute Ruheposition als auch das momentane Verharren der Hand in einer relativen Ruheposition bei zeitlich eng aufeinander folgenden Zeigegesten. Die Entwicklung der Erkener erfolgte anhand solcher Daten aus einer Untersuchung an 37 Versuchspersonen.



Abb. 2: Gestenerfassung mit Positions-/Bewegungssensoren an Handgelenken, Stereobrille und Rücken (Körperreferenz) sowie Datenhandschuhen, über die Fingerstellung und Handspannung ermittelt werden (Spracheingabe über Kopfmikrofon).

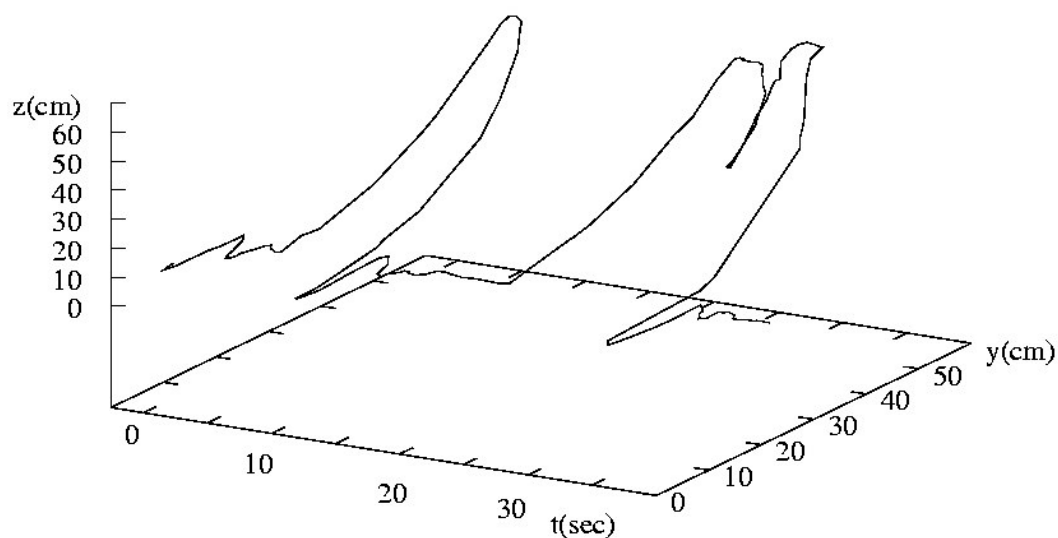


Abb. 3: Bewegungsverlauf der rechten Hand bei aufeinanderfolgenden Zeigegesten (aufgetragen sind hier die y- und z-Richtung über der Zeit).

Als weiterer Ansatz zur expliziten Gestenbeschreibung wird das aus der Gebärdensprachenlehre hervorgegangene "Hamburger Notationssystem" HamNoSys (Prillwitz et al. 1989) zugrunde gelegt, das Gesten auf Basis von atomaren Formelementen symbolisch-kompositionell beschreibt, mit einem Fokus auf der Darstellung der oberen Gliedmaßen. Gesten werden in HamNoSys als Wörter notiert, die aus Grundsymbolen zusammengesetzt sind; hiermit ist eine formalsprachliche Charakterisierung von Symbolstrukturen möglich, die maschinelle Verwendung ermöglicht. Mit den Kombinationsmöglichkeiten der Grundsymbole können statische Konfigurationen (Posturen) und dynamische Gestenteile (Aktionen) beschrieben werden. In Abb. 4 ist beispielsweise eine Rechtszeigegeste notiert, bestehend aus einem statischen Anteil der Hand-Arm-Konfiguration (Hand mit Zeigefinger nach vorn gestreckt, Handfläche nach links orientiert, Arm in Schulterhöhe ganz gestreckt) und einem dynamischen Anteil (Hand erst nach vorn, dann nach rechts bewegt). Die Gestenerkennung wird durch zeitliche Integration der nebenläufig im Kurzzeitspeicher auflaufenden Formelemente mit dafür entwickelten Symbol-Integratoren bewerkstelligt (Sowa et al. 1999).

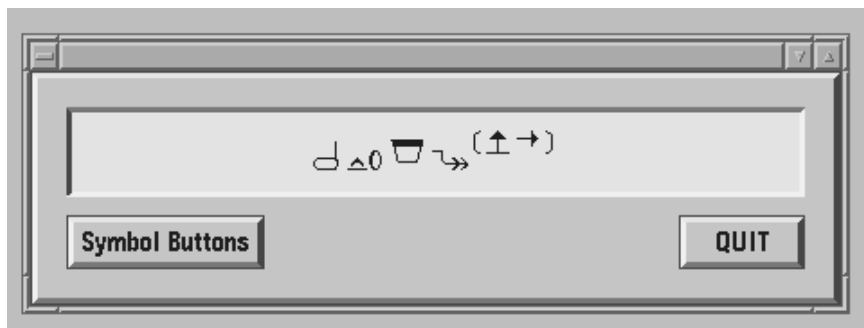


Abb. 4: HamNoSys-Notat einer "Nach-vorn-und-dann-rechts"-Zeigegeste, die in etwa der in Abb. 3 dargestellten zweiten (zusammengesetzten) Zeigebewegung entspricht.

Realisiert wurden zunächst Erkener für universelle Basisinteraktionen (Zeigen, Greifen, Loslassen, Rotation, Translation), die in HamNoSys ausgegeben können. Bei der sprachlich-gestischen Interaktion werden drei Typen kommunikativer Gesten ausgewertet: Deiktische Gesten ("nimm <Zeigegeste> dieses Teil") spezifizieren ein Objekt der virtuellen Umgebung oder auch einen Ort, mimetische Gesten ("drehe es <kreisender Zeigefinger> so herum") qualifizieren die Ausführung einer Aktion, und ikonische Gesten ("das so <Andeutung von Form/Lage mit den Händen> geformte Objekt ...") werden zur Objektreferenz verwendet (Abb. 5). Bei der Zusammenführung von Sprache und Gestik (multimodale Integration) ist das Korrespondenzproblem (Srihari 1994) zu lösen, d.h. die semantisch-pragmatische Zuordnung zeitlich paralleler sprachlicher und gestischer Äußerungssegmente. Dabei wird als Hauptkriterium zeitliche Nähe ausgewertet: Perzepte des gestischen und sprachlichen Kanals werden zeitgestempelt in einem Kurzzeitspeicher abgelegt, auf den der Integrationsprozess zugreift, und über ein erweitertes ATN einander zugeordnet (Latoschik 2000).

Die Interpretation der Gesten erfolgt grundsätzlich im Kontext einer sprachlichen Äußerung und vor dem Hintergrund des Anwendungsszenarios der Virtuellen Konstruktion. Die Umsetzung multimodaler Anweisungen kann dabei diskret oder kontinuierlich erfolgen (Latoschik et al. 1999). Sind alle Parameter einer Manipulation durch die Eingabe und den Kontext bestimmt, wird sie in einem diskreten Schritt durchgeführt; dies ist der Fall für alle gestischen Eingaben mit deiktischem Typ und im Grundansatz auch für ikonische Gesten angemessen. Mimetische Gesten (die gewünschte Objektmanipulationen vormachen) werden als kontinuierliche Interaktionen umgesetzt. Zur Abstraktion von den gemessenen Kör-

perdaten dienen sog. Aktuatoren als virtuelle Repräsentanten; sie binden während einer kontinuierlichen Interaktion an sog. Motion-Modifikatoren, die die übermittelten Daten im Hinblick auf die gestisch angedeutete Bewegung auswerten. Die so erhaltenen Bewegungsprimitive binden dann an Objekt-Manipulatoren, die wiederum die virtuellen Objekte modifizieren. Die Interpretation mimetischer Gesten wird unter Auswertung externer Wissensquellen durch die möglichen Lage- und Positionsänderungen der dargestellten Bauteile beschränkt. Unverbundene Teile lassen prinzipiell alle manipulativen Freiheitsgrade zu; nach einem Verbindungsschluss werden diese Freiheitsgrade entsprechend eingeschränkt.

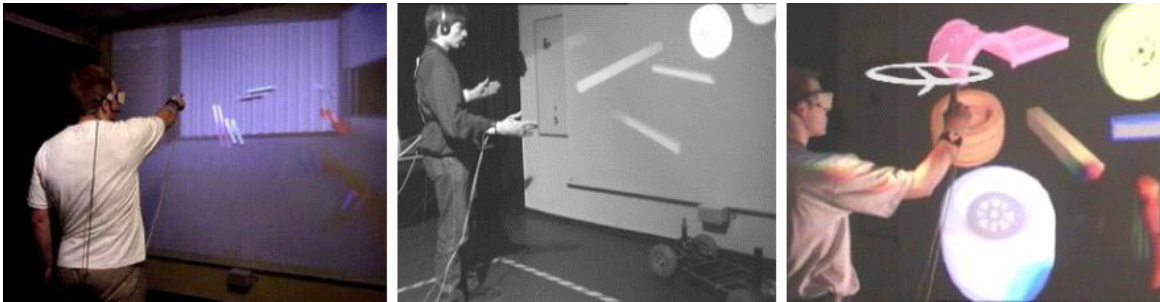


Abb. 5: "... stecke es an dieses Rohr...", "...das so liegende Objekt...", "... drehe es so herum..."

2.2 Multimodale, insbesondere gestische Ausgabe-Interaktion

Als Pendant zur multimodalen Eingabeverarbeitung werden die beschriebenen Ansätze auch zur Generierung von "lifelike"-Gesten eingesetzt, mit denen sich u.a. deiktische Referenzen im virtuellen Raum erzeugen und system- wie benutzerseitig auswerten lassen. Eine virtuelle anthropomorphe Figur – in Abb. 6 als kinematisches Skelett dargestellt – kann durch Kombination symbolischer und numerischer Techniken Gesten mit den oberen Gliedmaßen und Händen in natürlichem zeitlichen Ablauf generieren; sie wird derzeit für den simultanen Einsatz mit synthetischer Sprachausgabe vorbereitet. Abb. 6 zeigt Extrempunkte im Phasenablauf einer Geste des Heranziehens oder -winkens, die aus einer erweiterten HamNoSys-Beschreibung (Abb. 7) in Realzeit generiert wird (Kopp & Wachsmuth 2000).

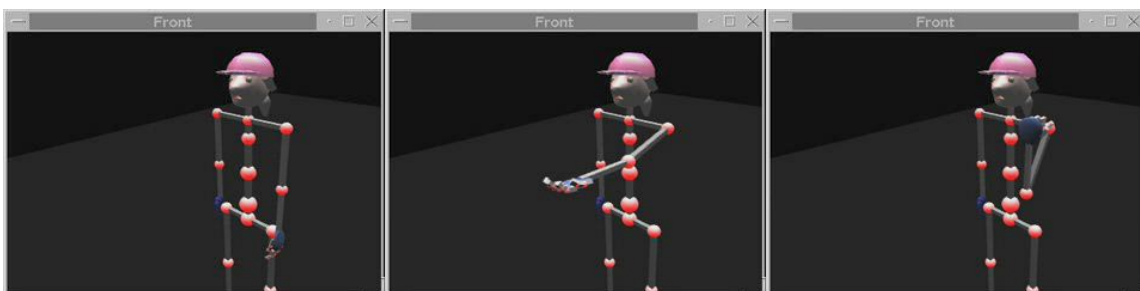


Abb. 6: "pull"-Geste eines artikulierte Kommunikators

Als ein Einsatzziel soll diese Figur als sog. artikulierter Kommunikator, gewissermaßen als Verkörperung des Systems, dem Benutzer bei virtuellen Konstruktionsabläufen assistieren und sich dabei multimodal äußern können (Kopp & Jung 2000).

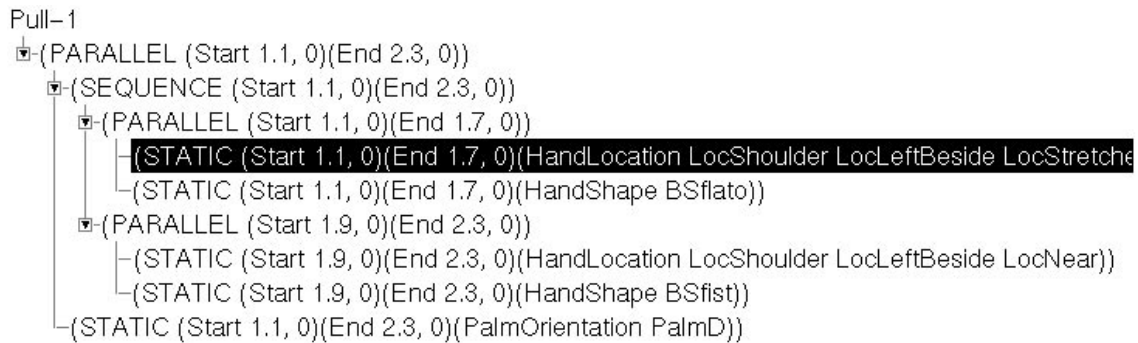


Abb. 7: Erweiterte HamNoSys-Beschreibung (hier in ASCII) der "pull"-Geste

3 Diskussion und Ausblick

Mit den hier beschriebenen Arbeiten wurden Grundlagentechniken für multimodale Interaktion entwickelt, die in folgender Hinsicht über bislang existierende Ansätze hinausgehen: Formen des natürlichen Zeigens und Vormachens können parallel zu gesprochenen Instruktionen echtzeitfähig zur Manipulation virtueller Objekte eingesetzt werden; die Interaktionen können lose gekoppelt (diskret) oder eng gekoppelt (kontinuierlich) erfolgen; die Generierung natürlicher Gestik durch einen synthetischen Agenten ist auf Basis "hochsprachlicher" Spezifikationen komfortabel möglich und zudem an Zeitvorgaben ausrichtbar, welche die Synchronisation mit Sprechausgaben vorbereiten. Da unsere Arbeiten auf die Erweiterung bisheriger Techniken abzielen, sind die implementierten Systeme Forschungsprototypen, deren Erfolgsmesspunkt die Erprobung durch Laborpersonal ist; die Systemperformanz ist für prototypische Beispiele wiederholbar und variierbar verlässlich. Weitergehende Evaluationen und Untersuchungen der Systemnützlichkeit sind derzeit noch nicht vorgesehen.

In unseren weiteren Arbeiten wollen wir Forschungsansätze aus den Bereichen Multimodale Interaktion und Virtuelles Konstruieren erweitern und derart zusammenführen, dass ihre realitätsnahe Erprobung auf einer generischen Handhabungsplattform demonstrierbar wird. Verstärkt sollen dabei zweihändige manipulative Gesten aufgegriffen werden, wobei auch Erkenntnisse aus den Humanwissenschaften über bimanuales Arbeiten (z.B. Guiard & Ferrand 1996) einbezogen werden sollen. Mit der Netzwerkverteilung und Installation der virtuellen Umgebung bei einem Kooperationspartner (GMD) soll zudem die Eignung der entwickelten Techniken für Anwendungen im Concurrent Engineering untersucht werden.

Die Integration einer virtuellen Figur in die virtuelle Umgebung lässt verschiedene Fortsetzungen der Arbeiten denkbar erscheinen. Einerseits kann die Figur das systemseitige Gegenüber des Benutzers verkörpern und mit ihm/ihr in beiderseitige multimodale Interaktion treten, als autonomer artikulierter Kommunikator. Andererseits erscheint es möglich, die körperlichen Äußerungen eines entfernten (menschlichen) Kooperationspartners in verteilter VR zu erfassen und aus entsprechenden Beschreibungen vor Ort mit dem artikulierten Kommunikator zu reproduzieren; die virtuelle Figur würde in diesem Fall den entfernten Partner als Avatar repräsentieren. Eine andere Fortsetzung finden unsere Arbeiten im neu angelaufenen DEIKON-Projekt ("Deixis in Konstruktionsdialogen"), in dem entwickelte Formen des Zeigens, die ikonische Anteile der Beschreibung von Formen und Orientierungen einschließen können – empirisch und in Simulationen mit einer virtuellen Figur – untersucht werden. Hierüber wird zu anderer Gelegenheit zu berichten sein.

Hinweise

Die Forschungsarbeiten im SGIM-Projekt wurden im Projektverbund "Multimedia NRW: Die Virtuelle Wissensfabrik" in Zusammenarbeit mit Partnerprojekten der Universität Bielefeld, des Europäischen Mechatronikzentrums Aachen und des Instituts für Medienkommunikation der GMD (GMD IMK-VMSD) Sankt Augustin durchgeführt und vom Land Nordrhein-Westfalen unterstützt. Die Arbeiten zum virtuellen Konstrukteur und im DEIKON-Projekt werden im Rahmen des Sonderforschungsbereichs 360 "Situierete Künstliche Kommunikatoren" von der Deutschen Forschungsgemeinschaft gefördert.

Literatur

- Barfield, W. & Furness, T.A. (Eds.) (1995): *Virtual Environments and Advanced Interface Design*, Oxford University Press.
- Böhm, K., Hübner, W. & Väänänen, K. (1992): GIVEN: Gesture driven interactions in virtual environments: A toolkit approach to 3D interactions. *Proc. Conf. Interface to Real and Virtual Worlds* (Montpellier, France, March 1992).
- Burdea, G. & Coiffet, P. (1994): *Virtual Reality Technology*. New York: Wiley.
- Cavazza, M., Pouteau, X. & Pernel, D. (1995): Multimodal communication in virtual environments. In Y. Anzai et al. (Eds.): *Symbiosis of Human and Artifact*, pp. 597-604. Elsevier Science B.V.
- Cruz-Neira, C., Sandin, D.J., DeFanti, T.A. (1993): Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE, *Computer Graphics Vol. 27* (Proc. SIGGRAPH 93), 135-142.
- Cutler, L. D., B. Fröhlich, P. Hanrahan (1997): Two-Handed Direct Manipulation on the Responsive Workbench. *Proceedings of the 1997 Symposium on Interactive 3D Graphics*. ACM.
- Dai, F. (1998): *Virtual Reality for Industrial Applications*. Berlin u.a.: Springer.
- Dai, F. & M. Göbel. Virtual Prototyping – An approach using VR-techniques (1994): In *Proc. of the 14th ASME Int. Computers in Engineering Conference*, Minneapolis, September 11-14.
- Drews, P. & M. Weyrich (1998): Interactive Functional Evaluation in Virtual Prototyping Illustrated by an Example of a Construction Machine Design. In *IECON'98 – Proceedings of the 24th Annual Conference of the IEEE Industrial Electronics Society, Vol 4*, IEEE, 2143-2145.
- Fellner, D.W. & Hopp, A. (1997): MRT-VR Multi-User Virtual Environment, *Proc. AAAI '97*.
- Fink, G.A., C. Schillo, F. Kummert & G. Sagerer (1998): Incremental speech recognition for multimodal interfaces. In *Proc. 24th Annual Conference of the IEEE Industrial Electronics Society*, 2012-2017.
- Fröhlich, M. & I. Wachsmuth (1998): Gesture recognition of the upper limbs - from signal to symbol. In I. Wachsmuth and M. Fröhlich (eds.): *Gesture and Sign Language in Human-Computer Interaction* (pp. 173-184). Berlin: Springer-Verlag (LNAI 1371).
- Gausemeier, J., M. Grafe & R. Wortmann (1998): Interactive Planning of Manufacturing Systems with Virtual Construction Sets. In *IECON'98 – Proceedings of the 24th Annual Conference of the IEEE Industrial Electronics Society, Vol 4*, IEEE, 2146-2151.
- Guiard, Y. & T. Ferrand (1996): Asymmetry in Bimanual Skills. In *Manual Asymmetries in Motor Performance*, CRC Press.
- Gupta, R., Whitney, D. and Zeltzer, D. (1997): Prototyping and Design for Assembly Analysis Using Multimodal Virtual Environments. *Computer-Aided Design*, 29(8): 585-597.
- Hauptmann, A.G. & P. McAvinney (1993): Gestures with speech for graphic manipulation. *International Journal of Man Machine Studies*, 38: 231-249.
- Jayaram, S., Connacher, H.I. & Lyons, K.W. (1997): Virtual Assembly Using Virtual Reality Techniques. *Computer-Aided Design*, 29(8): 575-584.
- Jörding, T. & I. Wachsmuth (1996): An Anthropomorphic Agent for the Use of Spatial Language, *Proceedings of ECAI'96-Workshop "Representation and Processing of Spatial Expressions"* (pp. 41-53), Budapest.
- Jung, B. & J.T. Milde (Hrsg.) (2000): *KI – Künstliche Intelligenz 2/00*, Themenheft zum Schwerpunkt "Intelligente virtuelle Umgebungen".

- Jung, B. & I. Wachsmuth (1998): Integration of Geometric and Conceptual Reasoning for Interacting with Virtual Environments. *Proc. AAAI '98 Spring Symposium on Multimodal Reasoning*, pp. 22-27.
- Jung, B., S. Kopp, M.E. Latoschik, T. Sowa & I. Wachsmuth (2000): Virtuelles Konstruieren mit Gestik und Sprache. *KI – Künstliche Intelligenz 2/00*, Themenheft zum Schwerpunkt "Intelligente virtuelle Umgebungen", 5-11.
- Kopp, S. & B. Jung (2000): An Anthropomorphic Assistant for Virtual Assembly: Max. In Working Notes Workshop "Communicative Agents in Intelligent Virtual Environments", Autonomous Agents 2000.
- Kopp, S. & I. Wachsmuth (2000): A Knowledge-based Approach for Lifelike Gesture Animation. In W. Horn (ed.): *ECAI 2000 - Proceedings of the 14th European Conference on Artificial Intelligence*, IOS Press, 663-667.
- Krüger, W., Bohn, C.A., Fröhlich, B., Schüth, H., Strauss, W., Wesche, G. (1995): "The Responsive Workbench: A Virtual Work Environment", in: *IEEE Computer*, Vol. 28, No. 7.
- Latoschik, M. E. (2000): Multimodale Interaktion in Virtueller Realität am Beispiel des virtuellen Konstruierens. Eingereichte Dissertationsschrift, Technische Fakultät/Universität Bielefeld, Okt. 2000.
- Latoschik, M. E., M. Fröhlich, B. Jung & I. Wachsmuth (1998): Utilize Speech and Gestures to Realize Natural Interaction in a Virtual Environment. *IECON'98 - Proceedings of the 24th Annual Conference of the IEEE Industrial Electronics Society*, Vol. 4, IEEE, 2028-2033.
- Latoschik, M. E., B. Jung & I. Wachsmuth (1999): Multimodale Interaktion mit einem System zur Virtuellen Konstruktion. In K. Beiersdörfer, G. Engels & W. Schäfer (Hrsg.): *Informatik '99, 29. Jahrestagung der Gesellschaft für Informatik*, Paderborn. Berlin: Springer-Verlag, 88-97.
- Lenzmann, B. & I. Wachsmuth (1997): Contract-Net-Based Learning in a User-Adaptive Interface Agency. In Gerhard Weiss (ed.): *Distributed Artificial Intelligence Meets Machine Learning. Learning in Multi-Agents Environments*, (pp. 202-222), Berlin: Springer-Verlag (LNAI 1221).
- Lucente, M., Zwart, G. & George, A. D. (1998): Visualization space: A testbed for deviceless multimodal user interface. In *Intelligent Environments Symposium*, American Assoc. for Artificial Intelligence Spring Symposium Series.
- Luck, M. & R. Aylett (Eds.) (2000): *Applied Artificial Intelligence 14(1)*, Special Issue "Intelligent Virtual Environments".
- Mapes, D.P. & J.M. Moshell (1995): A Two-Handed Interface for Object Manipulation in Virtual Environments. *Presence*, 4(4): 403-416.
- Metaxas, D. M. (1996): *Physics-Based Deformable Models. Applications to Computer Vision, Graphics and Medical Imaging*. Boston: Kluwer Academic Publishers.
- Milne, M.R. (1997): ISAAC: A Meta-CAD System for Virtual Environments. *Computer-Aided Design*, 29(8): 547-553.
- Mine, M., F.P. Brooks & C. Sequin (1997): Moving Objects in Space: Exploiting Proprioception in Virtual-Environment Interaction. *Proceedings SIGGRAPH'97*, ACM, 19-26.
- Prillwitz, S., R. Leven, H. Zienert, T. Hanke, and J. Henning (1989): HamNoSys Version 2.0: Hamburg Notation System for Sign Languages: An Introductory Guide. *International Studies on Sign Language and Communication of the Deaf*, Vol. 5. Hamburg: Signum Press.
- Shaw, C. & M. Green (1994): Two-Handed Polygonal Surface Design. *Proceedings of the ACM Symposium on User Interface Software and Technology*, 205-212.
- Sowa, T., M. Fröhlich & M. Latoschik (1999): Temporal Symbolic Integration Applied to a Multimodal System Using Gestures and Speech. In A. Braffort et al. (Eds.), *Gesture-based Communication in Human-Computer Interaction – Proceedings GW'99* (pp. 291-302). Berlin: Springer-Verlag (LNAI 1739).
- Sparrell, C. J. & Koons, D. B. (1994): Interpretation of coverbal depictive gestures. In *AAAI Spring Symposium Series* (pp. 8-12), Stanford University.
- Srihari, R.K. (1994): Computational models for integrating linguistic and visual information: A survey. *Artificial Intelligence Review*, 8: 349-369.
- van Dam, A. (1997): Post-WIMP User Interfaces, *Communications of the ACM*, Vol. 40 (2), 63-67.
- Wachsmuth, I. (1998): Das aktuelle Schlagwort: Virtuelle Realität, *KI – Künstliche Intelligenz 98/1*, 34.
- Wachsmuth, I. & Cao, Y. (1995): Interactive Graphics Design with Situated Agents. In W. Strasser & F. Wahl (eds.): *Graphics and Robotics* (pp. 73-85). Berlin: Springer-Verlag.
- Weimer, D. & S.K. Ganapathy (1989): A synthetic visual environment with hand gesturing and voice input. In *CHI 89 Conference Proceedings, Human Factors in Computing Systems* (pp. 235-240).