Situierte Künstliche Kommunikatoren

SFB 360

Darf ich dich Marvin nennen? Instruktionsdialoge in einem Wizard-of-Oz-Szenario: Szenario-Design und Auswertung

C. Brindöpke, J. Häger, M. Johanntokrax,
A. Pahde, M. Schwalbe und B. Wrede

Report 95/16

Universität Bielefeld

Darf ich dich Marvin nennen?
Instruktionsdialoge in einem Wizard-of-Oz Szenario:
Szenario-Design und Auswertung.

Inhalt

1 Einleitung

Simulierte Mensch-Maschine-Kommunikation bietet seit Anfang der 80er Jahre unter anderem die Möglichkeit, sprachverstehende Systeme bereits während ihrer Entwicklungsphase zu evaluieren. Sie ist seither unter dem Stichwort „Wizard-of-Oz-Experimente“ in zunehmendem Maße für die Entwicklung und das Design von sprachverstehenden Systemen eingesetzt worden.

Für das Szenario des SFB-360 (Instruktionsdialoge in der Baufixdomäne) wurde die Simulation eines sprachverstehenden Systems mit akustischer Spracheingabe und -ausgabe entwickelt, das in einem Instruktionsdialog der Baufixdomäne die Rolle des Konstrukteurs übernimmt. Zu einer erfolgreichen Simulation des Konstrukteurs gehören die Simulation sprachlicher Reaktionen und die Simulation der Bauleistung eines maschinellen Konstrukteurs. Ersteres wurde durch die Simulation eines sprachverstehenden Systems gelöst, letzteres durch die Übermittlung von Standbildern der aktuellen Bauaggregate auf den Bildschirm der Versuchsperson. Die Versuchsperson ging davon aus, daß es sich bei den Standbildern um eine Computergrafik handelte, die die noch nicht fertiggestellte Robotikkomponente des Systems ersetzte.

Die entwickelte Simulation war die Basis eines Szenarios für die Aufnahme von simulierter Mensch-Maschine- und Mensch-Mensch-Kommunikation (Kontrollgruppe) in der SFB-Domäne. Aufgenommen wurden insgesamt 54 Dialoge. Davon wurden 4 in einer Trainingsphase für simulierte Mensch-Maschine-Kommunikation aufgenommen und 10 in einer Kontrollgruppe, d.h. diese 10 Versuchspersonen wurden darüber informiert, daß sie mit einem Menschen und nicht mit einer Maschine kommunizierten. Ansonsten unterschieden sich die Bedingungen für die Kontrollgruppe nicht von denen der 40 Versuchspersonen der simulierten Mensch-Maschine-Kommunikation.

Insgesamt umfassen die Aufnahmen ca. 30 Stunden (simulierter) Mensch-Maschine-Kommunikation und ca. 5 Stunden Mensch-Mensch-Kommunikation (inklusive der Pausen für Systemreaktionszeiten). Damit sind für die im SFB-Szenario erstellten Instruktionsdialoge direkte Vergleiche von Mensch-Mensch- und Mensch-Maschine-Kommunikation möglich. Im Rahmen des SFB bieten die Daten insbesondere die Möglichkeit, die Anforderungen an einen künstlichen Kommunikator unter den Bedingungen des SFB-Szenarios zu untersuchen.

In den Kapiteln 2 und 3 werden zunächst Forschungsstand und methodische Fragen von Systemsimulationen diskutiert. Das Kapitel 4 stellt die Szenarien auf der Grundlage der entwickelten Simulation vor, in denen die Sprachaufnahmen durchgeführt wurden. Kapitel 5 beschäftigt sich mit der Erstellung des Sprachmaterials in den Szenarien. Dazu werden die

Aufnahmebedingungen (vgl. Kapitel 5.1), die Sprecherauswahl (vgl. Kapitel 5.2) und die Durchführung der Sprachaufnahmen (vgl. Kapitel 5.3) beschrieben. Das Kapitel 6 ist der Korpusbeschreibung gewidmet und umfaßt neben einem kurzen Überblick über die bisher geleistete Korpusbeschreibung die Auswertung einiger Aspekte der Wizard-of-Oz-Experimente: Zum einen wird der Aufbau des die Aufnahmen begleitenden Fragebogens erläutert und zum anderen werden erste Ergebnisse im Hinblick auf die Bewertung der Softwareergonomie des (simulierten) Systems, die allgemeine Akzeptanz des (simulierten) Systems und die Expertise der Versuchspersonen im Umgang mit Rechnern dargestellt. Kapitel 7 faßt die wichtigsten Ergebnisse der vorliegenden Arbeit kurz zusammen und gibt einen Ausblick auf sich anschließende Fragestellungen. In den Anhängen A und B befinden sich die Szenario-Spezifikationen im Überblick, sowie weitere Materialien (schriftliche Versuchseinführung, Fragebogen etc.) zu den einzelnen Szenarien. Anhang C befaßt sich ausführlich mit der technischen Realisierung der Systemsimulation.

2 Überblick über den Forschungsstand

Die grundlegende Idee bei einem Wizard-of-Oz-Experiment besteht darin, daß Versuchspersonen mit einem Computer kommunizieren, ohne zu wissen, daß die Aktionen des Computers von einem Menschen (dem sogenannten *Wizard*) gesteuert werden. Die Bezeichnung „Wizard-of-Oz“ geht auf ein Kinderbuch von L. Frank Baum (Baum 1900) zurück, in dem der „große und schreckliche“ Zauberer von Oz nichts weiter ist, als eine Maschine, die von einem Mann hinter einem Vorhang bedient wird. Diese Art von Experimenten wird auch als PNAMBIC-Technik (Pay No Attention to the Man BehInd the Curtain) bezeichnet.

Das Ziel von Wizard-of-Oz-Simulationen ist die Gewinnung von Informationen, die für das Design und die Entwicklung eines geplanten Systems von Bedeutung sind. Dabei ist eine derartige Simulation ein geeignetes Mittel, das Systemdesign über die Introspektion einzelner Systementwickler hinauszuheben. Die Wizard-of-Oz-Methode beinhaltet jedoch immer den folgenden Zirkel: man benötigt Informationen über eine Mensch-Maschine-Interaktion, die aber nicht stattfinden kann, da das System, mit dem sie stattfinden könnte, noch nicht realisiert ist. Das Design einer Wizard-of-Oz-Simulation kann den Ablauf der simulierten Mensch-Maschine-Interaktion und daher auch das spätere Systemdesign entscheidend prägen. Daher ist die Frage eines sorgfältigen methodischen Vorgehens bei der Entwicklung einer Simulation von besonderer Bedeutung (vgl. Kapitel 3).

Simulationen von Systemen unterscheiden sich durch technische Aspekte wie Ein- und Ausgabeverhalten, sowie durch ihre Aufgabenstellung bezogen auf eine bestimmte Domäne. Fraser/Gilbert (1991) bemühen sich um eine Teilklassifikation von Wizard-of-Oz-Simulationen unter dem Gesichtspunkt ihres jeweiligen Ein- und Ausgabeverhaltens. Dabei wird zunächst zwischen Simulationen, die mit natürlicher Sprache („speech“, „typing“) und solchen, die nicht mit natürlicher Sprache („symbol manipulation“) arbeiten, unterschieden. Abbildung 1 zeigt die Klassifikation von Fraser/Gilbert (1991) in einer etwas modifizierten Form.

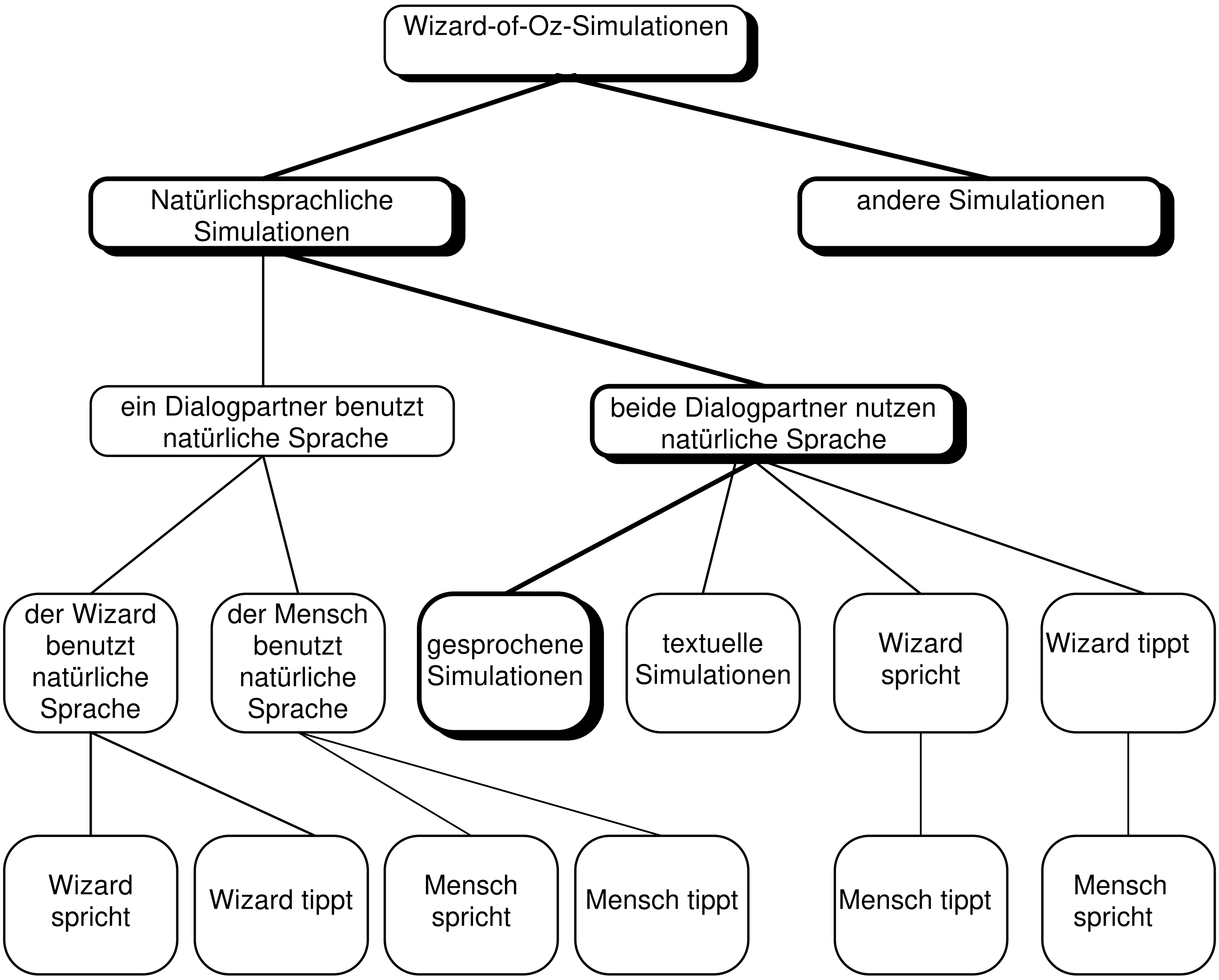

Abb. 1: Klassifikation von Wizard-of-Oz-Simulationen anhand ihres Ein- und Ausgabeverhaltens: Die für die hier vorgestellte Simulation zutreffenden Kategorien sind durch eine dickere Umrandung gekennzeichnet.

Die Simulationen, in denen natürlicher Sprache verwendet wird, teilen sich zunächst danach auf, ob sowohl die Versuchsperson als auch der Wizard natürliche Sprache benutzen, oder ob dies nur für die Versuchsperson (z.B. Hauptmann 1989) bzw. nur für den Wizard (z.B. Labrador/Dinesh 1984) zutrifft.

Wird ausschließlich mit natürlicher Sprache gearbeiten, wird weitergehend danach unterteilt, ob die Versuchsperson bzw. der Wizard eine textuelle oder eine akustische Sprachausgabe verwendet. Hier finden sich Wizard-of-Oz-Experimente, bei denen die Versuchsperson und der Wizard eine textuelle

Sprachausgabe benutzen (z.B. Dahlbäck/Jönsson 1989, Whittaker/Stenton 1989), die Versuchsperson eine akustische und der Wizard eine textuelle Ausgabe verwendet (z.B. Hauptmann/Rudnicky 1988) und Simulationen, in denen von beiden mit einer akustischen Sprachausgabe gearbeitet wird (z.B. Dybkjaer/Bernsen/Dybkjaer 1993, Luzzatti/Neel 1989, MacDermid 1993). Die im Rahmen des SFB entwickelte Simulation (vgl. Kapitel 4, Anhang C) hat folgende Eigenschaften:

- Beide Dialogpartner (Wizard und Nutzer) benutzen natürliche Sprache.
- Die Simulation hat eine akustische Spracheingabe und -ausgabe
- Die Simulation verfügt über eine Bildausgabe.

Damit gehört die Simulation zum einen unter die Kategorie „gesprochene Simulationen" und fällt zum anderen wegen ihrer Bildausgabe zum Teil auch unter die Kategorie „andere Simulationen". In der Literatur sind keine Beispiele für natürlichsprachliche Simulationen bekannt, in denen der Wizard spricht und der Mensch eine textuelle Eingabe vornimmt.
Für eine weitere Klassifikation von Wizard-of-Oz-Simulationen kommen zusätzlich Aspekte wie die genaue Beschreibung der Tätigkeiten des Wizard (vgl. Kapitel 3), sowie die spezielle Aufgabenstellung bzw. die Domäne des simulierten Systems hinzu. Häufig verwendete Domänen für Simulationen sind die Bereiche Zugauskunft (Delomier/Meunier/Morel 1989), Anwendung von Statistiksoftware (Guindon 1987), Telephoninformationsservice (Andernach/Deville/Mortier 1993) und Flugreservierung (Dalsgaard/Baekgaard 1994).
Die in Wizard-of-Oz-Experimenten erstellten Daten werden nicht nur in bezug auf das Design bzw. die Ergonomie des geplanten Systems ausgewertet, sondern es werden in der Regel weitergehende Untersuchungen zur Akzeptanz des geplanten Systems und Analysen zum sprachlichen Verhalten der Versuchspersonen durchgeführt.

3 Methodisches Vorgehen

Die Simulation eines sprachverstehenden Systems mit akustischer Spracheingabe und -ausgabe beinhaltet eine Reihe von Bedingungen, die von Simulation zu Simulation verschieden sein können. Diese Bedingungen können je nach Szenario Variablenstatus (abhängig/unabhängig) haben oder zu den allgemeinen Versuchsbedingungen gehören.

Abbildung 2 zeigt die Bereiche, die die Variablen einer Wizard-of-Oz-Simulation bilden können. Bezogen auf das simulierte System treten die Variablen zum einen als definierte Variablen des simulierten Systems auf und zum anderen als schwer erfaßbare Einflüsse, die aus dem persönlichen

Kommunikationsverhalten der Person resultieren, die das System simuliert bzw. die Systemsimulation bedient (Wizard). Letztere Einflüsse werden in der Tabelle nicht berücksichtigt, da sie beim jetzigen Forschungsstand nicht quantifizierbar sind.

	Versuchsperson	Simuliertes System (Wizard)
Erkennung	Akustisch Lexikalisch	Akustisch Lexikalisch Syntaktisch Pragmatisch
Produktion	Dialekt Wortwahl etc.	Antwortzeit Sprachausgabe Diskursstil etc.
Wissen	Domänenwissen allgemeines Weltwissen Systemerfahrung (u.a. Wissen über das simulierte System) etc.	Domänenwissen etc.
Vorbereitung	Versuchseinführung	Training

Abb. 2: Variablen in einem Wizard-of-Oz-Szenario

Die Variablen, die sich auf die Versuchspersonen beziehen, können bis auf wenige Ausnahmen, wie z.B. die Versuchseinführung oder durch Auswahlkriterien erzielte Einschränkungen (z.B. die Muttersprache), erst nach Abschluß der Sprachaufnahmen durch die Auswertung orthographischer Transkriptionen oder zusätzlichen Materials (z.B. Fragebögen) untersucht werden. Damit müssen für die Durchführung von Sprachaufnahmen in einem Wizard-of-Oz-Szenario und für die Realisierung einer Simulation zunächst die Variablen bezogen auf das simulierte System spezifiziert werden.

Die Variablen des zu simulierenden Systems können explizit durch die Algorithmisierung des Ein- und Ausgabeverhaltens des angestrebten Systems, bezogen auf die in Abbildung 2 angesprochenen Bereiche, definiert werden. Dabei ist zu beachten, daß nicht jedes Maschinenverhalten von einem Menschen (Wizard) simuliert werden kann. So ist es z.B. nicht möglich, ohne die Integration einer spracherkennenden Komponente ins simulierte System, in der dann Lexikon, Syntax etc. modelliert sind, die Variablen bezogen auf die Erkennung genau festzulegen. Um eine maschinelle Erkennungsleistung adäquat zu simulieren, müßte der Wizard allein bezogen auf die Worterkennung in dialogfähiger Zeit eine ganze Reihe von Entscheidungen treffen:

- Sind alle Wörter der von der Versuchsperson geäußerten Wortkette in einem vorab festgelegten Lexikon (in einem Umfang von 300 bis 1000 Wörtern) enthalten oder nicht?
- Welche Erkennungsfehler treten bei bekannten Wörtern aufgrund der Unsicherheiten der automatischen Spracherkennung auf?

Ähnliche Entscheidungsprozesse müssen bei der Simulation einer maschinellen Erkennungsleistung auch bezogen auf die Bereiche Syntax und Semantik erfolgen. Da diese Anforderungen selbst von einem gut geschulten Muttersprachler nicht zu erfüllen sind, ist bei Wizard-of-Oz-Simulationen im Bereich der Erkennung zunächst das natürliche, persönlich geprägte Kommunikationsverhalten des Wizard maßgeblich. Eine solche Vorgehensweise bei der Festlegung der Variablen bezogen auf die Erkennung soll im folgenden als eine implizite Variablenfestlegung bezeichnet werden. Erst ein vollständig entwickeltes System, das an die Stelle des simulierten Systems tritt, bietet eine explizite Festlegung aller Systemvariablen.

Für die Festlegung aller Systemvariablen wird in der Wizard-of-Oz-Methodologie eine schrittweise Annäherung an das geplante Zielsystem vorgeschlagen. Dabei werden die zunächst vom Wizard simulierten Systemfähigkeiten in bezug auf Produktion, Erkennung und Wissen sukzessive durch fertiggestellte Systemeinheiten ersetzt. Diese iterative Methode, bereits bei Kelley (1983/84) und Fraser/Gilbert (1991) diskutiert, wird auch für die Entwicklung der hier verwendeten Szenarien zugrundegelegt. Bei einer sorgfältigen Dokumentation der einzelnen Szenario-Iterationen (z.B. durch orthographische Transkriptionen und deren Analyse) bietet diese Vorgehensweise u.a. den Vorteil, daß das Verhältnis der zunächst implizit einfließenden Einflüsse der Person des Wizard zu den späteren expliziten Variablendefinitionen berücksichtigt werden kann.

Abbildung 3 setzt die iterative Entwicklung von Wizard-of-Oz-Szenarien und Zielsystem in Beziehung zu den drei Phasen der Erstellung von Sprachaufnahmen (vgl. auch Gibbon 1995).

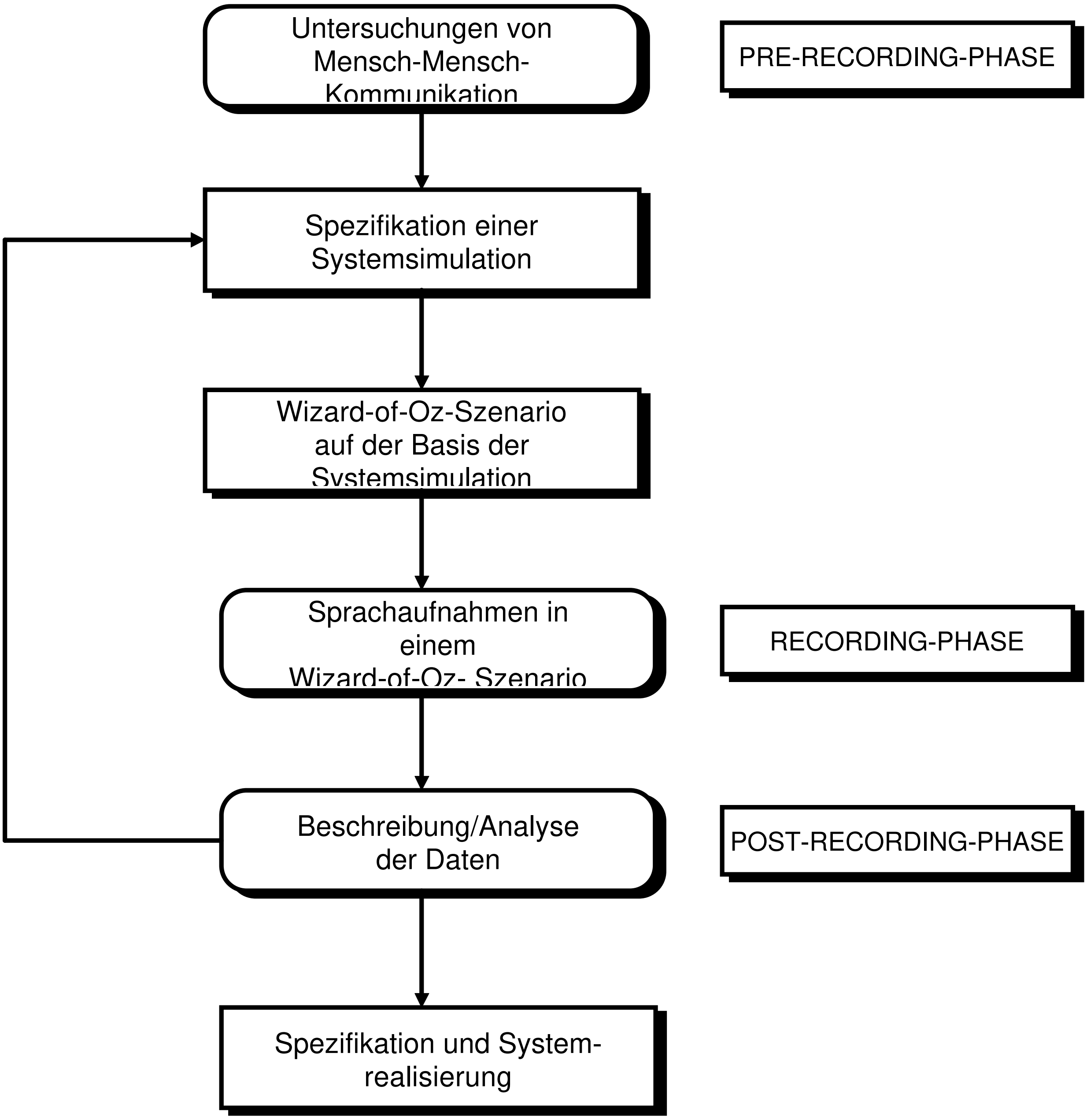

Abb.3: Methodisches Vorgehen bei der Szenarioentwicklung: Gerundete Boxen zeigen Untersuchungen und Experimente an; rechteckige Boxen entsprechen Produkten. Die Phasen der Erstellung von Sprachaufnahmen finden sich in den Boxen rechts.

- **Arbeitsphase 1:** ***Voruntersuchungen und Szenario-Entwicklung***
  Um ein angemessenes Eingabe- und Ausgabeverhalten der Systemsimulation definieren zu können, wurden zunächst 22 Mensch-Mensch-Dialoge (Sagerer et al. 1994) der SFB-Domäne untersucht. Diese Untersuchungen umfaßten die Erstellung von Wortlisten zur ersten Bestimmung des späteren Zielwortschatzes für Produktion und Erkennung, die Untersuchung der in den Dialogen etablierten Handlungsabläufe und des sprachlichen Verhaltens des Konstrukteurs. Letzteres wurde insbesondere für die Entwicklung der Sprachausgabe der Systemsimulation gebraucht, ersteres vor allem für die Spezifikation der Anforderungen an die Simulation der Bautätigkeit .
  Daraufhin wurden Benutzerschnittstellen des zu simulierenden Systems konzipiert und implementiert, sowie eine Aufnahmeumgebung (Szenario) für die Aufnahme von (simulierter) Mensch-Maschine-Kommunikation und einer Kontrollgruppe geschaffen (vgl. Kapitel 4.2). Desweiteren wurde ein die Sprachaufnahmen begleitender Fragebogen entwickelt (vgl. Kapitel 6).

- **Arbeitsphase 2:** ***Training***
  In einer anschließenden Trainingsphase (Sprachaufnahmen, Transkription und Analyse) wurden mit 4 Testpersonen Aufnahmen simulierter Mensch-Maschine-Kommunikation durchgeführt. Die Daten wurden orthographisch transkribiert und analysiert. Ziele dieser Trainingsphase waren die Überprüfung der aus den Analysen von Mensch-Mensch-Interaktion gewonnenen Erkenntnisse über die Anforderungen an das simulierte System, die Überprüfung des technischen Versuchaufbaus sowie das Training der Projektmitarbeiter, die später die Rolle des Wizard und des Versuchsleiters einnahmen (vgl. Kapitel 4.2).
- **Arbeitsphase 3:** ***Durchführung der Sprachaufnahmen***
  Auf der Grundlage der in der Trainingsphase gewonnenen Erkenntnisse wurde die Systemsimulation verbessert. Es wurde das endgültige Szenario für die Aufnahmen (simulierter) Mensch-Maschine-Kommunikation (Szenario I) und die Aufnahmen von Mensch-Mensch-Dialogen einer Kontrollgruppe (Szenario II) erstellt (vgl. Kapitel 4.3). In dem Szenario der Kontrollgruppe wurden Dialoge vom Typ der Mensch-Mensch-Kommunikation unter denselben Bedingungen wie die Dialoge der simulierten Mensch-Maschine-Kommunikation aufgenommen (d.h. kein persönlicher Kontakt zum Konstrukteur, Sprachübermittlung via Mikro und Sprachsynthetisator etc., vgl. Kapitel 5).
- **Arbeitsphase 4:** ***Korpusbeschreibung und -analyse***
  In dieser Arbeitsphase wurde der größte Teil des aufgenommenen Korpus orthographisch transkribiert. Erste Analysen des Wortschatzes u.a. im

Hinblick auf das Vorkommen von Partikeln wurden durchgeführt. Desweiteren wurde der Fragebogen in zentralen Teilaspekten ausgewertet (vgl. Kapitel 6).

4 Korpusdesign

Insgesamt wurden drei leicht variierende Szenarien definiert: ein Szenario für die Aufnahmen simulierter Mensch-Maschine-Kommunikation in der Trainingsphase (Szenario 0), eines für die Aufnahmen simulierter Mensch-Maschine-Kommunikation (Szenario I) und eines für die Aufnahmen von Mensch-Mensch-Kommunikation in der Kontrollgruppe (Szenario II). Das Szenario der Trainingsphase (Szenario 0) ist eine Vorversion des Szenarios I für die (simulierte) Mensch-Maschine-Kommunikation. Das Szenario der Kontrollgruppe (Szenario II) unterscheidet sich vom Szenario I nur durch die Informationen, die den Versuchspersonen gegeben wurden.

4.1 Voruntersuchungen

Den Voruntersuchungen liegen 22 Mensch-Mensch-Dialoge (Sagerer et al. 1994) in der SFB-Domäne zugrunde. In diesen Dialogen weist eine Testperson (Instrukteur) eine zweite Person (Konstrukteur) an, ein Baufixflugzeug zusammenzubauen. Die Dialoge unterscheiden sich durch die Sichtbedingungen für die Versuchsteilnehmer (blockiert, eingeschränkt, volle Sicht) sowie durch die Art der für die Instruktionen verwendeten Vorlage (Zeichnung, nachgebautes Modell).

Für die Szenario-Entwicklung wurden die Dialoge, in denen die Sicht blockiert war, untersucht. Der Verzicht auf mimische Information oder deiktische Ausdrücke, der durch diese Versuchsanordnung gefördert wird, ermöglicht es, die inhaltlichen und sprachlichen Anforderungen an den Konstrukteur herauszuarbeiten.

Minimale Anforderung an einen erfolgreichen Instruktionsdialog ist die sprachliche Etablierung von einer im Hinblick auf das zu erreichende Ziel geeigneten Handlungsabfolge. Für die Baufixdomäne bedeutet dies die Etablierung einer Handlungsabfolge der Handlungen vom Typ NEHMEN und POSITIONIEREN und zwar in einer für das Erreichen der Zielvorgabe, Konstruktion eines Baufixflugzeuges, geeigneten Reihenfolge. Das Befestigen von Bauteilen ist dabei eine Ausprägung von POSITIONIEREN. Die Etablierung der geeigneten Reihenfolge (Plan) kann natürlich auch im Dialog explizit thematisiert werden, soll hier aber wie auch die sprachliche Etablierung weiterer inhaltlicher Dialogebenen oder dialogstrukturierender Funktionen zunächst nicht weiter berücksichtigt werden. Ein vereinfachtes allgemeines Modell[1] für die Abfolge der im Dialog etablierten Handlungen (Handlungssequenzen) kann wie folgt aussehen:

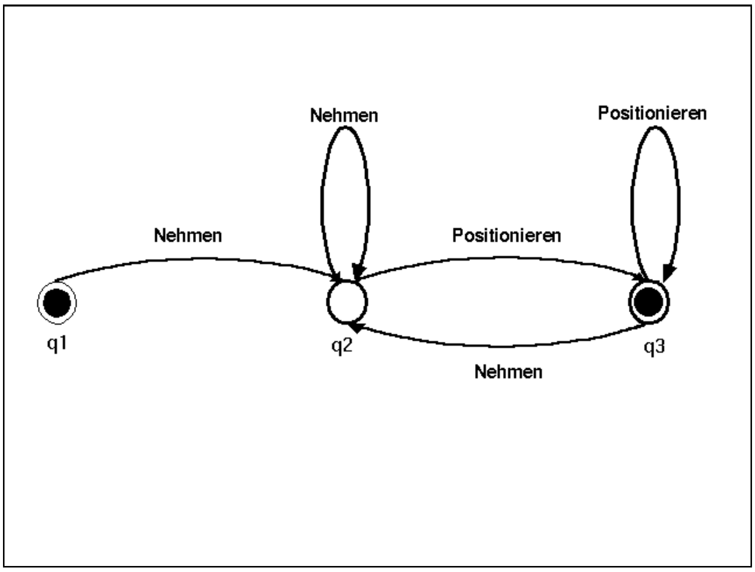

Abb. 4: Zentrale Handlungssequenzen in den Baufix-Dialogen als deterministischer endlicher Automat

Stark vereinfacht gesehen, werden die beiden Handlungen NEHMEN und POSITIONIEREN über eine Abfolge von Aufforderungs- und Bestätigungshandlungen etabliert. Aufforderungshandlungen beziehen sich dabei in der Regel auf die Handlungsschritte NEHMEN und POSITIONIEREN, die Bestätigungshandlungen sind Reaktion auf die Aufforderungshandlungen zu den Handlungsschritten. Die Beispiele 1-3 zitieren eine kurze Abfolge von Aufforderungshandlungen aus dem neunten Dialog (Sagerer et al. 1994). Die Schreibweise gibt zuerst den Typ der sprachlichen Handlung an, dann die inhaltliche Handlung (in Großbuchstaben), auf die sich die sprachliche Handlung bezieht und schließlich die gesprochene Wortkette in Anführungszeichen. T-66 entspricht der Seitenzählung des Materialbandes. Da sich sprachliche Handlungen aufeinander beziehen können, entstehen Schachtelungen wie in 2.

1. Aufforderung a_i
   (NEHMEN n_j

[1] Definition des Modells in Abbildung 4 als deterministischer endlicher Automat: Q = {q1, q2, q3}, Σ = {Nehmen, Positionieren}, Startzustand = q1 ε Q, Endzustand = q3 ε Q, δ = {<q1, Nehmen, q2>, <q2, Nehmen, q2>, <q2, Positionieren, q3>, <q3, Positionieren, q3>, <q3, Nehmen, q2>}.

("Nimmst den roten Würfel.")) T-66

2. Bestätigung b_k
(Auforderung a_i
(NEHMEN n_j
("Ja."))) T-66

3. Aufforderung a_{i+1}
(POSITIONIEREN p_j
("Legst ihn auf Stelle sechs drauf.")) T-66

Klärungsdialoge werden hauptsächlich über die sprachlichen Handlungen Frage und Antwort, die sich aufeinander sowie auf Aufforderungs- und Bestätigungshandlungen beziehen können, etabliert. Bestätigungshandlungen, die sich auf Frage und Antwort beziehen, gehören ebenfalls zu den Klärungsdialogen. Antizipation, d.h. eine sprachliche Äußerung, die sich auf eine kooperierende Vorwegnahme eines Dialogziels bezieht, soll hier nicht berücksichtigt werden, da das Phänomen für die Entwicklung der sprachlichen Ausgabe der Systemsimulation nicht unmittelbar von Bedeutung ist. Die sprachliche Gestaltung der jeweiligen Aufforderungs- und Bestätigungshandlungen, sowie von Frage und Antwort ist vielfältig und unterscheidet sich z.B. durch Wortwahl oder verwendete syntaktische Muster. Die Beispiele 4-6 zeigen einen Klärungsdialog, der sich direkt an die Beispiele 1-3 anschließt.

4. Frage a_i
(Aufforderung a_{i+1}
(POSITIONIEREN p_j
("Von oben neben die Schrauben, ja?"))) T-66

5. Antwort a_j
(Frage a_i
(Aufforderung a_{i+1}
(POSITIONIEREN p_j
("Neben die Schrauben genau.")))) T-66

6. Bestätigung b_{k+1}
(Antwort a_j
(Frage a_i
(Aufforderung a_{i+1}
(POSITIONIEREN p_j
("Ja."))))) T-66

Für die Definition der Systemvariablen des Wizard-of-Oz-Szenarios gilt, daß das sprachliche Verhalten des Konstrukteurs entscheidend ist für die Definition der Produktionsvariablen, während das sprachliche Verhalten des Instrukteurs wichtig ist für die Anforderungen an die Erkennungsleistungen des simulierten Systems.
Hauptsächlich bestehen die sprachlichen Handlungen des Konstrukteurs zum einen aus Bestätigungen, die in den Mensch-Mensch-Dialogen häufig als Interjektionen (z.B. mhm), Partikeln (z.B. ja) oder Reformulierungen der Aufforde-

rungen realisiert sind, und zum anderen aus der Initiierung von Klärungsdialogen durch Fragen, die sich auf die Aufforderungshandlungen des Instrukteurs beziehen.
Die sprachlichen Handlungen des Instrukteurs bestehen hauptsächlich aus Aufforderungshandlungen bezogen auf die Handlungsschritte NEHMEN und POSITIONIEREN, sowie aus Antworten und Fragen bezogen auf die Klärungsdialoge. Die Analyse der Dialoge zeigt, daß in der Regel mehrere Baufixteile (ca. 2 bis 4) in den Aufforderungshandlungen zu NEHMEN eingeführt werden, die dann sukzessive positioniert werden. Aufforderungshandlungen können sich auch speziell auf die Elizitierung von Bestätigungshandlungen des Konstrukteurs richten. Dies ist allerdings eine Besonderheit, die für die Modellierung der Dialogkomponente eines Systems zunächst nicht von Bedeutung ist.

4.2 Trainingsphase: Szenario 0

Von den Variablen des zu simulierenden Systems wurden die Bereiche, die die sprachliche Produktion des simulierten Systems betreffen, festgelegt. Dazu wurde auf der Basis der Voruntersuchungen eine Menge von geeigneten Textstrings (Textbausteine) definiert, die über Mouse-Klick auf ein Bedienfeld eines Wizard-Kontrollbildschirms an ein Sprachsynthesegerät übertragen und von diesem in eine akustische Sprachausgabe, die die Testperson über ihren Kopfhörer hören konnte, umgesetzt wurden. Die akustische Komponente der Produktion des simulierten Systems wurde damit durch die Verwendung eines Sprachsynthesegerätes definiert, Lexikon und Syntax sowie Diskursstil (höflich, nicht höflich) durch die vorgegebenen Textbausteine (vgl. Anhang A,B). Der pragmatische Aspekt der Produktion kann durch ein Dialogmodell geklärt werden.

Die Anwortzeit des simulierten Systems setzt sich additiv aus den menschlichen Reaktionszeiten und den Systemreaktionszeiten (Digitalisierung und Übertragung von Sprach- und Bilddaten) zusammen und ist damit länger als die menschliche Reaktionszeit. Letzteres wird häufig als wichtiges Indiz für die Glaubwürdigkeit von simulierten Systemen gesehen. Trotzdem sollte die Reaktionszeit noch einen Dialog anstelle einer Abfolge von Einzeläußerungen ermöglichen.

Die Szenario-Variablen, die die Erkennung betreffen, konnten wegen der in Kapitel 3 bereits diskutierten begrenzten menschlichen Fähigkeiten zur Simulation von Maschinenverhalten für diese erste Iteration der Methode nicht systematisch spezifiziert werden. An die Stelle definierter Erkennungsleistungen traten Verabredungen für das Verhalten der Personen, die die Rolle des Wizard übernahmen, die während einer Trainingsphase mit ihnen eingeübt wurden. Abbildung 5 zeigt einen schematischen Überblick über den Szenario-Aufbau.

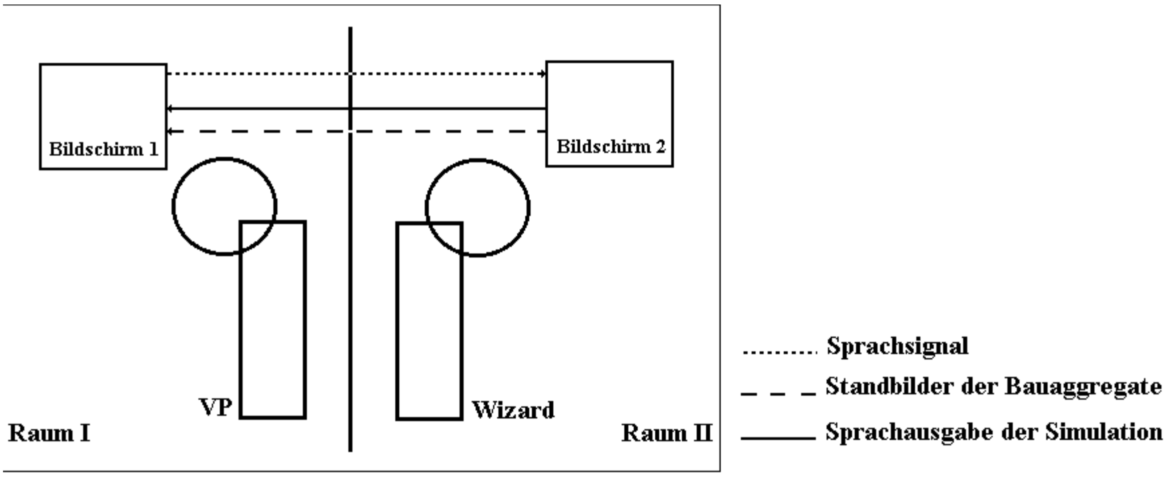

Abb.5: Schematischer Überblick über den Szenario-Aufbau

Für die Trainingsphase war eine vorläufige Systemsimulation und ein Szenario für die Durchführung der Sprachaufnahmen entwickelt worden. Die sprachliche Produktion des simulierten Systems in dem Szenario der Trainingsphase umfaßte:

- Fragen für die Initiierung von Klärungsdialogen
- Fragen bezogen auf die Spezifikation der Objekte (Form, Farbe und Anzahl der Löcher in einer Lochleiste)
- Fragen bezogen auf Ortsangaben
- Bestätigungshandlungen
- Äußerungen mit metakommunikativen/dialogstrukturierenden Funktionen wie z.B. „Der Arbeitsschritt ist beendet. Bitte fahren Sie fort."

Diese Textstrings (Set A) waren auf der Grundlage der Voruntersuchungen entwickelt worden. Für eine Auflistung aller Textbausteine aus Set A, sowie eine komplette Szenariospezifikation siehe Anhang A. Außer diesen vorab festgelegten Textbausteinen war die freie Eingabe von Textstrings zulässig, die dann ebenfalls über den Sprachsynthetisator realisiert wurden.

Für Erkennungsleistung und Bautätigkeit des simulierten Systems wurden in der Trainingsphase mit den Mitarbeitern geeignete Konzepte eingeübt. Dazu gehörte die Festlegung auf eine einheitliche Interpretation von Raumrelationen und eine Reduktion der Kooperativität natürlicher Sprecher im Hinblick auf den reduzierten Kenntnis- und Fähigkeitsstand, den die Wizard-Rolle zur Simulation eines sprachverstehenden Systems erforderte. Abbildung 6 zeigt ein extrinsisches Koordinatensystem, das in den Dialogen für die Handhabung von Begriffen wie vorne, hinten, seitlich etc. zugrundegelegt wurde.

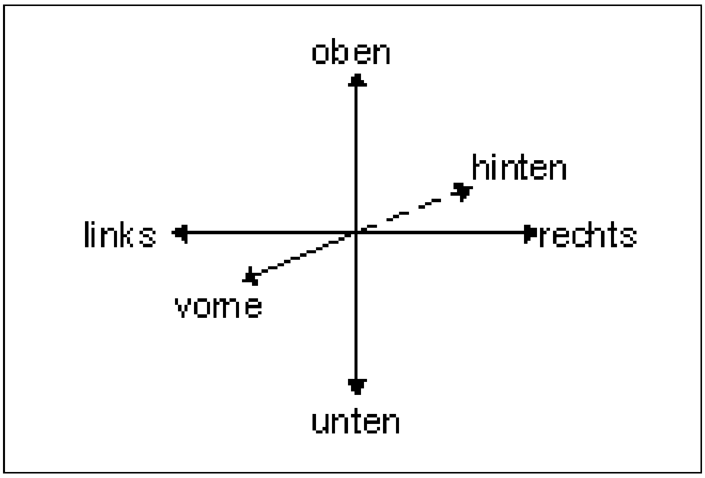

Abb. 6: Räumlichkeit im Wizard-of-Oz-Szenario

Die Reduktion der natürlichen Sprecherkooperativität, unter dem Aspekt eine Maschine zu simulieren, beinhaltete die folgenden Aspekte:

- das Zurückweisen einer bestimmten Anzahl von Äußerungen des Instrukteurs zur Simulation von Erkennungsfehlern
- eine Sensibilisierung für Wörter und Äußerungen, deren Interpretation für ein sprachverstehendes System zu weitreichend sind (eingeschränkte Kompetenz des simulierten Systems)
- nicht eindeutig spezifizierte Objekte wurden nicht in die Bautätigkeit einbezogen
- es wurde kein Gedächtnis des simulierten Systems vorausgesetzt, d.h. Anweisungen des Instrukteurs wie „Wiederhole diesen Arbeitsschritt." wurden abgelehnt
- zu globale Anweisungen wie „Baue das Flugzeug." wurden zurückgewiesen.

Die Sprachaufnahmen der Trainingsphase wurden transkribiert und im Hinblick auf die sprachliche Produktion des simulierten Systems und die Konsistenz des Wizard-Verhaltens untersucht. Die Auswertung der Transkriptionen der Voraufnahmen ergab in bezug auf die sprachliche Produktion, daß Fragen bezogen auf die Objektspezifikation, die in den untersuchten Mensch-Mensch-Dialogen relativ häufig auftraten, in der Mensch-Maschine-Kommunikation nicht genutzt worden waren. Darüber hinaus fehlten in den Textbausteinen dialogeinleitende und -abschließende Sequenzen. Im Anhang A findet sich eine Auflistung der Vorkommen der Textbausteine in den Vorversuchen, sowie die während der Versuche gemachten „freien" Texteingaben.
Die technischen Aspekte des Szenarios erwiesen sich vor allem im Hinblick auf die Reaktionszeiten des simulierten Systems als unbefriedigend und wurden für die beiden Folgeszenarien verbessert (vgl: Kapitel 4.3). Bezogen auf das Training der Mitarbeiter, die die Rechnersimulation bedienten (Wizard), fiel auf, daß für menschliche Sprecher systematische Verletzungen der Kooperativität, wie sie z.B.

nötig sind, um Erkennungsfehler eines Systems zu simulieren oder unvollständige Anweisungen zurückzuweisen, Schwierigkeiten bereiten.

4.3 Szenario I und II: simulierte Mensch-Maschine-Kommunikation und Kontrollgruppe

Die Szenarien I und II enthalten gegenüber dem Szenario 0 der Trainingsphase Verbesserungen technischer Details, die vor allem auf schnellere Reaktionszeiten des (simulierten) Systems abzielen. Darüberhinaus liegt der Sprachausgabe der Simulation eine revidierte Menge von Textstrings (Set B) zugrunde.

- **Szenario I:** Die Versuchspersonen wurden aufgefordert, einen Rechner anzuweisen, ein Baufixflugzeug zusammenzubauen.
- **Szenario II:** Die Versuchspersonen der Kontrollgruppe wurden aufgefordert, einen Menschen, der sich in einem anderen Raum befinde, anzuweisen, ein Baufixflug zusammenzubauen. Die Bedingungen der Kommunikation waren dieselben wie im Szenario I (kein persönlicher Kontakt zum Konstrukteur, Sprachübermittlung via Mikro und Sprachsynthetisator etc.). Dies wurde der Versuchsperson gegenüber mit dem Testen neuer Kommunikationskanäle begründet.

In beiden Szenarien bestand die Sprachausgabe der Systemsimulation aus einer Menge vorab definierter Textstrings, die über ein Sprachsynthesegerät realisiert wurden. Die inhaltliche Festlegung der Textstrings bezogen auf dialogische Funktionen und semantischen Gehalt erfolgte auf der Basis der Trainingsphase und der Voruntersuchungen (Set B). Dazu kam ein vorläufiges Modell für die Abfolge des Dialoges, das während der Aufnahmen jedoch nicht eingehalten werden mußte. Die technische Möglichkeit, freie Textstrings einzutippen und durch das Sprachsynthesegerät realisieren zu lassen, wurde in den Szenarien I und II nicht mehr genutzt, um die sprachlichen Reaktionen des simulierten Systems vergleichbar zu halten. Die Bildausgabe der Systemsimulation in den Szenarien I und II bestand wie im Trainingsszenario (Szenario 0) aus digitalisierten Einzelbildern der jeweils aktuellen Bauaggregate, die auf den Bildschirm der Versuchsperson gesendet wurden.

Die zeitliche Koordination der Sprach- und Bildausgabe sowie die Auswahl der jeweils passenden Textbausteine erfolgte, wie in der Trainingsphase eingeübt, durch einen Projektmitarbeiter. Für diese Aufgabe stand ihm ein benutzerfreundlich gestalteter Kontrollbildschirm zur Verfügung, auf dessen Bedienfeldern die Zustände der Simulation und die Datenübertragungen (Sprache, Bild und Steuersignale) hauptsächlich via Mouse-Klick gesteuert werden konnten (vgl. Anhang C). Das Zusammenbauen des Baufixflugzeugs wurde von einem weiteren Mitarbeiter manuell ausgeführt. Eine schwarze Box mit eingezeichneten Koordinatensystemen erleichterte das „maschinengerechte“ Ausrichten der Bauaggregate für die Bildaufnahme. (s. Abb. 7 für die Szenario-Spezifikation).

SET-UP:

Geräte:	drei Unix-Workstations (Dec-Workstation 5000), ein PC mit angeschlossenem Bildsequenzsystem, eine Stereokamera (3-Chip-RGB-CCD), zwei Gradient Desklab 216, ein Sprachsynthetisator SVS 3105, Ethernet; Mischpult, Dat-Rekorder, Vorverstärker, Headset, Kopfhörer; zwei Räume (Raum der Versuchsperson, Raum des Wizards); zwei Baufixset, Fragebögen etc.
Personen:	Versuchsleiter, Personen für die Bedienung des simulierten Systems (Wizard), Versuchspersonen

VERSUCHSLEITER:

Aufgabe:	Durchführung und Überwachung des Versuchsablaufs, Bedienung der Aufnahmegeräte
Personen:	ein Projektmitarbeiter

WIZARD:

Aufgabe1:	Zusammenbau des Baufixflugzeuges nach den Anweisungen der Versuchsperson
Aufgabe2:	Koordination der Bild-und Sprachübermittlung an die Versuchsperson
Personen:	je ein Projektmitarbeiter für Aufgabe1 und 2
Tools:	Kontrollbildschirm zur Steuerung der Simulation, Software der Simulation; Baufixset
Produktion:	Sprache: festgelegte Textstrings, die über ein Bedienfeld auf dem Kontrollbildschirm ausgewählt und an das Sprachsynthesegerät weitergegeben werden Bild: Standbilder vom jeweiligen Bauzustand, die via Ethernet auf den Bildschirm der Testperson geschickt werden
Erkennung:	akustische Erkennung: menschliche Erkennungsleistung beeinflußt durch die Qualität der Sprachdatenübertragung Erkennung Lexikon/Syntax/Pragmatik: menschliche Erkennungsleistung eingeschränkt durch die im Training geübten Verhaltensweisen
Wissen:	s.Training
Training:	Handhabung der Simulation, Zusammenbau des Flugzeugs, Reduktion der natürlichen Sprecherkooperativität (Einbau von Erkennungsfehlern u.a.)

VERSUCHSPERSONEN:

Personen:	Studierende aller Fakultäten
Anzahl:	**Szenario I:** 40 **Szenario II**: 10 (Kontrollgruppe)
Aufgabe:	**Szenario I:** einen Rechner verbal anzuweisen, ein Baufixflugzeug zusammenzubauen; **Szenario II:** eine in einem anderen Raum befindliche zweite Person verbal anzuweisen, ein Baufixflugzeug zusammenzubauen (Kontrollgruppe)
Vorwissen:	einmaliger Zusammenbau des Flugzeugs nach der Bedienungsanleitung, schriftliche Versuchseinleitung, keine Erfahrung mit dem simulierten System

Abb. 7: Übersicht über die Szenarien I und II; Die beiden Szenarien unterscheiden sich in Anzahl und Aufgabenstellung für die Versuchspersonen

5 Korpusaufnahme

Die Trainingsphase (vgl Kapitel 4.2) wurden im Dezember 1994 durchgeführt, die 50 Aufnahmen, die in den Szenarien I und II durchgeführt wurden, fanden vom 1. bis 28 Februar 1995 statt. Sofern nicht explizit angegeben, gelten die folgenden Beschreibungen jeweils für die Szenarien I und II.

5.1 Aufnahmebedingungen

Die Versuche fanden in zwei getrennten Räumen der technischen Fakultät der Universität Bielefeld statt (D6-131, M5-106). Versuchsperson und Versuchsleiter befanden sich dabei in einem Büroraum (D6-131), in dem auch die Aufnahmen durchgeführt wurden. Die Projektmitarbeiter, die die Rolle des Konstrukteurs (simulierte Maschine bzw. menschlicher Konstrukteur) übernahmen, befanden sich in einem größeren Rechnerraum (M5-106).

Abbildung 8 zeigt eine vereinfachte Übersicht über die technische Ausstattung der Räume von Versuchsperson (Raum I) und Wizard (Raum II) sowie den Datenfluß zwischen den Räumen.

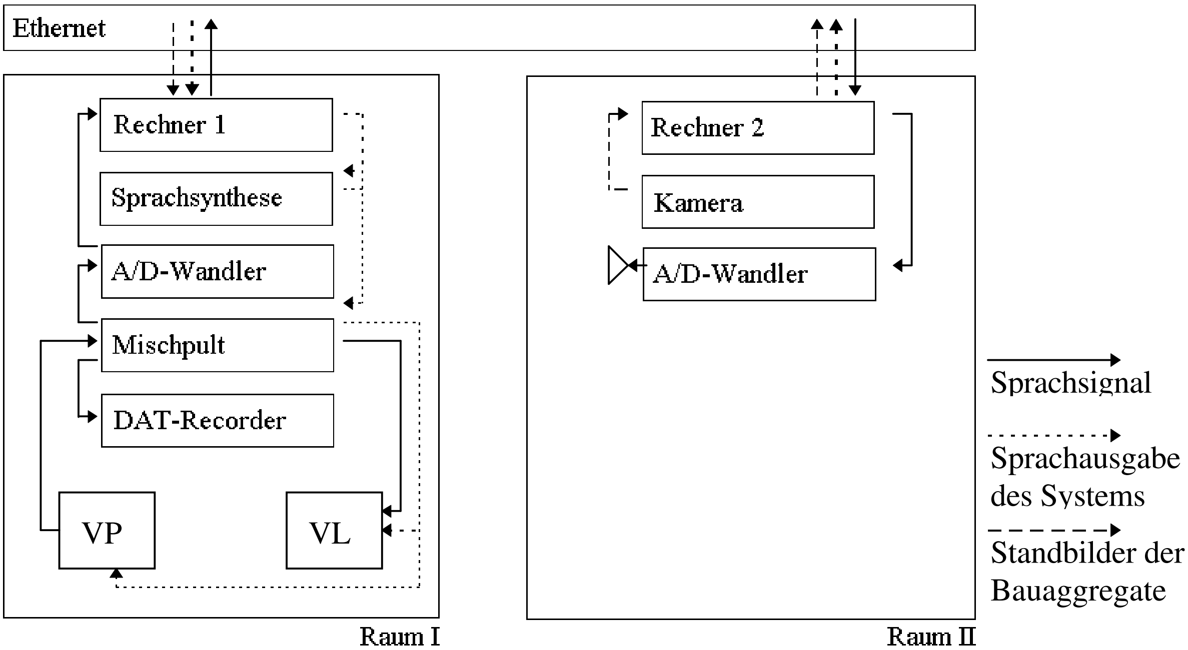

Abb.8: Überblick über die Raumausstattung und den Datenfluß

Im Raum der Versuchsperson (Raum I) befand sich eine Unix-Workstation mit hochauflösendem Bildschirm (8bit Farbtiefe), ein Sprachsynthetisator, ein Gradient Desklab und alle für die Aufnahmen erforderlichen Geräte (Dat-Rekorder, Mischpult, Vorverstärker, Headset etc.) Im Raum des Wizard (Raum II) befand sich ebenfalls eine Unix-Workstation, ein PC mit angeschlossenem Bildsequenzsystem, ein Gradient Desklab und eine Stereokamera. Vom Bildschirm des Wizard in Raum II konnte die gesamte Simulation kontrolliert und gesteuert werden (Kontrollbildschirm).

Die Sprachsignale sowohl der Versuchsperson als auch die vom System erzeugte Sprachausgabe wurden in Raum I mit einer Abtastfrequenz von 48 kHz auf unterschiedlichen Kanälen auf DAT-Band aufgenommen. Die Möglichkeit, das Sprachsignal der Versuchsperson nicht nur auf Dat-Rekorder aufzuzeichnen, sondern auch direkt auf die Festplatte eines Rechners, bestand, wurde allerdings wegen des sehr großen und im vorhinein schwer abzuschätzenden Speicherbedarfs für die Versuche nicht genutzt. Desweiteren wurde das Sprachsignal der Versuchsperson via Ethernetverbindung von Raum I nach Raum II übertragen (vgl. Anhang C), sowie auf den linken Kanal des Kopfhörers des Versuchleiters, damit dieser das Gespräch zwischen Versuchsperson und (simuliertem) System verfolgen konnte.

Die Sprachausgabe des (simulierten) Systems wurde durch Mouse-Klick auf entsprechende Bedienfelder des Kontrollbildschirms in Raum II gesteuert. Die Steuersignale gingen via Ethernet an den Sprachsynthetisator in Raum I, wo sie als Sprachsignal umgesetzt wurden, das dann über Mischpult auf den rechten Kanal des Kopfhörers der Versuchsperson und den rechten Kanal des Kopfhörers des Versuchsleiters geschickt wurde.

Die Standbilder der in Raum II nach den Anweisungen der Versuchsperson gebauten Flugzeugaggregate wurden digitalisiert und dann via Ethernet auf den Bildschirm der Versuchsperson weitergeleitet. Die Realisierung der drei Aufgaben der Simulation, nämlich Übertragung der Sprachdaten der Versuchsperson von Raum I nach Raum II, die Realisierung der Sprachausgabe des (simulierten) Systems und die Übertragung der Bilddaten wird in Anhang C ausführlicher erläutert.

5.2 Sprecherauswahl

Bei den 50 Sprechern in Szenario I und II handelt es sich um Studierende unterschiedlicher Fakultäten im Alter von 20 bis 37 Jahren, die über Aushänge in der Universität gesucht wurden. Auf diesen Aushängen wurden Versuchspersonen zur Überprüfung eines sprachverstehenden Systems gesucht (Szenario I) bzw. im Fall der Kontrollgruppe (Szenario II) wurden Versuchspersonen zum Testen neuer Kommunikationskanäle gesucht. Voraussetzung für die Teilnahme an den Versuchen war Deutsch als Muttersprache.

Die Anzahl der weiblichen Sprecher lag ca. bei einem Drittel in beiden Szenarien. Abbildung I zeigt die Verteilung von Alter und Geschlecht der 50 Versuchspersonen in den beiden Szenarien. Die Sprechereigenschaften Alter und Geschlecht sind nicht nur phonetisch interessant, sondern können auch das sprachliche Verhalten sowie die Bewertung des simulierten Systems (vgl. Kapitel 6.2) beeinflussen.

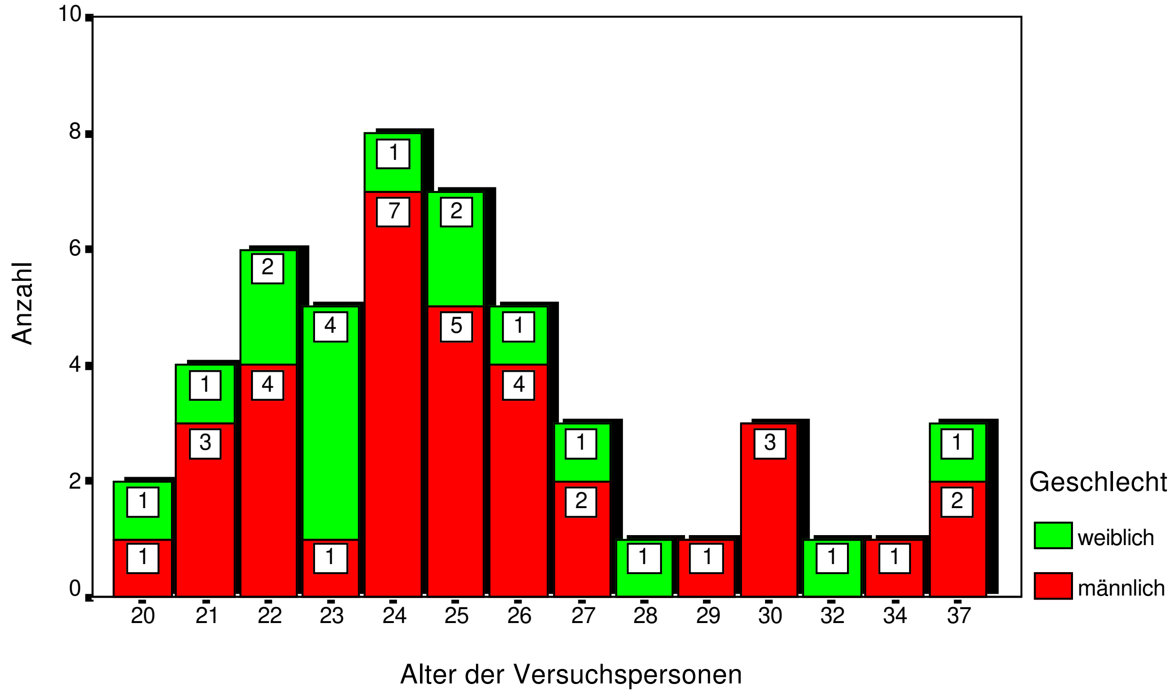

Abb. 9: Verteilung von Alter und Geschlecht der Versuchspersonen.

5.3 Durchführung der Aufnahmen

Der Ablauf der Dialogaufnahmen gliederte sich in beiden Szenarien in drei Phasen, die für die Versuchsperson jeweils auch räumlich voneinander abgegrenzt waren. Phase 1 und 3 fanden an einem Tisch in Raum I, auf dem sich die schriftliche Versuchseinleitung, sowie das Baufixset befanden, statt. Die zweite Phase fand an einem Schreibtisch ebenfalls in Raum I statt, auf dem sich der Bildschirm, sowie die entsprechenden Geräte für die Aufnahme der Daten befanden (DAT-Rekorder, Mikrophonvorverstärker etc.). Abbildung 10 zeigt die entscheidenden Interaktionen in den verschiedenen Phasen des Versuches. In Phase 1 und 3 interagieren der Versuchsleiter und die Versuchperson (s.u.), während in Phase 2 die Versuchsperson mit dem simulierten System interagiert (s.u.). Dabei kann der Versuchsleiter die stattfindende Kommunikation verfolgen, ist aber selbst nicht direkt beteiligt.

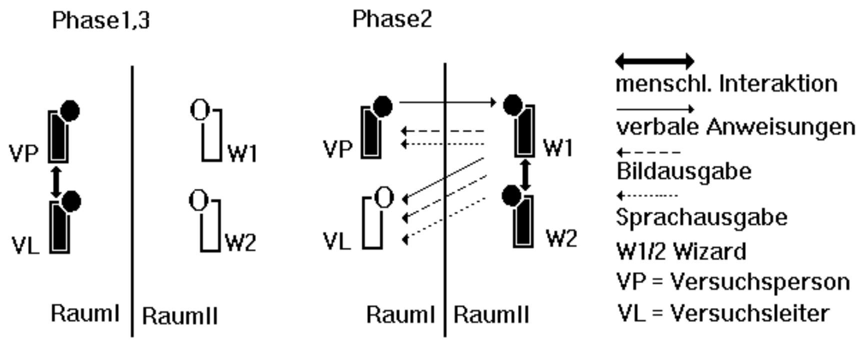

Abb. 10: Schwerpunkte der Interaktion zwischen den beteiligten Personen in den verschiedenen Phasen des Versuches

Phase1:

In der ersten Phase wurden zunächst vom Versuchsleiter phonetisch relevante Sprecherdaten (wie Alter, Geschlecht etc.) erfaßt. Im Anschluß daran las die Versuchsperson die schriftliche Versuchseinleitung (s. Anhang B) und baute dann das Baufixflugzeug anhand der dem Bausatz beiliegenden Bauanleitung zusammen. Stimmte das zusammengebaute Flugzeug in Form und Farbe nicht mit der Abbildung auf der Bauanleitung überein, wurde die Versuchsperson vom Versuchsleiter darauf aufmerksam gemacht und gebeten, die entsprechenden Korrekturen vorzunehmen.

Für die beiden Projektmitarbeiter, die die Systemsimulation bedienten, war die Phase1 eine Wartephase: Der Kontrollbildschirm, von dem aus die Simulation gesteuert wurde, war bereits eingerichtet und die Übertragung der Sprachdaten von Raum I nach Raum II bestand, so daß die Mitarbeiter verfolgen konnten, wann die zweite Phase, in der sie das System zu simulieren hatten, anfing.

Phase2:

Zu Beginn der zweiten Phase setzte sich die Versuchsperson an den Schreibtisch, auf dem der Bildschirm stand, auf welchen später während des Dialogs die vom Konstrukteur durchgeführten Bauschritte als Standbild übermittelt wurden. Zur Erhöhung der Glaubwürdigkeit des Szenarios stand zu Beginn der Mensch-Maschine-Interaktion eine fingierte Systemmeldung als Standbild auf dem Bildschirm.

Die Versuchsperson wurde aufgefordert, während des Dialogs auf keinen Fall mit dem Versuchsleiter zu sprechen und ihr Mikrophon nicht zu berühren. Dann wurde sie gebeten, das (simulierte) System mit dem Befehl „Weiter bitte." zu starten.

Nach dem Starten des (simulierten) Systems hörte die Versuchsperson eine durch das Sprachsynthesegerät realisierte Versuchseinweisung, bei der es sich inhaltlich um eine verkürzte Version der schriftlichen Versuchseinleitung handelt (s. Anhang B). Ziel dieser Versuchseinführung durch den Rechner war, daß die Versuchsperson sich in die Qualität des Synthesegerätes einhören konnte, damit akustische Verständnisschwierigkeiten auf Seiten der Versuchsperson während des nun folgenden Dialoges möglichst reduziert wurden. Die durch das Sprachsynthesegerät realisierte Versuchseinweisung, ebenso wie die fingierte Systemmeldung entfielen in den Dialogen der Kontrollgruppe, um deren Annahme, daß sie mit einer zweiten Person über etwas ungewöhnliche Kommunikationswege kommunizierten, nicht unnötig zu widerlegen.

Nach der Gewöhnung an die akustische Qualität des Sprachsynthesegerätes wies die Versuchsperson den Konstrukteur (Rechner/bzw. Mensch) entweder nach der Original-Bauanleitung oder nach dem von ihr zusammengebauten Flugzeug über Mikrophon an, das Flugzeug zusammenzubauen. Während dieser Zeit übernahm einer der Mitarbeiter in Raum II die manuelle Bautätigkeit, während der zweite Projektmitarbeiter über die Bedienung des Kontrollbildschirms die zeitliche Koordination der Sprach- und Bildausgabe des (simulierten) Systems übernahm. Welche Äußerungen der Versuchsperson als „nicht verstanden“ zurückgewiesen wurden und welche Textbausteine als sprachlicher Output des Systems zurückgegeben wurden, wurde von beiden Mitarbeitern gemeinsam entschieden. Es ergaben sich während der 50 Versuche in den Szenarien I und II neun verschiedene Mitarbeiterkonstellationen für die Bedienung der Systemsimulation. Damit wurden individuelle Faktoren der Konstellationen weitgehend nivelliert.

Phase3:

Nach Abschluß der Dialogaufnahme füllte die Versuchsperson einen Fragebogen aus, in dem ihre Expertise im Umgang mit Rechnern, sowie ihre Einschätzungen des simulierten Systems ermittelt wurden. Dann wurde die Versuchsperson gefragt, ob sie während des Versuches davon ausgegangen sei, mit einem Rechner zu kommunizieren. Im Anschluß daran wurde die Versuchsperson im Szenario I darüber aufgeklärt, daß es sich bei dem sprachverstehenden System um eine Simulation gehandelt habe. Ihr wurden die Ziele des Versuches erklärt, und es wurde ihr Stillschweigen für die Dauer von einigen Wochen erbeten, damit die Versuche im Szenario I nicht durch „vorgewarnte" Versuchsteilnehmer verfälscht würden. Die Versuchspersonen beider Szenarien wurden gegen Quittung für ihre Teilnahme am Experiment entlohnt.

6 Korpusbeschreibung

Insgesamt umfassen die Sprachaufnahmen (ohne die Aufnahmen der Trainingsphase) ca. 30 Stunden (simulierter) Mensch-Maschine-Kommunikation und ca. 5 Stunden Mensch-Mensch-Kommunikation. Die Stundenzahlen beinhalten die Pausen für die Systemreaktionszeiten. Die Sprachdaten liegen z.Z. auf Audio-DATs (Dialoge mit Konstrukteur- und Instrukteuräußerungen) und DAT-Streamer-Bändern (nur die Instrukteuräußerungen) vor.

6.1 Transkriptionen

Aus den insgesamt aufgenommenen 50 Dialogen wurden zunächst 22 Dialoge simulierter Mensch-Maschine-Kommunikation und alle 10 Dialoge der Kontrollgruppe orthographisch transkribiert. Die Verschriftlichung dieser 32 Dialoge erfolgte nach der Konvention von Fink/Johanntokrax/Schraffanietz (1995). Um die Anzahl nicht korrekter Zuweisungen bei den Transkriptionen möglichst gering zu halten, wurde jede erstellte Transkription anhand der Sprachaufnahmen nochmals überprüft und gegebenenfalls korrigiert (Brindöpke et al. 1995a). Gleichzeitig wurde eine Lexikondatenbank aufgebaut, die u.a. für jeden Eintrag die Standardaussprache vermerkt.

6.2 Fragebogenauswertung

Der Fragebogen, den die Versuchsperson im Anschluß an die Sprachaufnahmen ausfüllten, gliederte sich in die Teile A, B und C. Bei den meisten Fragen entschied sich die Versuchsperson für einen Wert auf einer Skala von 0 bis 4 zwischen „trifft zu" und „trifft nicht zu". Einige Fragen erforderten schriftlich formulierte Antworten der Versuchsperson.

Ziel des A-Teils war die Erfassung der Expertise der Versuchsperson im Umgang mit Rechnern. Der Teil B umfaßte Fragen zur Softwareergonomie des simulierten Systems (Fehlerfreundlichkeit, Güte der Grafikausgabe, Antwortzeiten etc.) und allgemeine Fragen zur Mensch-Maschine-Kommunikation, sowie Fragen nach den Annahmen, die die Versuchspersonen über die Fähigkeiten von sprachverstehenden Systemen machen. Die Fragen des B-Teils wurden in randomisierter Reihenfolge dargeboten und traten jeweils doppelt formuliert auf, um mögliche Links-/Rechtseffekte beim Ankreuzen zu vermeiden. Im Teil C des Fragebogens wurde die Versuchsperson zu ihren persönlichen Eindrücken bzw. ihrer Akzeptanz des simulierten Systems befragt. Zum einen konnte sie sich zwischen Gegensatzpaaren wie „langweilig/anregend“ auf einer Skala von -2 bis 2 entscheiden, zum anderen wurde sie gebeten, schriftlich einige Fragen zu beantworten wie z.B. „Was gefällt Ihnen an dieser Art der Interaktion mit einem Computer?" (Muster des Fragebogens s. Anhang B). Der für die Kontrollgruppe (Szenario II) verwendete Fragebogen umfaßte die Teile A und C, wobei der C-Teil in den Formulierungen an das Szenario II angepaßt wurde. Die Fragen des B-Teiles mußten entfallen, da sie das Vorhandensein eines sprachverstehenden Systems voraussetzen. Im folgenden soll eine Teilauswertung der Fragebögen bezogen auf einige wichtige Teilaspekte vorgestellt werden.

a) Glaubwürdigkeit der Simulation

Von besonderer Bedeutung für die weitere Bearbeitung und Untersuchung des Sprachmaterials ist die Frage, ob die Simulation gelungen ist, d.h. ob die Versuchspersonen im Szenario I davon ausgegangen sind, mit einem Rechner zu kommunizieren. Gehen die Versuchspersonen nicht davon aus, mit einem Rechner zu kommunizieren, so kann das Material nicht als Mensch-Maschine-Interaktion bewertet werden. Um diese Frage zu klären, wurden die Versuchspersonen im Anschluß an das Ausfüllen des Fragebogens befragt. Von den 40 Versuchspersonen in Szenario I waren nur drei Personen (7,5%) der Ansicht, nicht mit einem Rechner gesprochen zu haben. Damit kann das aufgenommene Material als erfolgreich simulierte Mensch-Maschine-Interaktion betrachtet werden. Aufgrund der relativ kleinen Anzahl von Versuchspersonen, die nicht davon ausgegangen sind, mit einem Rechner zu sprechen, können weitere diese Gruppe betreffende Fragestellungen, z.B. die Frage, ob sich ihr sprachliches Verhalten von dem der anderen Gruppe unterscheidet, oder ob sie prinzipiell andere ethische Bewertungen von Mensch-Maschine-Kommunikation zugrundelegen, statistisch nicht bearbeitet werden[2].

b) Expertise der Versuchspersonen

Die Expertise der Versuchspersonen bezogen auf ihren allgemeinen Umgang mit Rechnern war sehr unterschiedlich. Das Histogramm zeigt eine zweigipflige Verteilung: Entweder fallen die Versuchspersonen in die Kategorie Experte oder Novize. Die drei Personen, die während des Versuchs nicht davon ausgingen, mit einem Rechner zu kommunizieren, gehörten in die Gruppe der Novizen.

[2] Auffällig war, daß alle drei Probanden Soziologie studierten und wenig Expertise im Umgang mit Rechnern besaßen.

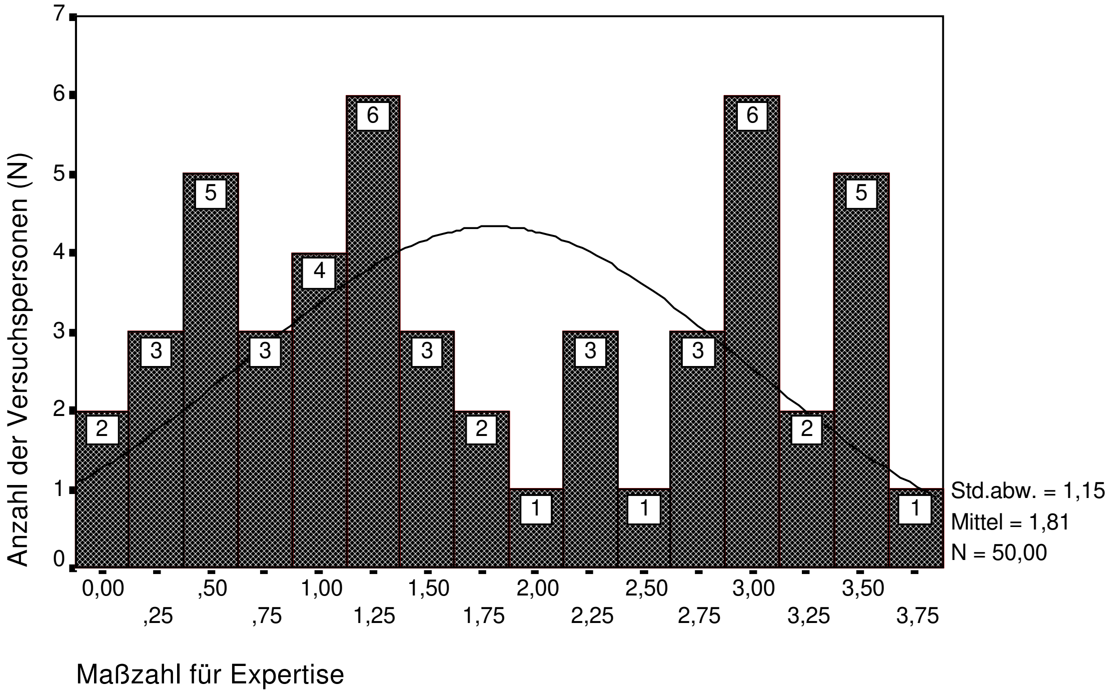

Abb. 11: Das Histogramm wurde berechntet über die Punktzahl, die die jeweilige Versuchsperson im A-Teil des Fragebogens erreicht hat, geteilt durch die Anzahl der vorgelegten Fragen (Maßzahl für Expertise).

c) Softwareergonomie des simulierten Systems

Die Auswertung der Fragen zur Softwareergonomie im Szenario I zeigen insgesamt eine leicht positive Bewertung des simulierten Systems durch die Versuchspersonen (Abb. 12). So wurde die Grafikausgabe im allgemeinen für eine gelungene technische Leistung gehalten, ebenso wurde das Verhalten des simulierten Systems bezogen auf die Verstehensleistung positiv bewertet. Die Antwortzeiten und der Beitrag des Systems zur Problemlösung (Zusammenbau des Baufixflugzeuges) wurden durchgängig als schlecht bewertet. Die drei Versuchspersonen, die während des Versuches nicht davon ausgingen, mit einem Rechner zu kommunizieren, wurden für die Berechnung der Abbildung nicht berücksichtigt. Bei weiteren Iterationen des Szenarios sind in der Beurteilung der softwareergonomischen Fähigkeiten des simulierten Systems Schwankungen zu erwarten. Z.B. würde bei Einsatz eines realen spracherkennenden Systems in einer weiteren Iteration die Beurteilung der sprachverstehenden Leistung schlechter werden. Ziel bei einer weiteren Ersetzung des simulierten Systems durch realisierte Systemeinheiten sollte jedoch eine möglichst hohe Bewertung in den Fragen zur Softwareergonomie sein.

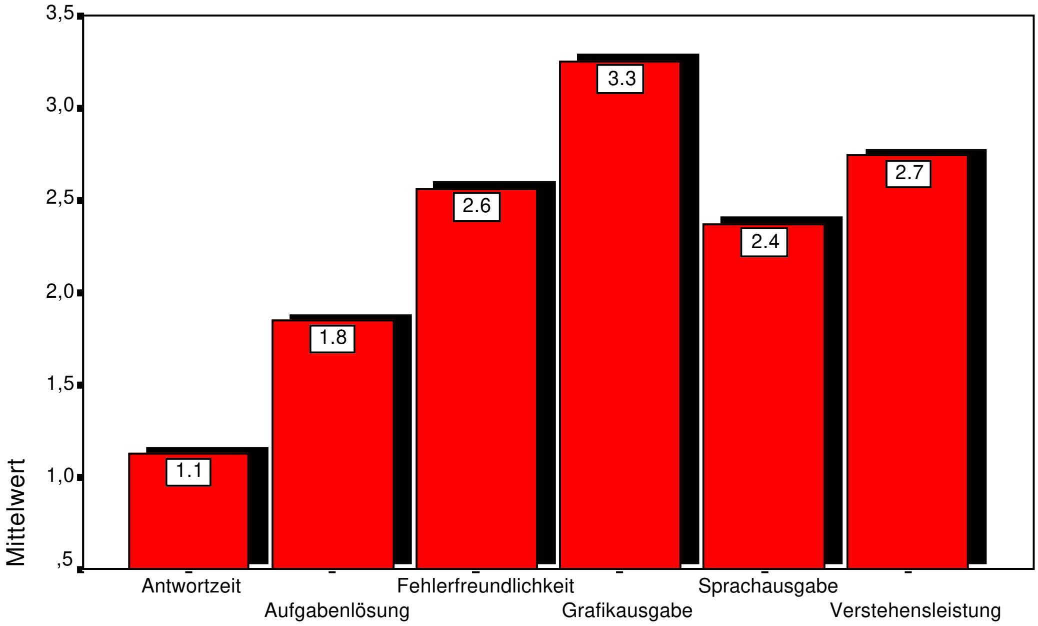

Abb. 12: Softwareergonomische Eigenschaften der Systemsimulation

d) Akzeptanz und Beurteilung des simulierten Systems

Die Akzeptanztests im C-Teil weisen für die 37 Versuchspersonen der Mensch-Maschine-Kommunikation, die während der Versuche davon ausgegangen sind, mit einem Rechner zu kommunizieren, folgende Mittelwerte bei der Bewertung des simulierten sprachverstehenden Systems auf (Abb. 13):

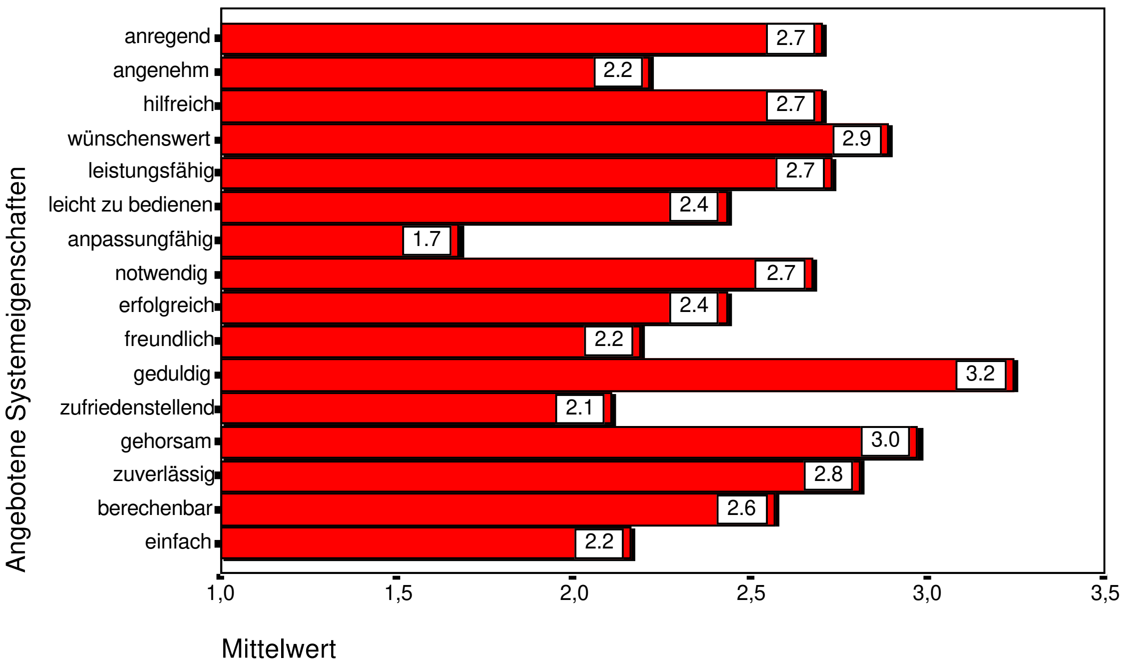

Abb. 13: Einschätzungen des simulierten sprachverstehenden Systems in Szenario I

Die Anpassungsfähigkeit des Systems wird relativ schlecht bewertet, besonders auffällig ist die deutliche Bewertung des Systems als „geduldig“.
Die Bewertungen der Systemeigenschaften durch die 10 Versuchspersonen der Mensch-Mensch-Interaktion unterscheiden sich von den Bewertungen der Versuchspersonen in der Mensch-Maschine-Interaktion (Abb.14).

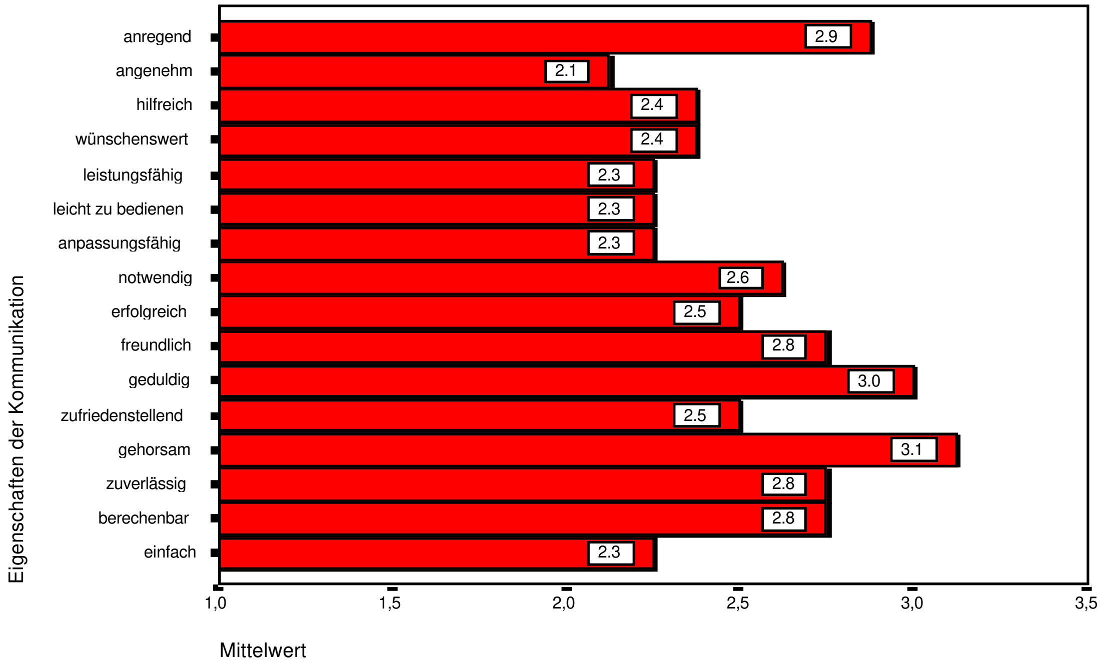

Abb. 14: Einschätzungen des Kommunikationsweges durch die Kontrollgruppe

So werden von der Kontrollgruppe die Eigenschaften „hilfreich“, „wünschenswert“ und „leistungsfähig“ deutlich weniger zutreffend empfunden als im Mensch-Maschine-Kontext. Gleichzeitig bewerteten die Versuchspersonen die Kommunikation stärker als „freundlich“ und „zufriedenstellend“. Diese Unterschiede weisen darauf hin, daß sich nicht nur das sprachliche Verhalten in Mensch-Mensch- und Mensch-Maschine-Interaktion unterscheidet, sondern auch die Bewertungsmaßstäbe, die die Probanden an die Mensch-Maschine-Kommunikation anlegten, sich von ihren Bewertungsmaßstäben für Mensch-Mensch-Kommunikation unterscheiden. Dabei wird derselbe Stil von Kommunikation unter dem Etikett der Mensch-Mensch-Kommunikation offensichtlich in manchen Aspekten strenger beurteilt. Das läßt auf einen gewissen „Maschinenbonus“ bei der Einschätzung der Mensch-Maschine-Kommunikation schließen.

7 Schlußbemerkungen

Die Entwicklung der Szenarien erfolgte nach einem klar definierten methodischen Vorgehen (vgl. Kapitel 3). Dabei waren bereits die Vorannahmen bezogen auf die Variablenfestlegung (Produktion, Erkennnung etc.) des zu simulierenden Systems durch Voruntersuchungen an Mensch-Mensch-Dialogen derselben Domäne, sowie der Durchführung und Auswertung von Sprachaufnahmen in einer Trainingsphase stark empirisch fundiert (vgl. Kapitel 4.1, 4.2).

Die entwickelte Systemsimulation beinhaltete die Benutzerschnittstelle eines Dialogsystems mit akustischer Spracheingabe und -ausgabe, sowie eine Bildkomponente für die Simulation des Bauvorganges eines Baufixflugzeuges und integrierte damit anders als bisherige natürlichsprachliche Simulationen Bild und Sprache (vgl. Kapitel 4.2, 4.3, Anhang C)

Die Auswertungen der Fragebögen in Kapitel 6.2 zeigen, daß in den Szenarien auf der Basis der Simulation eine erfolgreiche Simulation von Mensch-Maschine-Interaktion gelungen ist. Damit ist das Sprachmaterial geeignet für Untersuchungen der Mensch-Maschine-Interaktion in der vorliegenden Domäne. Aufgrund der Menge des Sprachmaterials (ca. 4700 Turns der Instrukteure) bietet es sich insbesondere für das Training statistischer Spracherkenner sowie für Untersuchungen mit statistischen Methoden an. 32 der insgesamt 50 Dialoge, darunter die 10 Dialoge der Kontrollgruppe, liegen bereits als Transkripte in einem Materialband vor (Brindöpke et al. 1995a).

Eine Weiterentwicklung der Simulation im Hinblick auf die schrittweise Erstellung eines Zielsystems erfordert auf einer konzeptuellen Ebene die explizite Definition weiterer Variablen des Wizard-of-Oz-Szenarios. Dazu sind genauere Untersuchungen des sprachlichen Verhaltens in Mensch-Maschine-Interaktion der Baufixdomäne sowie weitere Überlegungen zu den Prinzipien des Systemdesigns notwendig. Eine genauere Auswertung des sprachlichen Verhaltens von simuliertem System und Instrukteur, insbesondere auch im Vergleich zum sprachlichen Verhalten der Mensch-Mensch-Kontrollgruppe oder in bezug auf geschlechtsspezifische Merkmale wird zur Zeit erarbeitet. In bezug auf die Gestaltung des sprachlichen Verhaltens von Dialogsystemen wird das Spannungsverhältnis zwischen der vorherrschenden Strategie, den Mensch-Maschine-Dialog soweit wie möglich durch das System „führen“ zu lassen und dem Ideal eines flexiblen und sich an das sprachliche Verhalten des Nutzers anpassenden Dialogsystems näher untersucht werden müssen. Dies könnte auch die weitergehende Untersuchung von Unterschieden in der Bewertung von Mensch-Mensch und Mensch-Maschine-Kommunikation umfassen.

Desweiteren ist die technische Realisierung von Teilmodulen eines künstlichen Kommunikators oder deren Integration in das simulierte System notwendig. So könnte die Erkennungsleistung des (simulierten) Systems durch die Integration eines Spracherkenners spezifiziert werden. Außerdem könnte die sprachliche Produktion flexibler gestaltet werden z.B. mittels geeigneter Produktionsregeln unter Zuhilfenahme einer Wissensbasis sowie eines weiterentwickelten Dialogmodells. Eine Integration von realisierten Teilmodulen eines künstlichen Kommunikators erfordert darüberhinaus eine weitere Verbesserung des simulierten Systems bezogen auf seine Reaktionszeiten.

8 Literaturhinweise

Brindöpke, C./Johanntokrax, M./Pahde, A./Wrede, B: (1995a): *Darf ich dich Marvin nennen? Instruktionsdialoge in einem Wizard-of-Oz Szenario: Materialband*. Report 95/7, SFB 360 „Situierte Künstliche Kommunikatoren“, Universität Bielefeld.

Andernach, T./Deville, G./Mortier, L. (1993): The design of a real world Wizard of Oz experiment for a speech driven telephone directory information system, in: *Proceedings of the 3rd European Conference on Speech Communikation and Technology*, pp. 1165-1168.

Baum L.F. (1900): *The Wonderful Whizard of Oz*. Chicago: G.M.Hill.

Dahlbäck, N./Jönsson, A. (1989): Emirical studies of discourse representations for natural language interfaces, in: *Proceedings of the 4th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 291-298.

Dalsgaard, P./Baekgaard, A. (1994): Spoken Language Dialogue Systems, in: Niemann, H./de Mori, R./Hanreider, G. (eds.): *Progress and prospects in speech and language technology*. Sankt Augustin: Infix, pp. 178-191.

Delomier, D./Meunier, A./Morel, M.A. (1989): Linguistic features of human-machine oral dialogue, in: *Proceedings of the 3rd European Conference on Speech Communikation and Technology.*

Dybkjaer, H./Bernsen, N.O./Dybkjaer, L. (1993): Wizard-of-Oz and the trade-off between naturalness and recogniser constraints, in: *Proceedings of the 4th European Conference on Speech Communikation and Technology*, pp. 947-950.

Fink, G.A./Johanntokrax, M./Schaffranietz, B. (1995a): A Flexible Formal Language for the Orthographic Transcription of Spontaneous Spoken Dialogues, in: *Proceedings of the 4th European Conference on Speech Communication and Technology,* pp. 871-874.

Fraser, N.M./Gilbert, G.N. (1991): Simulating speech systems, in: *Computer Speech and Language*, pp. 81-99.

Gibbon, D. (1995): Empirical and Semiotic Foundations for Prosodic Analysis, in: Quasthoff, U.M. (ed.): *Aspects of oral communication*. Berlin: De Gruyter, pp.441-479.

Guindon, R.(1987): Grammatical and ungrammatical structures in user-adviser dialogues: Evidence for sufficiency of restricted languages in natural language interfaces to advisory systems, in: *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*, pp. 41-44.

Hauptmann, A.G. (1989): Speech and gestures for graphic image manipulation, in: *CHI ‘89*, pp. 241-245.

Hauptmann, A.G./Rudnicky, A.I. (1988): Talking to computers: an empirical investigation, in: *International Journal of Man-Machine Studies*, pp. 583-604.

Kelley, J.F. (1983): An empirical methodology for writing user-friendly natural language computer applications, in: *CHI ‘83*.

Kelley, J.F. (1984): An interactive design methodology for user-friendly natural language office information applications, in: *Association for Computing Machinery Transactions on Office Information Systems 2*, pp. 26-41.

Labrador, C./Dinesh P. (1984): Experiments in speech interaction with conventional data service, in: *Interact ‘84*, pp. 104-108.

Luzzatti, D./Neel, F. (1989): Dialogue behaviour induced by the machine, in: *Proceedings of the 3rd European Conference on Speech Communikation and Technology*, pp. 601-604.

MacDermid, C. (1993): Features of naive callers’ dialogues with a simulated speech understanding and dialogue system, in: *Proceedings of the 4th European Conference on Speech Communikation and Technology*, pp. 955-958.

Sagerer, G./Eikmeyer, H.J./Rickheit, G. (1994): *„Wir bauen jetzt ein Flugzeug“: Konstruieren im Dialog*. Arbeitsmaterialien. Tech. Rep., SFB 360 „Situierte Künstliche Kommunikatoren“, Universität Bielefeld.

Whittaker, S./Stenton, P. (1989): User studies and the design of natural language systems, in: *Proceedings of the 4th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 116-123.

Anhänge

Anhang A: Materialien zur Tainingsphase

Anhang B: Materialien zu den Szenarien I und II

Anhang C: Technische Realisierung der Systemsimulation

A1 Spezifikation Szenario 0

SET-UP:

Geräte:	drei Unix-Workstations (Dec-Workstation 5000), ein PC mit angeschlossenem Bildsequenzsystem, eine Stereokamera (3-Chip-RGB-CCD), zwei Gradient Desklab 216, ein Sprachsynthetisator SVS 3105, Ethernet; Mischpult, Dat-Rekorder, Vorverstärker, Headset, Kopfhörer; zwei Räume (Raum der Versuchsperson, Raum des Wizards); zwei Baufixset, Fragebögen etc.
Personen:	Versuchsleiter, Personen für die Bedienung des simulierten Systems (Wizard), Versuchspersonen

VERSUCHSLEITER:

Aufgabe:	Durchführung und Überwachung des Versuchsablaufs, Bedienung der Aufnahmegeräte
Personen:	ein Projektmitarbeiter

WIZARD:

Aufgabe1:	Zusammenbau des Baufixflugzeuges nach den Anweisungen der Versuchsperson
Aufgabe2:	Koordination der Bild-und Sprachübermittlung an die Versuchsperson
Personen:	je ein Projektmitarbeiter für Aufgabe1 und 2
Tools:	Kontrollbildschirm zur Steuerung der Simulation, Software der Simulation; Baufixset
Produktion:	Sprache: festgelegte Textstrings, die über ein Bedienfeld auf dem Kontrollbildschirm ausgewählt und an das Sprachsynthesegerät weitergegeben werden, sowie freie Texteingabe beliebiger Textstrings Bild: Standbilder vom jeweiligen Bauzustand, die via Ethernet auf den Bildschirm der Testperson geschickt werden
Erkennung:	akustische Erkennung: menschliche Erkennungsleistung beeinflußt durch die Qualität der Sprachdatenübertragung Erkennung Lexikon/Syntax/Pragmatik: menschliche Erkennungsleistung bei Reduktion der natürlichen Sprecherkooperativität
Wissen:	Handhabung der Simulation, Zusammenbau des Flugzeugs, Reduktion der natürlichen Sprecherkooperativität (Einbau von Erkennungsfehlern u.a.)
Training:	-

VERSUCHSPERSONEN:

Personen:	Studierende verschiedener Fakultäten
Anzahl:	4
Aufgabe:	einen Rechner verbal anzuweisen, ein Baufixflugzeug zusammenzubauen;;
Vorwissen:	einmaliger Zusammenbau des Flugzeugs nach der Bedienungsanleitung,keine Erfahrung mit dem simulierten System

Abb. A1: Szenario-Spezifikation für das Szenario der Trainingsphase (Szenario 0).

A2 Textbausteine (Set A) und ihr Vorkommen in der Trainingsphase:

TEXTBAUSTEIN	Vorkommen insgesamt	Anzahl der Gespräche
Welche Farbe hat die Schraube?	1	1
Welche Form hat die Schraube?	-	-
Welche Farbe hat der Würfel?	-	-
Wieviele Löcher hat die Lochleiste?	-	-
Welches Loch meinen Sie bitte?	3	1
An welcher Seite des Baufixteiles soll die Operation ausgeführt werden?	1	1
In welchem Winkel sollen die Baufixteile zueinander stehen?	6	3
Wie meinen Sie das bitte?	-	-
Einen Augenblick bitte.	95	4
Ich habe verstanden.	118	4
Der Arbeitsschritt ist beendet. Bitte fahren Sie fort.	8	4
Bitte formulieren Sie Ihre Anweisung neu.	1	1
Ihre Anweisung wird durchgeführt.	6	2
Ihre Anweisung wird noch bearbeitet.	20	3
Ihre Anweisung ist zu komplex. Bitte gehen Sie in kleineren Schritten vor.	3	2
Ihre Anweisungen sind zu ungenau. Bitte geben Sie konkretere Instruktionen.	10	3
Ihre Anweisung kann nicht durchgeführt werden. Bitte formulieren Sie neu.	-	-

Gesamtzahl der Wizardäußerungen: 310; "Freie" Wizardäußerungen: 38 (ca. 12%)

A3 Auflistung der "freien" Wizard-Äußerungen der Trainingsphase:

Die "freien" Äußerungen sind in der Reihenfolge ihres Auftretens in den Gesprächen aufgelistet. Die beiden Großbuchstaben zu Beginn eines jeden Blockes sind die Sprecherkürzel.

MP:

Bitte sprechen Sie etwas lauter Punkt ö.
Bitte sprechen Sie lauter.
Bitte beginnen Sie jetzt mit Ihren Anweisungen.
Das ist die runde Schraube.
Mit welcher Schraube soll die Leiste befestigt werden.
Vielen Dank.

XY:

Bitte beginnen Sie jetzt mit Ihren Anweisungen.
Welche Fünflochleiste meinen Sie?
Mit welchem Loch?
Durch welche Löcher sollen sie verbunden sein?
Welche Leiste meinen Sie bitte?
Diese Operation kann nicht durchgeführt werden.
Mit welcher Schraube?
Diese Operation ist nicht möglich.
Es gibt nur eine kantige Schraube.
An welchen Seiten bitte?
Vielen Dank.

KB:

Bitte beginnen Sie jetzt mit ihren Anweisungen.
Bitte beginnen Sie jetzt mit ihren Anweisungen.
In das Loch welcher Leiste bitte?
Welches erste Loch meinen Sie bitte?
Welche Mutter meinen sie bitte?
Soll der grüne Klotz abmontiert werden?
Mit welcher Mutter bitte?
Da ist kein Loch frei.
Da ist kein blauer Klotz.
Es gibt keine rote eckige Mutter.
Es gibt keinen lila Plastikring.
Vielen Dank für Ihre Mitarbeit.

PS:

Bitte beginnen Sie mit Ihren Instruktionen.
Leiste kann nicht auf den Würfel gelegt werden.
Was meinen Sie mit Bild?
Bitte geben Sie genauere Anweisungen.
Ich habe keine sechseckige rote Schraube.
Diese Anweisung kann nicht durchgeführt werden.
Woran sollen die Leisten montiert werden?
Bitte fahren Sie fort.
Vielen Dank.

B1 Spezifikation der Szenarien I und II

SET-UP:

Geräte:	drei Unix-Workstations (Dec-Workstation 5000), ein PC mit angeschlossenem Bildsequenzsystem, eine Stereokamera (3-Chip-RGB-CCD), zwei Gradient Desklab 216, ein Sprachsynthetisator SVS 3105, Ethernet; Mischpult, Dat-Rekorder, Vorverstärker, Headset, Kopfhörer; zwei Räume (Raum der Versuchsperson, Raum des Wizards); zwei Baufixset, Fragebögen etc.
Personen:	Versuchsleiter, Personen für die Bedienung des simulierten Systems (Wizard), Versuchspersonen

VERSUCHSLEITER:

Aufgabe:	Durchführung und Überwachung des Versuchsablaufs, Bedienung der Aufnahmegeräte
Personen:	ein Projektmitarbeiter

WIZARD:

Aufgabe1:	Zusammenbau des Baufixflugzeuges nach den Anweisungen der Versuchsperson
Aufgabe2:	Koordination der Bild-und Sprachübermittlung an die Versuchsperson
Personen:	je ein Projektmitarbeiter für Aufgabe1 und 2
Tools:	Kontrollbildschirm zur Steuerung der Simulation, Software der Simulation; Baufixset
Produktion:	Sprache: festgelegte Textstrings, die über ein Bedienfeld auf dem Kontrollbildschirm ausgewählt und an das Sprachsynthesegerät weitergegeben werden Bild: Standbilder vom jeweiligen Bauzustand, die via Ethernet auf den Bildschirm der Testperson geschickt werden
Erkennung:	akustische Erkennung: menschliche Erkennungsleistung beeinflußt durch die Qualität der Sprachdatenübertragung Erkennung Lexikon/Syntax/Pragmatik: menschliche Erkennungsleistung eingeschränkt durch die im Training geübten Verhaltensweisen
Wissen:	s.Training
Training:	Handhabung der Simulation, Zusammenbau des Flugzeugs, Reduktion der natürlichen Sprecherkooperativität (Einbau von Erkennungsfehlern u.a.)

VERSUCHSPERSONEN:

Personen:	Studierende aller Fakultäten
Anzahl:	40 (**Szenario I**) 10 (Kontrollgruppe in **Szenario II**)
Aufgabe:	**Szenario I**: einen Rechner verbal anzuweisen, ein Baufixflugzeug zusammenzubauen; **Szenario II**: eine in einem anderen Raum befindliche Person anzuweisen, ein Baufixflugzeug zusammenzubauen;
Vorwissen:	einmaliger Zusammenbau des Flugzeugs nach der Bedienungsanleitung, schriftliche Versuchseinführung, keine Erfahrung mit dem simulierten System

Abb. B1: Spezifikation der Szenarien I und II

B2 Schriftliche Versuchseinführung Szenario I:

In diesem Versuch geht es um den Zusammenbau eines Baufixflugzeuges. Zunächst bauen Sie das Flugzeug anhand der Anleitung zusammen. Dabei achten Sie bitte auf eine Übereinstimmung in Form und Farbe mit der Anleitung.
Dann weisen Sie anhand Ihres zusammengebauten Flugzeuges oder der Anleitung ein sprachverstehendes Programm an, das Baufixflugzeug zusammenzubauen. Dabei erhalten Sie außer sprachlichen Systemäußerungen auch einzelne Bauabschnitte als Bild auf dem Bildschirm zur visuellen Kontrolle. Die Darstellung auf dem Bildschirm kann um 90° bzw. 180° gedreht werden.
Nachdem das System nach ihren Anweisungen das Baufixflugzeug zusammengebaut hat, füllen Sie einen Fragebogen zu Ihren gemachten Erfahrungen aus.
Bitte beachten Sie während der Phase des Versuchs, in der Sie den Rechner anweisen, das Baufixflugzeug zusammenzubauen, zwei Regeln:

I. **W I C H T I G** ist, daß Sie während des Zusammenbaus des Flugzeugs mit dem Rechner **auf keinen Fall** mit dem Versuchsleiter reden. Das Programm ist selbsterklärend und benutzerfreundlich. Sollten Verständigungsschwierigkeiten aufgrund von Fehlern des Systems entstehen, so können Sie diese selbst durch Wiederholung oder Umformulierung ihrer Anweisungen beheben. Wartezeiten zwischen Ihren Anweisungen und der Reaktion des Rechners sind normal.
II. Ebenso **W I C H T I G** ist, daß Sie während dieser Phase des Versuches **auf keinen Fall** ihr Mikrophon verrücken oder berühren. Beim Beachten dieser beiden entscheidenen Grundregeln ist Ihnen der Erfolg sicher.

B3 Schriftliche Versuchseinführung Szenario II

Testen neuer Kommunikationskanäle:

In diesem Versuch geht es um den Zusammenbau eines Baufixflugzeuges. Zunächst bauen Sie das Flugzeug anhand der Anleitung zusammen. Dabei achten Sie bitte auf eine übereinstimmung in Form und Farbe mit der Anleitung.
Dann weisen Sie anhand Ihres zusammengebauten Flugzeuges oder der Anleitung eine zweite Person an, das Baufixflugzeug zusammenzubauen. Sie tun dies, indem Sie in das Mikro Ihres Headsets sprechen. Ihr Partner/Ihre Partnerin kann Ihnen die durchgeführten Bausschritte auf den Bildschirm zur visuellen Kontrolle senden. Außerdem erhalten Sie sprachliches Feedback über Ihren Kopfhörer.
Nachdem Sie mit Ihrem Partner/Ihrer Partnerin das Flugzeug zusammengebaut haben, füllen Sie einen Fragebogen zu Ihren gemachten Erfahrungen aus.
Bitte beachten Sie während der Phase des Versuchs, in der Sie mit Ihrem Partner/Ihrer Partnerin das Flugzeug zusammenbauen, zwei Regeln:

I. **W I C H T I G** ist, daß Sie, während Sie das Flugzeug mit Ihrem Partner/Ihrer Partnerin zusammenbauen, **auf keinen Fall** mit dem Versuchsleiter reden.
II. Ebenso **W I C H T I G** ist, daß Sie während dieser Phase des Versuches **auf keinen Fall** ihr Mikrophon verrücken oder berühren.

B4 Akustische Versuchseinführung durch das simulierte System

Guten Tag! Sie haben sich freundlicherweise bereiterklärt, bei der Erstellung einer Datenbasis zur Weiterentwicklung eines sprachverstehenden Systems mitzuwirken. Dieses System benötigt eine große Datenmenge als Training, um sprecherungabhängig arbeiten zu können. Es ist Teil eines Projektes zur Entwicklung eines sprachgesteuerten

Roboters. Bisher ist nur die Sprachsteuerung realisiert worden. Die Aufgabe des Roboters übernimmt eine Computersimulation, deren aktuelle optische Ausgabe Sie jeweils auf dem Bildschirm vor sich sehen werden. Ihre Aufgabe wird sein, den Computer verbal zu instruieren, das Flugzeug nachzubauen. Um sich mit der Konstruktion vertraut zu machen, haben Sie bereits das Flugzeug zusammengebaut. Achten Sie bei Ihren Anweisungen an den Rechner bitte darauf, daß die Farben und Formen der Einzelteile mit der Bauanleitung übereinstimmen. Bitte beginnen Sie jetzt mit den Bauanweisungen.

B5 Textbausteine (Set B) für die Szenarien I und II

Auf der linken Seite finden sich die Kürzel für die einzelnen Äußerungen. Rechts findet sich die jeweilige Äußerung, die dann bei Auswahl durch den Projektmitarbeiter, der die Systemsimulation bediente (Wizard), vom Sprachsynthesegerät realisiert wurde.

Kürzel	Textstring
META I	Bitte beginnen Sie jetzt mit Ihren Anweisungen.
META II	Der Arbeitsschritte ist beendet. Bitte fahren Sie fort.
META III	Vielen Dank.
WARTEN	Einen Augenblick bitte.
LAUTER	Bitte sprechen Sie lauter.
VERSTEHEN/JA	Ich habe verstanden.
VERSTEHEN/NEIN	Ich habe Sie leider nicht verstanden.
ERKENNEN/NEIN.	Ich konnte Ihre Äußerung leider nicht verstehen.
HANDLUNG/JA.	Ihre Anweisung wird bearbeitet.
HANDLUNG/NEIN	Ihre Anweisung kann leider nicht durchgeführt werden.
ZU_UNPRÄZISE	Ihre Anweisung ist zu ungenau.
ZU_KOMPLEX	Ihre Anweisung ist zu komplex.
UNBEKANNTE_WÖRTER	Ihre Anweisungen enthalten unbekannte Wörter.
ÜBERPRÜFEN_SIE	Bitte kontrollieren Sie Ihre Anweisung noch einmal und machen dann weiter.
OBJEKTSPEZIFIKATION	Bitte benennen Sie die gewünschten Objekte genauer.
BAUVORGANG	Bitte beschreiben Sie den gewünschten Bauvorgang genauer.
KLEINERE_SCHRITTE	Bitte gehen Sie in kleineren Schritten vor.
NEU_FORMULIEREN	Bitte formulieren Sie Ihre Anweisung neu.
WIEDERHOLEN	Bitte wiederholen Sie Ihre Anweisung.

B6 Fragebogen zu Szenario I (Mensch-Maschine-Interaktion):

Dem Fragebogen waren einige Hinweise zum Ausfüllen des Fragebogens vorangestellt. Insbesondere wurde um Beantwortung aller Fragen des Fragebogens gebeten. Im folgenden findet sich die Auflistung der Fragen, dabei werden die Leerzeilen für die schriftliche Formulierung der Antworten nicht wiedergegeben.

Teil A

(1) Ich verfüge über Erfahrungen mit mindestens einem Textverarbeitungssystem oder habe schon mit einem bzw. mehreren Editoren gearbeitet.
trifft zu ____: ____: ____: ____: ____: trifft nicht zu

(2) Ich habe auch schon mit anderen Anwendungen als Textverarbeitung gearbeitet, z.B. mit Statistik- oder Graphikprogrammen oder Tools wie z.B. dem Norton Commander oder anderen Anwendungen.
trifft zu ____: ____: ____: ____: ____: trifft nicht zu
(3) Ich habe Kenntnisse in einer oder mehreren Programmiersprachen.
trifft zu ____: ____: ____: ____: ____: trifft nicht zu
Bitte geben Sie an in welcher Programmiersprache Sie über gründliche Kenntnisse verfügen:
(4) Ich habe mit der Installation von Software (z.B. Installation von Textverarbeitungsprogrammen oder Betriebssystemen) schon Erfahrungen gemacht.
trifft zu ____: ____: ____: ____: ____: trifft nicht zu
(5) Ich habe Erfahrungen mit kleineren Erweiterungen oder Reparaturen an einem Rechner, d.h. ich könnte z.B. zusätzliche Speichereinheiten in einen Rechner einfügen oder Kleinigkeiten löten und ähnliches.
trifft zu ____: ____: ____: ____: ____: trifft nicht zu
(6) Ich habe beruflich, bzw. in meinem Studium mit Rechnern zu tun.
trifft zu ____: ____: ____: ____: ____: trifft nicht zu
In welchem/en Fachbereich/en?
(7) Ich arbeite schon länger mit Textverarbeitung oder anderen Anwendungen.
trifft zu ____: ____: ____: ____: ____: trifft nicht zu
Bitte kreuzen Sie an, in welchen Bereichen Sie Erfahrungen haben und schreiben Sie auf die Linie dahinter, wie lange (1/2 Jahr, 2 Jahre, mehr als zwei Jahre).
Textverarbeitung __
Statistik-,Graphikprogramme und andere Anwendungen __
Rechnerreparatur, Löten etc. __
Programmierung __
Softwareinstallation/ Systembetreuung __
(8) Wenn ich am Rechner arbeite, benutze ich häufiger die Kommandosprache und Alt-Tastenkombinationen als Menüs.
trifft zu ____: ____: ____: ____: ____: trifft nicht zu
(9) Ich habe auch schon mal Hardware installiert (z.B. RAM, Karten, Drucker)
trifft zu ____: ____: ____: ____: ____: trifft nicht zu
(10) Ich kenne einige Werkzeuge, die man braucht, um einen Rechner zu reparieren, z.B. könnte ich jemandem erklären, was eine Abisolierzange ist.
trifft zu ____: ____: ____: ____: ____: trifft nicht zu

Teil B

(1) Es war eher schwierig, Fehler beim Zusammenbau des Flugzeugs oder Fehler in der Verständigung zu korrigieren.
trifft zu ___: ___: ___: ___: ___: trifft nicht zu
(2) Das Programm hat die Lösung der gestellten Aufgabe, nämlich den Rechner anzuweisen, das Baufixflugzeug zusammenzubauen, für mich leichter gemacht.
trifft zu ____: ____: ____: ____: ____: trifft nicht zu
(3) Ich würde eine solche Konstruktionsaufgabe lieber mit einen Menschen in der Rolle des Konstrukteurs lösen.
trifft zu ___: ___: ___: ___: ___: trifft nicht zu
Warum hätte Sie es lieber, wenn ein Mensch die Rolle des Konstrukteurs übernimmt bzw. warum nicht?
(4) Die Graphikausgabe des Programms war meiner Ansicht nach eine technisch gute Leistung.
trifft zu ___: ___: ___: ___: ___: trifft nicht zu
(5) Die Antwortzeiten des Systems waren häufig so kurz, daß der Dialog mit dem System zügig verlaufen konnte.

trifft zu ___: ___: ___: ___: ___: trifft nicht zu

(6) Mit dem Rechner konnte man wie mit einem Menschen kommunizieren.
trifft zu ___: ___: ___: ___: ___: trifft nicht zu
Bitte begründen Sie warum Sie finden, daß Sie mit dem Rechner wie mit einem Menschen kommunizieren können bzw. warum dies für Sie weniger zutreffend ist:

(7) Die Graphikausgabe sah für eine Computergraphik eher ungewöhnlich aus.
trifft zu ___: ___: ___: ___: ___: trifft nicht zu

(8) Manchmal konnte ich die Antworten des Rechners akustisch schlecht verstehen.
trifft zu ___: ___: ___: ___: ___: trifft nicht zu

(9) Es war leicht, Fehler wie z.B. Baufehler oder Verständigungsfehler zu finden und zu korrigieren.
trifft zu ___: ___: ___: ___: ___: trifft nicht zu

(10) Ein sprachverstehendes System, das menschliche Bauanweisungen verstehen kann, ist das Höchste dessen, was ein Rechner leisten kann.
trifft zu ___: ___: ___: ___: ___: trifft nicht zu

(11) Die Lösung einer solchen Aufgabe wäre mit einem Menschen anders verlaufen.
trifft zu ___: ___: ___: ___: ___: trifft nicht zu
Bitte begründen Sie in welcher Hinsicht sich bei der Lösung der Konstruktionsaufgabe mit einem Menschen etwas geändert hätte bzw. inwiefern sich nichts geändert hätte:

(12) Die Situation hat auf mich manchmal einen unrealistischen Eindruck gemacht.
trifft zu ___: ___: ___: ___: ___: trifft nicht zu

(13) Sprachverstehende Systeme können noch viel mehr, als nur nach sprachlichen Anweisungen Baufixflugzeuge zusammenbauen.
trifft zu ___: ___: ___: ___: ___: trifft nicht zu

(14) Häufig mußte ich sehr lange auf eine Antwort bzw. Reaktion des Systems warten, so daß der Dialog mit dem System oft nur sehr schleppend voranging.
trifft zu ___: ___: ___: ___: ___: trifft nicht zu

(15) Die Verstehensleistung des Systems war zufriedenstellend.
trifft zu ___: ___: ___: ___: ___: trifft nicht zu

(16) Die Lösung dieser Konstruktionsaufgabe wäre mit einem Menschen in der Rolle des Konstrukteurs einfacher gewesen.
trifft zu ___: ___: ___: ___: ___: trifft nicht zu
Bitte begründen Sie, inwiefern die Lösung der Aufgabe mit einem Menschen als Konstrukteur einfacher gewesen wäre, bzw. warum dies nicht zutrifft:

(17) Das System hat die Lösung der Aufgabe eher erschwert bzw. verzögert.
trifft zu ___: ___: ___: ___: ___: trifft nicht zu

(18) Ich habe mich während des Versuchs oft gefragt, ob es ein sprachverstehendes System geben kann, das nach Instruktionen ein Baufixflugzeug zusammenbauen kann.
trifft zu ___: ___: ___: ___: ___: trifft nicht zu

(19) Ein sprachverstehendes System, das die Anweisungen eines Menschen interpretieren kann, ist noch lange nicht die Höchstleistung, die ein Rechner erbringen kann.
trifft zu ___: ___: ___: ___: ___: trifft nicht zu

(20) Ich hätte angenommen, daß sprachverstehende Systeme noch nicht so weit entwickelt sind, daß sie bereits menschliche Anweisungen ausführen können.
trifft zu ___: ___: ___: ___: ___: trifft nicht zu

(21) Die Sprachausgabe des Rechners war gut verständlich.
trifft zu ___: ___: ___: ___: ___: trifft nicht zu

(22) Ich mußte meine Sprechweise stark den Fähigkeiten des Systems anpassen.
trifft zu ___: ___: ___: ___: ___: trifft nicht zu

(23) Ich wußte bereits, bzw. hatte schon davon gehört, daß es sprachverstehende Programme gibt, die auch menschliche Anweisungen durchführen können.
trifft zu ___: ___: ___: ___: ___: trifft nicht zu

(24) Ein sprachverstehendes System zu entwickeln, das menschliche Anweisungen ausführen kann, ist bei dem heutigen Stand der Technik nicht sonderlich schwierig.
trifft zu ___: ___: ___: ___: ___: trifft nicht zu

Teil C

Im folgenden Teil des Fragebogens geht es vor allem um Ihre persönliche Meinung. Bitte kreuzen Sie das Kästchen an, das Ihre Meinung am ehesten widerspiegelt und beantworten Sie dann die untenstehenden Fragen. Wenn Sie der Ansicht sind, daß weder die linksstehende Charakterisierung noch die rechtstehende auf das System zutreffen, so kreuzen Sie das mittlere Kästchen an z.B.:

musikalisch ___: ___: _X_: ___: ___: unmusikalisch

(1) Bitte beschreiben Sie, was Sie von einem solchen sprachverstehenden System halten?

anregend	___: ___: ___: ___: ___:	langweilig
angenehm	___: ___: ___: ___: ___:	unangenehm
störend	___: ___: ___: ___: ___:	hilfreich
wünschenswert	___: ___: ___: ___: ___:	abzulehnen
leistungsfähig	___: ___: ___: ___: ___:	unbrauchbar
schwer zu bedienen	___: ___: ___: ___: ___ :	leicht zugänglich
anpassungsfähig	___: ___: ___: ___: ___:	unflexibel
überflüssig	___: ___: ___: ___: ___:	notwendig
erfolgreich	___: ___: ___: ___: ___:	ineffektiv
unfreundlich	___: ___: ___: ___: ___:	freundlich
ungeduldig	___: ___: ___: ___: ___:	geduldig
frustrierend	___: ___: ___: ___: ___:	zufriedenstellend
gehorsam	___: ___: ___: ___: ___:	rechthaberisch
zuverlässig	___: ___: ___: ___: ___:	unzuverlässig
unberechenbar	___: ___: ___: ___: ___:	berechenbar
kompliziert	___: ___: ___: ___: ___:	einfach

(2) Was gefällt Ihnen an dieser Art der Interaktion mit einem Computer?
(3) Was mißfällt Ihnen an dieser Art der Interaktion mit einem Computer?
(4) Welche Vorschläge könnten Sie machen zur Verbesserung des Systems?
(5) Können Sie beschreiben, welche besonderen Probleme Sie während der Lösung der Aufgabe hatten?
(6) Wo glauben Sie, könnte eine solche Anwendung nützlich sein?

B7 Fragebogen zu Szenario II (Mensch-Mensch-Interaktion)

Der Fragebogen für die Mensch-Mensch-Interaktion bestand aus dem A-Teil und einem leicht modifizierten C-Teil des Fragebogens für die Mensch-Maschine-Interaktion. Statt nach der Bewertung des Systems wurde im C-Teil nach der Bewertung des Kommunikationsweges gefragt.

(1) Bitte beschreiben Sie, was Sie von einem solchen Kommunikationsweg halten?
(2) Was gefällt Ihnen an dieser Art der Interaktion?

(3) Was mißfällt Ihnen an dieser Art der Interaktion?
(4) Welche Vorschläge könnten Sie machen zur Verbesserung?
(5) Können Sie beschreiben, welche besonderen Probleme Sie während der Lösung der Aufgabe hatten?
(6) Wo glauben Sie, könnte eine solche Anwendung nützlich sein?

C1 Aufgabenstellung

Für die Realisierung der Systemsimulation in den Szenarien 0, I und II mußten jeweils drei Aufgabenbereiche gelöst werden:

(1) die Bildübertragung des simulierten Systems
(2) die Sprachausgabe des simulierten Systems
(3) die Übertragung der Sprachdaten der Versuchsperson vom Raum der Versuchsperson in den Raum des Wizard

Im folgenden sei Raum II der Raum, in dem sich der Wizard befindet, der Raum der Versuchsperson sei Raum I. Die Steuerung der drei Aufgabenbereiche erfolgt über einen Kontrollbildschirm in Raum II (vgl. Kapitel C4.1).

C2 Technische Vorraussetzungen

C2.1 Hardware

Für die Realisierung wurden drei Unix-Workstations (DEC-Workstation 5000), zwei Gradient Desklab 216 und ein Sprachsynthetisator vom Typ SVS 3105 sowie eine Stereokamera (3-Chip-RGB-CCD) und ein PC mit angeschlossener Bildsequenzanlage, sowie das Ethernet verwendet. Zwei der Unix-Workstations wurden als Frontend für die Versuchsperson und den Wizard benötigt (vgl. C3.2, C3.3), die dritte wurde für die Bildübertragung benötigt (vgl. C3.1).

C2.2 Software

Die Bildübertragung wurde innerhalb einer KBV-Umgebung realisiert. Bei KBVision handelt es sich um eine Programmierumgebung, die in erster Linie zur Entwicklung von Programmen im Rahmen der automatischen Bilderkennung, und -verarbeitung genutzt wird. Bei der Erstellung unseres Szenarios ist KBVision speziell zur Bereitstellung und Übertragung von Bildern genutzt worden. Programmiersprache unter KBV ist C.

Die Umsetzung der Steuerung des Sprachsynthetisators erfolgte unter TCL-TK. Es handelt sich hierbei um zwei Softwarepakete, die zusammen ein System zur Entwicklung und Einsatz graphischer Benutzerschnittstellen bilden. TCL bietet einen in C-Code implementierten, und durch eigene Befehle erweiterbaren Befehlssatz, während TK ein Toolkit für das X-Window System zur Verfügung stellt. Dieses ermöglicht es, erstens Anwendungen interaktiv zu erstellen, und zweitens weitgehend unabhängig von der Basissprache C auf einem hohen Niveau programmieren zu können.

C3 Realisierung

Die Abbildung C1 gibt einen Überblick über den gesamten Simulationsaufbau: die verwendeten Rechner sowie die verwendete Software. Der Rechner aidec506 bildet das Frontend für die Versuchsperson und befindet sich dementsprechend in Raum I. Die Rechner aidec514, aidec511 und aibild sowie die Sequenzanlage werden für die Bilddigitalisierung bzw. -übertragung benötigt und befinden sich in Raum II. Neben dem Datenfluß (Sprachsignal der Versuchsperson, Standbilder der Bauaggregate, Sprachausgabe des Systems) sind die verwendeten Programme in der Abbildung vermerkt, letztere werden in den folgenden Abschnitten näher erläutert.

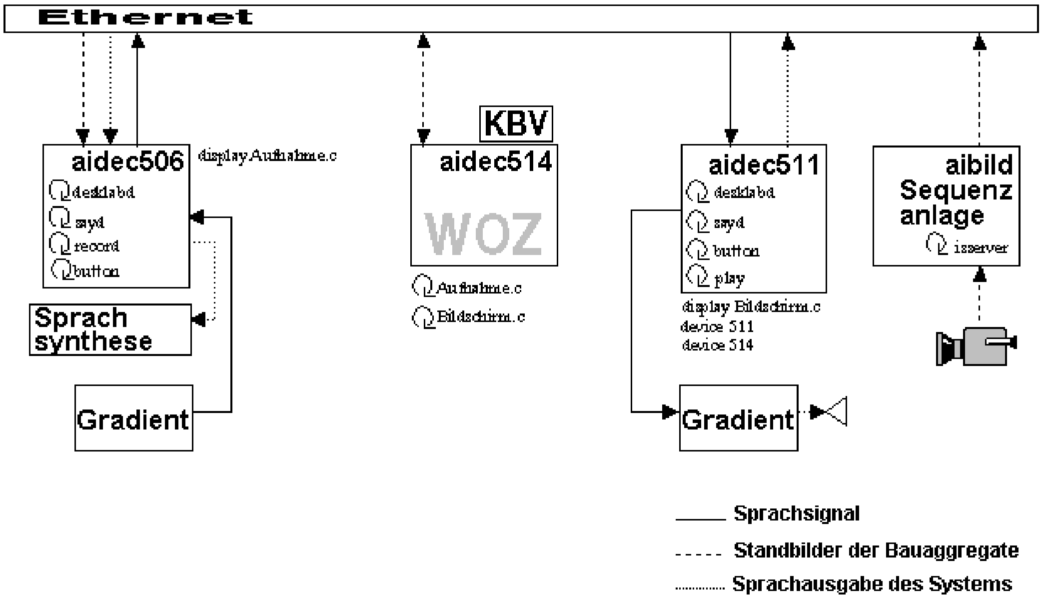

Abb. C1: Schematischer Überblick über den Aufbau der Simulation

C3.1 Realisierung der Bildaufnahme, -übertragung

Die Aufgabe besteht aus der digitalen Bildaufbereitung zur Bereitstellung eines Bildes und der Übertragung des Bildes auf den Bildschirm der Versuchsperson in Raum I.

Zur Ausführung dieser Aufgaben ist eine KBVision-Anwendung (Task) unter dem Namen *aufnahme* geschrieben worden. Die Steuerung dieser Task wurde durch eine Buttonmenüleiste auf dem Bildschirm des Wizards in einer zweiten Task *bildschirm* unter KBV implementiert. Diese Trennung der Tasks war notwendig, weil KBV die Verwaltung zwei verschiedener Bildschirme innerhalb einer Task nicht unterstützt. Die Kommunikation zwischen den beiden Tasks *aufnahme* und *bildschirm* wurde über eine shared-memory-Variable realisiert. Eine shared-memory-Variable ist eine Variable, auf die mehrere Programme gleichzeitig schreibend und lesend zugreifen können. Die beiden Tasks wurden zu einer Compounded-Task unter dem Namen *Wizard* zusammengefaßt.

C3.1.1 Digitale Bildaufbereitung

Für diese Aufgabe wurde das Tool *SnapStore* verwendet und unseren Erfordernissen angepaßt. *SnapStore* weist den Rechner, der die Bildsequenzanlage ansteuert, an, den Inhalt einer oder mehrerer Bildkanäle einer Videokamera (s/w bzw rot/grün/blau) in den Speicher eines gewünschten Rechners zu schreiben. Es wurden zwei Modifikationen mit dem Ziel, die übertragene Datenmenge zu reduzieren, vorgenommen:

- zum einen wurde die Zwischenspeicherung auf Platte des digitalisierten Bildes und die Neuinitialisierung des Buffers, in den das Bild eingelesen wird aufgehoben,
- zum anderen wurde die Auflösung des Bildes von 512*512 auf 256*256 heruntergerechnet.

Durch diese Veränderungen konnten Verbesserungen in der Übertragungszeit des Bildes von ca. 20 Sekunden erreicht werden.

C3.1.2 Bildübertragung

Die Bildübertragung ist in Form von einzelnen Standbildern realisiert worden, die nach den jeweils neuesten Anweisungen der Versuchsperson aktualisiert wurden. Beim Starten der Bildübertragungstask *aufnahme* wurde zunächst eine fingierte Systemmeldung auf dem Bildschirm im Raum der Versuchsperson ausgegeben, um die Illusion eines betriebsfähigen Worterkennungssystems zu steigern. Diese Systemmeldung wird gelöscht, sobald ein neues Bild vom Wizard zur Versuchsperson geschickt wird. Zur Beschleunigung des Steuerungsablaufs wurden die Bildaufbereitung und die Bildübertragung unter dem Befehl *Aufnahme* zusammengefaßt. Dieser Befehl stellt die Schnittstelle zwischen den beiden KBV-Tasks dar, indem er über das Setzen der shared-memory-variablen von der Kontrolltask *bildschirm* auf die Bildbearbeitungstask *aufnahme* umschaltet. Für die Dauer der Bildaufnahme und der Bildübertragung ist die Kontrolltask nicht aktiv. Dieser Zeitraum wird durch ein rotes Flag auf dem Bildschirm des Wizards (Kontrollbildschirm, vgl. Kapitel C4.2) markiert.

C3.2 Realisierung der Sprachausgabe der Simulation

Der Name des Programms zur Steuerung des Sprachsynthetisators ist *button*. Ein wichtiger Aspekt bei Erstellung diese Programmes war die Forderung, die Rückmeldungen des simulierten Systems zu standardisieren. Zu diesem Zweck wurden vordefinierte Äußerungen (siehe Anhang A/B) als Textbausteine abgelegt. Durch Auswahl eines Buttons mit einem bestimmten Textbausteinkürzel konnte die entsprechende Äußerung über den Sprachsynthetisator im Raum der Versuchsperson ausgegeben werden. Das Programm *button* ist in TCL-TK geschrieben worden, weil hier die Verwendung von Buttons unterstützt wird. Darüber hinaus wurde noch ein Sonderfeld zur freien Texteingabe implementiert, um auch auf unvorhergesehene Situationen reagieren zu können. Zur Datensicherung wurden die im Laufe eines Dialoges benutzten Textbausteine bzw. freien Texteingaben jeweils in einer Datei gespeichert.

C3.3 Realisierung der Sprachdatenübertragung

Die Übertragung der Sprachdaten der Versuchsperson von Raum I in den Raum, in dem sich die Mitarbeiter, die die Systemsimulation bedienten, befanden, wird durch folgende drei Programme realisiert:

(1) record
(2) play
(3) control

Das Programm *record* läuft auf dem Rechner der Versuchsperson (aidec506) und hat die Aufgabe, die Sprache aufzunehmen, an das zweite Programm *play* über das Netzwerk zu senden und gegebenenfalls auf Platte zu speichern. Das zweite Programm *play* läuft auf dem Rechner des Wizards und spielt dort die Daten ab, die es durch *record* empfängt. Die Steuerung von *record* geschieht entweder durch das Programm *control*, oder durch ein, mit den Routinen aus der Bibliothek *voice* ausgestattetes, Programm.

C3.3.1 *record*

record ist als Client des Servers *play* ausgelegt. Es überträgt die aufgenommenen Sprachdaten in Blöcken zu 16kByte (ca 0,5 sec) an *play* und erhält als Antwort den aktuellen Status des Duos *record/play*. Über diesen Status wird sowohl die Aufnahme auf Platte, als auch das Ende der Übertragung gesteuert. Da die Erfahrung gezeigt hat, daß die Übertragung eines einzelnen Blocks gelegentlich länger als 0,5 sec dauert, gleichzeitig aber die Verzögerung zwischen Aufnahme und Abspielen 1 sec nicht überschreiten soll, werden bei bestehender Verzögerung von über einer Sekunde „stille“ Blöcke, d.h. Blöcke, in denen die Lautstärke des Signals unter einem bestimmten Dezibelwert liegt, gesucht, und dann nicht übertragen. Dieses Vorgehen ist parametrisiert, so daß der gewünschte Dezibelwert bzw. die Länge der Blöcke oder die zulässige Verzögerung eingestellt werden können.

C3.3.2 *play*

play empfängt als Server die Blöcke von *record* und gibt sie über das Gradient DeskLab 216 aus. Darüber hinaus verwaltet *play* den Status des Duos *record/play* und gibt ihn auf dem Bildschirm des Wizards aus (vgl. Kapitel C4.2).

C3.3.3 *control*

Das Steuerprogramm *control* hat die Aufgabe die Übertragung und die Aufzeichnung auf Platte zu starten, bzw zu beenden. Dies geschieht durch Änderung des Status im Server *play.* Darüberhinaus bietet es die Möglichkeit, das Programm *play* abzubrechen.

C3.3.4 Bibliothek *voice*

Die Bibliothek *voice* enthält die wesentlichen Steuerroutinen von *control* und ermöglicht die Steuerung von *record* durch fremde Programme.

C4 Bedienung der Simulation

C4.1 Starten der Simulation

Die KBVision-Tasks verfügen über Parameter zur Benennung der Displaydevices des Kontrollbildschirms und des Bildschirms der Versuchsperson, die vor dem Start der Umgebung gesetzt werden müssen. Die folgende Startbeschreibung bezieht sich auf die Belegung der Parameter in den KBV-Tasks für Szenario I und II. Um dieses Set-up auf anderen Maschinen laufen zu lassen, müssen dort entsprechende Änderungen vorgenommen werden.

- Einloggen auf aidec506 (befindet sich im Raum des Wizard)
- Über telnet Fenster auf folgenden Maschinen öffnen:
  ein Fenster auf der aidec514 (KBV-Maschine),
  zwei Fenster auf der aidec511 (VP-Maschine in Raum I).
- Mit dem Befehl *xhost aidec506* die jeweiligen Maschinen für Befehle von aidec506 öffnen.
- Die Aufnahme- bzw. Wiedergabeparameter des Gradient DeskLab 216 an der aidec506 und an der aidec511 mit dem Befehl *gclear* bzw. mit dem Shellscript *gset* einstellen.
- Öffnen der KBV-Umgebung und Starten der KBV-Task *Wizard* auf aidec514. Es erscheint jetzt auf der aidec506 das Buttonmenu zur Steuerung der Bildübertragung und auf der aidec511 eine Systemmeldung des simulierten Systems.
- Starten des Programms *record* in einem Fenster der aidec511 mit dem Befehl *record*.

- Starten des Programms *play* auf der aidec506 mit dem Befehl *play*. Nach dem Starten des Duos *record/play* ist über den Gradient eine einseitige akustische Verbindung vom Raum der Versuchsperson zum Raum des Wizard hergestellt.
- Aufrufen der TKL-TK Shell mit *wish* und starten des TKL-TK Skripts zur Steuerung der Sprachausgabe im zweiten Fenster der aidec511 mit dem Befehl *source button*. Auf dem Bildschirm der aidec506 erscheint jetzt das Buttonmenü zur Kontrolle des Sprachsynthetisators. Dieses Skript ist erst lauffähig, wenn die Datei zur Sicherung der Textbausteine über den dafür vorgesehenen Button *file* geöffnet wurde.

C4.2 Der Kontrollbildschirm des Wizard

Auf dem Kontrollbildschirm (Abb. C2), von dem aus der Wizard die Simulation bedient, sind fünf Fenster und zwei Buttonmenüleisten geöffnet. Fenster und Menüleisten gehören zu drei verschiedenen Rechnern, einmal zum Rechner des Wizard (aidec511), zum Rechner der Versuchsperson (aidec506) und zum Rechner für die Bildverarbeitung (aidec514). In der Abbildung C2 befinden sich hinter den Rechnerkürzeln Nummern für eine leichtere eindeutige Zuordnung der Fenster.

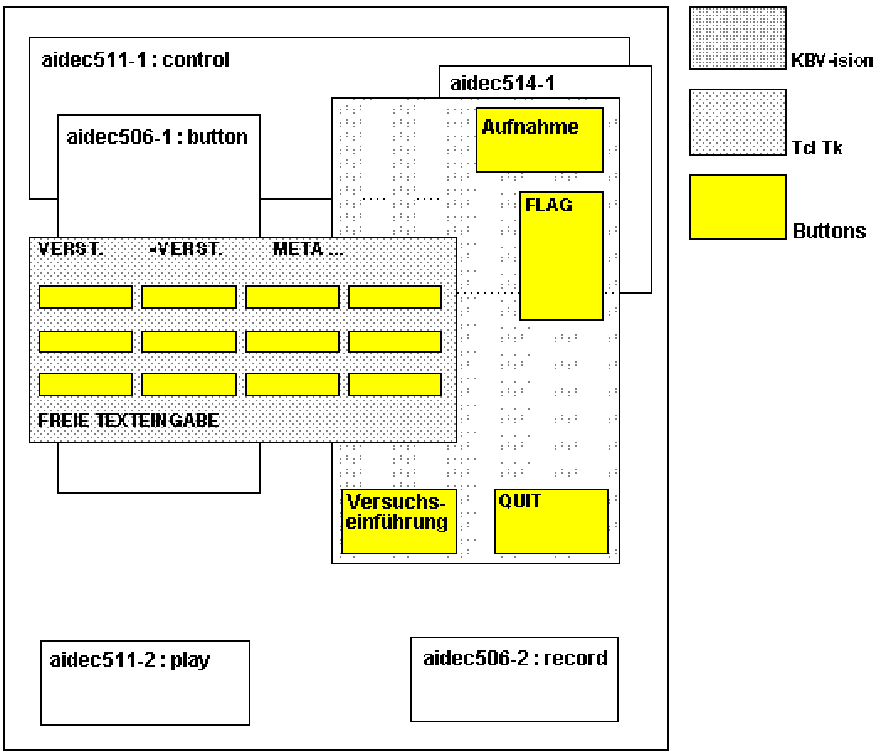

Abb. C2: Schematischer Überblick über den Kontrollbildschirm des Wizard

Die Fenster aidec511-1 und aidec511-2:
Oben auf dem Bildschirm befindet sich das Konsolenfenster der aidec511 (aidec511-1), von dem aus über das Programm *control* die Übertragung des Sprachsignals bzw. die Aufnahme des Sprachsignals auf die Festplatte des Rechners gesteuert werden kann. Unten links befindet sich das zweite Fenster (aidec511-2) auf der aidec511. Es dient der visuellen Kontrolle der Abspielroutine *play*, die auch hier gestartet wird. Für die visuelle Kontrolle wird in diesem Fenster die Größe des übertragenen Sprachsignals in KiloByte und die Verzögerung bei der Übertragung in Sekunden fortlaufend angezeigt.

Die Fenster aidec506-1, aidec506-2, Buttonmenüleiste1:
Von dem Rechner der Versuchsperson sind auf dem Kontrollbildschirm zwei Fenster sowie eine Buttonmenüleiste geöffnet. Im Fenster unten rechts auf dem Bildschirm (aidec506-2) wird die Aufnahmeroutine *record* für die Übertragung der Sprachdaten gestartet. Ähnlich wie bei der Routine *play* wird während des Versuchs hier die jeweils aktuelle Statusmeldung über das Senden und Empfanden des Sprachsignals ausgegeben (gesendete Bytes, Verzögerung in Sekunden, abgespielte Bytes). Im Fenster oben links auf dem Bildschirm (aidec506-1) wird das Programm *button* gestartet, das die Buttonmenüleiste1 zur Steuerung der Systemantworten öffnet. Diese Menüleiste besteht aus nach inhaltlichen Kriterien spaltenweise angeordneten Buttons zur Steuerung der Systemsprachausgabe (Textbausteine, die an das Sprachsynthesegerät gesendet werden) und des Programmablaufs (Speichern der in einem Versuchsdurchlauf gesendeten Textbausteine, Beenden des Programms).

Das Fenster aidec514, Buttonmenüleiste2:
Oben rechts auf dem Kontrollbildschirm befindet sich das Fenster aidec514, von dem das Programm *aufnahme* gestartet wurde. Dieses Programm öffnet auf dem Kontrollbildschirm des Wizard-Rechners (aidec511) ein Bedienfeld zur Steuerung und Kontrolle der Bildaufnahme und -übertragung und führt die Bildübertragung zum VP-Rechner (aidec506) aus. Die Buttonmenüleiste2 umfaßt drei Buttons und ein Flag:

- Oben rechts in der Menüleiste befindet sich der Button zum Auslösen der Bildübertragung (AUFNAHME).
- Darunter befindet sich ein Flag, das die Dauer der Systemeinführung bzw. der Bildübertragung farblich anzeigt (FLAG).
- Links unten in der Menüleiste befindet sich der Button für das Starten der durch das System realisierten Versuchseinführung der Versuchsperson (VERSUCHSEINFÜHRUNG).
- Rechts unten in der Menüleiste befindet sich der Button zum Beenden des Programms A*ufnahme* (QUIT).