

Eine kognitiv motivierte Architektur für einen anthropomorphen künstlichen Kommunikator

Ipke Wachsmuth und Nadine Leßmann

AG WBS-KI, Technische Fakultät

33594 Bielefeld, Germany

E-mail {ipke, nlessman}@techfak.uni-bielefeld.de

Abstract

Für den im Teilprojekt D3 des SFB 360 verfolgten Auftrag der Systemintegration stellt sich mit der Zusammenführung perceptiver, kognitiver und aktorischer Komponenten die Aufgabe der Gesamtkonzeption der Architektur eines Situierten Künstlichen Kommunikators. In dem vorliegenden Beitrag wird dies am Beispiel des anthropomorphen Agenten „Max“ exploriert, der in der virtuellen Realität verkörpert ist. Dabei wird ein besonderer Schwerpunkt auf die kommunikativen Fähigkeiten des Agenten unter Berücksichtigung seiner Körperlichkeit gelegt. Vorgestellt werden die Kernideen einer kognitiv motivierten Hybrid-Architektur für einen kommunizierenden Agenten und erste Realisierungsansätze.

1 Einleitung

Körperlich verankerte Kognition (embodied cognition) spielt heute in zahlreichen Gebieten eine zunehmend wichtige Rolle, sei es in der Modellierung und dem Test von Berechnungsmodellen in der Kognitionswissenschaft, in der Entwicklung Mensch-zentrierter robotischer Systeme oder für virtuelle Humanoide in Szenarien der virtuellen Realität. Sie ist auch ein zentraler Aspekt im Sonderforschungsbereich 360 „Situierte Künstliche Kommunikatoren“, wo situierte Kommunikationsfähigkeiten künstlicher Systeme in einer Mensch-Maschine-Kooperation erforscht werden.

Ein erklärtes Ziel des SFB 360 ist die Konstruktion maschineller Systeme, die das Verhalten und die Kompetenz natürlicher Kommunikatoren in relevanten Aspekten rekonstruieren. Da dies eine ausgesprochen komplexe Aufgabe ist und die meisten kognitiven Leistungen zudem situationsabhängig sind, wird von einem begrenztem Basis-Szenario ausgegangen. Gegenstand des Szenarios ist ein aufgabenorientierter Dialog, in dem Sprache, Gestik, Wissen, Planung, Handlung und Sensomotorik auf natürliche Weise miteinander verknüpft sind. Bei der kooperativen Bewältigung einer Montageaufgabe – der Konstruktion eines Flugzeugmodells aus Baufixteilen – übernehmen die Dialogpartner Rollen als Instrukteur und Konstrukteur und verhandeln über die einzelnen

Konstruktionsschritte. In unserem Ansatz können beide Rollen sowohl vom Menschen als auch von der Maschine eingenommen werden.

Als eine exemplarische Realisierung eines künstlichen Kommunikators wird mit Beiträgen verschiedener Teilprojekte im SFB 360 der anthropomorphe Agent „Max“ entwickelt, der eine solche Rolle des maschinellen Dialogpartners in virtueller Realität übernimmt (Abb. 1). Max verfügt bereits über Grundfähigkeiten der multimodalen Interaktion, von der eingabeseitigen Erfassung von Bewegungen und sprachlicher Instruktionen des Menschen bis zur ausgabeseitigen Synthese integrierter Sprache, Mimik und Gestik. Die sichtbare Physis von Max besteht aus einer mehrteiligen Hülle, die durch ein unterliegendes kinematisches Skelett kontrolliert wird. Das Skelett besteht aus 18 kinematischen Ketten, die in Segmente aufgeteilt sind wie Oberarm und Unterarm etc. Es enthält insgesamt 68 Segmente und 103 Freiheitsgrade in 57 Gelenken, davon allein 25 Freiheitsgrade (16 Gelenke) in den Fingern jeder Hand. Die Darstellung von Bewegungen des Körpers von Max erfolgt durch Echtzeit-Computeranimation [7].



Abbildung 1. Interaktion mit Max.

Die Produktion multimodaler kommunikativer Akte auf der einen Seite, sowie die Umweltwahrnehmung und Interpretation multimodaler Äußerungen auf der anderen Seite erfordern ein komplexes Zusammenspiel sensorischer, kognitiver und aktorischer Fähigkeiten. In dem hier vorgestellten Ansatz einer kognitiv motivierten Architektur für den anthropomorphen künstlichen Kommunikator Max fließen Ansätze aus verschiedenen Hintergründen ein, wie Modelle kommunikations- und handlungsfähiger Agenten aus KI und Kognitionswissenschaft, die sich zum Teil aus empirischen Daten psycholinguistischer Untersuchungen ableiten und daran ebenfalls evaluiert werden sollen (experimentell-simulative Methode). Auf der anderen Seite geht es um die Konzeption eines robusten informatischen Systems, das reaktive und deliberative (planerisch-zielverfolgende) Verhalten im aufgabenorientierten Dialog stimmig und glaubhaft abwickeln kann.

2 Anforderungen und Ausgangspunkte

Der situierte Agent sollte befähigt sein, seine Umwelt und insbesondere seinen menschlichen Kommunikationspartner visuell und akustisch wahrzunehmen und das Wahrgenommene situativ und kognitiv zu verarbeiten. Situierte Kommunikation im Handlungskontext geht über reine Instruierbarkeit hinaus; sie impliziert Dialogfähigkeiten des künstlichen Systems, das in der Lage sein muß, Sprache situationssensitiv zu verstehen und zu generieren. Um dies zu ermöglichen, ist es notwendig, Diskurswissen in das System zu integrieren und Mechanismen des Abwechselns im Dialog (Turntaking) vorzusehen. Zudem soll der Agent in der Lage sein, sein aktuelles Teilziel zu beschreiben, Aussagen über den perzeptuellen Status zu formulieren und seine Aktionen handlungsübergreifend zu kommentieren.

Die Abwicklung eines Dialoges zwischen Instrukteur und Konstrukteur erfordert weiterhin die Fähigkeit des Systems, ein explizites Partnermodell aufzubauen und dieses über den Zeitraum eines Diskurses hinweg zu verwalten. Eine entsprechend aufgebaute Gedächtnis-komponente soll dabei die aktuelle Situation erfassen, eingeführte Diskursreferenten speichern, den Status der aktuellen Teilhandlung wiedergeben und schließlich mögliche Intentionen und Überzeugungen des Dialogpartners repräsentieren können.

Von einem anthropomorphen Agenten wird erwartet, daß er sich möglichst natürlich bewegt, sich menschähnlich verhält und über mehrkanalige Ausdrucksmöglichkeiten verfügt. Die Körperlichkeit des Agenten spielt dabei eine essentielle Rolle. Durch Mimik und Körpersprache könnte er Auskunft über seinen internen Zustand geben und gleichzeitig in verbalen Äußerungen Auskünfte über Konstruktionschritte erteilen. Multimodalität in der Form

sprachbegleitender Gestenproduktion stellt einen weiteren Kanal dar, der zusätzliche Informationen liefern kann, zum Beispiel wenn beim Sprechen auf ein Szenenobjekt gezeigt wird. Betreffs der kommunikativen Kompetenz des Agenten im Konstruktionsszenario ergeben sich weiterhin die folgenden Fragen:

- Welche Art von Tätigkeiten und kommunikativen Aufgaben sind auszuführen?
- In welchem Ausmaß können dabei Störungen auftreten?
- Ist die Handlungsplanung schon zu Beginn der Interaktion präzise festlegbar, oder ergibt sie sich erst im Interaktionsvollzug?
- Welche Restriktionen gelten betreffs gemeinsamen Wissens über die Interaktionssituation?

Das gewählte SFB-Basisszenario schränkt zwar das benötigte Hintergrundwissen und die Interaktionssituation ein, jedoch soll die Handlungs- und Äußerungsplanung keinen starren Skripten folgen, sondern sich dynamisch der Situation anpassen. Das heißt, wenn auch ein Planungssystem mit Hilfe einer Wissensbasis einen konkreten Plan für ein momentanes Konstruktionsziel erstellen kann, soll der Agent dennoch in der Lage sein, vom Plan abzuweichen, sofern es die Situation erfordert, und ein robustes Verhalten aufweisen. Störungen wie beispielsweise Unterbrechungen seitens des Instrukteurs sollen im Rahmen von Turntaking abgefangen und verarbeitet werden können. Auf Anfrage soll der situierte künstliche Kommunikator in der Lage sein, Baupläne und Konstruktionen auszuführen bzw. zu veranlassen, zu beschreiben und zu erklären. Ferner soll er mit dem Benutzer in einen Dialog treten können, Fragen beantworten und gegebenenfalls von sich aus verständigungssichernde Rückfragen stellen.

2.1 Ausgangspunkte in der Architekturforschung

Die deliberativen Ansätze der „klassischen“ Architekturforschung gingen zunächst von symbolbasierten Planungs- und Schlußfolgerungsmechanismen aus, die auf einem internen Weltmodell operieren. Allerdings wiesen derart konzipierte Systeme Defizite in Bezug auf robustes, adaptives und flexibles Verhalten in dynamischen Umwelten auf. Geprägt von biologisch-verhaltensbasierten Ansätzen verwenden neuere reaktive Architekturen darum keine explizite Modellierung von Weltwissen, sondern legen die Annahme zugrunde, daß sich angepaßtes Verhalten eines Agenten allein aus der Interaktion einfacher Strategien mit einer komplexen Umwelt ergibt. Zeichnen sich reaktive Architekturansätze durch hohe Robustheit aus, fehlt allerdings derart realisierten Agenten die Fähigkeit zur Meisterung komplexer Situationen, in denen mehrere Schritte im Voraus überblickt und geplant werden müssen (variable Zielgerichtetheit).

Die mittlerweile verbreiteten Hybrid-Architekturen kombinieren deliberative und reaktive Komponenten, die in engem Zusammenspiel zum Gesamtverhalten eines Agenten beitragen [9], [3]. Hinzu kommen Ansätze, die elementare Mechanismen des menschlichen Verarbeitungssystems beschreiben, welche kognitiven Prozessen zugrunde liegen, und für Modellierungsansätze nutzbar zu machen. Dabei werden auch mentalistische Kategorien wie Überzeugungen, Wünsche und Absichten im Hinblick auf intentionales rationales Verhalten einbezogen. Als Spezialfälle einer kognitiv motivierten Modellierung der Intentionalität und Zielgerichtetheit haben sich *Belief-Desire-Intention*-(BDI)-Architekturen verbreitet, die von Rao und Georgeff [11] in einer operationalisierbaren Formalisierung für technische Systeme vorgeschlagen wurden. Zu den bekanntesten Vertretern zählen das Procedural Reasoning System (PRS) von Georgeff und Lansky [4] und sein Nachfolger JAM [5], welches PRS in einigen Punkten erweitert.

Neben diesen Ansätzen, die für autonome, intelligent handelnde Agenten entwickelt wurden, befassen sich Dialogmanagement-Architekturen speziell mit der Modellierung kommunikativer Fähigkeiten. Oft geschieht das in der Form aufgabenorientierter, system- oder benutzergeleiteter Dialoge mit vorhersagbaren Abläufen, in deren Verlauf eine Frame-artige Repräsentation gefüllt wird [8]. In *mixed-initiative*-Dialogen mit Anforderungen an Asynchronität, Initiativwechsel, Offenheit und Unvorhersehbarkeit des Diskurses rücken konversationale Aspekte des Dialoges in den Fokus.

Dialogmanagement wird in gegenwärtigen Systemen oft mit agentenbasierten Architekturen realisiert. Die dominierenden planbasierten Ansätze betrachten Sprechhandlungen als intentionale Akte, die formal als Operatoren mit Rollen, Vorbedingungen, Constraints und Effekten dargestellt werden. Weitere Ansätze basieren auf linguistischen Dialogtheorien, nach denen die kooperative Dialogstrategie dynamisch im Verlauf der Interaktion entsteht; das System realisiert konversationale Akte im Sinne der Sprechakttheorie [13] und überwacht seinen eigenen mentalen Zustand; z.B. Conversational Agency von Traum [14].

Aufbauend auf der Verallgemeinerung der Sprechakttheorie auf multimodale Äußerungen (Communicative Act Theory von Poggi und Pelachaud [10]) schlagen Cassell et al. [2] in dem FMTB-Modell die Abstraktion konversationaler Funktionen und die getrennte Verarbeitung von propositionalen und interaktionalen Aspekten vor. Der dort eingeführte *Embodied Conversational Agent* (ECA) konzentriert sich auf konversationale Fähigkeiten eines verkörperten Agenten, wobei sich Deliberation auf ein zentrales Dialogsystem beschränkt. In der zugrunde liegenden FXPAL-Referenzarchitektur ist zusätzlich eine

Schicht angelegt, welche den reaktiven konversationalen Aspekten Rechnung trägt.

Die Synthese einer kognitiv motivierten Modellierung von Handlung und Problemlösungskompetenz mit konversationalem Verhalten in einer integrierten Architektur für verkörperte Agenten wurde bislang nicht umfassend realisiert. Erste Ansätze für eine entsprechende deliberative Komponente wurden in der Conversational Agency [14] (auf Basis der BDI-Architektur) und dem virtuellen Tutor STEVE [12] (als Erweiterung von SOAR) vorgeschlagen. Defizite bestehen insbesondere beim integrativen Einbezug der Physis.

3 Kernideen einer Architektur für Max

Das für den anthropomorphen Agenten Max konzipierte Kernsystem eines situierten künstlichen Kommunikators integriert symbolverarbeitende und verhaltensbasierte Ansätze in einer hybriden Systemarchitektur, die Wahrnehmung und reaktives Verhalten, höhere mentale Prozesse wie Schlußfolgern und planvolles Handeln bis hin zum Einbezug von Aufmerksamkeit und motivationaler Handlungsbewertung (über „Desires“) betreffen. Ihre wichtigsten Prinzipien seien vorab erwähnt:

- nebenläufige (nicht sequentielle) Realisierung von Wahrnehmungs-, Schlußfolgerungs- und Handlungskomponenten (Perceive, Reason, Act)
- parallele Verarbeitung durch eine reaktive und eine deliberative Komponente
- ständiger Rückfluß von Informationen und Feedbackmechanismen in einer kognitiven Schleife, aber auch Feedback an aktive reaktive Behaviors
- BDI-Kernel mit dynamischen, eigenständig operierenden Planern für Deliberationen
- Berücksichtigung der Körperlichkeit (Physis) des Agenten, Multimodalität
- Aktionsausführung der reaktiven wie auch der deliberativen Komponente mittels Behaviors
- eigenbestimmte Prioritätswerte als Entscheidungsgrundlage für die Schlichtung von reaktivem und deliberativem Verhalten
- modulare und komponentenbasierte Struktur
- Sprechakte bzw. allgemeiner kommunikative Akte als intentionale Handlungen
- Integration eines Diskursgedächtnisses und eines Turntakingmoduls

Mit diesem Ansatz wird nicht allein eine Konzeption für einen humanoiden Dialogpartner verfolgt, sondern es soll mittelfristig auch ein theoretisches Modell für die Integration verschiedener Ansätze kognitiver Modellierung entwickelt werden, das Körperlichkeit und Kognition stimmig zusammenführen kann.

3.1 Struktureller Aufbau

Die Entwicklung einer kognitiv motivierten Architektur für einen konversationellen Agenten wirft als erstes die Frage des strukturellen Aufbaus auf (Abb. 2). Der Kreis, unterteilt in eine *Perceive-Reason-Act-Triade*, stellt die interne Verarbeitung des Agenten dar und grenzt ihn von seiner Umwelt ab. Dabei hebt die Dreiteilung die Verzahnung und das enge Zusammenspiel des klassischen *Perceive-Reason-Act-Zyklus* hervor. Der direkte Informationsfluß zwischen den Sektoren *Perceive* und *Act* berücksichtigt jedoch, daß reaktives Verhalten entstehen kann, ohne daß zuvor eine Deliberation stattgefunden haben muß und die kognitive Schleife durchlaufen wurde. Damit kann der vorliegende Ansatz als eine Hybrid-Architektur charakterisiert werden, die reaktives und deliberatives Verhalten in einer Struktur vereintigt.

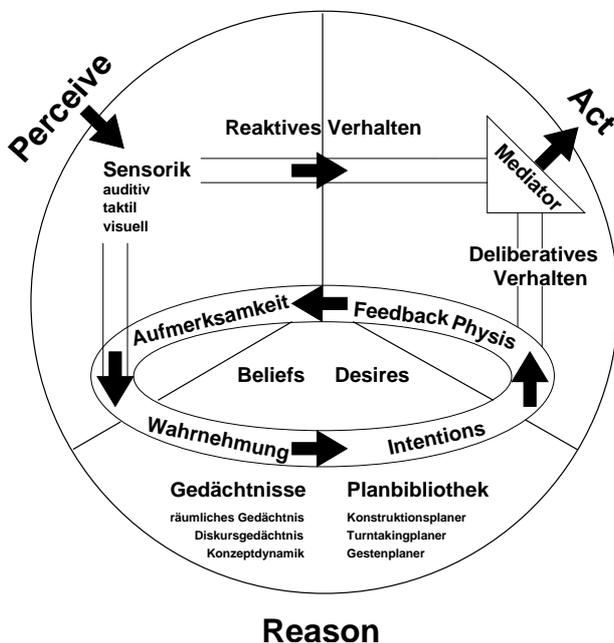


Abbildung 2. Architekturaufbau.

3.2 Zentraler Ablauf

Auf der einen Seite können die Sensordaten (von Spracheingabe/auditiv, Körpersensorik/taktil, Szenenwahrnehmung/visuell) direkt ein reaktives Verhalten auslösen, welches schnell und auf einem niedrigen Abstraktionsniveau operiert. Die deliberative Schleife präsentiert auf der anderen Seite einen Kreislauf, der die interne kognitive Verarbeitung des Agenten betrifft und das Wechselspiel zwischen Datenakquisition und Informationsverarbeitung aufzeigt.

Wahrnehmung besteht hier nicht aus der starren Erfassung sensorischer Daten, sondern aus einer situations-

sensitiven Verarbeitung perzipierter Sinneseindrücke. Kognition wird damit nicht als abgelöster interner Vorgang betrachtet, sondern vielmehr als stark an die Physis gekoppelter Prozeß mit einer stärkeren Betonung der prozeduralen Komponente. Die Sensordaten finden Eingang in die kognitive Schleife, werden dabei durch eine *Aufmerksamkeitssteuerung* gefiltert und wechselwirken in Form einer interpretierten und analysierten Wahrnehmung mit verschiedenen Gedächtnissen. Diese sind auf unterschiedliche Aspekte spezialisiert und arbeiten auf verschiedenartigen Repräsentationen, legen aber alle jeweils relevante Fakten auf einem hohen Abstraktionsniveau in den *Beliefs* ab, die das Arbeitsgedächtnis des Agenten darstellen.

Nach Ermittlung einer *Intention* und der Schlichtung konfligierender Verhaltensweisen der deliberativen und reaktiven Komponente im *Mediator* kommt es zur Ausführung einer Aktion. Sowohl die aktiv ausgeführten Intentionen als auch die aktuell anliegenden und möglicherweise konkurrierenden Verhaltensweisen werden bei den zurückfließenden *Feedbackinformationen* berücksichtigt. Die Rückkopplung der erfolgten Aktionen und Effektorzustände wirken sich wiederum in Form einer *Aufmerksamkeitssteuerung* auf die Sensorik und Wahrnehmung aus und schließen somit den Zyklus. Die Schleife verdeutlicht eine der zentralen Kernideen der Architektur, nämlich daß ein ständiger Strom von Informationen zwischen den Sektoren umläuft, der sowohl aktuelle Sensor- und Effektorinformationen als auch interne Zustände einbezieht, inklusive interner physischer Parameter.

3.3 Mediator

Die Kreissektoren *Perceive* und *Act* repräsentieren die Physis des Agenten. Durch seine Körperlichkeit ist der Agent in der Umwelt verankert, erhält er Weltbezug. Sie dient weiterhin als Ausdrucksmöglichkeit in Form der Multimodalität (Gestik, Sprechmimik wie auch emotionale Mimik). Die Sensorverarbeitung sowie die ausführende Effektorik sind durch körperliche Eigenschaften des Agenten geprägt. Dies wird auch bei der Modalitätenwahl einer auszuführenden Aktion berücksichtigt.

Der Mediator schlichtet zwischen den reaktiven und deliberativen Verhaltensweisen und zieht dabei in Betracht, welche Modalitäten gerade frei bzw. im Rahmen anderer Verhaltensweisen bereits im Einsatz sind. Die Entscheidungsgrundlage des Mediators besteht aus Prioritätswerten, die die Dringlichkeit und Angemessenheit eines Verhaltens in einer vorliegenden Situation ausdrücken und von den Verhaltensweisen und Intentionen selbst lokal berechnet werden [1]. Für das Schlichtungsverhalten des Mediators sollen in Zukunft verschiedene Algorithmen entwickelt und untersucht werden.

3.4 Reaktives Verhalten

Sowohl reaktives als auch deliberatives Verhalten wird durch Behaviors und Motorskills verschiedener Komplexitätsstufen umgesetzt.

Reaktive Verhaltensweisen zeichnen sich durch eine enge Kopplung an die Sensorik des Agenten aus und übernehmen zwei wichtige Funktionen. Sie operieren auf einer sehr niedrigen Abstraktionsebene und sind dadurch auch besonders schnell. In der Form von Reflexen mit hohen Prioritätswerten dienen sie zum einen unmittelbaren Systemreaktionen wie Ausweichverhalten bei drohender Kollision; im Bereich der Kommunikation treten reaktive Verhaltensweisen z.B. beim Turntaking auf.

Auf der anderen Seite wird der reaktiven Komponente die Aufgabe der *Secondary Behaviors* zuteil. Diese lassen den Agenten beispielsweise durch Augenzwinkern oder Atembewegungen natürlicher erscheinen und können ferner über seinen internen Zustand Aufschluß erteilen. *Secondary Behaviors* besitzen jedoch in der Regel niedrige Prioritäten, um deliberative geplante Verhaltensweisen nicht zu überstimmen.

3.5 Deliberatives Verhalten

Die deliberative Komponente wird im *Reason*-Bereich abgewickelt und ist durch die kognitive Schleife geprägt. Der Kern des deliberativen Moduls folgt der *Belief-Desire-Intention*-Philosophie und ist in Abb. 2 innerhalb der Schleife verortet.

Als innerer, verhaltensauslösender Antrieb dienen explizit repräsentierte Ziele (*Desires*), die sowohl durch interne Verarbeitung als auch von außen aufgeworfen werden können. Die Intentionbildung der kognitiven Schleife wird durch einen BDI-Interpreter vorgenommen, welcher aufgrund der vorliegenden *Beliefs*, den aktuellen Wünschen und Zielen sowie den alternativen Handlungsmöglichkeiten eine aktuelle *Intention* bestimmt.

Handlungsoptionen liegen in Form von Plänen vor, die durch Vorbedingungen, Kontextbedingungen, erreichbaren Konsequenzen und eine Prioritätsfunktion beschrieben werden. Die Planbibliothek besteht zum einen aus simplen Plankonstrukten, die einfache Aktionen direkt in entsprechende Behaviors umsetzen können. Zum anderen existieren jedoch auch dynamische eigenständige Planer, die mittels eines Prioritätswertes ihre Kompetenz, ein gewünschtes Ziel zu realisieren, mitteilen und die bei Bedarf angestoßen werden, um einen konkreten Plan auszuarbeiten. Für die Erstellung des Plans können die Planer auf die verschiedenen Gedächtnisse zugreifen, das heißt, sie sind nicht auf die *Beliefs* des Arbeitsgedächtnisses beschränkt. Die Gedächtnisse speichern nicht nur passiv Daten, sie können auch aufgrund ihrer Informationslage auch aktiv Anfragen bearbeiten.

Aus der Priorität des *Desires* sowie der Kompetenzbewertung des Planers und eventuell weiteren Parametern wird eine Gesamtpriorität bestimmt, mit der ein Plan darum konkurriert, aktiv zu werden. Verfügt er über die höchste Priorität, so wird er zur aktuellen Intention und erhält die Möglichkeit, interne Variablen und *Beliefs* zu beeinflussen sowie Behaviors zu instantiiieren, die dann wiederum im Mediator um den Zugriff auf die Effektoren „kämpfen“.

Planselektion findet somit einerseits auf der Ebene der kognitiven („bewußten“) Intentionbildung statt, andererseits und ebenfalls auch durch den Mediator auf der Ebene der direkten Planausführung, bei der die Plandringlichkeit und die Verfügbarkeit der Effektor-Modalitäten berücksichtigt werden.

3.6 Kommunikationsfähigkeiten

Die beschriebene Hybrid-Architektur mit BDI-Kernel findet für den anthropomorphen künstlichen Kommunikator *Max* Einsatz. Sie ermöglicht sowohl Fähigkeiten eines Dialoges mit geplanten Äußerungen, als auch die Fähigkeit zu spontaner reaktiver Äußerung, beispielsweise in Form von Turntaking- und Feedback-Signalen. Zusätzlich können verschiedene spezialisierte Planer – z.B. mit Wissen über die Konstruktion von Baufix-Flugzeugen – und spezialisierte Gedächtnisse – z.B. mit dynamisch aktualisierten Konzeptmodellen für strukturierte Aggregate („Konzeptdynamik“) – integriert und in die Kommunikation einbezogen werden.

Konzipiert wird die Kommunikation auf Basis der Sprechakttheorie nach Searle [13] bzw. der Theorie kommunikativer Akte nach Poggi und Pelachaud [10]. Das Dialogsystem ist planbasiert; kommunikative Akte werden als Aktion-Plan-Operatoren dargestellt. Dabei kann die kognitive Komponente für die Dialog- wie für die Handlungsplanung eingesetzt werden. Zur Dialogkontrolle zählen Turntaking, Reparatur- und Feedbackmechanismen. Es soll ein explizites Diskursgedächtnis entwickelt und repräsentiert werden.

4 Bisheriger Stand der Arbeiten

Die Architektur ist in ersten Schritten umgesetzt. Es existiert ein BDI-Interpreter, der auf JAM [5] aufsetzt und zum jetzigen Zeitpunkt über verschiedene Pläne verfügt, die es *Max* ermöglichen, auf Anfrage des Benutzers die Konstruktion verschiedener Aggregate zu erläutern.

Aktuell existieren zwei Einsatz-Szenarien. Zum einen kann der Benutzer sich von *Max* direkt zeigen lassen, wie eine Konstruktion vorgenommen wird. Das heißt *Max* beschreibt mittels seines Konstruktionswissens in synthetischer Sprache und unter Zuhilfenahme verschiedener Gesten, welche Baufixteile miteinander verbunden wer-

den müssen, und erläutert so schrittweise den gesamten Bauplan eines Aggregats.

Zum anderen ist Max auch in der Lage, interaktiv mit dem Benutzer zusammen ein Aggregat zu bauen. Dabei erklärt Max, welcher Konstruktionsschritt als nächster vollzogen werden soll, überläßt jedoch die Ausführung dem Benutzer, der entsprechende Instruktionen absetzen kann. Die Erkennung des Benutzers in der realen Welt (Bewegung, Blickrichtung, Gestik) erfolgt über getrackte Marker, Datenhandschuhe und durch ein Mikrofon, das Daten an eine Sprachverarbeitungs-komponente liefert.

Nach erfolgter Benutzeraktion liefert Max Feedback. Wurde die Konstruktion korrekt vorgenommen, so stimmt er zu und setzt seine Erläuterungen fort. Im Falle einer falschen Handlung jedoch macht er die Aktion des Benutzers rückgängig und erklärt den Konstruktionsschritt erneut.

Für die Beurteilung des Erfolgs der Benutzeraktion greift Max derzeit ausschließlich auf das Szenenwissen von COAR zurück [6]. Die schritthaltend aktualisierte COAR-Beschreibung der Szene enthält Informationen über die Objekte und deren eingegangene Verbindungen und weitere Objekteigenschaften. Zusätzlich soll Max in einer fortentwickelten Ausbaustufe eine visuell-räumliche Wahrnehmung erhalten (realisiert über View-Frustum), die auch Eingang in ein räumliches Gedächtnis finden wird. Auch der Benutzer soll in der virtuellen Szene registriert und mittels virtueller Sensorik wahrgenommen werden können.

Die Planstruktur des BDI-Moduls ermöglicht es, während der Ausführung einer Intention neue Ziele zu assertieren, die die aktuelle Intention ablösen können, sofern sie über höhere Priorität verfügen. Wird die vorherige Intention dabei nicht gezielt verworfen und gelten ihre Kontextbedingungen noch, so nimmt diese nach der Unterbrechung ihre Ausführung wieder auf. Es existiert weiterhin ein datengetriebener Plan mit einer hohen Priorität, welcher auf den Verlust des Turns reagiert.

Damit ist Turntaking auf zwei Ebenen vorgesehen, zum einen auf eine rein reaktive Art, so daß der Agent bei einer direkten Intervention zunächst seine Handlungen unterbricht, um zu analysieren, was der Auslöser war. Zum anderen findet Turntaking auf einer deliberativen Ebene statt, wobei in Bezug auf kleinere Unterbrechungen kontextabhängig agiert wird. Ein explizites Diskursgedächtnis ist vorgesehen. Bisher ist es nur rudimentär umgesetzt und verfolgt den Erhalt oder Verlust des Turns; in Zukunft soll es jedoch noch weiter ausgebaut werden und beispielsweise eine vollständigere Repräsentation des kommunikativen Gegenübers aufnehmen.

Auf einer weiteren Ebene sollen in Zukunft Ergebnisse eines Teilprojekts integriert werden, mit deren Hilfe

verständnis-sichernde Rückfragen abgewickelt werden können [15].

Reaktives Verhalten existiert in der aktuellen Version in Form von *Secondary Behaviors*, die durch Atmungsbewegungen und Augenblinzeln zu einem "lebendigen" Erscheinungsbild von Max beitragen. Ferner existiert ein Behavior der Fixation des Blickpunktes auf den Benutzer. Ist Max in einem Dialog begriffen und vollführt gerade keine Aktionen, bei denen er Sichtüberwachung benötigt, so er schaut er dem Benutzer in die Augen und verfolgt ihn mit seinem Blick. Sobald Max jedoch anfängt, etwas zu erklären, wird dieses Verhalten überstimmt und Max fixiert die Objekte, die er gerade referenziert. Sind alle benutzergetriebenen Intentionen abgewickelt, verharrt Max in *Secondary Behaviors*, zu denen auch Langeweile-Verhalten gehören.

Einige der skizzierten Verhaltensformen sind in der Ablaufbeschreibung einer Demonstration illustriert (siehe Anhang).

5 Zusammenfassung

In diesem Paper wurden die Kernideen und erste Realisierungsansätze einer kognitiv motivierten Architektur für einen anthropomorphen situierten Kommunikator vorgestellt. Dabei betont der derzeitige Ansatz die Integration von reaktiven und deliberativen Verhalten unter Berücksichtigung der Körperlichkeit des Agenten. Als nächste Schritte sollen verschiedene Algorithmen für die Schlichtung konfligierender Verhaltensweisen durch den Mediator untersucht werden. Die Prioritätsfunktionen und Feedbackmechanismen der einzelnen Behaviors sollen verbessert und erweitert werden. Weiterhin steht der Ausbau der Kompetenz des Turntaking durch einen Planer und ein erweitertes Diskursgedächtnis im Vordergrund der Arbeiten. Es soll umfassend untersucht werden, wie sich konversationale Fähigkeiten wie Turntaking, Mimik und Gestik zusammen mit intelligentem wissensbasiertem Verhalten sowie robusten, reaktiven Verhaltensweisen integrieren lassen.

References

- [1] Blumberg, B. M. und Galyean, T. A.: 1995, Multi-level direction of autonomous creatures for real-time virtual environments, *Computer Graphics* 29, 47–54.
- [2] Cassell, J., Bickmore, T., Campbell, L., Vilhjalms-son, H., & Yan, H.: 2000, Human conversation as a system framework: Designing embodied conversational agents, in Cassell, J., Sullivan, J., Prevost, S., und Churchill, E. (eds.): *Embodied Conversational Agents* (pp 29–63), The MIT Press, Cambridge (MA).

- [3] Gat, E.: 1997, On three-layer architectures, in D. Kortenkamp, R. P. Bonasso, und R. Murphy (eds.), *Artificial Intelligence and Mobile Robots.*, MIT/AAAI Press
- [4] Georgeff, M. & Lansky, A.: 1987, Reactive reasoning and planning, in *Proceedings of the Sixth National Conference of Artificial Intelligence*, pp 677–682, AAAI Press/MIT Press.
- [5] Huber, M.J.: 1999, JAM: A BDI-theoretic Mobile Agent Architecture, *Proceedings of the Third International Conference on Autonomous Agents (Agents'99)*, pp. 236-243, Seattle, WA.
- [6] Jung, B. & Wachsmuth, I.: 1998, Integration of Geometric and Conceptual Reasoning for Interacting with Virtual Environments. *Proc. AAAI'98 Spring Symposium on Multimodal Reasoning* (pp. 22-27).
- [7] Kopp, S. & Wachsmuth, I.: 2000, A Knowledge-based Approach for Lifelike Gesture Animation, in W. Horn (ed.), *ECAI 2000 Proceedings of the 14th European Conference on Artificial Intelligence*, pp 661–667, IOS Press, Amsterdam.
- [8] McTear, M.: 2002, Spoken dialogue technology: Enabling the conversational user interface, *ACM Computing Surveys* 34(1), 90–169.
- [9] Müller, J. P.: 1996, *The Design of Intelligent Agents: A Layered Approach*, Springer, LNAI 1177.
- [10] Poggi, I. & Pelachaud, C.: 2000, Performative facial expression in animated faces, in Cassell, J., Sullivan, J., Prevost, S., und Churchill, E. (eds.): *Embodied Conversational Agents* (pp 155–188), The MIT Press, Cambridge (MA).
- [11] Rao, A. & Georgeff, M.: 1991, Modeling rational behavior within a BDI-architecture, in *Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning*, pp 473–484.
- [12] Rickel, J. & Johnson, W.: 1999, Animated agents for procedural training in virtual reality: Perception, cognition, and motor control, *Applied Artificial Intelligence* 13, 343–382.
- [13] Searle, J. R. & Vanderveken, D.: 1985, *Foundations of Illocutionary Logic*. Cambridge UP.
- [14] Traum, D.: 1996, Conversational agency: the trains-93 dialogue manager, in S. LuperFoy, A. Nijholt, und G. van Zanten (eds.), *Proc. 11th Workshop on Language Technology: Dialogue Management in Natural Language Systems*, Universiteit Twente, Enschede, The Netherlands.
- [15] Voss, I.: 2001, Anticipation in construction dialogues. In J. Vanderdonckt, A. Blandford & A. Derycke (eds.) *Interaction without frontiers*, *Proceedings of Joint AFIHM-BCS Conference on Human-Computer Interaction IHM-HCI'2001 (Lille, France, Sept. 2001)*, Vol. II, pp. 189-190, Cépaduès-Editions, Toulouse.

Anhang: Demo-Skript



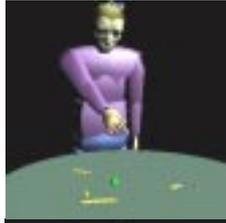
User: Erkläre mir den Propeller.



Max: Ich erkläre Dir jetzt wie man einen Propeller baut.

Max: Stecke diese Schraube in diese Leiste.

User: Stecke die gelbe Schraube in die Leiste.



Max: Nein. Die Aktion war falsch. Ich zeige es Dir noch einmal. Stecke diese Schraube in diese Leiste.

User: Verbinde die gelbe Schraube mit dem mittleren Loch der Dreierleiste.



Max: Ok. Und jetzt stecke diese Schraube in diese Leiste.

User: Stecke die gelbe Schraube in die Mitte der Dreierleiste.



Max: Ok. Und jetzt drehe diese Leiste quer zu der Leiste.

User: Drehe die Leiste quer zur Leiste.



Max: Super. Du hast so eben einen Propeller richtig zusammengebaut. Soll ich dir noch etwas zeigen oder erklären?

User: Ja, baue ein Leitwerk.



Max: Stecke diese Schraube in diese Leiste.

User: Stop.

Max: Ja bitte ?



User: Ok.

Max: Stecke diese Schraube in diesen Gewindeblock. So, das Leitwerk ist richtig zusammengebaut.

Start der Demo.

Max wartet auf eine Benutzereingabe.

Max erklärt die Aggregation einer gelben Schraube mit einer Dreilochleiste. Der Benutzer führt eine Aktion aus, bei der die unterspezifizierte Anweisung jedoch dazu führt, dass die Schraube in die Fünflochleiste gesteckt wird.

Max erkennt, dass die Schraube in der falschen Leiste steckt, negiert die Aktion und erklärt den Vorgang erneut.

Max beschreibt nach der richtigen Ausführung das weitere Vorgehen.

Max erklärt die weitere Konstruktion und der Benutzer agiert entsprechend.

Der Propeller ist fertig zusammengebaut. Max liefert positives Feedback und fragt nach weiteren Aufträgen. Der Benutzer fordert zu einer weiteren Aktion auf.

Max baut ein Leitwerk und wird dabei vom Benutzer unterbrochen. Max schaut den Benutzer an und wartet auf die Rückgabe des Turns. Der Benutzer hat Gelegenheit, den Konstruktionsstand zu inspizieren.

Nach der Rückgabe des Turns, stellt Max das Leitwerk fertig.

Durch Aufforderung des Benutzers wird das „Desire“, etwas zu erklären in Max erweckt. Die Planbibliothek verfügt über den Eintrag, eines Planers mit der Kompetenz der Konstruktion mit Baufixteilen.

In der vorliegenden Situation bestehen keine weiteren „Desires“ - Das Desire mit der höchsten Priorität gewinnt und wird zur Intention. Der eigenständiger Planer wird angeworfen und liefert einen Plan zurück, welcher sukzessive abgearbeitet wird.

Den einzelnen Planschritten folgend werden schrittweise Unterziele definiert, ein Objekt zu zeigen oder die Konstruktion zu erklären.

Hat der Benutzer einen Konstruktionsritt vorgenommen, überprüft Max mittels einer einfachen Wahrnehmung der Szene, ob die Aktion korrekt ausgeführt wurde und gibt Rückmeldung über den Erfolg der Konstruktion. Gegebenenfalls macht er die Benutzeraktion rückgängig und zeigt den Konstruktionsritt noch einmal.

Das zweite Beispiel präsentiert ein rudimentäres Turntaking-Verhalten. Bei der Konstruktion eines Leitwerks wird Max von dem Benutzer unterbrochen. Aufgrund des „daten-getriebenen“ Überwachens des Turns, reagiert Max, indem er den Benutzer anschaut und wartet bis er den Turn zurückerhält. Danach ist er in der Lage, seinen unterbrochenen Plan weiterfortzusetzen und das Leitwerk zu Ende zu bauen.