

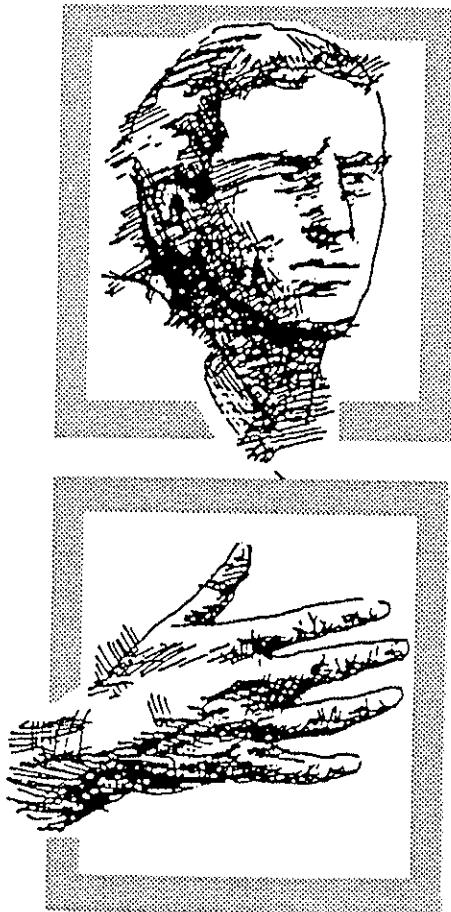
FORSCHUNGSNETZ

ANWENDUNGEN DER KI IN NRW

Ein anthropomorpher Interface-Agent für die Interaktion
mit einer virtuellen Umgebung

Tanja Jörding, Britta Lenzmann, Ipke Wachsmuth

Arbeitsgruppe Wissensbasierte Systeme
(Künstliche Intelligenz)
Technische Fakultät
Universität Bielefeld



Projekt
VIENA

KI-NRW 95 - 02

Ein anthropomorpher Interface-Agent für die Interaktion mit einer virtuellen Umgebung*

Tanja Jörding, Britta Lenzmann, Ipke Wachsmuth

Arbeitsgruppe Wissensbasierte Systeme
(Künstliche Intelligenz)
Technische Fakultät
Universität Bielefeld

1 Einleitung

Um die Bedienung komplexer technischer Systeme besser an menschengerechte Interaktionsformen anzupassen, werden vermehrt spezialisierte Software-Systeme eingesetzt, die als Mittler in der Kommunikation zwischen Menschen und technischen Systemen auftreten. Solche *Interface-Agenten* erleichtern und verbessern die Mensch-Maschine-Kommunikation z.B. dadurch, daß sie Wissen über die Tätigkeiten und Präferenzen ihrer Benutzer einbeziehen, um an deren Stelle Handlungen auszuführen [Maes et al., 1993].

In diesem Beitrag geht es um Interface-Agenten für die Interaktion mit einer virtuellen Umgebung. Ein erfolgreiches Interagieren mit solch dreidimensionalen Grafikwelten erfordert, daß Interface-Agenten sich auf implizite Annahmen und Voraussetzungen des Menschen beziehen können. Hierzu zählen u.a. Voraussetzungen wie die situierte Wahrnehmung und Handlung [Lobin, 1993] und Annahmen über die physikalische Modellierung der Szenenobjekte. Zudem sind bei der verbalen Interaktion kognitive Faktoren der Raumwahrnehmung einzubeziehen [Wachsmuth & Cao, 1995]. Um solchen Anforderungen gerecht zu werden, muß die analoge Grafikbeschreibung um symbolisches Wissen angereichert werden. Erst durch diese hybride Repräsentation wird der Interface-Agent in die Lage versetzt, gemäß den menschlichen Anforderungen zu handeln.

Im Bereich der Virtual Reality werden stark immersive Systeme entwickelt, wofür allerdings technisch aufwendiges Equipment, wie z.B. Datenhelm und Tracking-Systeme, benötigt wird. Hier hingegen wird ein durch natürliche Sprache bewegbarer Interface-Agent betrachtet mit dessen Hilfe sich der Designer "in der Szene einbringen" und aus wechselnden Perspektiven heraus handeln kann. Somit kann ein höheres Maß an "Situiertheit" in der Mensch-Maschine-Kommunikation ermöglicht werden, ohne daß gleich extreme Formen der Virtual Reality erforderlich sind.

Diese Aspekte werden nachfolgend anhand der Konzeption eines in der virtuellen Umgebung sichtbaren Interface-Agenten konkretisiert, der im Rahmen des Projektes VIENA (Abschnitt 2) entwickelt wird. Im Abschnitt 3 wird der anthropomorphe Agent vorgestellt und seine Funktionalitäten im virtuellen Raum beschrieben.

*Ausgearbeitete Version des Papiers zum Vortrag auf dem KI-95 Workshop 9: Hybride Systeme für die Mensch-Maschine-Kommunikation.

In Abschnitt 4 wird der Einsatz des Agenten und der damit verbundene Nutzen für die Bedienung des Grafiksystems diskutiert. Darunter fällt aus ergonomischer Sicht die Verbesserung der Explorationen im virtuellen Raum, die zum einen durch den Perspektivenwechsel auf die involvierte Sicht des Agenten erreicht wird, zum anderen durch die mögliche Bezugnahme auf anthropometrische Merkmale. Aus psychologischer Sicht wird der Agent als personifiziertes Gegenüber betrachtet, wodurch der Gebrauch natürlicher Sprache motiviert werden kann. Schwerpunktmaßig wird schließlich der erweiterte Gebrauch situierter Sprache beschrieben, wobei das Problem der verschiedenen Referenzsysteme betrachtet und im weiteren die mögliche Verwendung der Adverbien "hier" und "da" vorgestellt wird.

Abschließend enthält Abschnitt 5 eine Zusammenfassung über die vorgestellten Ergebnisse und ein Ausblick auf weitere Forschungsvorhaben.

2 Das Projekt VIENA

Die globale Zielsetzung im Projekt VIENA¹ ist die Entwicklung einer intelligenten Kommunikation mit einer virtuellen Umgebung im Bereich des interaktiven Grafikdesigns [Wachsmuth & Cao, 1995]. Als Beispielszenario dient eine Anwendung aus der Innenarchitektur. Statt der relativ abstrakten Eingabe mit Maus und Menüs können Manipulationen

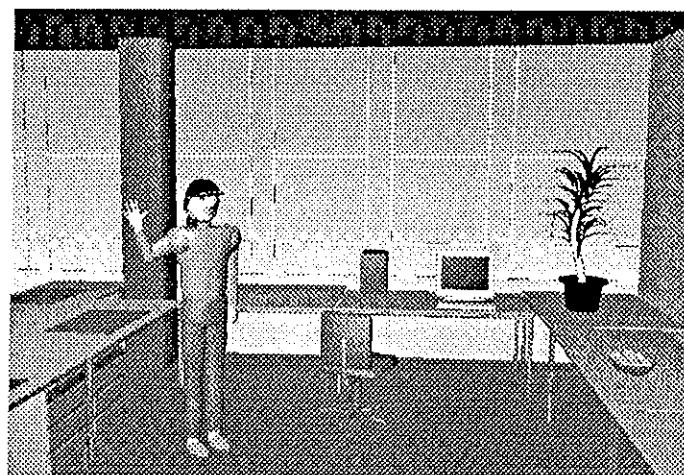


Abbildung 1: Der anthropomorphe Agent im virtuellen Büroraum

in der virtuellen Umgebung einfach durch verbale Anweisungen vorgenommen werden. Dabei übernimmt ein Agentensystem die Rolle eines intelligenten Mittlers, indem qualitative (auch unpräzise) verbale Eingaben in quantitative interne Kommandos übersetzt werden, mit denen die visualisierte Szene aktualisiert wird.

Im ersten Projektabschnitt wurde eine in der Szene nicht vergegenständlichte Interface-Agentur entwickelt. Darauf aufbauend wird in dem seit 1995 laufenden Projektabschnitt

¹VIENA ist ein Projekt im KI-NRW Forschungsschwerpunkt "Künstliche Intelligenz und Computergrafik" an der Universität Bielefeld und wird vom Ministerium für Wissenschaft und Forschung in Nordrhein-Westfalen unter der Förder-Nr. IVA3-107 007 93 unterstützt.

ein virtueller Interface-Agent (VIA) entwickelt, der in die künstliche Szene einbezogen ist. So können Explorationen und Manipulationen in der virtuellen Umgebung vorgenommen werden, die auf anthropomorphe Merkmale des Menschen bezugnehmen. Abbildung 1 zeigt den vergegenständlichten VIA in einer Beispielszene eines virtuellen Büroraums aus der VIENA-Testumgebung.

Erste Schritte in dieser Richtung wurden in Forschungsarbeiten mit computergrafischen Animationen humanoider Körper vorgenommen, die u.a. ergonomische Faktoren des Aufenthalts und der Bewegung von Menschen in engen Fahrerkabinen von Zweckfahrzeugen an CAD-Modellen dieser Fahrzeuge demonstrieren [Badler et al., 1991]. Diese Arbeiten beschränkten sich allerdings auf die rein bildhafte Darstellung, ohne daß weitergehende Eingriffe des menschlichen Betrachters in der virtuellen Umgebung möglich sind. Der VIA soll aber verbesserte bzw. erweiterte interaktive Manipulationen erlauben, indem er als personifiziertes Gegenüber dem Benutzer einerseits als Explorationshilfe im Raum dient, vor allem aber flexiblere Kommunikationsformen eröffnet, in denen verschiedene situierte Referenzsysteme berücksichtigt werden können.

3 Anthropomorpher Interface-Agent

Der virtuelle Agent ist mit Hilfe der Modellierungssoftware SOFTIMAGE entwickelt worden. Da wir keine animierten Bewegungen betrachten, war es nicht notwendig, dementsprechende funktionale Einzelheiten zu modellieren. Jedoch sollte ohne zu hohen Aufwand eine möglichst menschenähnliche Figur entstehen. Der Agent, der in unserem System *Hamilton* genannt wird, ist mit zwei Zeigearmen ausgerüstet, da eine wesentliche Aufgabe in der Einbringung von Zeigegesten in die natürlichsprachliche Interaktionen liegt. Zopf und Schirmmütze des Agenten sowie die Schriften auf dem Overall dienen zur besseren Unterscheidbarkeit von Vorder- und Rückseite. Ein weiterer Wunsch bei der Modellierung war, die Figur androgyn zu halten, um eine Identifikation durch Benutzerinnen und Benutzer gleichermaßen zu ermöglichen.

Der anthropomorphe Agent kann im virtuellen Raum in vielfältiger Weise agieren:

- **Translationen und Rotationen**

Der Agent kann als Ganzes im Raum verschoben und gedreht werden. Die Translationen sind wegen Berücksichtigung eines durch Schwerkraft gekennzeichneten Bezugssystems nur in der horizontalen Ebene möglich, die Rotationen nur um die vertikale Achse. Auf diese Weise bewegt sich der Agent analog zu menschlichen Bewegungsmöglichkeiten, und der Anwender kann durch ihn den virtuellen Raum von allen Standorten und Perspektiven aus wahrnehmen. Bei allen Transformationen wird die Undurchdringbarkeit von Gegenständen berücksichtigt (Kollisionskontrolle).

- **Änderung der Blickrichtung**

Zur weiteren Verbesserung der Wahrnehmungsfähigkeit kann der Kopf des Agenten um zwei Achsen gedreht werden, so daß er nach rechts, links, oben und unten sehen kann. Die maximalen Winkel liegen bei Rotationen um die vertikale Achse bei 45 Grad und bei Rotationen um die horizontale Achse bei 20 Grad. Diese Grenzwerte ergeben sich aus dem optischen Bild des Agenten (der Kopf sollte eine feste Verbindung zum Hals behalten und der Zopf darf nicht den Körper durchdringen).

Zudem besteht bei größeren Winkeln die Gefahr, daß der Anwender wegen des eingeschränkten Blickfeldes die Orientierung verliert. Wird der Kopf des Agenten durch eine Benutzeranweisung gedreht, so wird diese Änderung in der nächsten Visualisierung automatisch wieder rückgängig gemacht, um den Anwender ein explizites Zurücksetzen der Blickrichtung zu ersparen. Als zusätzlicher Vorteil dieser Realisierung ergibt sich, daß kein zusätzliches unabhängiges Referenzsystem des Kopfes berücksichtigt werden muß.

- **Zeigegesten**

Eine weitere Funktion des Agenten ist die Erleichterung der Kommunikation durch Zeigegesten. Von den vielen möglichen Arten der Körpersprache ist bei Hamilton eine Zeigegeste mit ausgestrecktem Arm und dem Zeigefinger als weitere Verstärkung realisiert (Abb. 2). Dieses ist im Vergleich zu anderen Körpergesten und zur Mimik ein sehr präziser Hinweis, der auch noch aus weiterer Entfernung verstanden werden kann. Die Arme des Agenten sind dazu um alle drei Achsen rotierbar, wodurch er (z.Z. nur mit dem rechten Arm) auf beliebige Positionen (Objekte) im Raum verweisen kann. Diese Gesten sind so implementiert, daß sich der Körper des Agenten

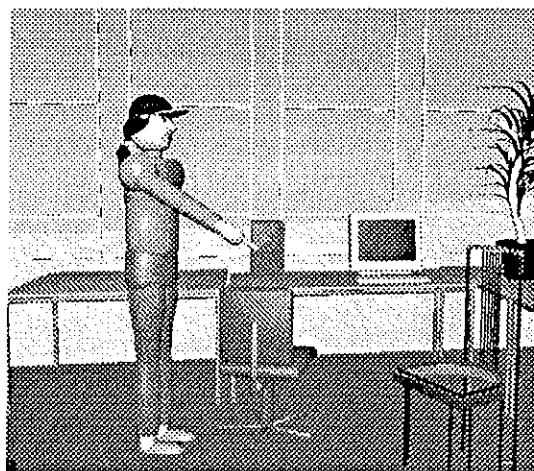


Abbildung 2: "Point to the designer-chair."

immer erst zu dem jeweiligen Objekt dreht. Daher ergeben sich keine Schwierigkeiten bezüglich Kollision zwischen Arm und Körper. Der Zeigearm wird in einer nachfolgenden Visualisierung ebenfalls automatisch wieder zurückgesetzt. Weitere richtungszeigende Gesten (z.B. Augenbewegungen) finden zur Zeit keine Anwendung, da sie für den Benutzer oft zu schwierig zu erfassen sind.

- **Größenveränderung**

Weitere Flexibilität bei Explorationen im Raum ist durch die Größenveränderung von Hamilton gegeben. Mit den Befehlen *be smaller/taller* kann der Agent um einen bestimmten Prozentwert wachsen oder schrumpfen.

- **Hello**

Zusätzlich zu den zwei Zeigearmen wurde noch ein Arm modelliert, der sich in Grüßstellung befindet. Dieser kann als Reaktion auf ein *Hello* vom Benutzer für eine vorgegebene Zeitspanne gegen den rechten Zeigearm ausgetauscht werden, um dem Benutzer vor dem Bildschirm zuzuwinken (Abb. 1). Dazu dreht sich Hamilton zu der Kamera und sieht den Benutzer an.

- **Perspektivenwechsel**

Mit Hilfe des Befehls *change the view* kann die Kamera nun geeignet im Kopf von Hamilton positioniert werden. Dabei wird der Schirm der Kappe am oberen Bildrand sichtbar, jedoch nur insoweit, daß nicht zu viel vom Sichtfeld verdeckt wird. Um den Wechsel des Standortes und der Orientierung darüber hinaus zu verdeutlichen, ist an der Kappe der Schriftzug "Hamilton View" angebracht, der nur bei Einnahme von Hamiltons Sicht zu erkennen ist.

Die folgende Tabelle liefert einen Überblick über die möglichen Anweisungen, mit denen der Anwender den anthropomorphen Agenten steuern kann.

Translationen im Raum	Go [a bit/a little bit] right/left Go [a bit/a little bit] forwards/backwards Go to the <object> Go to the left/right of the <object> Go to the front/back of the <object>
Rotationen des Körpers	Turn (a bit/a little bit) right/left Turn to the <object>
Veränderung der Blickrichtung	Look [a bit/a little bit] right/left Look [a bit/a little bit] up/down
Gesten ausführen	Point to the <object> Hello
Auf Gesten bezugnehmen	Go there Come here Move the <object> there
Größenveränderungen	Be smaller/taller
Perspektivenwechsel	Change the view

Die Implementierung dieser Befehle ist auf einfachen Grundfunktionen aufgebaut. Diese können auf verschiedene Weisen gekoppelt werden, wodurch sich eine Vielzahl von Realisierungsmöglichkeiten ergibt und das System somit leicht zu modifizieren bzw. zu erweitern ist.

4 Der Einsatz des anthropomorphen Agenten

Benutzer von interaktiven Grafiksystemen stammen häufig aus Anwendungsfeldern, in denen mehr Kreativität als technisches Verständnis gefordert ist. Durch das Sichtbarmachen eines mit menschenähnlichen Merkmalen ausgestatteten Agenten möchte man daher ein personalisiertes Gegenüber schaffen, das dem ungeübten Benutzer die Kommunikation mit

dem System erleichtern kann. Der anthropomorphe Agent ist der sichtbare Teil des Virtuellen Interface-Agenten (VIA). In ihm sieht der Anwender das Wissen des technischen Systems vertreten, so daß er in seiner Vorstellung mit einem menschlichen Gegenüber kommuniziert. Menschen neigen dazu, Gegenstände zu personifizieren. Das Erkennen eines menschenähnlichen Wesens kommt dem Benutzer daher entgegen und motiviert zur Verwendung der natürlichen Sprache [Laurel, 1990].

Im folgenden wird nun betrachtet, wie der anthropomorphe Agent dem Anwender zwei Sichten auf den virtuellen Raum ermöglicht und wie Explorationen damit in verschiedenen Formen durchgeführt werden können. Das Hauptaugenmerk richtet sich jedoch dann im zweiten Unterkapitel auf die erweiterten Möglichkeiten für den situierten Sprachgebrauch. Eine ausführlichere Darstellung findet man in [Jörding, 1995].

4.1 Unterstützung bei der Exploration

Die Existenz eines anthropomorphen Agenten ermöglicht erweiterte Formen der Explorationen im virtuellen Raum. Bisher konnte sich der Anwender durch Änderung der Position und des Point of Interest der Kamera durch das visualisierte Büro bewegen. Dabei ist die Änderung des Standortes nur durch die Undurchdringbarkeit von Objekten eingeschränkt. Im übrigen kann die Kamera in Form eines 1cm^3 -großen unsichtbaren Würfels frei im virtuellen Raum bewegt werden. Diese unabhängigen Kamerabewegungen sind bereits im ersten Projektabschnitt (1993/94) realisiert worden. Durch den Perspektivenwechsel hat der Anwender jetzt zwei Möglichkeiten, den Raum zu erkunden:

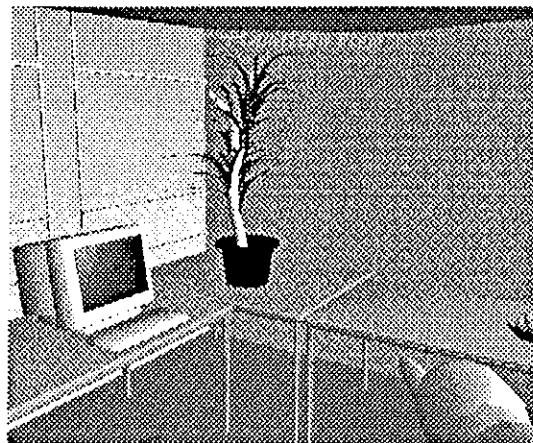


Abbildung 3: Sicht durch Hamiltons Augen bei der Ausführung einer Zeigegeste

- **Die involvierte Sicht:** Durch Hamilton steht eine körperliche Hülle zur Verfügung, in die der Benutzer schlüpfen kann. Auf diese Weise kann er virtuell mit den gewohnten menschlichen Bewegungsmöglichkeiten den Raum erkunden, was sich zum Beispiel in einer konstanten Blickhöhe bemerkbar macht (die allerdings durch "be smaller/taller" veränderbar ist). Der Anwender kann so stärker in den virtuellen Raum involviert werden und Objekte aus seiner individuellen Perspektive wahrnehmen. Die Sicht, die der Mensch im realen Leben auf Räume und Gegenstände hat,

kann so auf den virtuellen Raum übertragen werden. Abbildung 3 zeigt den Blick durch Hamiltons Augen bei der Ausführung einer Zeigegeste.

- **Die externe Sicht:** Betrachtet man auf der anderen Seite die externe Sicht auf den anthropomorphen Agenten, so hat man durch ihn eine Personifizierung eines allozentrischen, durch den Standort des Agenten variabel verankerten Bezugssystems. Der Benutzer kann Hamilton im Raum auf verschiedene Arten bewegen, wodurch er einen zweiten Standort und eine zweite Orientierung erhält, die er dann einnehmen kann. Durch diese externe Sicht auf den Agenten kann der Anwender Objekte bezogen auf anthropometrische Merkmale wahrnehmen. Ergonomische Gesichtspunkte, wie zum Beispiel die Höhe einer Arbeitsplatte, können durch den Vergleich mit dem anthropomorphen Agenten besser beurteilt werden.

Die Explorationen im Raum können auf zwei Arten durchgeführt werden: man kann von außen den menschenähnlichen Agenten sehen, zudem aber auch in dessen Körper schlüpfen und den virtuellen Raum von innen beurteilen.

Diese Erkundungen können durch eine Größenveränderung von Hamilton noch flexibler gestaltet werden. Der Designer hat dadurch die Möglichkeit, das Zimmer aus mehreren Blickhöhen zu betrachten, um so zum Beispiel auch die Perspektive eines Kindes berücksichtigen zu können. Einfach realisierbar wäre auch eine Anpassung an die individuelle Benutzergröße.

4.2 Die Verwendung situierter Sprache

Wenn man im täglichen Leben versucht, einen Weg zu erklären oder eine Zimmereinrichtung zu beschreiben, so entstehen bei Sprecher und Hörer oft Ausdrucks- bzw. Verständigungsschwierigkeiten. Da die Kommunikationspartner meist unterschiedliche Standpunkte und Orientierungen einnehmen, ist es möglich, räumliche Aussagen in verschiedenen Bezugssystemen auszudrücken. Dies führt zu einem Blickpunktproblem, d.h. die Gesprächspartner müssen Kenntnis von den möglicherweise unterschiedlichen Perspektiven haben [Herrman & Grabowski, 1987].

Eine weitere Schwierigkeit ist durch die strukturelle Verschiedenheit von Sprache und Raum gegeben. Der Raum ist dreidimensional, die Sprache jedoch beruht auf der Eindimensionalität der Zeit [Friederici, 1989]. Dies bedeutet, daß das Nebeneinander des Raumes auf das geordnete Nacheinander der Sprache abgebildet werden muß. Dadurch ergibt sich für den Sprecher ein Linearisierungsproblem [Levelt, 1986].

Wie also lösen wir diese Probleme in der täglichen Kommunikation? Besonders in der gesprochenen Sprache fällt zunächst der starke Kontextbezug auf. Hört man einzelne Sätze eines Gesprächs außerhalb ihres Kontextes, so hat man in den meisten Fällen große Verständnungsschwierigkeiten. Das begründet sich z.B. in der vielfältigen Verwendung deiktischer Verweise. Das Wort Deixis leitet sich von einem griechischen Wort ab, welches 'hinweisend' oder 'anzeigend' bedeutet. Man unterscheidet zwischen der Personaldeixis für die Identifikation der Kommunikationspartner (ich, du, ihr), der Zeitdeixis zur temporalen Einordnung (hier, jetzt, heute) und der Ortdeixis zur Lokalisation (da, hier, vorne, rechts von). Mit den deiktischen Verweisen kann eine Beziehung zwischen den Personen, Objekten, Ereignissen und dem augenblicklichen Raum-Zeit-Kontext hergestellt werden [Bühler, 1965].

Die direkte Kommunikation ist ein komplexer Vorgang, der nicht nur aus dem Sprechen und Hören besteht, sondern auch Augenkontakt, Mimik und Gestik beinhaltet. Diese Körpersprache kann eine Vielzahl von Empfindungen verraten, wie Interesse, Verwirrung, Zuneigung oder Abneigung. Mit Nicken, Schulterzucken, fragendem Blick oder zustimmenden Geräuschen signalisiert der Hörer sein Verständnis oder Unverständnis. Der Sprecher kann auf diese Hörersignale eingehen und so in seinen Erklärungen das passende Maß an Granularität erreichen [Dix et al., 1993]. In vielen Situationen sind Mimik und Gestik redundant, sie erleichtern jedoch durch die zusätzliche Betonung das Verständnis des Hörers und bringen Lebendigkeit in die Kommunikation. In anderen Fällen sind Gesten oder Blickrichtungen zum Verständnis unbedingt notwendig, da sie oft den Fokus der Konversation verdeutlichen. Dazu verwendet der Mensch meist ein Deuten mit dem Kopf, verbunden mit einem kurzen Blick, oder verschiedene Arten von Zeigegesten mit Arm und Hand, die entweder ungenau in eine Richtung weisen oder aber gezielt lokalisieren. Dennoch kommt es in alltäglichen Gesprächen durchaus oft zu Mißverständnissen. Im Vergleich zu anderen Kommunikationsformen können wir diese im direkten Gespräch jedoch meist schnell erkennen und durch flexible Sprecherwechsel auch unverzüglich korrigieren.

Als ein erster Vorstoß wird im weiteren nun die Ortdeixis betrachtet, die sich in dimensionale und positionale Deixis aufteilen lässt. Dabei wird das Ziel verfolgt, diese kontextbezogenen Kommunikationsformen durch den anthropomorphen Agenten in größerem Maße für die Mensch-Maschine-Kommunikation zu modellieren.

4.3 Dimensionale Deixis

Die dimensionale Deixis ist ein System aus den drei Paaren oben/unten, vorne/hinten und rechts/links, mit denen auf die drei Dimensionen des Raumes verwiesen wird. Die so aufgebaute räumliche Beziehung ist sowohl von der Position als auch von der Orientierung der Kommunikationspartner abhängig

Die Wahrnehmung der Dreidimensionalität wird durch physikalische und biologische Faktoren bestimmt. Die vertikale Achse ist durch die Gravitation und durch die Asymmetrie des menschlichen Körpers gegeben, die vorne/hinten-Achse durch die Ausrichtung unserer Wahrnehmungsorgane und ebenfalls durch die Asymmetrie des menschlichen Äußeren. Verwechslungen treten hingegen häufig bei der zweiten Horizontalen auf, da der Mensch bezogen auf die rechts/links-Achse äußerlich fast symmetrisch ist.

Referenzsysteme

Die eindeutige Interpretation räumlicher Ausdrücke erfordert die Festlegung eines Referenzrahmens. Bezogen auf diese implizit oder explizit festgelegte Orientierung können die Positionen und Richtungen, die vom Sprecher benutzt werden, auch vom Hörer eingeordnet und verstanden werden. Als Referenzrahmen stehen dem Sprachverwender mehrere Bezugssysteme zur Verfügung: die Koordinaten der eigenen Person, die des Hörers, die von dritten Personen oder die der physikalischen Umwelt [Friederici, 1989].

In der Literatur findet man mehrere Unterteilungen, die verschiedene Aspekte dieser Referenzarten unterschiedlich stark betonen. Von der *intrinsischen* Perspektive spricht man, wenn räumliche Ausdrücke relativ zur Orientierung des Bezugsobjektes verhandelt werden. In Abb. 4 liegt der Ball, bezogen auf die Vorderseite des Schreibtisches, an seiner rechten Seite. Die Verwendung der intrinsischen Perspektive erfordert jedoch Bezugsobjek-

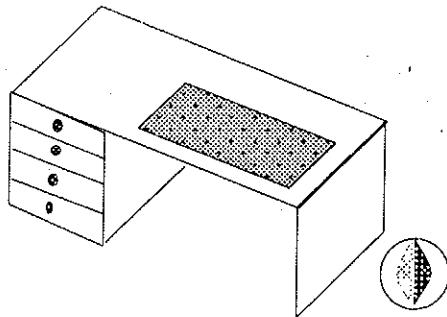


Abbildung 4: Der Ball liegt rechts vom Schreibtisch

te, die eine ausgezeichnete Ausrichtung haben. Diese Ausrichtung erhalten Objekte durch die Art, wie sie üblicherweise verwendet werden.

Wenn ein Bezugsobjekt keine intrinsische Orientierung hat oder diese nicht benutzt werden soll, entscheidet der situative Kontext über den Referenzrahmen, und man spricht dann von der *extrinsischen* Perspektive [Retz-Schmidt, 1988]. Diese kann zum Beispiel durch die aktuelle Bewegungsrichtung des Objektes gegeben sein oder oft auch durch die Gravitation. In Abb. 5 wird durch die Fahrtrichtung des Autos die Verwendung von *vor* motiviert. Die dritte Möglichkeit für den Referenzrahmen ist durch die *deiktische*

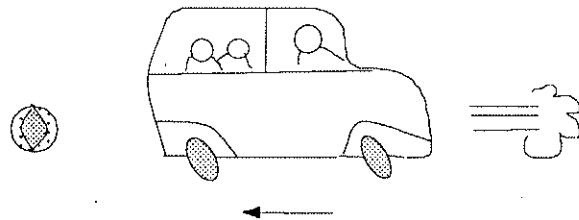


Abbildung 5: "Der Ball rollt *vor* das Auto (welches gerade rückwärts fährt)"

Perspektive gegeben. Dabei wird der Ursprung des Koordinatensystems durch die Position des Menschen festgelegt und die drei Achsenrichtungen durch die Art der menschlichen Raumwahrnehmung. In Abb. 6 stehen zwei Menschen in unterschiedlicher Position zu zwei Objekten, wodurch sich aus ihrer jeweiligen Sicht verschiedene Beschreibungen des Ballortes ergeben.

Zusätzlich zu dem Standpunkt des Menschen sind auch noch durch Augen, Kopf und Oberkörper eigene Referenzsysteme gegeben, da diese Körperbereiche in eine jeweils eigene Richtung gedreht sein können. Im allgemeinen sind jedoch die Standpunktkoordinaten für die Art der Referenzierung maßgebend [Bühler, 1965]. Deiktische Referenzsysteme können beim Sprecher, Hörer oder bei dritten Personen verankert sein (Abb. 6).

Diese Art der Einteilung findet man sehr häufig in verschiedenen Literaturquellen ([Ehrich, 1992], [Sichelschmidt, 1989] und [Retz-Schmidt, 1988]), wobei zum Teil nur die intrinsische und die deiktische Perspektive betrachtet werden, da diese im Sprachgebrauch

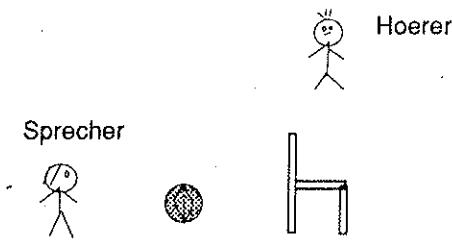


Abbildung 6: Aus meiner Sicht liegt der Ball *vor* dem Stuhl, von dir aus liegt er *rechts* vom Stuhl

überwiegen.

Eine andere Art der Unterteilung, die sich auf den Ort des Referenzrahmens bezieht, wird in [Rauh, 1983] beschrieben: das *egozentrische* Bezugssystem, definiert durch die drei Körperachsen des Sprechers und das *allozentrische* Bezugssystem, welches ebenfalls durch drei senkrecht aufeinanderstehende Achsen gegeben ist, die jedoch unabhängig vom Sprecher sind. In [Wachsmuth & Cao, 1995] wird darüber hinaus beschrieben, wie diese Systeme in virtuellen Umgebungen übernommen werden können. Verglichen mit den oben eingeführten Perspektiven, findet man das egozentrische Bezugssystem in der sprecherdeiktischen Perspektive wieder und das allozentrische System in allen übrigen Orientierungen.

Eine Einordnung, die sich auf das benötigte Wissen bezieht, findet man in [Saile, 1984]. Er unterscheidet zwischen der *objektiven* und der *subjektiven* Interpretation. Unter die objektive Interpretation fällt die Benutzung der intrinsischen Perspektive. Für diese spielen die Standorte des Sprechers und des Hörers keine Rolle, dafür benötigt man jedoch Wissen über das Objekt, wie zum Beispiel die Kenntnis über die ausgezeichnete Vorderseite beim Schreibtisch [Lang, 1989]. Die subjektive Interpretation wird bei dem deiktischen Referenzrahmen benötigt. Hier werden räumliche Ausdrücke bezogen auf den Sprecher oder auf die eigene Person interpretiert. Dazu benötigt man Wissen über Orte und Perspektiven der betreffenden Personen. Wenn der Sprecher einen Raum beschreibt, ohne im Sichtbereich des Hörers zu sein, so kann dieser die Beschreibungen aus sprecherbezogener deiktischer Sicht nur verstehen, wenn er Kenntnis über die Position und Orientierung des Sprechers hat.

In Abb. 7 wird versucht, über die vorgestellten Einteilungsformen eine Übersicht zu geben. Die verschiedenen Referenzsystemarten sind dazu als Mengen dargestellt, um Zugehörigkeiten zu verdeutlichen.

In der natürlichen Sprache kann das verwendete Referenzsystem explizit zum Ausdruck gebracht werden (*von dir aus gesehen ...*). Von dieser Möglichkeit wird jedoch meist nur Gebrauch gemacht, wenn den Kommunikationspartnern die Mehrdeutigkeit bewußt ist. Das kann zum Beispiel geschehen, wenn sie sich spiegelbildlich gegenüber stehen, so daß eine rechts/links-Verwechslung naheliegt.

Häufig hilft jedoch das Wissen über die jeweilige Situation, räumliche Beschreibungen

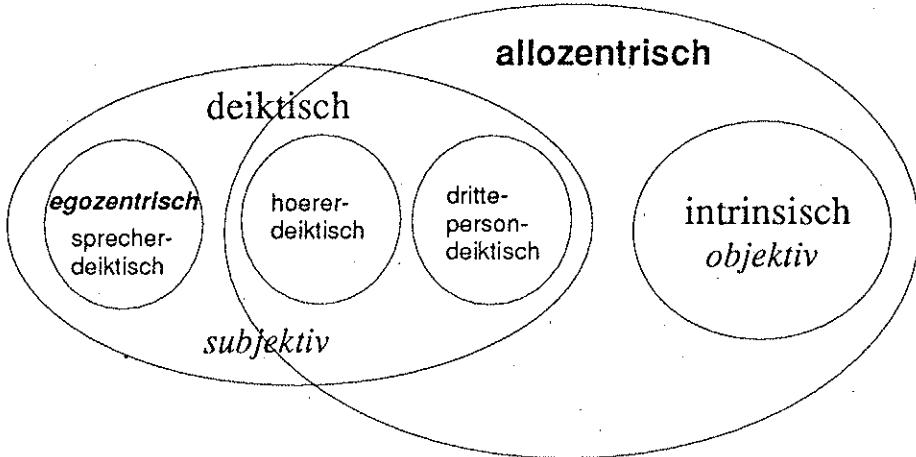


Abbildung 7: Zusammenhang zwischen den unterschiedlichen Einteilungsformen der Referenzsysteme

richtig zu verstehen. Wenn bei *schiebe die Pflanze nach rechts* nur auf einer Seite genügend Platz vorhanden ist, treten Mißverständnisse selten auf, da die andere Seite erst gar nicht in Betracht gezogen wird. Kommt es dennoch zu Verständnisschwierigkeiten, so werden diese in der direkten Kommunikation meist umgehend erkannt und können durch den schnell möglichen Sprecherwechsel behoben werden.

Wie aber löst man dieses Problem in einem interaktiven Grafiksystem? Im folgenden werden nun die Referenzsysteme und deren Einsatz im virtuellen Raum beschrieben.

Verwendung im VIENA-System

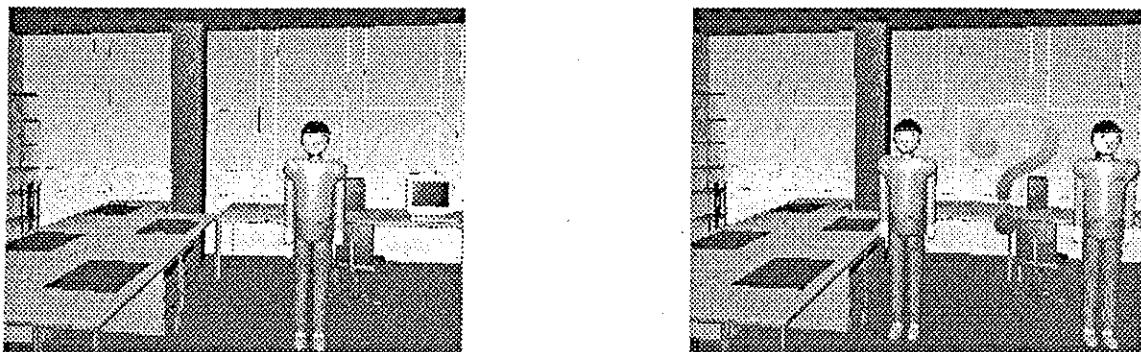


Abbildung 8: "Go left"

In dem VIENA-System können verschiedene Bezugssysteme verwendet werden. Bewegungen des Hamilton-Agenten können sowohl durch dessen eigene Sicht spezifiziert sein, als auch durch die Blickrichtung des Benutzers. Diese verschiedenen Referenzsysteme beeinflussen vorwiegend die Translationen im Raum. Am ausgeprägtesten tritt dieses zutage,

wenn der virtuelle Agent dem Benutzer gegenüber steht. Abb. 8 zeigt für diesen Fall die zwei Realisierungsmöglichkeiten auf die Anweisung *go left*. Nimmt der Benutzer mit Hilfe des Befehls *change the view* dagegen Hamiltons Blickrichtung ein, so fallen die beiden Referenzsysteme zusammen. Bei Objektransformationen ergeben sich bei Gegenständen mit

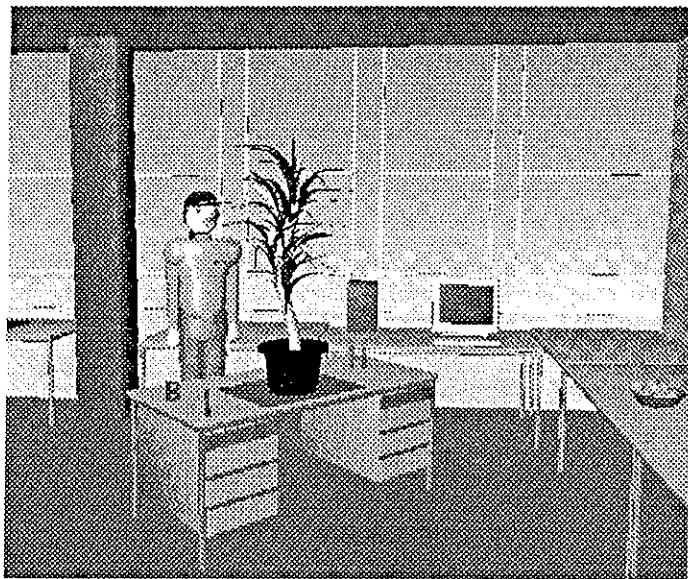


Abbildung 9: "Move the palmtree to the left." I: intrinsisch, B: Benutzer-deiktisch, H: Hamilton-deiktisch

ausgezeichneten Seiten drei mögliche Bezugssysteme: das intrinsische, welches Verschiebungen bezogen auf die Ausrichtung des Referenzobjektes durchführt, das Hamilton-deiktische und das benutzer-deiktische. In Abb. 9 sind die Auswirkungen der verschiedenen Referenzsysteme in einer Beispielanweisung (*move the palmtree to the left*) eingezeichnet. In dem intrinsischen Koordinatensystem würde die Schale auf die Position I geschoben, aus Hamiltons Sicht auf die Position H und aus Sicht des externen Benutzers auf die Position B.

Empirische Untersuchung

Die Auswahl des jeweils gültigen Referenzsystems ist von sehr vielen unterschiedlichen Faktoren abhängig. Darunter fallen sowohl geometrische Aspekte, wie Abstand und Winkel zwischen Benutzer, Hamilton und Objekt, optische Faktoren, wie Sichtbarkeit und Zugänglichkeit, als auch der situative Kontext, durch den das Anwendungsgebiet oder auch die soziale Situation bestimmt ist. Um zunächst einen Eindruck zu erhalten über die Verwendung verschiedener Referenzsysteme, wurde eine einfache Untersuchung mit 62 zufällig ausgewählten Versuchspersonen durchgeführt (Abb. 10). Dazu mußten sich die Versuchspersonen einzeln im Abstand von ca. 2 Metern und einem Winkel von 45 Grad zu einem Schreibtisch stellen. In der Mitte des Schreibtisches stand eine Kaffeekanne. Die Aufgabe bestand darin, auf die Anweisung *Schiebe die Kaffeekanne nach links* die Kanne auf dem Schreibtisch zu einer neuen Position zu bewegen.

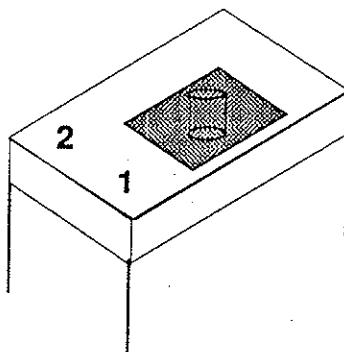


Abbildung 10: "Schematische Skizze der Versuchsanordnung"

Position 1 (intrinsisch)	Position 2 (deiktisch)
22 Vpn (35.5 %)	40 Vpn (64.5 %)

Für dieses Versuchsszenario war die Auswahl des Referenzsystems durch die Versuchsperson also keineswegs eindeutig, sondern signifikant von individuellen Präferenzen getragen. Weiterhin interessant war die Beobachtung, daß unter den Versuchspersonen, die die intrinsische Perspektive bevorzugten, weitaus mehr Frauen als Männer zu finden waren (Dieses Ungleichgewicht fiel erst im Laufe der Untersuchung auf und wurde daher nicht als Datum erhoben). Für das Anwendungsfeld des virtuellen Designs zog man den Schluß, daß beide Referenzsysteme relevant sind und in den Interaktionen berücksichtigt werden müssen. Zur Frage der Bevorzugung von Referenzsystemen wurde deshalb zunächst relevante Fachliteratur ausgewertet.

Betrachtet man die darin enthaltenen Experimente, so stellt man fest, daß die Aussagen sehr vage sind und sich zum Teil widersprechen. [Miller & Johnson-Laird, 1976] postulieren, daß die intrinsische Perspektive der deiktischen überlegen ist. [Ehrich, 1992] erhält bei Raumbeschreibungen überwiegend die Verwendung des deiktischen Bezugssystems. [Wunderlich, 1981] beobachtet bei statischen Situationen eine Mehrheit für die intrinsische Perspektive, in unserer eigenen Untersuchung wird jedoch die deiktische Sicht bevorzugt. Die Wahl des gültigen Referenzsystems scheint daher sehr stark vom situativen Kontext abhängig zu sein. Dies führt dazu, daß man eine Vielzahl von Einflußfaktoren berücksichtigen müßte, um exakte und aussagekräftige Ergebnisse zu erhalten. Die persönliche Wahrnehmung der jeweiligen Situation und der individuelle Sprachgebrauch spielen eine entscheidene Rolle. Die Daten, die man erhalten würde, sind daher nur in der jeweiligen Versuchssituation aussagekräftig. Eine Verallgemeinerung, die man bei einer Regelextraktion vornehmen würde, wäre vermutlich wenig aussagekräftig. Aus diesem Grund wurde davon abgesehen, einen Entscheidungsprozeß durch Regelwissen herbeizuführen.

Realisierung

Sowohl in der eigenen Untersuchung, als auch bei Wunderlich erkennt man, daß die Entscheidung für die deiktische oder intrinsische Sicht selten mit überwiegender Mehrheit geschieht. Beide Bezugssysteme sind also relevant. Zudem liefert die Untersuchung von

[Graf & Herrmann, 1989] noch Hinweise für die Hörerabhängigkeit. Um den individuellen Präferenzen des Benutzers bei der Wahl des Referenzsystems Rechnung zu tragen, benötigt man ein System, welches die verschiedenen Sichten flexibel verwenden kann. In VIENA werden daher alle drei Referenzsysteme (sprecher-deiktisch, hörer-deiktisch und intrinsisch) berücksichtigt und in Form von verschiedenen Agenteninstanzen dem Benutzer zur Verfügung gestellt.

Auf diese Weise kann der Benutzer nach einer durchgeführten Transformation das System mit der Anweisung "wrong" korrigieren und erhält daraufhin einen Realisierungsvorschlag, der ein alternatives Referenzsystem verwendet. Darauf aufbauend wird in aktuellen Arbeiten des Projektes an einem adaptiven Verhalten des Systems gearbeitet [Lenzmann & Wachsmuth, 1995]. Auf eine Anweisung, wie zum Beispiel *move the palm-tree left* wird zu Beginn der Interaktionen das sprecher-deiktische Bezugssystem verwendet. Stimmt diese Transformation nicht mit den Vorstellungen des Anwenders überein, so kann er das System korrigieren und erhält dann einen anderen Realisierungsvorschlag. Die Agenteninstanzen führen sogenannte credit values mit, die eine Bewertung für ihren Zuverlässigkeitgrad darstellen. Entsprechend der jeweilig durchgeführten Transformation werden die credits der verschiedenen Agenten erhöht oder erniedrigt. Bei einer neuen Benutzeranweisung erhält dann der Agent mit dem höchsten credit value den Auftrag. Auf diese Weise erfolgt während einer Session eine Adaption an die Präferenzen des jeweiligen Anwenders. Die zur Zeit noch relativ einfachen credit-Funktionen sollen in zukünftigen Arbeiten weiter ausgebaut werden, zum Beispiel dadurch, daß das System über verschiedene Auswahlkriterien Statistiken erstellt.

Ein weiterer Vorteil der Agenteninstanzen liegt in der Verbesserung der Handlungsfähigkeit des Systems. Falls der zuerst ausgewählte Agent eine Transformation nicht ausführen kann, so könnte er der anderen Instanz des Agenten seinen Auftrag übergeben. Durch diese Kommunikation der Agenten untereinander könnte bei der Wahl des gültigen Referenzsystems räumliches Wissen berücksichtigt werden. Ist auf einer Seite nicht genügend Platz vorhanden, so würde auf diese Weise dem Anwender automatisch die andere Seite vorgeschlagen und eine entsprechende Lösung vorgelegt.

Anhaltspunkte für die Auswahl eines Referenzsystems können in Zukunft auch durch die Gestenerkennung gegeben sein. Wenn der Anwender in der Situation der Abb. 9 zusätzlich zu der Anweisung *move the palm-tree left* eine Geste nach rechts ausführt, so könnte sich seine Äußerung auf das Hamilton-deiktische Referenzsystem beziehen. In Situationen dieser Art können Ambiguitäten somit vorzeitig durch die Gestik des Benutzers aufgelöst werden.

4.4 Positionale Deixis

Die positionale Raumdeixis ist im Deutschen ein dreigliedriges System, gegeben durch *hier*, *da* und *dort* [Ehrich, 1992]. Durch diese Adverbien werden Regionen im Raum bestimmt, wobei sich die Lokalisation fast immer auf die Koordinaten des Sprechers bezieht. Die Worte *hier*, *da* und *dort* können in sehr vielfältiger und komplexer Weise verwendet werden, wodurch mehrere Probleme bei der Interpretation der angegebenen räumlichen Bereiche entstehen [Klein, 1978]. Als erstes muß das gültige Bezugssystem herausgefunden werden. Dieses kann dadurch erschwert werden, daß Sprecherort und -zeit und Hörerort und -zeit nicht übereinstimmen. Durch Zeigegesten oder sprachliche Formulierungen kann der Koordinatenursprung verschoben werden, es kann auf abstrakte Räume verwiesen wer-

den, und die Bereiche können unterschiedlich weit ausgedehnt sein. Zusätzlich können sich die Bereiche überlappen. In manchen Situationen ist nur eines der Adverbien möglich, in anderen können sie gleichwertig benutzt werden [Klein, 1978]. Bei [Bühler, 1965] findet man folgende Charakterisierung:

- *Hier* bezeichnet eine den Sprechort umgebene Region.
- *Da* bezeichnet eine den Sprechort ausschließende Region, die im Interaktionsbereich des Sprechers liegt.
- *Dort* bezeichnet eine den Sprechort ausschließende Region, die jedoch außerhalb des Interaktionsbereiches liegt.

Interpretationsmöglichkeiten für das VIENA-System

In unserem Szenario beschränkt sich die Kommunikation auf Anweisungen, die Ortsveränderungen von Objekten zur Folge haben sollen. Beim Verschieben von Einrichtungsgegenständen wird vorwiegend auf bestimmte Regionen in einem speziellen Zimmer verwiesen, meist sogar nur auf Positionen auf dem Boden oder auf Einrichtungsgegenstände. Positionen außerhalb des Raumes oder Verweise in abstrakte Räume finden in der VIENA-Testumgebung keine Anwendung. Dieser eng gefaßte Kontext des virtuellen Designs schränkt somit die Verwendungsmöglichkeiten für die positionale Deixis ein und ermöglicht überhaupt erst eine Interpretation der angesprochenen Raumbereiche. Die Verwendung von *dort* ist in einem eng begrenzten Raum selten. Im folgenden werden für das VIENA-System daher nur Interpretationsmöglichkeiten für die *hier*- und *da*-Bereiche aufgeführt². Diese sind als Vorschläge des Systems zu verstehen, die vom Anwender abgelehnt bzw. korrigiert werden können.

“Stelle den Stuhl hierhin!”

- Aus Sicht des Benutzers entspricht das Zentrum des *hier*-Bereiches üblicherweise seinem Standpunkt. In dem VIENA-System hat der Benutzer jedoch die Möglichkeit, zwischen verschiedenen Perspektiven zu wechseln. Er hat eine externe Sicht auf den Raum, die durch eine virtuelle Kamera gegeben ist, und er kann die Sicht des anthropomorphen Agenten einnehmen. Dies führt dazu, daß nicht eindeutig bestimmt ist, wo der Anwender seine Position in der virtuellen Umgebung empfindet. Hat er die externe Sicht auf den Raum und den Agenten, so ist sein Standpunkt formal durch die Kameraposition gegeben. Auf der anderen Seite ist es jedoch auch denkbar, daß sich der Anwender mit dem anthropomorphen Agenten identifiziert. Somit kann der *hier*-Bereich je nach Wahl des gültigen Bezugssystems durch den Positionsvektor der Kamera bzw. des anthropomorphen Agenten gegeben sein. Nimmt der Anwender die Sicht des Agenten ein, so fallen die beiden Bereiche zusammen. Im Kontext des oben aufgeführten Satzes bedeutet es, daß der Stuhl entweder in die Nähe des anthropomorphen Agenten gestellt wird oder aber in unmittelbarer Nähe der virtuellen Kamera, die die Position des externen Betrachters repräsentiert (d.h. vorne im Bild).

²Diese Vereinfachung ergibt sich für VIENA ohnehin durch den gegenwärtigen Gebrauch der englischen Sprache, da hier die positionale Raumdeixis nur zweigliedrig ist (*here* und *there*).

- Zusätzlich besteht die Möglichkeit, diesen Bereich durch eine Zeigegeste zu verlagern, wobei der Ort von *hier* dann in der Verlängerung des Zeigearms zu erwarten ist (“Ein verschobenes *hier* ersetzt ein unverschobenes *da*” [Ehrich, 1992]).

“Stelle den Schreibtisch dahin!”

- Im täglichen Leben werden derartige Aussagen häufig durch ein vielfältiges Repertoire von Gesten unterstützt, wie zum Beispiel ein Deuten mit dem Kopf oder Zeigegesten mit Arm und Hand. In unserem System haben wir die Möglichkeit, den anthropomorphen Agenten mit seinem (rechten) Arm auf Objekte zeigen zu lassen. Folgt darauf dann ein Befehl mit der Ortsangabe *da*, so kann man diesen Ort wiederum in der Verlängerung des Zeigearms erwarten. Eine weitere Flexibilität könnte dem Benutzer in Zukunft durch einen mausgesteuerten Zeigearm geboten werden. Zur Zeit werden die Zeigehandlungen durch eine sprachliche Eingabe ausgelöst, so daß die Vorteile der Gesten, wie das Vermitteln räumlicher Informationen, nicht ausreichend genutzt werden können. Könnte man jedoch die Gestik unabhängig von der sprachlichen Aussage verwenden, so würde dem Benutzer das schwierige und unnatürliche Beschreiben von Positionen im dreidimensionalen Raum erspart.
- Wenn keine Zeigegeste vorangegangen ist und der anthropomorphe Agent für den Benutzer sichtbar ist, so kann sich *da* auf die Position von Hamilton beziehen (“Sprecher-da ist oft Hörer-hier” [Bühler, 1965]). Diese Situation kann man sich anhand folgender Befehlssequenz vorstellen:

„Hamilton, gehe nach rechts,
ein bißchen weiter,
stelle den Schreibtisch dahin.“

- Eine dritte Möglichkeit kann durch die Blickrichtung des Benutzers gegeben sein, wenn er die Sicht des anthropomorphen Agenten einnimmt und auf diese Weise sich virtuell in der visualisierten Szene bewegt. Statt eine Kamera zu dirigieren, nimmt der Benutzer den virtuellen Raum aus einer anthropomorphen Hülle wahr.

„Hamilton, gehe ein wenig zurück,
schau nach links,
stell den Schreibtisch dahin.“

Diese Interpretation von *da* spezifiziert die Position für das Objekt jedoch nur sehr ungenau, da sich unter Umständen ein großer Teil des Raumes im Sichtbereich des Benutzers befinden kann. In der Realität würde dieser Bereich durch ein kurzes Nicken oder Deuten mit Kopf oder Augen eingeschränkt. Diese feinen Gesten des Menschen sind jedoch nur sehr schwer auf ein Grafiksystem zu übertragen, da sie nicht realistisch genug modelliert werden können (z.B. Augenbewegungen) und daher für den Benutzer nur schwer zu erkennen sind.

Die hier aufgeführten Möglichkeiten geben Position an, an die sich die Lokalisation der Adverbien richten kann. Offen bleibt hier noch, wie ausgedehnt die angegebenen Bereiche sein können und wo genau die Objekte positioniert werden. Da dieses wiederum von vielen Faktoren (z.B. Objektgröße, Sichtbarkeit) abhängen kann, auf der anderen Seite durch bereits belegten Raum die Wahl der Zielpositionen eingeschränkt ist, wird es hier nicht näher betrachtet.

Realisierung

Das Problem bei der Realisierung von Befehlen, wie *Go there* oder *Come here* bestand darin, daß benötigtes Wissen auf mehrere Agenten verteilt ist. Über vorangegangene Zeigegesten kann nur der Buchhalter Auskunft geben, räumliches Wissen ist hingegen nur in den Raum- und Hamilton-Agenten vorhanden.

Wie bereits zu Beginn beschrieben, liegt der Hamilton-Agent in zwei Instanzen vor: eine führt Transformationen aus der Sicht des Benutzers durch, die andere verwendet die Sicht des anthropomorphen Agenten. Diese zwei Agenten stellen bei der Anweisung *Come here* jeweils eine Anfrage an den Buchhalter. Dieser gibt dem Agenten bei einer vorangegangenen Zeigegeste die Daten des Objektes, auf das gezeigt wurde. Die Agenten realisieren dann die Befehlsform *Go to the < object >* mit den übergebenen Daten. *Hier* und *da* werden somit derzeit als Nullobjekte betrachtet, die Orte im Raum repräsentieren, aber keine Ausdehnung haben. Liegt keine Zeigegeste vor, so bekommt der benutzer-deiktische Agent für die Lokalisation von *hier* die Daten der Kamera, der Hamilton-deiktische Agent erhält die Daten seiner eigenen Position.

Bei Befehlen, die *da* verwenden, wird vom Buchhalter ebenfalls zunächst überprüft, ob eine Zeigegeste vorangegangen ist. Ist dies nicht geschehen, so übergibt der Buchhalter zur räumlichen Bestimmung von *da* die Daten von Hamilton. Aus Sicht des Hamilton-deiktischen Agenten ist somit *da* nur durch eine Zeigegeste spezifiziert. Fehlt diese, so stellt er eine Nachfrage an den Benutzer: "Wo ist *da*"? Die Adverbien *hier* und *da* sind im System quasi Objektvariablen, die zur Zeit mit *hamilton*, der *camera* oder mit *referenzierten Objekten* gefüllt werden können. Die genaue Positionierung der Objekte ist durch die verwendete Funktion des Raumagenten determiniert und kann in [Otto, 1994] nachgelesen werden.

Der Buchhalter hat somit Funktionen (*hier_fkt()*, *da_fkt()*), die als Argumente die Perspektive der nachfragenden Agenteninstanz (benutzer-deiktisch oder Hamilton-deiktisch) und das Prädikat *geste* erhalten, welches angibt, ob in der vorangegangenen Interaktion eine Zeigegeste durchgeführt wurde. Ergebnis der Funktionen ist ein Objekt, welches in dem angegebenen Kontext den Bezug für *hier* bzw. *da* darstellt.

```
hier_fkt (perspektive, geste)
= referenziertes_Objekt, if geste = TRUE;
  camera,
    if geste = FALSE und
      perspektive = benutzer-deiktisch;
  hamilton,
    if geste = FALSE und
      perspektive = Hamilton-deiktisch;

da_fkt (perspektive, geste)
= referenziertes_Objekt, if geste = TRUE;
  hamilton,
    if geste = FALSE und
      perspektive = benutzer-deiktisch;
  unbestimmt,
    if geste = FALSE und
      perspektive = Hamilton-deiktisch;
```

Analog zu der hier vorgestellten Realisierung der Hamilton-Agenten kann eine Erweiterung der Raumagenten erfolgen, um Benutzeranweisungen der Art *put the bowl here* und

move the chair there zu ermöglichen.

5 Zusammenfassung und Ausblick

In dieser Arbeit wurde ein anthropomorpher Agent für ein interaktives Grafiksystem mit dem Ziel entwickelt, die Bedienung des Systems auch für Menschen mit geringer technischer Vorbildung zu vereinfachen und den Nutzen für das jeweilige Anwendungsfeld zu steigern. Es wurde dargestellt, wie der anthropomorphe Agent als personifiziertes Gegenüber psychologisch motiviert ist und wie durch ihn eine stärkere Involviertheit des Menschen in der virtuellen Umgebung erreicht werden kann.

Der Schwerpunkt dieser Arbeit lag in der Betrachtung der natürlichen Kommunikation über den dreidimensionalen Raum. Mit dem anthropomorphen Agenten als virtuellem Gesprächspartner entstanden neue Möglichkeiten bei der Verwendung deiktischer Ausdrücke.

Bei den Adverbien der dimensionalen Deixis standen die verschiedenen Referenzsysteme im Mittelpunkt der Betrachtung. Hier wurden die Schwierigkeiten bei der Auswahl des jeweils gültigen Bezugssystems verdeutlicht, was zu dem Lösungsansatz der verschiedenen Agenteninstanzen führte. Die Agentenprogramme liegen in mehreren Instanzen vor, die sich in ihrer Funktionalität dahingehend unterscheiden, daß sie verschiedene Referenzsysteme für ihre Berechnungen verwenden. Präferenzen des Benutzers können durch Korrekturen und daraufhin generierte Alternativvorschläge berücksichtigt werden.

Für Adverbien der positionalen Deixis ergaben sich mehrere Lokalisationsmöglichkeiten durch die Positionen der kooperierenden Gesprächspartner und durch die Zeigegesten des anthropomorphen Agenten. Auch wenn damit keineswegs alle Interpretationsvarianten erschöpft werden konnten, sind geläufige Verwendungsformen der deiktischen Ausdrücke nun in dem interaktiven Grafiksystem anwendbar.

Abschließend sollen derzeit diskutierte Erweiterungen angesprochen werden, die zum Teil schon in den vorherigen Kapiteln kurz erwähnt wurden:

- **Parallele Visualisierung zweier Sichten**

Die Explorationen im Raum können durch die Existenz des anthropomorphen Agenten auf zwei Arten durchgeführt werden: durch die externe Sicht, die durch die virtuelle Kamera gegeben ist, und durch eine involvierte Sicht, die dadurch erreicht wird, daß Hamilton die Kamera "auf die Nase gesetzt" wird. Um diese beiden Perspektiven in höherem Maße ausnutzen zu können, ist es wünschenswert, beide Sichten parallel zu verfolgen. Dazu muß lediglich eine zweite Kamera verwaltet werden. Der Bildschirm könnte dann in der Mitte geteilt und beide Sichten in jeweils einer Hälfte visualisiert werden. Möchte man einen Wechsel in dem Fensterformat vermeiden, so ist auch denkbar, eine Sicht in einem kleinen übergelagerten Fenster zu visualisieren.

- **Maussteuerung des Zeigearms**

Problematisch ist zur Zeit, daß man die Zeigegesten nur mit der sprachlichen Eingabe auslösen kann. Eine wirkliche Gestik, die eine Ergänzung zur Sprache darstellt, findet noch nicht statt. Eine nächste Zielsetzung ist daher, den Zeigearm mit Hilfe der Maus zu positionieren, um so mit einer nicht-sprachlichen Handlung die sprachliche Instruktion um räumliches Wissen zu erweitern. So kann der Benutzer mit Anweisungen, wie *Schiebe den Schreibtisch dahin* auf sehr einfache Art einen beliebigen Ort lokalisieren. Ebenso hilfreich wäre diese unabhängige Gestik bei der Referenzierung

von Objekten. *Dieser Stuhl* könnte mit einer Zeigegeste identifiziert werden, ohne daß er zuvor angesprochen werden muß.

- **Zeigegesten von außen**

Alternativ dazu könnte der Benutzer von außen mit Hilfe eines geeigneten Eingabegerätes (experimentiert wird derzeit mit dem PowerGlove) eine Position im Raum bedeuten, die von Hamilton durch entsprechendes Zeigen mit seinem Arm beantwortet und vom System als räumliches Wissen aufgenommen werden kann.

- **Beantwortung von Fragen**

Um die Funktion des anthropomorphen Agenten als Gesprächspartner für den Menschen noch zu steigern, ist eine weitergehende Idee die Beantwortung von Fragen. Auf diese Weise könnte der Benutzer auf einfache und natürliche Art Wissen von dem System vermittelt bekommen. Dabei wäre sowohl räumliches Wissen interessant, wie z.B. *Wo steht der Designer-Stuhl?* als auch Fachwissen über Objekte im Raum: *Welche Maße hat die Arbeitsplatte?* Hamilton könnte den Charakter eines Experten erhalten, der den Anwender mit fachlichen Ratschlägen durch den virtuellen Raum führen kann. Die fraglichen Informationen (z.B. auch die genauen metrischen Angaben) sind aus der erweiterten Datenbank exakt ermittelbar und könnten durch Zeigegesten oder durch eine verbale Ausgabe beantwortet werden.

- **Adaptives Verhalten**

Ein wichtiger Schritt zu einem benutzerfreundlichen System ist die Entwicklung von adaptivem Verhalten [Lenzmann & Wachsmuth, 1995]. Dieses ist durch die Vergabe von credit-values an die Agenteninstanzen bereits in einfacher Form entwickelt worden. In zukünftigen Arbeiten soll dieser Aspekt z.B. durch das Erstellen von Statistiken in dem Agentensystem weiter verstärkt werden, um so Benutzerpräferenzen bezogen auf situative Kontextfaktoren berücksichtigen zu können.

Mit Hilfe dieser Erweiterungen möchten wir dem Ziel der Vereinfachung und Verbesserung der Interaktion mit virtuellen Umgebungen einen weiteren Schritt näher kommen.

Literatur

- [Badler et al., 1991] Badler, N.I., Barsky, B.A., Zeltzer, D. (1991). *Making Them Move - Mechanics, Control, and Animation of Articulated Figures*. San Mateo (Cal.): Morgan Kaufmann.
- [Bosshardt, 1986] Bosshardt, H.-G. (1986). *Perspektiven auf Sprache*. Berlin: DeGruyter.
- [Bühler, 1965] Bühler, K. (1965). *Sprachtheorie: Die Darstellungsfunktion der Sprache*. Stuttgart: G.Fischer.
- [Dix et al., 1993] Dix, A., Finlay, J., Abowd, G., Beale, R. (1993). *Mensch Maschine Methodik*. Prentice Hall.
- [Ehrich, 1992] Ehrich, V. (1992). *Hier und Jetzt, Studien zur lokalen und temporalen Deixis im Deutschen*. Linguistische Arbeiten. Tübingen: Niemeyer.

- [Friederici, 1989] Friederici, A. D. (1989). Raumreferenz unter extremen perzeptuellen Bedingungen: Perzeption, Repräsentation und sprachliche Abbildung. In Ch. Habel, M. Herweg, K. Rehkämper (Hrsg.): *Raumkonzepte in Verstehensprozessen*, 17-33. Tübingen: Niemeyer.
- [Graf & Herrmann, 1989] Graf, R., Herrmann, T. (1989). *Zur sekundären Raumreferenz: Gegenüberobjekte bei nicht-kanonischer Betrachterposition*. Bericht Nr. 11, Arbeiten aus dem SFB 245 "Sprechen und Sprachverstehen im sozialen Kontext", Heidelberg/Mannheim.
- [Herrman & Grabowski, 1987] Herrmann, T., Grabowski, J. (1987). *Sprechen: Psychologie der Sprachproduktion*. Spektrum Akademischer Verlag
- [Jörding, 1995] Jörding, T. (1995). Ein kommunikationsfähiger anthropomorpher Agent in einem virtuellen Raum. Diplomarbeit an der Technischen Fakultät der Universität Bielefeld, 1995.
- [Klein, 1978] Klein, W. (1978). Wo ist hier? In *Linguistische Berichte* 58: 18-40.
- [Lang, 1989] Lang, E. (1989). Primärer Orientierungsraum und inhärentes Proportionsschema: interagierende Kategorisierungsraster bei der Konzeptualisierung räumlicher Objekte. In Ch. Habel, M. Herweg, K. Rehkämper (Hrsg.): *Raumkonzepte in Verstehensprozessen*. Tübingen: Niemeyer.
- [Laurel, 1990] Laurel, B. (1990). Interface agents: Metaphors with character. In B. Laurel (Ed.): *The art of human-computer interface design*, 355-366, Addison-Wesley.
- [Lenzmann & Wachsmuth, 1995] Lenzmann, B. & Wachsmuth, I. (1995). A User-Adaptive Interface Agency for Interaction with a Virtual Environment. Working Notes zum IJCAI-Workshop *Adaptation and Learning in Multi-Agent-Systems*, 43-46. Montreal 1995.
- [Levelt, 1986] Levelt, W. (1986). Zur sprachlichen Abbildung des Raumes: Deiktische und intrinsische Perspektive. In H.-G. Bosshardt (Hrsg.): *Perspektiven auf Sprache*, 187-211. Berlin: DeGruyter.
- [Lobin, 1993] Lobin, H. (1993). Situiertheit. *KI 1/1993*, S. 61.
- [Maes et al., 1993] Maes, P., Kozierok, R. (1993). Learning Interface Agents. In *Proceedings of the Eleventh National Conference on Artificial Intelligence (AAAI-93)*, 459-465.
- [Miller & Johnson-Laird, 1976] Miller, G.A., Johnson-Laird, P.N. (1976). *Language and Perception*. Cambridge University Press, 1976.
- [Otto, 1994] Otto, K. (1994). Ein Raumagent für Objekttransformationen in einer virtuellen Umgebung. Diplomarbeit an der Technischen Fakultät der Universität Bielefeld, 1994.
- [Rauh, 1983] Rauh, G. (1983). Essays on Deixis, 9-60. Tübingen: Gunter Narr Verlag

- [Retz-Schmidt, 1988] Retz-Schmidt, G. (1988). Various Views on Spatial Prepositions. In *AI Magazine*, 9(2), 95-105.
- [Saile, 1984] Saile, G. (1984). *Sprache und Handlung*, Schriften zur Linguistik 10. Braunschweig: Vieweg.
- [Sichelschmidt, 1989] Sichelschmidt, L. (1989). Wo hier dort ist - primär- und sekundärdeiktische Raumreferenz. In Ch. Habel, M. Herweg, K. Rehkämper (Hrsg.): *Raumkonzepte in Verstehensprozessen*. Tübingen: Niemeyer.
- [Wachsmuth et al., 1994] Wachsmuth, I., Lenzmann, B., Siekmann, N., Cao, Y. (1994). Systeme von Interface-Agenten als Mittler in der Mensch-Maschine-Kommunikation. *KI-NRW Report 94-16*.
- [Wachsmuth & Cao, 1995] Wachsmuth, I., Cao, Y. (1995). Interactive Graphics Design with Situated Agents. In W. Strasser, F. Wahl (eds.): *Graphics and Robotics*, 73-85, Springer.
- [Wunderlich, 1981] Wunderlich, D. (1981). Linguistic Strategies. In F. Coulmas (ed.): *A Festschrift for Native Speaker*.