Erkennung von Arbeitsgedächtnisbelastung und Zeitdruck in Dialogen

Empirie und Modellierung mit Bayesschen Netzen

Diplomarbeit
nach einem Thema von Prof. Dr. W. Wahlster
Fachbereich 14 – Informatik
der Universität des Saarlandes

vorgelegt von

Thomas Weyrath

Saarbrücken

August 1998

Eidesstattliche Erklärung

Ich erkläre an Eides Statt, daß ich diese Diplomarbeit selbständig und unter ausschließlicher Verwendung der angegebenen Quellen und Hilfsmittel angefertigt habe.

Saarbrücken, den 26. August 1998

Thomas Weyrath

Danksagung

Mein Dank gilt Herrn Prof. Dr. Wolfgang Wahlster für die Vergabe der Diplomarbeit und die Themenstellung. Außerdem danke ich Herrn Prof. Dr. Wolfgang Wahlster für die Möglichkeit der Mitarbeit im READY-Projekt.

Für die hervorragende Betreuung und Unterstützung beim Anfertigen dieser Arbeit bedanke ich mich bei Dr. Anthony Jameson.

Ich danke auch meinen Kollegen im READY-Projekt, Ralph Schäfer, Thomas Weis und André Berthold, für ihre Unterstützung, Hilfsbereitschaft und die sehr gute Zusammenarbeit.

Ich danke auch Markus Hüttner, Steffen Krause und Reiner Merkle für das Korrekturlesen meiner Arbeit und die konstruktiven Kommentare. Ganz besonders bedanke ich mich bei Patricia Riviere.

Ich danke meinen Eltern, die mich stets unterstützt und gefördert haben.

Inhaltsverzeichnis

Kapitel 1

Einleitung und Übersicht der Fragestellungen

Man stelle sich folgende Situation vor: Ein Autofahrer $\mathcal{P}$ hat eine Autopanne auf einer einsamen Landstraße. Des weiteren kann er mit seinem Mobiltelefon einen Kraftfahrzeugmechaniker $\mathcal{S}$ anrufen, der ihm bei der Behebung der Störung behilflich ist. Wie würde der Dialog zwischen den beiden verlaufen? $\mathcal{P}$ ist möglicherweise aufgeregt und in Eile wegen eines bevorstehenden Termins. Auch müßte $\mathcal{P}$ während der Reparatur des Wagens Instruktionen und Fragen von $\mathcal{S}$ verstehen und darauf reagieren können.

Viele Dialogsituationen finden unter derartigen Umständen statt, d.h. sie sind geprägt durch *Ressourcenbeschränkungen* des Dialogpartners. Menschliche Dialogteilnehmer passen sich in ihrem Dialogverhalten daran an. Dabei versuchen sie, sich ein Bild davon zu machen, inwieweit der Dialogpartner überfordert ist oder wie hoch sein Grad der Eile ist. In unserem Fall bedeutet das, daß $\mathcal{S}$ Rücksicht auf die gegenwärtige Situation von $\mathcal{P}$ nehmen sollte. $\mathcal{S}$ wird wahrscheinlich bei seinen Anweisungen darauf achten, daß $\mathcal{P}$ vielleicht aufgeregt oder mit einer Nebentätigkeit beschäftigt ist.

Könnte man diese Situation einem natürlichsprachlichen Dialogsystem anvertrauen? Ein solches System müßte jedenfalls ähnliche Fähigkeiten haben, wie unser fiktiver Mechaniker $\mathcal{S}$. D.h. das System müßte evtl. Ressourcenbeschränkungen des Dialogpartners einschätzen und sein Dialogverhalten daran anpassen können. Es müßte sich also *ressourcenadaptiv* verhalten können.

1.1 Zu berücksichtigende Ressourcenbeschränkungen

Im Rahmen dieser Arbeit werden zwei Typen von Ressourcenbeschränkungen untersucht:

(1) *Zeitbeschränkungen.*
Der Dialogpartner $\mathcal{P}$ kann mehr oder weniger starkem *Zeitdruck* unterliegen. Dieser Zeitdruck von $\mathcal{P}$ hat auch Auswirkungen auf das Dialogverhalten von $\mathcal{S}$. Ist $\mathcal{P}$ in Zeitdruck, sollte $\mathcal{S}$ z.B. seine Instruktionen kürzer fassen.

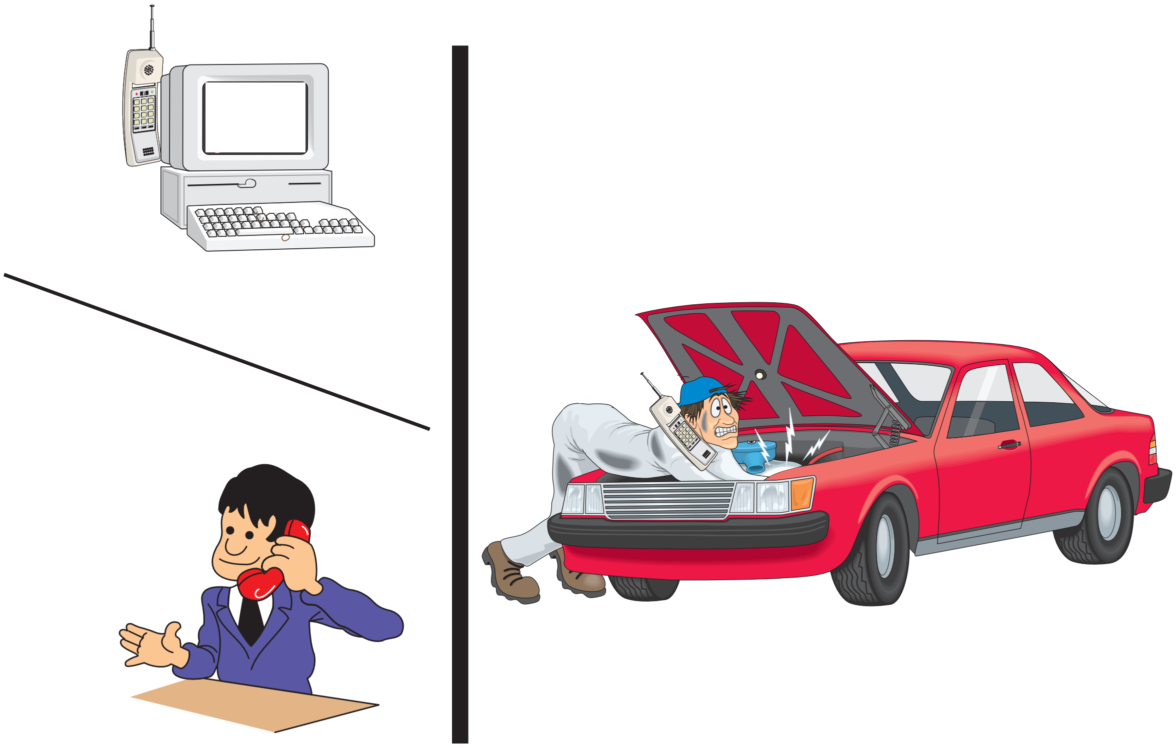

Abb. 1.1: *Beispielszenario des Systems* READY.

(2) *Beschränkungen des Arbeitsgedächtnisses.*
Arbeitsgedächtnisbelastung können merklich werden, wenn $\mathcal{P}$ mehr Informationen bearbeiten muß, als $\mathcal{P}$ in diesem Moment verarbeiten kann. Dieser Effekt wird dadurch verstärkt, daß $\mathcal{P}$ möglicherweise mit mehreren Aufgaben gleichzeitig beschäftigt ist, z.B. der Beantwortung einer Frage von $\mathcal{S}$ und der Reparatur des Wagens.

1.2 READY – Ressourcenadaptives Dialogsystem

Diese Arbeit entstand im Rahmen des von der Deutschen Forschungsgemeinschaft (DFG) geförderten Sonderforschungsbereich 378 (kurz SFB 378) „Ressourcenadaptive kognitive Prozesse". Das Thema des SFB 378 wird interdisziplinär in unterschiedlichen Fachbereichen untersucht: Künstliche Intelligenz, Psychologie, Computerlinguistik und Philosophie.

Die vorliegende Arbeit ist Bestandteil des Projektes READY (**Re**ssourcen**a**daptives **D**ialogs**y**stem), in dem untersucht wird, wie ein Dialogsystem seine Äußerungen an die Ressourcenbeschränkungen eines Dialogpartners anpassen kann. READY soll dabei die Rolle des Kraftfahrzeugmechanikers $\mathcal{S}$ im oben skizzierten Szenario übernehmen (siehe auch Abb. 1.1). Schäfer, Weis, Weyrath und Jameson (1997) geben eine kurze Übersicht des Projektes.

1.3 Die zu untersuchenden Fragestellungen

Die vorliegende Arbeit untersucht, wie READY Arbeitsgedächtnisbelastung und Zeitdruck eines Benutzers erkennen und einschätzen kann. Die hier beschriebenen Resultate werden in dem Benutzermodell von READY benutzt. Folgende Fragestellungen werden dazu untersucht:

(1) *Welche Inferenzen ziehen Menschen über Arbeitsgedächtnis und Zeitdruck des Dialogpartners?*
In der Psychologie lassen zahlreiche experimentelle Studien erkennen, welche Typen von Sprachproduktionsaufgaben das Arbeitsgedächtnis des Sprechers belasten und welche Merkmale sprachlicher Äußerungen sich negativ oder positiv auf das Sprachverstehen bzw. das Arbeitsgedächtnis des Hörers auswirken (siehe Abschnitt 2.1). Doch wurde bisher wenig untersucht, wie Sprecher Ressourcenbeschränkungen der Hörer, wie z.B. Arbeitsgedächtnisbelastung oder Zeitdruck, erkennen bzw. einschätzen.

(2) *Welche Inferenzen kann ein Dialogsystem über Arbeitsgedächtnis und Zeitdruck des Dialogpartners ziehen?*
Die Einschätzungen des Benutzermodells über den Grad der Arbeitsgedächtnisbelastung bzw. des Zeitdrucks des Dialogpartners sollen mit den Einschätzungen der Menschen vergleichbar sein. So kann sichergestellt werden, daß die Inferenzen des Dialogsystems auch nachvollziehbar bleiben. Um diese Vergleichbarkeit zu erhalten, sollen die Annahmen des Benutzermodells bzgl. den *Ursachen* und *Folgen* von Arbeitsgedächtnisbelastung und Zeitdruck auf dem (größtenteils implizitem) Wissen basieren, das Menschen in alltäglichen Dialogen anwenden.

(3) *Wie kann man die Unsicherheit und Veränderlichkeit der einzuschätzenden Benutzereigenschaften modellieren?*
Die Einschätzung der Arbeitsgedächtnisbelastung und des Zeitdrucks eines Dialogpartners sind mit *Unsicherheit* behaftet. Diese Tatsache stellt an den Inferenzmechanismus besondere Anforderungen. Zusätzlich sind die einzuschätzenden Benutzereigenschaften *zeitveränderlich*. Während eines Dialogs kann $\mathcal{P}$s Arbeitsgedächtnisbelastung mal höher, mal geringer sein. Dasgleiche gilt für den Zeitdruck von $\mathcal{P}$. Rückt der bevorstehende Termin immer näher, steigt der Zeitdruck. $\mathcal{P}$s Zeitdruck kann auch abnehmen, wenn z.B. aufgrund der Reparaturdauer der Termin sich nicht mehr pünktlich einhalten läßt.

1.4 Angewandte Methoden

1.4.1 Empirische Grundlagen

Für die Bearbeitung von Teil (1) der Fragestellungen wurden zwei explorative empirische Untersuchungen durchgeführt. Die Studien dienen als empirische Grundlage für die Modellierung der behandelten Ressourcenbeschränkungen des Dialogpartners. Dieser Aspekt ist im Hinblick auf die Entwicklung von READY von Bedeutung, da die Erfahrung menschlicher Dialogteilnehmer als wichtige Quelle von Einsichten dienen können.

1.4.2 Methoden aus der Künstlichen Intelligenz

Die Teile (2) und (3) der o.g. Fragestellungen beziehen sich auf die Realisierung der ressourcenbezogene Benutzermodellierung. Dazu wird der Inferenzmechanismus der Bayessche Netze verwendet. Für die Implementierung der Benutzermodelle wurde die Programmiersprache COMMONLISP/CLOS und die Tools IDEAL (Srinivas & Breese, 1992) bzw. NETICA (Netica97, 1997) verwendet. Beide Tools stellen den Inferenzmechanismus der Bayessche Netze zur Verfügung. Für die Entwicklung von READY wurde NETICA aufgrund des besseren Laufzeitverhaltens gewählt.

1.5 Aufbau der Arbeit

In Kapitel 2 werden die Grundlagen für diese Arbeit gelegt. Der erste Teil gibt einen Überblick über die relevanten einschlägigen bisherigen Forschungen aus der kognitiven Psychologie. Im zweiten Teil wird erklärt werden, inwieweit die dort entwickelten Modelle des Arbeitsgedächtnisses in die Mensch-Computer-Interaktion übernommen werden können. Im letzten Teil dieses Kapitels werden verschiedene Ansätze der Benutzermodellierung auf Möglichkeiten der Behandlung von Unsicherheit und Zeitveränderlichkeit überprüft.

In Kapitel 3 wird die explorative empirische Untersuchung beschrieben, die in der Feuerwehr Saarbrücken durchgeführt worden ist. Die Versuchspersonen waren Mitarbeiter der Feuerwehrnotrufzentrale, die regelmäßig Notrufe entgegennehmen. Die Analyse der Äußerungen der Versuchspersonen lieferte ein erstes Bild von Kausalzusammenhängen bzgl. der Arbeitsgedächtnisbelastung von Dialogteilnehmern.

In Kapitel 4 werden die Kausalzusammenhänge der Feuerwehrstudie mit dynamischen Bayesschen Netzen modelliert und getestet. Dabei wird auch beschrieben werden, wie man mit Hilfe von Modellierungsschemata dynamische Bayessche Netze einfacher entwickeln kann.

Kapitel 5 beschreibt die zweite empirische Studie, die in der READY-Domäne durchgeführt worden ist. In dieser Studie sollten speziell Ursachen und Folgen von Arbeitsgedächtnisbelastung bzw. Zeitdruck ermittelt werden, die in dem o.g. Szenario auftauchen können.

Im Anschluß daran wird in Kapitel 6 der Teil des Benutzermodells von READY beschrieben, der für die Erkennung und Einschätzung der Ressourcenlage des Benutzers verwendet wird. Für die Entwicklung des Benutzermodells werden die bis dahin gewonnenen Erkenntnisse genutzt. Das Modell wird am Schluß mittels eines Beispieldialogs aus dem Szenario getestet.

Zum Abschluß dieser Arbeit werden in Kapitel 7 die Ergebnisse zusammengefaßt und einen kurzen Ausblick auf mögliche Weiterentwicklungen aufgezeigt.

Kapitel 2

Einschlägige bisherige Forschung

Wichtig für diese Arbeit ist der Begriff des „Arbeitsgedächtnisses“. Dieser Begriff wurde innerhalb der kognitiven Psychologie geprägt und bezeichnet generell ein System, das für die Ausführung kognitiver Prozesse und die kurzfristige Speicherung von Informationen zuständig ist (vgl. Eysenck, 1994). Dieses System ist in seiner Kapazität begrenzt, so daß nicht eine beliebige Anzahl kognitiver Prozesse nebeneinander ausgeführt werden kann.

Viele Forschungen in der kognitiven Psychologie beschäftigen sich damit, welche Ursachen eine Belastung des Arbeitsgedächtnisses hervorrufen und welche Auswirkungen diese Belastung auf die kognitiven Prozesse hat. Kennt man solche Zusammenhänge, lassen sich kognitive Leistungen von Menschen besser beschreiben und vorhersagen. Auf der Basis von solchen Vorhersagen ist dann auch eine Anpassung möglich.

Der erste Teil dieses Kapitels beginnt mit einem kurzen Überblick über Forschungen im Bereich des Arbeitsgedächtnisses, die in der kognitiven Psychologie durchgeführt wurden. Im Anschluß daran wird eine Übersicht über eine Reihe von psychologischen Arbeiten gegeben, in denen Ursachen und Folgen von Arbeitsgedächtnisbelastung untersucht wurden. Für die vorliegende Arbeit ist insbesondere der Einfluß der Arbeitsgedächtnisbelastung auf die Sprachverarbeitung wichtig.

Danach wird exemplarisch gezeigt, inwiefern die Forschungen aus der kognitiven Psychologie in dem Bereich der Mensch-Computer-Interaktion angewendet werden. Da die meisten Modelle, die in der kognitiven Psychologie erarbeitet wurden, komplex sind, können die Modelle, die in der Mensch-Computer-Interaktion Anwendung finden, diese Beziehungen nur eingeschränkt umsetzen. Transparenz, Einfachheit und insbesondere Schnelligkeit stehen hier im Vordergrund.

Der letzte Teil dieses Kapitels beschäftigt sich mit Ansätzen, die im Bereich der Künstlichen Intelligenz entwickelt wurden, und die es ermöglichen, ein Modell des Benutzers zu konstruieren und für andere Systemkomponenten verfügbar zu machen. Die Ansätze werden auf ihre Möglichkeiten hin untersucht, Unsicherheit und Zeitveränderlichkeit zu behandeln.

2.1 Beiträge aus der kognitiven Psychologie

In experimentellen Studien über das Gedächtnis fand man heraus, daß unterschiedliche Gedächtnissubsysteme unterschiedliche Aufgaben bearbeiten. Die wohl bekannteste Differenzierung des menschlichen Gedächtnisses ist die Unterscheidung in Langzeitgedächtnis und Kurzzeitgedächtnis. Eines der ersten Gedächtnismodellen wurde von Atkinson und Shiffrin (1968) entwickelt. Dort wird dem Kurzzeitgedächtnis eine besondere Rolle zugeschrieben. Ihrem Modell zufolge gelangen alle mit Aufmerksamkeit versehenen Informationen in ein zwischengeschaltetes Kurzzeitgedächtnis, bevor sie in ein relativ andauerndes Langzeitgedächtnis gelangen können (siehe Abb. 2.1).

Dem Kurzzeitgedächtnis werden eine Reihe kognitiver Aufgaben im Sinne eines zeitweiligen Speichern bzw. Verarbeitens von Informationen zugesprochen, z.B. das sogenannte „innere Sprechen“ (engl. rehearsal), mit dessen Hilfe Informationen in diesem Speicher gehalten werden. Bereits im Modell von Atkinson und Shiffrin wurde angenommen, daß das Kurzzeitgedächtnis eine begrenzte Kapazität besitzt.

So ließen sich die empirischen Beobachtungen erklären, daß die Informationsmenge, die Menschen zu einem bestimmten Zeitpunkt behalten können, beschränkt ist, die sogenannte *Gedächtnisspanne*. Miller (1956) begrenzte die Gedächtnisspanne auf sieben plus oder minus zwei Items, die unmittelbar in richtiger Reihenfolge wiederholt werden könnten. Diese magische Zahl hat bis heute überlebt. Dennoch wird die Begrenzung der Gedächtnisspanne heute in einem theoretischeren Rahmen und differenzierter als in Millers Originalartikel interpretiert (siehe Shiffrin & Nosofsky, 1994).

2.1.1 Arbeitsgedächtnismodell von Baddeley

Im Gegensatz zu der von Miller (1956) aufgestellten Theorie vertritt Baddeley (1986) die Ansicht, daß der Umfang der Gedächtnisspanne nicht von der Anzahl der Informationen abhängt, sondern von der Geschwindigkeit, mit der die Informationen behalten werden können. Ein deutlicher Beleg dafür scheint der *Wortlängeneffekt* zu sein (Baddeley, 1986): Eine Liste von vielsilbigen Wörtern läßt sich schwieriger behalten, als eine Liste mit gleichvielen ein- oder zweisilbigen Wörtern. Der entscheidende Faktor dafür scheint zu sein, wieviel Zeit für die Aussprache eines Wortes benötigt wird. Aufgrund dieser empirischen Befunde wurde von Baddeley ein verändertes Modell des Arbeitsgedächtnisses entwickelt, das das Kurzzeitgedächtnis aus mehreren Subkomponenten zusammensetzt (siehe Abb. 2.2).

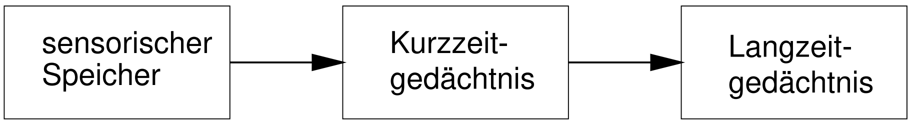

Abb. 2.1: *Gedächtnismodell von Atkinson und Shiffrin (aus: Anderson, 1995).*

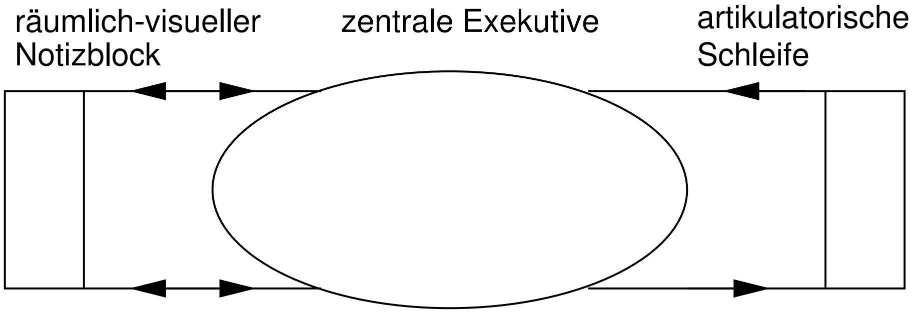

Abb. 2.2: *Baddeleys Theorie des Arbeitsgedächtnisses (aus: Baddeley, 1986).*

Im Hinblick auf sprachliche Informationen schlägt Baddeley eine *artikulatorische Schleife (articulatory loop)*[1] vor, in der Menschen so viele Informationen halten können, wie sie in einer bestimmten Zeitdauer aussprechen können. Diese Einschränkung erklärt den Wortlängeneffekt. Baddeley fand in Experimenten heraus, daß Menschen sich im Durchschnitt so viele Wörter unterschiedlicher Länge merken können, wie sie in etwa 1,5 Sekunden aussprechen können. Nach diesen 1,5 Sekunden zerfällt die Information in der artikulatorischen Schleife, wenn sie nicht durch den Rehearsal-Prozeß aufgefrischt wird.

Eine weitere Subkomponente, die Baddeley in seinem Modell vorschlägt, ist ein sogenannter *räumlich-visueller Notizblock (visuospatial sketchpad)* zum kurzen Behalten von Bildern. So ist es möglich, aus dem Gedächtnis einen Raum zu beschreiben, den man kurz zuvor betreten hat.

Die *zentrale Exekutive (central executive)*, als dritte Subkomponente, kontrolliert den Einsatz der verschiedenen Subsysteme, artikulatorische Schleife und räumlich-visueller Notizblock. Ihre Funktionsweise ist noch nicht so genau bekannt.

2.1.2 Das Doppelaufgabenparadigma

Wie das Modell von Baddeley (1986) zeigt, liegt der Fokus des Arbeitsgedächtnisses nicht allein auf der Speicherung von Informationen, sondern auch in der Ausführung von kognitiven Prozessen. Auch Baddeley postuliert, daß die Subkomponenten des Arbeitsgedächtnisses eine begrenzte Kapazität besitzen. Das Arbeitsgedächtnis kann also nicht beliebig viele kognitive Prozesse gleichzeitig ausführen. Nun stellt sich die Frage, wieviel *Arbeitsgedächtniskapazität* ein bestimmter Prozeß benötigt und welche Prozesse nebeneinander ausgeführt werden können, ohne daß eine Verringerung der Qualität der Ausführung die Folge ist. Diese Frage versucht das Doppelaufgabenparadigma zu beantworten.

In einen *Doppelaufgabenexperiment* werden die Versuchspersonen aufgefordert, zwei Aufgaben unterschiedlicher Priorität gleichzeitig auszuführen (siehe Eysenck, 1994, „dual-task performance"). Der Fokus liegt dabei auf der Messung der Arbeitsgedächtnisbelastung und der verfügbaren Arbeitsgedächtniskapazität. Kennt man die Leistung der Versuchspersonen beim Ausführen einer Aufgabe, ist es mit Hilfe dieser Technik möglich, die Arbeitsge-

[1] Die artikulatorische Schleife wird auch phonologische Schleife genannt.

dächtnisbelastung der Nebenaufgabe zu schätzen, wenn beide Aufgaben simultan ausgeführt werden. Da es nicht möglich ist, diese Arbeitsgedächtnisbelastung direkt zu messen und auch die übrig bleibende Arbeitsgedächtniskapazität unbekannt ist, wählt man die Nebenaufgabe so, daß eine Überbelastung des Arbeitsgedächtnisses entsteht. Die Folge ist ein Leistungsabfall bzgl. der Hauptaufgabe.

Power (1985) zeigte diesen Effekt in seinen Experimenten. Dort mußten die Versuchspersonen aus Wortpaaren ganze Sätze bilden. Die Versuchspersonen, die zusätzlich eine Nebenaufgabe (das Behalten von Zahlen) bewältigen mußten, brauchten für die Satzproduktion mehr Zeit und erzeugten häufiger kürzere Sätze. Aber auch bestimmte äußere Faktoren führen zu einer Verringerung der verfügbaren Arbeitsgedächtniskapazität, z.B. der Einfluß von Prüfungsangst (siehe Calvo & Eysenck, 1996, Abschnitt 2.1.3.2)

Im nächsten Abschnitt werden experimentelle Studien in der kognitiven Psychologie beschrieben, die Hinweise auf Einflüsse von Arbeitsgedächtnisbelastung auf die Sprachverarbeitung von Dialogteilnehmer geben können.

2.1.3 Einflüsse der Arbeitsgedächtnisbelastung auf die Sprachverarbeitung

Die Kommunikation mit dem System READY erfolgt durch natürliche Sprache. Aber auch andere Eingabemedien (z.B. Tastatur und handschriftliche Eingabe) und Ausgabemedien (z.B. Text und Grafik) sind denkbar. Auch hier spielt die Arbeitsgedächtnisbelastung eine besondere Rolle. Doch diese Interaktionsarten werden erst in einer folgenden Projektphase näher untersucht. Aus diesem Grund soll hier speziell auf den Aspekt der Sprachverarbeitung eingegangen werden.

Experimentelle Untersuchungen auf dem Gebiet der Sprachpsychologie liefern zahlreiche Teilantworten darauf, welchen Einfluß Arbeitsgedächtnisbelastung auf die Sprachverarbeitung hat und wie sich die Arbeitsgedächtnisbelastung auf das Verhalten der Dialogteilnehmer auswirkt. Folgende Fragen sind daher für READY von besonderem Interesse:

(1) *Was sind mögliche Ursachen hoher Arbeitsgedächtnisbelastung in Dialogsituationen?* Kennt READY solche Ursachen, kann das System besser einschätzen, wie stark der Dialogteilnehmer durch eine bestimmte Äußerung belastet wird.

(2) *Was sind typische Folgen hoher Arbeitsgedächtnisbelastung, die beim Sprechen und Hören auftreten?* Sichtbare Folgen können für READY als Symptome dienen, um die Arbeitsgedächtnisbelastung des Dialogpartners besser einzuschätzen.

Im nächsten Abschnitt wird eine Literaturübersicht der für READY relevanten Arbeiten gegeben, die bzgl. dieser Fragen untersucht werden. Da ein Dialogteilnehmer einerseits die Rolle des Sprechers und andererseits die Rolle des Hörers einnehmen kann, werden die Arbeiten in die Kategorien Sprachproduktion und Sprachverstehen unterteilt.[2]

[2] Die Gliederung der Literaturübersicht wurde von Jameson (1997b) übernommen.

Tabelle 2.1: *Ursachen von Arbeitsgedächtnisbelastung beim Sprecher*

Merkmale der sprachlichen Aufgabe:

- Fehlen von vorformulierten Äußerungen (Roßnagel, 1995b)
- Notwendigkeit, Informationen aus dem Langzeitgedächtnis abzurufen (Roßnagel, 1995a)
- Mehrfache Ziele (Greene, Lindsey & Hawn, 1990)
- Mangelnde Vorstrukturierung der Äußerungen sowie dadurch bedingte Länge und inhaltliche Komplexität (Oviatt, 1995)

Zusatzaufgaben:

- Von der sprachlichen Aufgabe unabhängige Nebenaufgaben (Power, 1985; Roßnagel, 1995a; Roßnagel, 1995b)

2.1.3.1 Arbeitsgedächtnisbelastung bei der Sprachproduktion

Im System READY treten Sprachproduktionsprozesse (seitens des Benutzers) auf, wenn der Benutzer eine Situationsbeschreibung gibt oder eine vom System gestellte Frage beantwortet. Probleme bei der Sprachproduktion werden mit Arbeitsgedächtnisbelastung assoziiert. Mit anderen Worten: Gibt der Benutzer eine schlechte Antwort bzgl. des Inhalts oder der Form der Äußerung, kann dies eine Folge hoher Arbeitsgedächtnisbelastung sein.

Die meisten Untersuchungen im Bereich der Sprachproduktion verwenden experimentelle Studien, um herauszufinden, inwieweit bestimmte Typen von Sprachproduktionsaufgaben das Arbeitsgedächtnis eines Sprechers belasten. Je nach Untersuchungsgegenstand verwendet man eher alltägliche Situationen (siehe Roßnagel, 1995a) oder künstliche Situationen, z.B. durch kleine Sprachspiele (siehe Power, 1985). Aber für alle Untersuchungen gilt, daß bestimmte Aspekte der sprachlichen Aufgabe variiert werden, die Auswirkungen in bezug auf die inhaltliche Qualität der Sprachproduktion oder auf die Zeit für das Lösen der Aufgabe erkennen lassen. Derartige Variationen sind z.B. das Hinzunehmen simultaner Nebenaufgaben, wie sie im Abschnitt 2.1.2 beschrieben wurden. So kann untersucht werden, inwiefern die Arbeitsgedächtnisbelastung eine mögliche Quelle für eine schlechtere Sprachproduktion ist. In den Tabellen 2.1 und 2.2 werden die Ursachen und Folgen hoher Arbeitsgedächtnisbelastung beim Sprecher zusammengefaßt, die in den nächsten beiden Abschnitten näher beschrieben werden.

Ursachen hoher Arbeitsgedächtnisbelastung beim Sprecher. Viele Untersuchungen zu diesem Thema beschäftigen sich mit Einflüssen auf das Sprechen als Prozeß für die Erzeugung von Sprache. Das Sprechen (also die Sprachproduktion) wird dort als kognitiver Problemlösungsprozeß aufgefaßt (vgl. Herrmann & Grabowski, 1994). Aus dieser Sicht führt eine zunehmende Übung in gleichartigen Problemsituationen zu einer Automatisierung des kognitiven Prozesses, ähnlich wie dies beim Lösen von Mathematikaufgaben geschieht. Auch im Alltag kann man diese Übungseffekte beim Sprechen beobachten: z.B. beim Reden vor einer Gruppe, bei mündlichen Prüfungen usw.

Roßnagel (1995b, 1995a) untersuchte Einflüsse von Übung auf die Sprachproduktion am Beispiel des monologischen Instruierens. Die Ergebnisse zeigen, daß geübte Sprecher deutlich schneller die Aufgabe ausführten als ungeübte Sprecher.[3] Des weiteren gab es bei geübten Sprecher weniger Störeffekte der simultanen Nebenaufgabe (das Behalten von Zahlen) auf die Hauptaufgabe (das Instruieren). Erklärt werden diese Effekte dadurch, daß geübte Sprecher auf weitgehend vorformulierte Äußerungen zurückgreifen. Derartige Äußerungen verringern den kognitiven Aufwand für das Erzeugen von Sprache.

Auch die Experimente von Power (1985) zeigen diesen Störeffekt. Durch den Einfluß von Nebenaufgaben verschlechterten sich die Leistungen bzgl. der Sprachproduktion der Probanden.

Andere Arbeiten beschäftigen sich mit den Planungsprozessen für die Sprachproduktion. Gerade schwierige und abstrakte Themen haben negative Auswirkungen auf das flüssige Sprechen. Eine Erklärung dafür ist, daß die Notwendigkeit, komplexere Pläne für die Sprachaufgabe zu erstellen, das spontane Sprechen beeinflußt. In Greene et al. (1990) wird zusätzlich dazu der Einfluß von mehrfachen (sozialen) Zielen beim monologischen Sprechen untersucht. Diese Experimente zeigten, daß solche Mehrfachziele Effekte auf die Sprachproduktion und die Arbeitsgedächtnisbelastung haben.

Auch im Bereich der Mensch-Computer-Interaktion werden positive und negative Einflüsse auf die Sprachproduktion untersucht. Angestrebtes Ziel dabei ist es, die Sprache als Eingabemedium in interaktiven Softwaresystemen optimaler einzusetzen. Konkret führt dies z.B. zur Entwicklung von Interfacetechniken, die fehlerhafte gesprochene Sprache als Eingabe minimieren sollen. Die Experimente von Oviatt (1995) zeigen, daß längere Äußerungen und schlecht strukturierte Ausgaben sich negativ auf die Spracheingabe und Texteingabe der Systembenutzer auswirken.

Folgen hoher Arbeitsgedächtnisbelastung beim Sprecher. Durch die Variierung einer sprachlichen Aufgabe, z.B. durch eine zusätzliche, simultane Nebenaufgabe, lassen sich Auswirkungen von Arbeitsgedächtnisbelastung auf das Sprechen untersuchen. Diese Auswirkungen zeigen sich dann durch sichtbare Symptome in bezug auf die Form und den Inhalt von Äußerungen. Tritt ein bestimmtes Symptom häufiger bei der sprachlichen Aufgabe mit Nebenaufgabe auf als bei der sprachlichen Aufgabe ohne Nebenaufgabe, verstärkt sich das Indiz dafür, daß die erhöhte Arbeitsgedächtnisbelastung des Sprechers (durch die Nebenaufgabe) Ursache für das vermehrte Auftreten des Symptoms ist. Mit anderen Worten: Das beobachtete Symptom ist eine Folge der hohen Arbeitsgedächtnisbelastung.

Die Experimente von Roßnagel (1995b) zeigten diesen Einfluß bei ungeübten und durch eine simultane Nebenaufgabe belasteten Sprechern deutlich. In bezug auf die Form von Äußerungen wurden längere Einsatzlatenzen, höhere Gesamtpausenanteile und mehr Fehlansätze beobachtet. In bezug auf den Inhalt von Äußerungen gaben die Probanden weniger Zusatzinformationen und erläuterten auch weniger Fachbegriffe.

[3] Der Untersuchungsbereich war die Sportart Karate, in der Teilnehmer gruppenweise instruiert werden. Die Übung im Instruieren bezieht sich auf die Erfahrung in der Leitung des Trainings.

Tabelle 2.2: *Folgen von Arbeitsgedächtnisbelastung beim Sprecher*

Folgen in bezug auf die Form von Äußerungen:

- Längere Einsatzlatenz (Roßnagel, 1995a, 1995b; Greene et al., 1990) bzw. kürzere Einsatzlatenz (Power, 1985)
- Höherer Gesamtpausenanteil, längere intraphrasale Pausen, mehr Fehlansätze (Roßnagel, 1995b)
- Mehr gefüllte Pausen und andere Unregelmäßigkeiten (Oviatt, 1995)
- Mehr Füllphrasen und Wiederholungen (Greene et al., 1990)

Folgen in bezug auf den Inhalt von Äußerungen:

- Geringere Auflösung von Anweisungen, weniger Zusatzinformationen, weniger Erläuterungen von Fachbegriffen (Roßnagel, 1995b)
- Geringere Hörerorientierung (Roßnagel, 1995a), vor allem bei geringer Motivation zur Hörerorientierung (Roßnagel, 1995a)
- Weniger kreative Konzeptualisierung (Power, 1985)
- Mehr Thematisierungen, die nicht aufgabenbezogen sind (Rummer, 1997)

Bei Power (1985) und Rummer (1997) zeigte sich der Einfluß der Nebenaufgabe dadurch, daß die Sätze weniger sorgfältig geplant wurden und mehr stereotype und informationslose Sätze geäußert wurden. Diese Ergebnisse lassen vermuten, daß eine Reduktion der semantischen Verarbeitung eintritt, wenn der Sprecher gleichzeitig eine simultane Nebenaufgabe mit kognitiven Aufwand ausführen muß.

Bei komplexen sprachlichen Äußerungen, die einen erhöhten Planungsaufwand der Sprachproduktion erfordern, treten mehr Füllphrasen auf, z.B. „ich vermute", „ok" (siehe Greene et al., 1990). Solche Füllphrasen geben dem Sprecher Zeit, weitere sprachliche Äußerungen zu planen, um sein Sprechen fortzusetzen. Ähnliche Konsequenzen hat eine mangelnde Vorstrukturierung der Äußerungen (siehe vorherigen Abschnitt). Eine Folge davon sind mehr gefüllte Pausen und andere sprachliche Unregelmäßigkeiten (siehe Oviatt, 1995, 1996). Eine hohe Vorstrukturierung der Informationen reduziert die Notwendigkeit, die Informationen selbst zu strukturieren, was zu einem flüssigeren Sprechen und weniger sprachlichen Fehler führt.

2.1.3.2 Arbeitsgedächtnisbelastung beim Sprachverstehen

Im System READY treten Sprachverstehensprozesse (seitens des Benutzers) auf, wenn READY dem Benutzer eine Frage stellt oder eine Anweisung gibt. Probleme beim Sprachverstehen werden auch hier mit Arbeitsgedächtnisbelastung assoziiert. Mit anderen Worten: Versteht der Benutzer eine bestimmte Frage nicht, kann dies eine Folge hoher Arbeitsgedächtnisbelastung sein.

Untersuchungen in Bereich Arbeitsgedächtnis und Sprachverstehen zielen längerfristig auf die Verbesserung der sprachlichen Verständigung ab. Das Forschungsinteresse richtet

Tabelle 2.3: *Ursachen von Arbeitsgedächtnisbelastung beim Hörer*

Individuelle Faktoren:

- Verhältnismäßig geringe Arbeitsgedächtniskapazität, z.B. im Zusammenhang mit hohem Alter (Kemper, Jackson, Cheung & Anagnopoulos, 1993)
- Geringes Vermögen, irrelevante Gedanken auszuschalten (Kemper et al., 1993)
- Emotionale Belastung, z.B. durch Examensangst (Calvo & Eysenck, 1996)

Merkmale der zu verstehenden Äußerungen:

- Unvorhersehbare Sprünge (Britton, Glynn & Smith, 1985; Kintsch & Vipond, 1979 – vgl. Grabowski, 1991)
- Notwendigkeit, inhaltliche Inferenzen zu machen (Kintsch & Vipond, 1979)
- Notwendigkeit, häufig Informationen aus dem Langzeitgedächtnis abzurufen (Britton et al., 1985; Kintsch & Vipond, 1979)
- Notwendigkeit, das mentale Modell des Textes häufig umzubauen (Kintsch & Vipond, 1979)
- Hohe propositionale Dichte; hohe Dichte neuer Konzepte (Kintsch & Vipond, 1979)
- Syntaktische Komplexität (Kemper et al., 1993)

sich dabei größtenteils auf die Untersuchung, wie sich sprachliche Äußerungen auf das Sprachverstehen des Hörers auswirken.[4] In den beiden folgenden Abschnitten werden die Ursachen und Folgen hoher Arbeitsgedächtnisbelastung beim Hörer beschrieben. Die Tabellen 2.3 und 2.4 fassen die Ergebnisse zusammen.

Ursachen von Arbeitsgedächtnisbelastung beim Hörer. Hier kann man zwischen individuellen Faktoren des Hörers und Merkmalen der sprachlichen Äußerungen unterscheiden.

Individuelle Faktoren beziehen sich auf die Leistungsfähigkeit des Hörers. Einerseits kann sie vom Alter abhängig sein (siehe Kemper et al., 1993), andererseits von der zeitweisen Aufnahmefähigkeit des Hörers. Einschränkungen der Aufnahmefähigkeit entstehen häufig aufgrund aufgaben-irrelevanten Wahrnehmungen, die den eigentlichen Sprachverstehensprozeß stören. Calvo und Eysenck (1996) untersuchten die Rolle des phonologischen Arbeitsgedächtnisses beim Lesen unter Einfluß von Prüfungsangst und beobachteten eine Verringerung der Verarbeitungskapazität des Arbeitsgedächtnisses bei ängstlichen Versuchspersonen.[5]

Im Unterschied zu den individuellen Faktoren beziehen sich die sprachlichen Merkmale auf den Verstehensgrad der Äußerung. Dies wurde im Bereich des Textverstehens sehr

[4] Es werden Aspekte des Sprachverstehens behandelt, die sowohl beim Hören als auch beim Lesen beteiligt sind. Im folgenden wird nicht zwischen Hörer und Leser unterschieden, da vielfach dieselben Faktoren eine Rolle spielen.

[5] Das Experiment zeigte, daß ängstliche Personen mehr auf die phonologische Kodierung und Speicherung zurückgreifen als nichtängstliche Personen, was sich durch erhöhtes offensichtliches Artikulieren (durch Lippenbewegungen) bemerkbar machte.

Tabelle 2.4: *Folgen von Arbeitsgedächtnisbelastung beim Hörer*

- Schlechtere Verstehens- und Behaltensleistungen (z.B. Kemper et al., 1993)
- längere Lesezeiten (z.B. Kemper et al., 1993; Calvo & Eysenck, 1996)
- Inadäquate bzw. uninformative Antworten (Anderson & Boyle, 1994)

gründlich untersucht. Angestrebtes Ziel dieser Untersuchungen ist es, die Verstehensleistung des Lesers und die Lesbarkeit von Texten zu erklären (siehe Kintsch & Vipond, 1979). Schwierig zu verstehende Texte führen danach zu der Notwendigkeit:

- viele inhaltliche Inferenzen zu machen, um eine adäquate Verbindung der Textstruktur zu erreichen (Kintsch & Vipond, 1979)
- häufig Informationen aus dem Langzeitgedächtnis abzurufen, um verlorene bzw. vergessene Inhalte wiederherzustellen (Kintsch & Vipond, 1979; Britton et al., 1985)
- die mentalen Modelle häufig umzubauen, um eine bestmögliche Organisation der mentalen Struktur zu erreichen (Kintsch & Vipond, 1979)

Auch komplizierte semantische Konstruktion, z.B. durch häufige Benutzung von Fremdwörter, und eine hohe syntaktische Komplexität (Kemper et al., 1993), z.B. durch tiefverschachtelte Teilsätze, erhöhen den Schwierigkeitsgrad der Lesbarkeit von Texten.

Da die o.g. Faktoren für die Schwierigkeit der Verständlichkeit von Äußerungen gut vorhersehbar sind, lassen sich in der Dialogplanung am ehesten Anpassungen an die Arbeitsgedächtnisbelastung des Hörers vornehmen.

Folgen von Arbeitsgedächtnisbelastung beim Hörer. Kemper et al. (1993) beobachtete in ihren Experimenten eine schlechtere Verstehens- und Behaltensleistung der Versuchspersonen, die durch eine Nebenaufgabe abgelenkt wurden. Auch längere Lesezeiten wurden beobachtet (Calvo & Eysenck, 1996).

Anderson und Boyle (1994) zeigten, daß Hörer, die ein Problem mit dem Verständnis einer Äußerung haben, mehr inadäquate Antworten („Sie haben was?“) oder informationslose Antworten („Mhm“) geben.

Die in diesem Abschnitt beschriebenen Arbeiten zu Ursachen und Folgen von Arbeitsgedächtnisbelastung werden später als Grundlage des Benutzermodells dienen, das $\mathcal{P}$s Arbeitsgedächtnisbelastung erkennen und einschätzen soll.

2.2 Die Berücksichtigung des Arbeitsgedächtnisses in der Mensch-Computer-Interaktion

In der kognitiven Psychologie wird das Konzept des Arbeitsgedächtnisses benutzt, um empirische Befunde, z.B. über die Behaltensleistung von Menschen, zu erklären. Die Mensch-Computer-Interaktion (engl. Human Computer Interaction, kurz HCI) beschäftigt sich damit, wie die Beiträge der kognitiven Psychologie beim Entwurf von Computersystemen berücksichtigt werden können. Im Zentrum stehen die kognitiven Fähigkeiten des Menschen, die sich durch seine Möglichkeiten, Informationen zu verarbeiten, ergeben: der Mensch als *Informationsverarbeiter*.

Ein guter Entwurf eines interaktiven Systems berücksichtigt den kognitiven Aufwand der Aufgaben, die zum Erreichen bestimmter Ziele in diesem System notwendig sind. Die Vorgehensweise ist eher pragmatisch, verglichen mit den in der kognitiven Psychologie aufgestellten Modellen und Theorien. Die Anwendung der kognitiven Psychologie in der HCI wird hier aus der Sicht eines Systemdesigners betrachtet, dessen Aufgabe es ist, die kognitiven Fähigkeiten von Menschen in der Systementwicklung zu berücksichtigen (siehe Baecker, Grudin, Buxton & Greenberg, 1995, Kapitel 9). Drei Punkte leiten diese Idee (siehe S.K. Card & Newell, 1980):

(1) Aufgabenanalyse. Die Benutzer eines Systems versuchen, sich beim Erreichen eines Zieles der Systemumgebung anzupassen. Die Aufgabenanalyse soll herausfinden, welche kognitiven Aufgaben der Benutzer ausführen muß, um das Ziel zu erreichen.

(2) Berechenbarkeit. Die kognitiven Aufgaben eines Benutzers müssen berechenbar sein, weil Berechnungen die Basis von technikorientierten Anwendungen sind.

(3) Approximation. Wenn die Berechnungen schnell ausgeführt werden müssen, ist es notwendig, sie zu vereinfachen.

Mit anderen Worten: die Berücksichtigung der feinkörnigen Modelle und Theorien der kognitiven Psychologie in der HCI führt zu gewissen Vereinfachungen.

Ein bekanntes Beispiel einer solchen Vereinfachung kognitiver Prozesse ist das *Modell des menschlichen Prozessors*, das von S.K. Card und Newell (1980) entwickelt wurde. Das Modell des menschlichen Prozessors ist ein idealisiertes Modell der Informationsverarbeitung von menschlicher Wahrnehmung, Motorik und kognitiven Systemen. Das Modell ist eine vereinfachte Architektur des Benutzers, das mit den Konzepten Speicher, Prozessoren und Operationsprinzipien beschrieben wird. Der Zweck dieses Modells ist es, dem Designer eines interaktiven Systems eine Technik in die Hand zu geben, die menschliche Leistung vorherzusagen: Wie schnell können Leute lesen, wie schnell können sie bestimmte Tasten drücken usw.

Auch in den später beschriebenen Benutzermodellen wird das Modell des Arbeitsgedächtnisses vereinfacht dargestellt. Das Arbeitsgedächtnis wird hier als homogener Speicher betrachtet, der eine gewisse Kapazität besitzt. Eine differenzierte Auffassung des Arbeitsgedächtnisses, die die Subkomponenten des Baddeleyschen Modells einbezieht, könnte z.B. die phonologische Schleife berücksichtigen. Beispielsweise könnte ein Wegauskunftssystem, das speziell die Behaltensleistung des Benutzers unterstützen will, seine Äußerungen so generieren, daß diese *bequem* durch Rehearsal-Prozesse in der phonologischen Schleife gehalten werden können. Eine solche Äußerung dürfte dann nicht länger als 1,5 bis 2 Sekunden sein. So wäre eine Instruktion „Gehe links, dann links, dann rechts!“ besser als die Instruktion „Gehen Sie bitte erst links, dann an der nächsten Kreuzung wieder links und dann biegen Sie rechts ab!“

2.3 Behandlung von Unsicherheit und Veränderung in der Benutzermodellierung

Die Entwicklung von intelligenten Benutzerschnittstellen (siehe Maybury & Wahlster, 1998) erfordert eine explizite Modellierung des Benutzers.

Im einfachsten Fall bestimmt der Programmierer beim Entwurf der Benutzerschnittstellen (z.B. Menüs, Dialogfenster), wie die Mensch-Computer-Interaktion verläuft. Die Interaktionsmöglichkeiten des Systems sind dabei nicht an einen individuellen Benutzer ausgerichtet, sondern werden vielmehr durch einen Durschnittsbenutzer bestimmt (vgl. Rich, 1983). Man spricht dabei von einem *kanonischen Benutzermodell*.

Intelligente Benutzerschnittstellen versuchen, das Dialogverhalten zwischen System und Benutzer flexibler und individueller zu gestalten. Sie unterhalten explizit ein *Benutzermodell*, d.h. eine Wissensquelle, in der explizite Annahmen über alle Aspekte des Benutzers enthalten sind, die für das Dialogverhalten des Systems relevant sind (siehe Wahlster & Kobsa, 1989). Der Teil eines Dialogsystems, dessen Funktion es ist, inkrementell ein Benutzermodell zu konstruieren, zu verwalten und andere Komponenten mit Annahmen über den Benutzer zu versorgen, nennt man *Benutzermodellierungskomponente* (siehe Wahlster & Kobsa, 1989).

Viele Arbeiten im Bereich der Benutzermodellierung beschäftigen sich mit der Anpassung des Systems an statische Benutzereigenschaften. Solche Eigenschaften sind z.B. der Wissensstand, die Interessen oder die Präferenzen des Benutzers (siehe Jameson, Paris & Tasso, 1997, Reader's Guide). Nur wenige Arbeiten beschäftigen sich damit, wie man in der Mensch-Computer-Interaktion die kognitive Belastung des Benutzers berücksichtigen kann. Zwei typische Anwendungsbeispiele sind:

- Präsentation von komplexen Informationen für zeitkritische Entscheidungen (siehe Horvitz & Barry, 1995)
- Präsentation von Informationen bei Instruktionen und Erklärungen in tutoriellen Systemen (siehe z.B. Kashihara, Hirashima & Toyoda, 1995)

Der Unterschied zu den statischen Benutzereigenschaften liegt darin, daß sich die kognitiven Belastungen des Benutzers im Laufe der Interaktion mit dem System ändern können. Man spricht hierbei von zeitveränderlichen Benutzereigenschaften, während die statischen Eigenschaften eher zeitlich unveränderlich sind.[6] Eine weitere Besonderheit von kognitiven Eigenschaften ist, daß ihre Einschätzung extrem mit Unsicherheiten behaftet ist. Es gibt kaum sichere Anzeichen für eine kognitive Belastung. Beide Aspekte, die Zeitveränderlichkeit und die Unsicherheit, stellen die Benutzermodellierung vor eine neue Herausforderung.

Im folgenden werden einige Benutzermodellierungsansätze beschrieben und auf folgende Fragenstellung hin untersucht:

(1) Was sind die Ansätze zur Berücksichtigung von Unsicherheit bei der Benutzermodellierung?

[6] Es gibt auch Anwendungen, in denen die o.g. statischen Benutzereigenschaften zeitveränderlich sind. Ein Beispiel dafür ist der Wissensstand des Benutzer in tutoriellen Systemen.

(2) Wie können die Ansätze zeitveränderliche Benutzereigenschaften repräsentieren?

Ein Vergleich der numerischen Ansätze zur Behandlung von Unsicherheit, Bayessche Netze, Dempster-Shafer-Theorie und Fuzzy-Logik, gibt auch Jameson (1996).

2.3.1 Logik-basierte Methoden

Eine wichtige Aufgabe der Benutzermodellierung ist es, den Benutzer bei der Ausführung komplexer Aufgaben zu unterstützen. Das Benutzermodell muß dazu in der Lage sein, die Ziele (*Goals*), Pläne und Annahmen (*Beliefs*) des Benutzers zu erkennen. Da Ziele und Weltzustände oft als prädikatenlogische Literale repräsentiert werden (siehe z.B. STRIPS), werden hier logik-orientierte Methoden bevorzugt.

Logik-basierte Ansätze der Benutzermodellierung werden z.B. in den Benutzermodellierungsshells UMT (siehe Brajnik & Tasso, 1992, 1994) und BGP-MS (Kobsa & Pohl, 1995) benutzt. Dieser Ansatz wird anhand BGP-MS näher erläutert.

2.3.1.1 BGP-MS: Benutzermodellierungsshell

BGP-MS[7] (siehe Kobsa & Pohl, 1995; Pohl & Höhle, 1997) ist ein unabhängiges System, das von einem Anwendungssystem als Benutzermodellierungskomponente verwendet werden kann. Neben Methoden zur Repräsentation von Annahmen verwendet BGP-MS auch Methoden zur Unterstützung der Wissensakquisition, wie z.B. Benutzerinterviews.

Aufbau des Benutzermodells. BGP-MS teilt die Wissensbasis des Benutzermodells in sogenannte Partitionen ein, die in einer Vererbungshierarchie organisiert sind. Jede Partition speichert Annahmen über den Benutzer bzgl. eines spezifischen Typs. Typische Partitionsnamen sind SBUB (System Beliefs User Beliefs), SBUW (System Beliefs User Wants) und SBMBUB (MB steht für gemeinsame Annahmen, engl. „Mutual Beliefs“). Zusätzlich macht BGP-MS Gebrauch von der Partition SB, die das Domänwissen des Systems verwaltet. Stereotypen werden auch in Partitionen modelliert: Für jeden Stereotyp gibt es eine Partition, deren Inhalt der Partition SBUB vererbt wird, wenn der Stereotyp auf den aktuellen Benutzer anwendbar ist.

Innerhalb einer Partition werden die Annahmen entweder in terminologischen SB-ONE-Strukturen oder durch prädikatenlogische Aussagen repräsentiert. Die Beziehung zwischen Inhalten einer BGP-MS-Wissensbasis und Ausdrücken wird durch eine multimodale Logik mit mehreren Akteuren hergestellt. Eine SB-ONE IS-A Relation zwischen zwei Konzepten „Laserjet“ und „Laserprinter“, die in der SBUB-Partition enthalten ist, kann durch die modallogische Formel $\diamond_{(B,S)} \diamond_{(B,U)} \forall x\ \mathit{laserjet}(x) \rightarrow \mathit{laserprinter}(x)$[8] dargestellt werden.

[7] BGP-MS = Belief, Goal and Plan Maintenance System

[8] lies $\diamond_{(B,S)} \diamond_{(B,U)}\ p$: das System glaubt, daß der Benutzer glaubt, daß p wahr ist

BGP-MS tauscht mit dem Anwendungssystem auf vielfältige Weise Informationen aus. Z.B. kann die Anwendung BGP-MS über beobachtete Benutzerannahmen oder -aktionen informieren. Im folgenden Beispiel wird BGP-MS mitgeteilt, daß das System glaubt, daß der Benutzer glaubt, daß alle Wale Fische sind:

```
(bgp-ms-tell (B S (B U all x (-> (WHALE x) (FISH x)))))
```

Des weiteren kann auch BGP-MS Fragen der Anwendung beantworten, z.B. ob das System und der Benutzer glauben, daß Peter Maria ein Buch gibt:

```
(bgp-ms-ask (B S (B M (GIVES PETER MARY BOOK))))
```

BGP-MS besitzt mehrere Inferenzmethoden, um neue Annahmen herzuleiten und beobachtete Ziele, Aktionen und Antworten des Benutzers zu verarbeiten. Dabei kann es passieren, daß alte Annahmen im Benutzermodell zu neuen Annahmen widersprüchlich sind. In diesem Fall müssen die widersprüchlichen Annahmen aufgehoben werden. Man spricht dabei von der Konsistenzerhaltung des Benutzermodells.

Behandlung von Unsicherheit und Veränderlichkeit. Da BGP-MS logische Inferenzmethoden benutzt, tritt keine explizite Unsicherheitsbehandlung auf. Es findet auch keine Modellierung von Zeit statt.

2.3.1.2 Diskussion

Ein Vorteil des logik-orientierten Ansatzes gegenüber numerischen ad hoc Methoden (siehe nächster Abschnitt) ist, daß die Schlußfolgerungen des Systems nachvollziehbar und somit für den Benutzer erklärbar sind. Für die Einschätzungen der kognitiven Belastung des Benutzers eignet sich aber ein logik-basierter Ansatz aufgrund der fehlenden Unsicherheitsbehandlung nicht.

Ein weiteres Problem des logik-basierten Ansatzes ist die Modellierung von zeitveränderlichen Benutzereigenschaften. In BGP-MS fehlt die Modellierung von Zeit völlig. Eine Möglichkeit wäre es, das System durch Inferenzenmethoden, die die Zeit berücksichtigen, zu erweitern. Giangrandi und Tasso (1997) stellen am Beispiel eines tutoriellen Systems eine logik-basierte Methode vor, mit dem das über die Zeit veränderliche Benutzerwissen beschrieben werden kann.[9]

2.3.2 Numerische ad hoc Methoden

Einige Benutzermodellierungsansätze verwenden numerische ad hoc Methoden, d.h. keine wohlfundierten Methoden, um Inferenzen abzuleiten. Ein bekanntes Beispiel ist das stereotypbasierte Bibliotheksauskunftssystem GRUNDY von Rich (1979). Aber auch in neueren

[9] Die Beschreibung des Domänwissens erfolgt noch z.Z. mit Hilfe der Aussagenlogik. Die Erweiterung auf Prädikatenlogik ist geplant.

Systemen werden solche ad hoc Methoden verwendet, wie z.B. in der Lernermodellierungskomponente CASTLE (siehe Murphy & McTear, 1997).

Ein Stereotyp repräsentiert eine Menge von Attributen, die bei einer Personengruppe oft vorkommt. Stereotypen spielen deshalb eine wichtige Rolle innerhalb der Benutzermodellierung, weil sie dem System es ermöglichen, auf der Basis einer kleinen Anzahl von Beobachtungen eine große Zahl von plausiblen Inferenzen zu ziehen. Der Stereotypenansatz wird anhand des Systems GRUNDY erläutert.

2.3.2.1 GRUNDY: Unsicherheitsbehandlung durch numerische ad-hoc Methoden

GRUNDY (siehe Rich, 1979, 1989) ist ein Dialogsystem, das ein Beratungsgespräch in einer Bibliothek simuliert. GRUNDY übernimmt dabei die Rolle eines Bibliothekars, der den Benutzer bei der Auswahl seiner Lektüre behilflich ist. Um zu entscheiden, welche Bücher der Benutzer gerne liest, verwendet GRUNDY stereotypisches Wissen über den Benutzer, das im Laufe eines Dialogs zu einem individuellen Benutzermodell verfeinert wird.

Aufbau des Benutzermodells. Die Stereotypen sind in einer Generalisierungshierarchie angeordnet. Der generellste Stereotyp beschreibt einen kanonischen Benutzer (ANY-PERSON). Spezifischere Stereotypen (z.B. WOMAN, EDUCATED-PERSON) beschreiben Unterklassen von Benutzern, die Eigenschaften von ihren Oberklassen erben. Am Anfang des Dialogs wird der ANY-PERSON-Stereotyp aktiviert und alle Eigenschaften, die von ihm vorhergesagt werden, in das Benutzermodell geschrieben. Durch Benutzeräußerungen, wie „I am unconventional, open, direct and humorous“, werden weitere Stereotypen aktiviert, die das Benutzermodell aktualisieren. Die Aktualisierung des Benutzermodells geschieht dadurch, daß alle Eigenschaften, die von den aktivierten Stereotypen vorhergesagt werden, in das Modell eingetragen werden, wenn sie noch nicht vorhanden sind. In Konfliktfällen müssen Widersprüche aufgelöst werden.

GRUNDY versucht im Laufe des Dialogs, ein Persönlichkeitsprofil des Benutzers zu erstellen, in dem seine charakteristischen Eigenschaften gespeichert sind. Jede dieser Eigenschaften (*Facet*) wird durch einen Ausprägungsgrad (*Value*) und einer Sicherheitsrate (*Rating*) beschrieben. Der Ausprägungsgrad charakterisiert den Benutzer bzgl. dieser Eigenschaft und die Sicherheitsrate repräsentiert den Glauben des Systems an das Zutreffen der Ausprägung (z.B. Facet: Sportlichkeit Value:3 Rating:600). Ändert sich das Wissen über den aktuellen Benutzer, z.B. durch Aktivierung eines neuen Stereotyps oder durch die Auflösung eines Konflikts, werden im Benutzermodell Ausprägungsgrade und Sicherheitsraten modifiziert. Die Modifizierung erfolgt durch Konfliktauflösungsstrategien.

Behandlung von Unsicherheit und Veränderlichkeit. Unsicherheiten in den Vorhersagen der Benutzereigenschaften werden durch Sicherheitsraten für das Zutreffen einer bestimmten Ausprägung repräsentiert. Die Veränderung dieser Sicherheitsraten wird anhand von Formeln durchgeführt, die keine wahrscheinlichkeitstheoretische Grundlage besitzen.

So wird z.B. die Sicherheitsrate eines Stereotyps etwas herabgesetzt, wenn das neue Wissen die Vorhersage des Stereotypen nicht unterstützt. Die Sicherheitsraten werden vom System insofern berücksichtigt, daß sie die Menge der Eigenschaften, die für die Buchauswahl herangezogen werden, auf die mit den höchsten Sicherheitsraten einschränkt.

Des weiteren ist der Stereotypenansatz für die Modellierung statischer Benutzereigenschaften konzipiert. Zeitveränderliche Variablen können mit dieser Methode nur umständlich bzw. nur durch eine Erweiterung des Ansatzes repräsentiert werden.

2.3.2.2 Diskussion

Ein Vorteil dieses Ansatzes ist, daß bereits nach wenigen Interaktionen mit dem Benutzer viele charakteristische Benutzereigenschaften dem Modell zur Verfügung stehen. Der Ansatz kann besonders gut dort eingesetzt werden, wo anhand eines stereotypischen Wissens passende Informationen aus einer großen Wissensbasis gefiltert werden sollen. Der Nachteil des Ansatzes ist, daß man die Inferenzen aufgrund der numerischen ad hoc Methoden schlecht nachvollziehen kann.

2.3.3 Fuzzy-Logik

Das Konzept der Fuzzy-Logik wurde eingeführt, um unpräzises oder vages Wissen zu verarbeiten. Da vages Wissen oder vage Konzepte in vielen Bereichen anzutreffen sind, hat diese Art der Unsicherheitsmodellierung weite Verbreitung gefunden, z.B. in Expertensystemen oder in der Steuerungselektronik. Der Kerngedanke der Fuzzy-Logik ist, ein Objekt nur zu einem gewissen Grad einer Menge zuzuordnen, im Gegensatz zur klassischen Mengentheorie, in der ein Objekt entweder zu einer bestimmten Menge gehört oder nicht.

Ein weiteres Konzept der Fuzzy-Logik ist das der *linguistischen Variablen*. Sie können Werte annehmen, die linguistischen Termen entsprechen. Z.B. kann eine linguistische Variable, die beschreibt, ob ein Benutzer ein Begriff kennt, die Werte *falsch*, *unwahrscheinlich*, *unsicher*, *wahrscheinlich* und *sicher* annehmen.

Die Verwendung der Fuzzy-Logik für die Benutzermodellierung wird anhand des Benutzermodellierungskomponente KNOME beschrieben.

2.3.3.1 KNOME: Benutzermodellierungskomponente von UC

KNOME[10] (siehe Chin, 1989) ist die Benutzermodellierungskomponente des UNIX-Consultant (UC), einem natürlich-sprachlichen Hilfesystem für UNIX. Anhand von Fragen des Benutzers, wie z.B. „How can I delete a file?“ schließt KNOME auf den Wissenstand des Benutzers und verwaltet ein Benutzermodell bzgl. seines UNIX-Wissens.

[10] KNOME = KNOwledge Model of Expertise

Tabelle 2.5: *Scalierung der Sicherheitsgraden und Wahrscheinlichkeitsänderungen*

False	Very unlikely	Unlikely	Somewhat unlikely	Uncertain	Somewhat likely	Likely	Very likely	True
-4	-3	-2	-1	0	+1	+2	+3	+4
Impos-sible	Much more unlikely	More unlikely	Somewhat more unlikely	Equally likely	Somewhat more likely	More likely	Much more likely	Certain

Aufbau des Benutzermodells. KNOME verwendet den Doppel-Stereotypen-Ansatz, der die oben beschriebenen Stereotypen erweitert. Die Benutzer werden nach ihrem Wissensstand in vier Klassen (NOVICE, BEGINNER, INTERMEDIATE, EXPERT) eingeteilt. Zusätzlich werden die UNIX-Kommandos und -Konzepte nach ihrem Schwierigkeitsgrad (SIMPLE, MUNDANE, COMPLEX, ESOTERIC) kategorisiert. Ein Doppel-Stereotyp entsteht durch die Verknüpfung zwischen Benutzergruppe und Kategorie. Der Aufbau des Benutzermodells erfolgt dadurch, daß der Benutzer den möglichen vier Benutzerklassen mit einer gewissen Wahrscheinlichkeit zugeordnet wird. Ausgehend von dieser Zuordnung kann KNOME vorhersagen, ob der Benutzer ein bestimmtes UNIX-Konzept kennt oder nicht. Dazu werden Vorhersageregeln (*Prediction rules*) benutzt, wie z.B.:

- IF User $\mathcal{U}$ is a BEGINNER AND the concept $\mathcal{C}$ is SIMPLE, THEN it is LIKELY that $\mathcal{U}$ knows $\mathcal{C}$.[11]

Beobachtet das System, daß der Benutzer ein bestimmtes UNIX-Konzept kennt oder nicht kennt, wird die Zuordnung des Benutzers bzgl. der Klassen neu ermittelt. Dazu dienen Interpretationsregeln (*Updating rules*), wie z.B.:

- IF the concept $\mathcal{C}$ is SIMPLE AND $\mathcal{U}$ knows $\mathcal{C}$ THEN it now seems MORE LIKELY that $\mathcal{U}$ is an EXPERT.

MORE LIKELY bezeichnet dabei eine Wahrscheinlichkeitsänderung relativ zur aktuell angenommenen Wahrscheinlichkeit, daß der Benutzer ein Experte ist. Die neue Wahrscheinlichkeit für die Zuordnung des Benutzers zu einer Gruppe ergibt sich durch die Addition der beteiligten Werte (siehe Tabelle 2.5). Wurde z.B. der Benutzer vor der Beobachtung mit UNLIKELY (-2) zur Gruppe der Experten zugezählt, wird er danach aufgrund der Verschiebung MORE LIKELY (+2) mit UNCERTAIN (0) zur Gruppe der Experten gezählt.

Behandlung von Unsicherheit und Veränderlichkeit. Die Unsicherheitsbehandlung erfolgt in KNOME nicht durch numerische Wahrscheinlichkeitsraten, sondern durch eine linguistische Variable, die die Werte TRUE, VERY-LIKELY, LIKELY, SOMEWHAT-LIKELY, UNCERTAIN, SOMEWHAT-UNLIKELY, UNLIKELY, VERY-UNLIKELY und FALSE annehmen kann.

[11]Diese Darstellung der Vorhersageregel wurde von Jameson (1996) entnommen.

Zeitveränderliche Benutzereigenschaften werden in KNOME nicht modelliert. Hat das System einen Benutzer mit TRUE einer Benutzerklasse zugeordnet, erfolgt auch keine Änderung dieser Einschätzung mehr. Ein Beispiel, wie dieser Fall auftreten kann, wird von Jameson (1996) beschrieben.

2.3.3.2 Diskussion

Die Art, wie Fuzzy-Logik mit Unsicherheiten umgeht, ist intuitiv und daher auch leicht nachvollziehbar. Da vage Konzepte auch oft von Menschen benutzt werden, um Unsicherheiten auszudrücken, ist diese Art der Unsicherheitsbehandlung für den Systemdesigner und den Benutzer einfach zu verstehen.

2.3.4 Dempster-Shafer-Theorie

Eine alternative Technik, Schlußfolgerungen unter Unsicherheit zu ziehen, verfolgt die Dempster-Shafer-Theorie (DST) (siehe z.B. Gordon & Shortliffe, 1984). Ihr Einsatz eignet sich besonders in den Fällen, in denen die Unsicherheit durch unvollständiges Wissen entsteht und nicht z.B. durch unsicheres Wissen über Kausalzusammenhänge (siehe Abschnitt 2.3.5, Bayessche Netze). Typisch für die Behandlung dieser Unsicherheit ist es, daß eine Evidenz Mengen von Hypothesen betreffen kann. Dagegen wird in einem Bayesschen Netz eine Evidenz nur einer Hypothese zugeordnet.

Die DST verwaltet für jede Hypothesenmenge drei Zahlen ($\in [0,1]$). Jeder Hypothesenmenge (H), die mit der Evidenz kompatibel ist, wird eine *Basiswahrscheinlichkeit* (engl. belief mass, kurz $m(H)$) zugeordnet. Bei völliger Unwissenheit wird der allgemeinsten Hypothesenmenge (der Menge, die aus allen Hypothesen besteht) die Basiswahrscheinlichkeit 1 zugeordnet. Da die Summe der Basiswahrscheinlichkeiten sich zu 1 addiert, werden demzufolge allen anderen Hypothesenmengen die Wahrscheinlichkeit 0 zugeordnet. Die Auswertung der Evidenz in DST kann betrachtet werden, als die Ermittlung der spezifischsten Hypothesenmenge, die die Evidenz unterstützt. Zusätzlich zu der Basiswahrscheinlichkeit wird jeder Hypothesenmenge eine *totale Wahrscheinlichkeit* (engl. total belief, kurz $Bel_m(H)$) und einen *Plausibilitätsfaktor* (engl. plausibility, kurz $Pl_m(H)$) zugeordnet, die wie folgt berechnet werden:

$$Bel_m(H) = \sum_{A \subseteq H} m(A), \qquad Pl_m(H) = \sum_{A \cap H \neq \emptyset} m(A).$$

Das Intervall $[Bel_m(H), Pl_m(H)]$ wird *Vertrauensintervall* genannt. $Bel_m(H)$ gibt an, wie hoch die Wahrscheinlichkeit ist, daß die wahre Hypothese in der Menge H liegt. $Pl_m(H)$ gibt den Vertrauensgrad an, zu dem die wahre Hypothese in der Hypothesenmenge H enthalten sein könnte.

Evidenzen aus einer Folge von Beobachtungen werden mit Hilfe der *Dempsterschen Regel* kombiniert. Diese Regel wird analog zu der Bayesschen Regel in den Bayesschen Netzen zur Interpretation nach erfolgter Evidenz eingesetzt. Im Gegensatz zur Fuzzy-Logik oder den Bayesschen Netzen erlaubt die DST aber nur diese Art von (diagnostischen) Inferenzen.

Verschiedene Benutzermodellierungskomponenten machen von der DST Gebrauch. Zum Beispiel verwendet Carberry (1990) die DST in einem Beratungssystem für Studenten. Dieses System versucht, anhand der Belegung von Kursen eines Studenten herauszufinden, welches Hauptfach (Mathematik oder Informatik) der Student hat. Ein andere Anwendung der DST ist das PHI-System, das im folgenden näher beschrieben wird.

2.3.4.1 PHI: Kombination von vergangenen und zukünftigen Evidenzen

Bauer (1996) benutzt die DST für die Erkennung von Plänen in dem intelligenten Hilfesystem PHI. Das System versucht herauszufinden, welchen Plan der Benutzer bei der Interaktion mit einem UNIX-Email-System verfolgt.

Aufbau des Benutzermodells. Bauer (1996) unterscheidet zwischen abstrakten Plänen und Basisplänen, die in einer Planhierarchie angeordnet sind. Jeder abstrakte Plan (z.B. Process-mail) kann somit aus einem der subsumierten Basisplänen (z.B. Read-mail) verwirklicht werden. Die Idee dieser Abstraktionshierarchie ist, daß es eine Menge von Plänen gibt, die denselben abstrakten Plan verwirklichen. Diese Mengen von Plänen werden als Hypothesenmengen angesehen. Die möglichen Aktionen des Benutzers werden als Evidenzen benutzt.

Hat der Benutzer bereits an mehreren Sitzungen mit dem System teilgenommen, werden die Basiswahrscheinlichkeiten der Hypothesenmengen gemäß der relativen Häufigkeit vergangener Aktionen des Benutzers initialisiert. Durch diese a-priori Einschätzung kann schon am Anfang der Interaktion das tendenzielle Verhalten des Benutzers bei der Planerkennung berücksichtigt werden. Abbildung 2.3 (A) zeigt ein Beispiel des Initialisierungsprozesses. In früheren Sitzungen hat der Benutzer tendenziell seine Nachrichten speichert. Die a-priori-Einschätzung spiegelt dieses tendenzielle Verhalten des Benutzers wider.

Werden Aktionen des Benutzer beobachtet, versucht das System den aktuellen Plan des Benutzers neu einzuschätzen. Diese Einschätzung erfolgt dadurch, daß die Aktion möglicherweise mehreren Hypothesenmengen gewichtet zugeordnet wird. Diese Gewichtung wird anhand von Beobachtungsregeln bestimmt, wie z.B.:

- WENN $\mathcal{U}$ eine Nachricht löscht DANN ordne dem Plan $\{$Read-delete-all-msgs$\}$ 0,25 und dem Plan $\Omega \setminus \{$Send-mail$\}$ das Plangewicht 0,75 zu.

Danach wird die Evidenz normal mit Hilfe der Dempsterschen Regel propagiert. Abbildung 2.3 (B) zeigt eine Fortsetzung des obigen Beispiels: Das System beobachtet, daß der Benutzer den Befehl „delete 5“ eingibt.

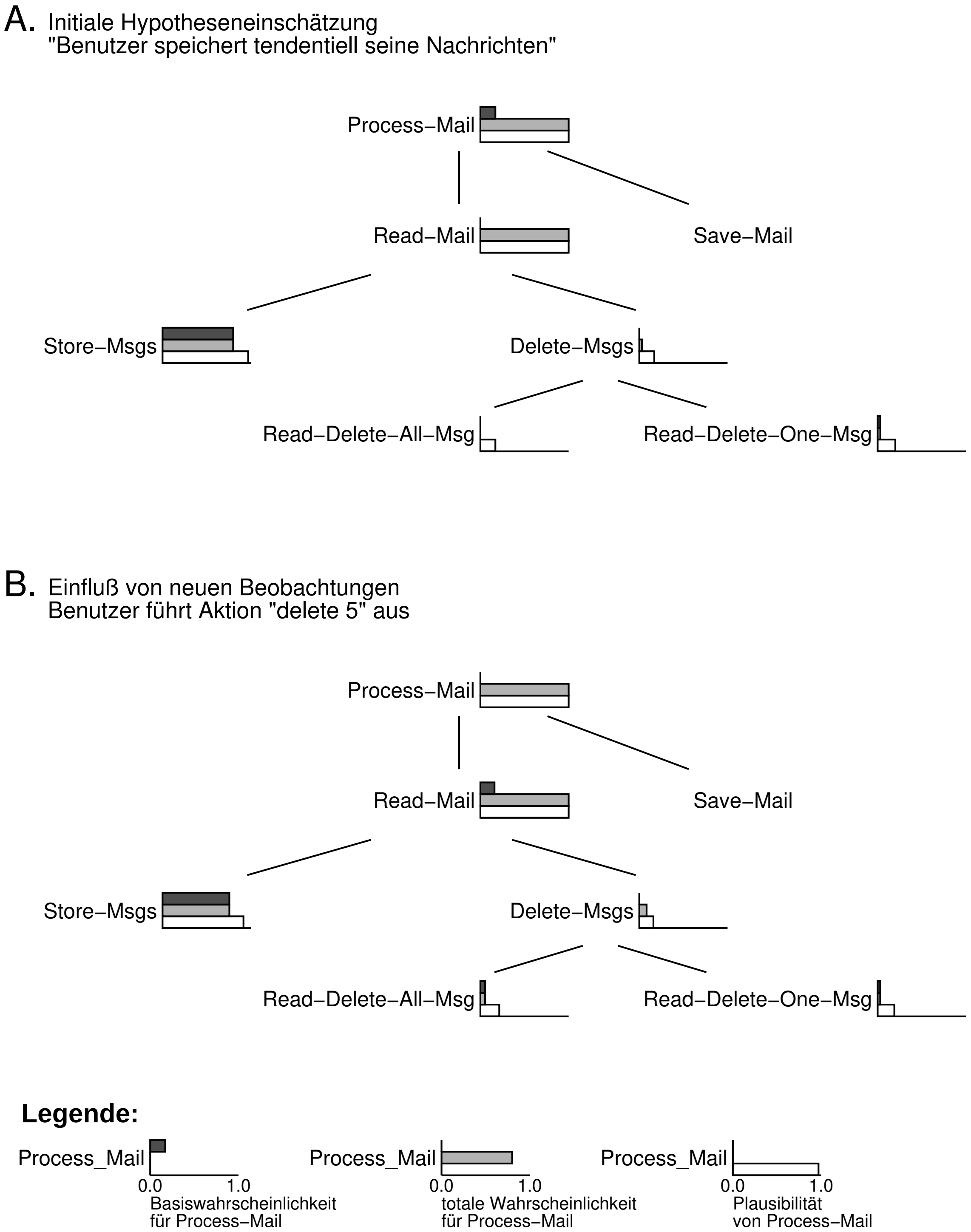

Abb. 2.3: *Verarbeitung von Benutzeraktionen mit DST, um Benutzerpläne zu erkennen (Ausschnitt der Planhierarchie). Die Notation wurde auf Jameson (1996) übernommen.*

Behandlung von Unsicherheit und Veränderlichkeit. PHI verarbeitet unsicheres und unvollständiges Wissen bzgl. des Benutzerverhaltens. Unsicheres Wissen entsteht durch die Annahmen, die über das Langzeitverhalten des Benutzers aufgestellt werden. Das Langzeitverhalten wird in Form von relativen Häufigkeiten bzgl. der Menge aller möglichen Pläne, die der Benutzer in der Vergangenheit ausgeführt hat, numerisch bewertet. Diese Bewertung wird dann als Beobachtungsregeln für die Beurteilung des zukünftigen Verhaltens herangezogen.

Unvollständiges Wissen resultiert dadurch, daß das System lediglich einen Ausschnitt des Benutzerverhaltens beobachten kann. Es besteht also Unwissenheit bzgl. aller Pläne und Aktionen, die der Benutzer in der Vergangenheit noch nicht ausgeführt hat, die allerdings auch nicht ausgeschlossen werden dürfen. Der Grad der Unwissenheit über eine Hypothesenmenge (H) wird in der DST durch die Differenz $Bel_m(H) - Pl_m(H)$ berechnet.

In PHI wird vereinfacht angenommen, daß sich die Interessen, Ziele oder Aufgaben des Benutzers nicht vollkommen ändern werden. Somit wird auch kein zeitveränderliches Verhalten des Benutzers berücksichtigt. Eine Konsequenz des zeitveränderlichen Verhaltens ergäbe eine andere Gewichtung der einzelnen Pläne.

2.3.4.2 Diskussion

Die DST erlaubt es, auf natürlichem Wege mit Evidenzen umzugehen. Z.B. kann eine Aussagen, wie „Der Plan, den $\mathcal{U}$ verfolgt, ist höchstwahrscheinlich nicht Send-Mail", mit der DST gut verarbeitet werden. Außerdem bietet die DST eine Möglichkeit, mit einer Folge von Evidenzen adäquat umzugehen. Nachteilig ist aber, daß eine Entscheidung, welche Hypothese der Menge der Evidenzen am ehesten entspricht, schwer zu treffen ist. Hierzu müssen meist zusätzliche Annahmen getroffen werden.

2.3.5 Bayessche Netze

Ein Bayessches Netz (für eine Einführung siehe Pearl, 1988 und Charniak, 1991) ist ein Inferenzformalismus, der auf der Wahrscheinlichkeitstheorie basiert. Im Unterschied zu den oben genannten Inferenzformalismen bezieht sich die Unsicherheit auf den Grad der Zusammenhänge in einem kausalen Modell, nicht so sehr auf die Unschärfe von Begriffen bzw. die Unvollständigkeit des Wissens, für deren Behandlung sich eher die Fuzzy-Logik bzw. die Dempster-Shafer-Theorie eignen.

Ein Bayessches Netz ist ein gerichteter azyklischer Graph, dessen Knoten Zufallsvariablen repräsentieren. Eine Zufallsvariable besteht aus einer Menge von Hypothesen, die sich gegenseitig ausschließen und alle möglichen Zustände der Zufallsvariable erschöpfend beschreiben. Für jede Hypothese wird die Wahrscheinlichkeit für ihr Zutreffen verwaltet. Alle diese Wahrscheinlichkeiten eines Knotens repräsentieren eine Wahrscheinlichkeitsverteilung, die der *Belief* eines Knotens genannt wird. Die Verbindungen zwischen den Knoten repräsentieren (kausale) Abhängigkeiten zwischen den Hypothesen. Eine solche Abhängigkeit wird durch eine Tabelle ausgedrückt, die für jeden möglichen Zustand des Knotens eine

Menge von bedingten Wahrscheinlichkeiten verwalten, gegeben jedes möglichen Zustandes der Elternknoten. Für Wurzelknoten (d.h. Knoten ohne Elternknoten) müssen a-priori Wahrscheinlichkeiten definiert werden.

Der Inferenzmechanismus der Bayesschen Netze erlaubt es, einerseits in kausale Richtung (von den Elternknoten zu den Kindknoten) als *Vorhersage* und andererseits in diagnostische Richtung (von den Kindknoten zu den Elternknoten) als *Interpretation* nach erfolgter Evidenz zu inferieren. Durch die Vorhersage können Annahmen über Benutzereigenschaften formuliert werden, die für die Interaktion mit dem Benutzer nützlich sind. Nach erfolgter Evidenz durch beobachtete Symptome werden die Einschätzungen der Knoten entsprechend der vorliegenden Situation angepaßt.

Ein Beispiel einer Anwendung von Bayesschen Netzen in der Benutzermodellierung ist das Dialogsystem PRACMA.

2.3.5.1 PRACMA: Bewertungsorientierte, nichtkooperative Dialoge

Das Dialogsystem PRACMA[12] (siehe Jameson, Schäfer, Simons & Weis, 1995) modelliert bewertungsorientierte Dialoge am Beispiel von Verkaufsgesprächen. Die Domäne ist der Gebrauchtautomarkt. Dabei wird von einer nicht ausschließlich kooperativen Dialogsituation ausgegangen: Der Käufer möchte dem Verkäufer möglichst detaillierte Informationen entlocken und das Auto zu einem möglichst geringen Preis erwerben. Der Verkäufer möchte negative Fakten über das Auto verschweigen und das Auto möglichst positiv dem Kaufinteressenten darbieten. PRACMA kann jeweils eine der Dialogrollen (Käufer/Verkäufer) übernehmen.

Aufbau des Benutzermodells. PRACMA versucht, das Wissen und das Interesse des Benutzers an bestimmten Bewertungsdimensionen eines Objekts einzuschätzen. Für die Gebrauchtwagendomäne werden z.B. die Dimensionen SICHERHEIT, SPORTLICHKEIT, KOMFORT und ZUVERLÄSSIGKEIT verwendet. In der Rolle des Verkäufers muß das System die Reaktionen des Benutzers auf die vom System gemachten Äußerungen vorhersagen und interpretieren. Abbildung 2.4 gibt ein Beispiel eines solchen Vorhersage- und Intepretationsprozesses. Das Bayessche Netz wird verwendet, um $\mathcal{P}$s Bewertung der Wichtigkeit des Kilometerstandes einzuschätzen.

Behandlung von Unsicherheit und Veränderlichkeit. In der Domäne von PRACMA resultiert die Unsicherheit daraus, daß einerseits keine sicheren Einschätzungen über die Interessen und das Wissen des Benutzers vorgenommen werden können und andererseits auch anhand von Benutzeräußerungen nicht sicher auf dessen Bewertungskategorien geschlossen werden kann.

Zeitveränderliche Benutzervariablen werden in PRACMA nicht modelliert. Es wird davon ausgegangen, daß in einem kurzen Dialog, die für PRACMA relevanten Benutzereigenschaften, wie z.B. das Interesse, konstant bleiben.

[12] **PR**ocessing **A**rguments between **C**ontroversially-**M**inded **A**ctors

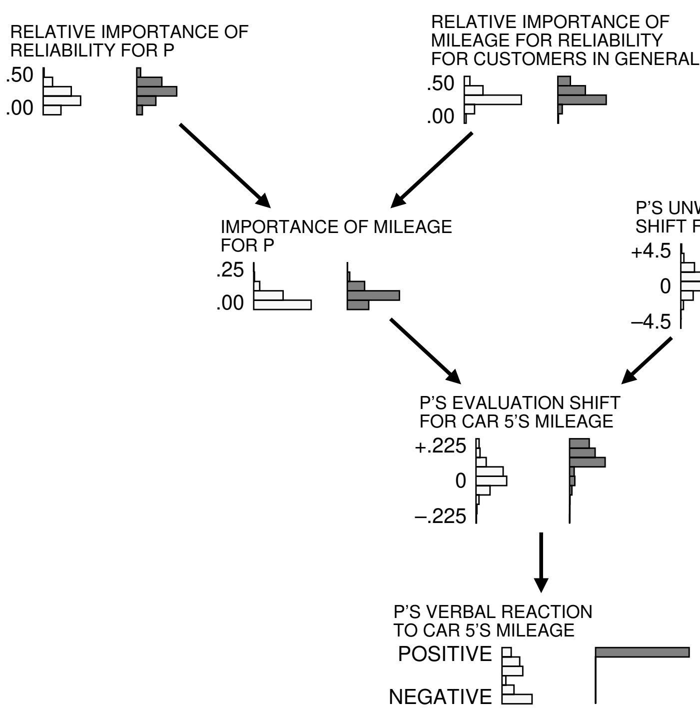

Abb. 2.4: *Bayessches Netz von* PRACMA *zur Vorhersage (helle Histogramme) und Interpretation (dunkle Histogramme) von $\mathcal{P}$s Reaktion auf den Kilometerstand.*

2.3.5.2 Dynamische Bayessche Netze

Die dynamischen Bayesschen Netze (DBNs) (siehe Dagum, Galper & Horvitz, 1992; Kjærulff, 1992; Nicholson & Brady, 1992 und Tawfik & Neufeld, 1994) stellen eine Erweiterung der Bayesschen Netze dar, mit denen sich auch zeitveränderliche Variablen modellieren lassen. Dagum et al. (1992) benutzen z.B. DBNs, um den Verkauf von amerikanischen Autos in Japan vorherzusagen. Die zeitveränderliche Variable dabei ist das Angebot amerikanischer Autos auf dem japanischen Markt. Eine weitere Anwendung ist die Lokalisierung eines Fahrzeuges (siehe z.B. Nicholson & Brady, 1992).

In einem DBN wird der zeitliche Verlauf durch *Zeitscheiben* repräsentiert. Jede Zeitscheibe entspricht einem Zeitraum, während dessen die zeitlichen Veränderungen der einzelnen Variablen vernachlässigt werden. Zeitveränderliche Eigenschaften werden als *dynamische Knoten* repräsentiert, die über die Zeitscheiben hinweg miteinander verbunden sind. Zeitlich unveränderliche Variablen werden entweder als statische oder temporäre Knoten repräsentiert. *Statische Knoten* repräsentieren statische Benutzereigenschaften, wie z.B. das Wissenniveau oder die Interessen des Benutzers. Sie sind in allen Zeitscheiben vorhanden, werden aber nur einmal definiert. *Temporäre Knoten* werden für Variablen verwendet, die nur zu bestimmten Zeitpunkten zusätzliche Informationen liefern. Zu ihnen gehören z.B. Variablen, die für beobachtete Symptome stehen.

Gewöhnlich repräsentieren einige Zeitscheiben die Vergangenheit, eine Zeitscheibe die Gegenwart (Zeitscheibe t) und einige Zeitscheiben die Zukunft (siehe Abb. 2.5). Die Zeit-

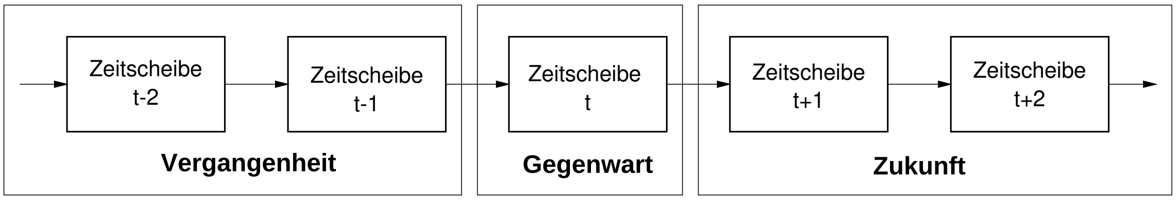

Abb. 2.5: *Einteilung des zeitlichen Verlaufs in Zeitscheiben.*

scheiben können entweder eine konstante oder variable Zeitdauer darstellen. Jensen (1995) benutzt z.B. für die Vorhersage über die Entwicklung des Winterweizens Zeitscheiben mit konstanter Zeitdauer von einer Woche. Das Hinzufügen einer neuen Zeitscheibe kann auch durch unregelmäßig auftretende Ereignisse geschehen, so daß die Zeitscheiben unterschiedliche Längen haben. Eine derartige Technik verwenden z.B. Nicholson und Brady (1994) in ihrem Roboter-Überwachung-System, in dem die Ereignisse Veränderung des Roboterzustandes (z.B. die Änderung der Bewegungsrichtung) repräsentieren.

Eine Beispielanwendung eines DBN ist die Planerkennung in einem Multi-User Dungeon Spiel. Dieses System wird im folgenden beschrieben.

2.3.5.3 Planerkennung in einem Multi-User Dungeon (MUD) Spiel

Ein MUD ist ein textbasiertes virtuelles Spiel, in dem Spieler gegeneinander oder miteinander antreten, die vorher festgelegten Ziele zu erreichen. Ein solches Ziel kann z.B. „die Rettung“ eines Teddy-Bären sein. Die Spieler kennen normalerweise das Ziel, wissen aber nicht, welche Aktionen benötigt werden, um das Ziel zu erreichen. Das System von Albrecht, Zukerman, Nicholson und Bud (1997) versucht, so früh wie möglich das Ziel eines Spielers zu erkennen und seine nächsten Aktionen vorherzusagen.

Aufbau des Benutzermodells. Das Benutzermodell wird durch ein DBN dargestellt, dessen Aufgabe es ist, die nächsten Aktionen und Aufenthaltsorte des Spielers in dem MUD vorherzusagen.[13] Darauf basierend kann auf das Ziel des Spielers zurückgeschlossen werden. Das Ziel des Spielers wird in dem Modell (siehe Abb. 2.6) als zeitlich unveränderlich angenommen, d.h. es wird angenommen, daß das aktuelle Ziel des Spielers sich nicht ändert.

Aktionen und Aufenthaltsorte sind hingegen zeitveränderlich. Die Aktion (A_t) und der Aufenthaltsort (L_t) in Zeitscheibe t hängen von dem aktuellen Ziel (Q) und von der vorangegangenen Aktion bzw. dem vorangegangenen Aufenthaltsort in $t - 1$ ab. Die Zahlen für die qualitativen Abhängigkeiten zwischen den Zusammenhängen wurden anhand von Trainingsläufen ermittelt.

Behandlung von Unsicherheit und Zeitveränderlichkeit. Der System erhält seine Informationen hauptsächlich durch die Tastaturkommandos des Benutzers, was dazu führt,

[13] Das MUD besitzt über 4700 Aufenthaltsorte, über 7200 Aktionen und 20 verschiedene Ziele.

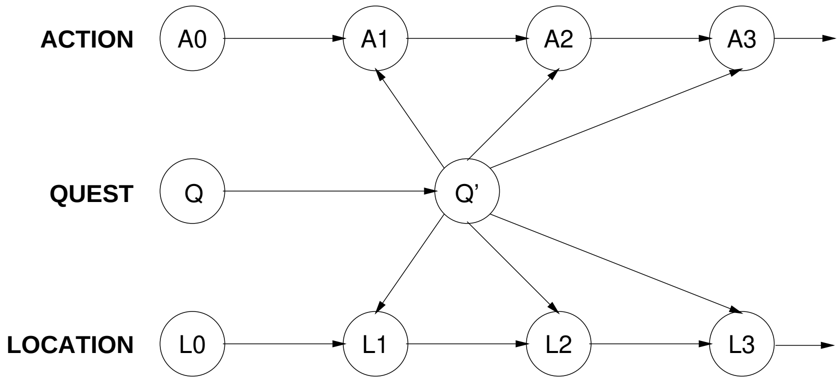

Abb. 2.6: *Modell eines dynamisches Bayessches Netz für die Vorhersage des Ziel eines Spielers in einem MUD.*

daß das System nur über ein begrenztes und daher unsicheres Wissen bzgl. des Benutzers verfügt. Außerdem sind die Ergebnisse von Aktionen unsicher, da die Ausführung einer Aktion nicht unbedingt ausreicht, den beabsichtigten Effekt der Aktion herbeizuführen (z.B. aufgrund anderer Spieler, die auch Aktionen ausführen können).

2.3.5.4 Weitere Anwendungen von Bayesschen Netzen

Bayessche Netze werden in verschiedenen Systemen zur benutzerorientierten Präsentation von Informationen eingesetzt. Das multimodale Präsentationssystem PPP benutzt Bayessche Netze in seiner Benutzermodellierungskomponente PEDRO (siehe van Mulken, 1996), um die Verständlichkeit der Präsentation technischer Anleitungen vorherzusagen. Horvitz und Barry (1995) verwenden Bayessche Netze in dem System VISTA-III, um komplexe Informationen für zeitkritische Entscheidungen darzustellen. Das System unterstützt einen Techniker im NASA-Kontrollzentrum, der die Antriebsaggregate des Space Shuttles überwacht.

Ein weiteres Beispiel, in dem Bayessche Netze für die Benutzermodellierung eingesetzt werden, ist das Lumiere-Projekt (siehe Horvitz, Breese, Heckerman, Hovel & Rommelse, 1998). In diesem Projekt wird ein Assistent entwickelt, der den Benutzer bei der Benutzung seiner Software unterstützt. Der Prototyp dieses Assistenten dient als Basis für den *Office Assistant* im Microsoft Office Paket.

Des weiteren werden Bayessche Netze in einigen technischen Anwendungen eingesetzt. Nicholson und Brady (1994) beschreiben die Entwicklung eines Systems zur Überwachung eines mobilen Roboters. Das System benutzt sensorische Beobachtungsdaten diskreter Ereignisse, um ein DBN zu konstruieren. Forbes, Huang, Kanazawa und Russell (1995) verwenden DBNs in ihrem BAT-System (**B**ayesian **A**utomated **T**axi). Das System hat die Aufgabe, ein autonomes Fahrzeug in einem normalen Straßenverkehr zu führen. Anhand von Sensoren, die Informationen aus der Umgebung aufnehmen, werden Informationen über Position, Geschwindigkeit und aktuelle Intention der Fremdfahrzeuge abgeleitet.

2.3.5.5 Diskussion

Im folgenden werden die Vorteile und Nachteile des Bayesschen Ansatzes beschrieben.

Nachteile des Bayesschen Ansatzes

(1) *Wo kommen die Zahlen her?*
Ein viel diskutiertes Problem der Bayesschen Netze ist die Herkunft der Zahlen, die für die Beschreibungen der quantitativen Zusammenhänge verwendet werden. Oftmals basieren die definierten Abhängigkeiten auf dem intuitiven Urteil des Entwicklers eines Bayesschen Netzes. Selten werden empirische Untersuchungen für die quantitative Modellierung hinzugenommen. In der vorliegenden Arbeit konnten die explorativen empirischen Untersuchungen hauptsächlich zur Klärung der qualitativen Kausalzusammenhängen benutzt werden. Die quantitativen Abhängigkeiten mußten auch hier mehr oder weniger so gewählt werden, daß sie den Aussagen der Versuchspersonen Rechnung tragen können.
Zu beachten ist aber, daß die Zahlen, die für die quantitativen Zusammenhänge benutzt werden, nicht im Sinne einer frequentistischen Interpretation der Wahrscheinlichkeit gemeint sind. Die frequentistische Interpretation der Wahrscheinlichkeit betrachtet die Wahrscheinlichkeit aus der Sicht von relativen Häufigkeiten eines Zufallsexperimentes, was voraussetzt, daß man eine bestimmte Anzahl von Ereignissen und ihre Ergebnisse beobachtet hat. Bei vielen Zusammenhängen ist eine solche Beobachtung gar nicht möglich. Der Bayessche Ansatz verfolgt vielmehr die logische Konzeption der Wahrscheinlichkeit (vgl. Spies, 1993). Sie interpretiert die Wahrscheinlichkeit als Maß für den Grad des Glaubens einer Person in bezug auf das Zutreffen einer Aussage.

(2) *Problem der mehrfachverbundenen Netze.*
Die Berechnung der Beliefs für ein einfachverbundenes Netz kann effizient durch einen Propagierungsalgorithmus (siehe Pearl, 1988) durchgeführt werden. Wenn das Netz mehrfachverbunden[14] ist, ist eine einfache Belief-Propagierung nicht mehr möglich. Für derartige Fälle existieren Inferenzalgorithmen, die auf Clustering, Conditioning oder stochastischer Simulation basieren. Das Problem dabei ist, daß diese Algorithmen unter Umständen eine exponentielle Laufzeit bzgl. der Anzahl der Knoten haben können (siehe Cooper, 1990).

(3) *Rollup.*
Verstreicht eine gewisse Zeit, so wird eine weitere Zeitscheibe an die bereits existierenden angehängt. Auf diese Weise vergrößert sich ein DBN im Laufe der Zeit. Dabei enstehen zwei Probleme: nach einigen Zeitscheiben wird das Netz sehr unhandlich und die Laufzeit des Evaluationsalgorithmus wächst kontinuierlich. Durch Anwenden einer *Rollup-Prozedur*, bei der alte Zeitscheiben gelöscht werden, kann das Netz vereinfacht werden. Ein Problem dabei ist, daß ein Rollup nicht für jede Netzstruktur ohne Infor-

[14]Ein Netz ist mehrfachverbunden, wenn zwei Knoten durch mehr als einen Pfad miteinander verbunden sind.

mationsverlust durchgeführt werden kann. Forbes et al., 1995 zeigen einfache Fälle, in denen ein Rollup ohne Informationsverlust anwendbar ist.

Vorteile des Bayesschen Ansatzes

(1) *Behandlung von Unsicherheit.*
Viele Zusammenhänge in der Welt lassen sich anhand von Ursache-Wirkungs-Beziehungen beschreiben, von denen man aber kein sicheres Wissen bzgl. ihres Grades des Zusammenhangs hat. Man spricht davon, daß der Zusammenhang nicht deterministisch, sondern eher probabilistisch ist. Gerade diese Art der Unsicherheit lassen sich mit Hilfe der Bayesschen Netze gut beschreiben.

(2) *Behandlung von Zeitveränderlichkeit.*
Wie in Abschnitt 2.3.5.2 beschrieben wurde, können mit Hilfe von DBNs sehr gut zeitliche Zusammenhänge beschrieben werden. Zeitveränderliche Benutzervariablen werden in den DBNs als dynamische Knoten repräsentiert, die zusätzlich zu ihren zeitlich unveränderlichen Einflüssen von ihrer eigenen früheren Einschätzung abhängen.

(3) *Modellierungsschemata.*
Wie sich später zeigen wird, sind zeitveränderliche Variablen schwieriger zu modellieren als zeitlich unveränderliche, insbesondere dann, wenn der dynamische Knoten, der diese zeitveränderliche Variable repräsentiert, von einem anderen dynamischen Knoten oder von einem statischen Knoten beeinflußt wird. Aus diesem Grund wurden Modellierungsschemata für zeitveränderliche Variablen entwickelt, die die Modellierung vereinfachen. Diese Schemata werden in Abschnitt 4.2 beschrieben.

Zusammenfassend läßt sich sagen, daß DBNs die vorliegenden Kriterien, Behandlung von Unsicherheit (kausaler Zusammenhängen) und Zeitveränderlichkeit, am besten entsprechen. Andere Inferenzmethoden, wie z.B. die DST oder Fuzzy-Logik, können zwar auch gut Unsicherheiten verarbeiten, die Modellierung zeitveränderlicher Variablen erfordert aber oft zusätzliche Erweiterungen. Aus diesem Grund wurde in der vorliegenden Arbeit der Inferenzmechanismus der Bayessche Netze verwendet.

Kapitel 3

Empirie: Einschätzung von Arbeitsgedächtnisbelastung in Notrufdialogen

Wie in Abschnitt 2.1 gezeigt wurde, können viele Aspekte des Sprechens auf die Arbeitsgedächtnisbelastung des Dialogpartners hindeuten. Dennoch stellt sich die Frage, inwiefern solche Hinweise in alltäglichen Dialogsituationen für die Dialogführung berücksichtigt werden. Eine explorative empirische Studie soll diesen Punkt herausstellen. Als Domäne wurde eine Feuerwehrnotrufzentrale (FNZ) gewählt. Mitarbeiter der FNZ, die regelmäßig Notrufe entgegennehmen, dienten dabei als Versuchspersonen. Obwohl die Notrufdomäne von dem eigentlichen READY-Szenario abweicht, ergänzt diese Studie die gestellte Thematik insofern, als die Ressourcenbeschränkungen der Anrufer hier noch stärker ausgeprägt sind als in der READY-Domäne und die Versuchspersonen besonders viel Erfahrung im Umgang mit solchen Ressourcenbeschränkungen haben. Es ist also damit zu rechnen, daß die Versuchspersonen brauchbare Zusammenhänge zwischen Ursachen und Symptomen von Arbeitsgedächtnisbelastung nennen. Des weiteren ist zu untersuchen, inwiefern die von den Mitarbeitern der FNZ genannten Zusammenhänge mit den psychologischen Ergebnissen aus Abschnitt 2.1 übereinstimmen.

In bezug auf Arbeitsgedächtnisbelastung und Zeitdruck sind folgende Aspekte von Bedeutung:

(1) Notrufe werden meistens in großer Eile geführt. Die wenigsten Anrufer bereiten sich auf das Gespräch vor. Daher ist mit einem hohen Grad an Spontansprache zu rechnen.

(2) Da die Anrufer zusätzlich durch die Vorfälle, die um sie herum passieren, abgelenkt sind, können selbst einfache Fragen hohe Anforderungen an das Arbeitsgedächtnis des Anrufers stellen.

Im diesem Kapitel wird die explorative empirische Studie näher beschrieben. Die Ergebnisse, die aus der Analyse der Äußerungen der Versuchspersonen gewonnen wurden, liefern ein Bild von qualitativen Kausalzusammenhängen, die im Kapitel 4 Grundlagen für die Modellierung bilden. Eine erste Auswertung wurde von Schäfer und Weyrath (1996) vorgenommen.

Tabelle 3.1: *Transliteration des Notrufdialogs: Frau in St. Johann*

1	$\mathcal{P}$:	<#Stöhnen>.
2	$\mathcal{S}$:	**Feuerwehr.**
3	$\mathcal{P}$:	Ja, hier ist Frau Müller, Schopenhauerstraße 10. Bei mir brennt der Badeautomat. Da war heute morgen noch ein Installateur da. <lautes Atmen>. Jetzt brennt das Ding. <#Atmen> [1][2]
4	$\mathcal{S}$:	**Schopenhauerstraße 10?**
5	$\mathcal{P}$:	Schopenhauerstraße 10.
6	$\mathcal{S}$:	**In welchem Stadtteil?** [3][4]
7	$\mathcal{P}$:	Bitte? [5]
8	$\mathcal{S}$:	**In welchem Stadtteil?**
9	$\mathcal{P}$:	Ei, das is' hier in Saarbrücken <Z> in St. Johann, in der Nähe der LVA. [6][7]
10	$\mathcal{S}$:	**Welche Etage?** [8]
11	$\mathcal{P}$:	<%> hoch [9]
12	$\mathcal{S}$:	**Welche Etage is denn das?**
13	$\mathcal{P}$:	Äh, die erste. <#Stöhnen> [10]
14	$\mathcal{S}$:	**mache, mache, <P>**
15	$\mathcal{P}$:	<#leise> oah schnell! <#lautes Atmen>
16	$\mathcal{S}$:	**joh, mache 'se die Tür zu, wir kommen sofort, ne?**
17	$\mathcal{P}$:	ja, es iss gut.
18	$\mathcal{S}$:	**ja, tschüß.** [11][12][13]

Anmerkung: $\mathcal{P}$ = Anrufer, $\mathcal{S}$ = Mitarbeiter der FNZ, <P> = Pause, <Z> = Zögern, <#> = nichtartikulartorisches Geräusch, % = Schwerverständliches, <%> = Unverständliches, [n] = Band während der Untersuchung gestoppt und Frage n gestellt (siehe auch Anhang A.1).

3.1 Beschreibung des Szenarios

Die Aufgabe eines Mitarbeiters der FNZ während eines Notrufgespräches ist, in möglichst kurzer Zeit Fakten über Ort und Art des Ereignisses herauszufinden, um adäquat handeln zu können. Tabelle 3.1 zeigt einen solchen Notrufdialog. Das Problem besteht darin, daß die Anrufer meistens unter Zeitdruck und mentaler Belastung stehen. Daher kann eine angemessene Kooperation des Anrufers nicht vorausgesetzt werden. Um die notwendigen Informationen zu erhalten, muß der Mitarbeiter Fragen über das Ereignis stellen. Wichtig dabei ist eine richtige Einschätzung der Anforderung, die die Äußerung an das Arbeitsgedächtnis des Dialogpartners stellt. Ein Aspekt der Untersuchung ist daher, herauszufinden, inwieweit sich der Mitarbeiter ein Bild davon macht, wie sehr das Arbeitsgedächtnis des Anrufers belastet ist.

Tabelle 3.2: *Transliteration des Notrufdialogs: Mann in der Gaststätte*

1	$\mathcal{S}$:	<#>
2	$\mathcal{P}$:	Bitte?
3	$\mathcal{S}$:	**Feuerwehr.**
4	$\mathcal{P}$:	Dieter Müller is mein Name. Un zwar <P> in de <#> <;redet mit jemanden anderes> direkt do vorne. nä, nä, do hinne, <P> Trifelsstroß. <;redet jetzt mit Feuerwehrmann> Äh, Trifelsstroß, owwe im Dachgeschoß is e Feuer. [1][2][3]
5	$\mathcal{S}$:	**Trifelsstraße?**
6	$\mathcal{P}$:	Ja, <P>
7	$\mathcal{S}$:	**Was für Hausnummer denn?** [4]
8	$\mathcal{P}$:	12.
9	$\mathcal{S}$:	**Wie war noch ihr Name?**
10	$\mathcal{P}$:	Dieter Müller, 47081511 DRK-Nummer. [5][6]
11	$\mathcal{S}$:	**Geben's mir noch ihre Telefonnummer!**
12	$\mathcal{P}$:	Äh, die Telefonnummer von hier? <%><P> Ich ruf aus einer Gaststätte an.
13	$\mathcal{S}$:	**Aus einer Gaststätte. Genn 'se mir die Nummer von der Gaststätte!** [7][8]
14	$\mathcal{P}$:	<;Zur Wirtin> Könn 'se mir gerade die Gaststättennummer äh äh die Nummer? <P> Bitte? <P> 25 <P> <;Im Hintergrund werden Zahlen genannt>
15	$\mathcal{S}$:	**27.**
16	$\mathcal{P}$:	25 <P>
17	$\mathcal{S}$:	**27.** [9][10]
18	$\mathcal{S}$:	**Also in der Trifelsstraße Nummer 12?**
19	$\mathcal{P}$:	Bin vorm Haus dann.
20	$\mathcal{S}$:	**Im Dachgeschoß?**
21	$\mathcal{P}$:	Im Dachgeschoß owwe.
22	$\mathcal{S}$:	**Ja, wir kommen sofort.**
23	$\mathcal{P}$:	Alles klar. [11][12]

Anmerkung: $\mathcal{P}$ = Anrufer, $\mathcal{S}$ = Mitarbeiter der FNZ, <P> = Pause, <Z> = Zögern, <#> = nichtartikulartorisches Geräusch, % = Schwerverständliches, <%> = Unverständliches, [n] = Band während der Untersuchung gestoppt und Frage n gestellt (siehe auch Anhang A.1).

3.2 Methode

3.2.1 Versuchspersonen

An der Studie nahmen 11 männliche Mitarbeiter der Feuerwehr Saarbrücken, die in der Notrufzentrale arbeiten, teil.[1]

[1] Ich bedanke mich bei den Mitarbeitern der Feuerwehr Saarbrücken für die Mithilfe bei der empirischen Studie, insbesondere denen, die sich als Versuchspersonen zur Verfügung stellten.

3.2.2 Material

In der Orientierungsphase der Studie wurden 15 Notrufdialoge auf Band aufgenommen. Aus diesen wurden zwei Dialoge ausgewählt, die für die Untersuchung am geeignetesten erschienen. Die Auswahl traf zwei Notrufe, in denen die Anrufer wenig Erfahrung mit Notrufgesprächen hatten (siehe Tabellen 3.1 und 3.2).[2] Für die Studie wurden zwei Tonbandgeräte benutzt, eines zum Abspielen der Notrufdialoge und ein weiteres zum Aufnehmen des Gesprächs. Die Bandaufzeichnungen wurden später transliteriert.

3.2.3 Durchführung

Jedes Gespräch dauerte ca. 30-40 Minuten. Jede Versuchsperson bekam beide Notrufdialoge von Band vorgespielt. An bestimmten Stellen wurde das Band gestoppt und Fragen dazu gestellt (siehe Anhang A.1), die spontan, d.h. ohne lange nachzudenken, beantwortet werden sollten. Die getroffenen Einschätzungen sollten möglichst mit denen vergleichbar sein, die der Mitarbeiter in einer echten Notrufsituation gemacht hätte. Unmittelbar danach wurde die Versuchsperson aufgefordert, im Sinne des retrospektiven lauten Denkens (siehe Ericsson & Simon, 1993) ihre Gedächtnisinhalte zu verbalisieren. Die Verbalisierung wurde durch die Instruktion „Was haben Sie dabei gedacht?“ zusätzlich motiviert. Da die Technik des retrospektiven lauten Denkens für die Versuchspersonen neu war, wurde dieser Aspekt ausführlich erklärt und an einem Beispiel erprobt. Dazu wurde ein dritter (hier nicht weiter erwähnter) Notrufdialog benutzt.

Zwei Fragetypen standen in der Untersuchung im Vordergrund: Fragen nach der Einschätzung einer Situation durch eine Vorhersage (z.B. „Wie schnell und deutlich wird sie die Frage beantworten können?“) und Fragen nach den Symptomen für evtl. Ressourcenbeschränkungen des Anrufers (z.B. „Wie gut hat sie die Frage beantwortet?“). Am Ende jedes Notrufdialogs wurden zusätzlich Fragen nach der Einschätzung von Aufgabenschwierigkeiten gestellt (z.B. „Stellen Sie sich vor, der Kollege fragt nach der Telefonnummer der Wohnung. Die Anruferin ist aber Gast in der Wohnung und muß daher die ihr nicht bekannte Nummer erst erfragen. Wieviele Ziffern hätte wohl die längste Nummer, die sie jetzt richtig durchgeben könnte?“)

Ein wichtiger Aspekt der Untersuchung war es, zu dem Ergebnis auch den Prozeß der Antwort aufzuzeigen. Dies wurde durch die retrospektive Verbalisierung erreicht, die die Gedankenabläufe der Versuchsperson, die sie während der zu treffenden Einschätzung hatte, wiedergeben soll. Das Herauslocken derartiger Einschätzungen durch die retrospektive Verbalisierung ist schwierig und nicht immer von Erfolg gekrönt. Oft werden nicht die zur Einschätzung passenden Gedankenabläufe verbalisiert, sondern z.B. generelle Äußerungen gegeben. Hier gilt es, die Spreu vom Weizen zu trennen. Ericsson und Simon (1993) beschreiben mehrere Ursachen, die zu Problemen bzgl. der retrospektiven Verbalisierung führen können (siehe Anhang A.2).

[2]Unter den 15 Notrufdialogen gab es auch einige, die von Polizeibeamten aufgegeben wurden. Hier waren weniger oder keine Ressourcenbeschränkungen zu erkennen.

Die aufgezeichneten Aussagen der Versuchspersonen wurden später transliteriert und bei der Analyse nach (z.T. impliziten) Verweisen auf Ursachen und Symptomen von Arbeitsgedächtnisbelastung gesucht. Die Ergebnisse dieser Analyse werden im nächsten Abschnitt erläutert.

3.3 Ergebnisse

3.3.1 Umschreibung des Arbeitsgedächtnisses

Im ersten Schritt der Datenauswertung wurde untersucht, wie die Mitarbeiter der FNZ den Begriff Arbeitsgedächtnis verwenden. Arbeitsgedächtnis bzw. verfügbare Arbeitsgedächtniskapazität sind Begriffe der kognitiven Psychologie und wurden daher von den Versuchspersonen in den Gesprächen nicht benutzt. Dennoch wurden sie durch alltagssprachliche Ausdrücke umschrieben. Von den Mitarbeitern der FNZ wurde z.B. der Begriff *Konzentriertheit* benutzt. Dieser Begriff wird hier mit Arbeitsgedächtnisbelastung assoziiert. Das bedeutet: wenn ich schreibe, „die Versuchspersonen sehen X als Folge oder Ursache von Arbeitsgedächtnisbelastung", meine ich „wenn sie den Begriff kennen würden, würden sie dieser Annahme zustimmen". Diese Vereinfachung ist insofern sinnvoll, weil die Versuchspersonen keine direkten Hinweise auf das Arbeitsgedächtnis und die Arbeitsgedächtnisbelastung gaben. Das Modell des Arbeitsgedächtnisses, das dieser Untersuchung zugrunde gelegt wird, ist folglich sehr einfach und undifferenziert. Beispiele von Aussagen über das Arbeitsgedächtnis sind:

> Vp1: *[...] daß die Frau, wie gesagt, sehr unter Streß ist, sehr ängstlich ist und sich wahrscheinlich auf eine Telefonnummer nicht konzentrieren kann.*
>
> Vp3: *Hat sich ja auch ziemlich gefangen, beim letzten Teil von der Antwort, die sie gegeben hat. Dann wird sie das da auch ruhig beantworten können.*
>
> Vp2: *Er hat sich offenbar inzwischen gefaßt und er weiß, was er von sich gibt. Er konzentriert sich auf das Gespräch [...]*
>
> Vp3: *Er war also ziemlich unkonzentriert und die Aussage war ziemlich oberflächlich.*

Das Verständnis der Terminologie der Versuchspersonen ist für die detaillierte und genaue Analyse der Gespräche wichtig (siehe Scott, Clayton & Gibson, 1991, Chapter 12 „Knowledge Acquisition and Interviewing"). Die Identifikation der Begriffe in den Aussagen der Versuchspersonen helfen, die Symptome und Ursachen von Arbeitsgedächtnisbelastung zu erkennen. Diese werden in den nächsten Abschnitten näher beschrieben.

3.3.2 Ursachen hoher Arbeitsgedächtnisbelastung

Die Mitarbeiter der FNZ haben drei unmittelbare Ursachen hoher Arbeitsgedächtnisbelastung genannt. Die Zusammenhänge werden beispielhaft anhand von Aussagen der Mitarbeiter belegt.

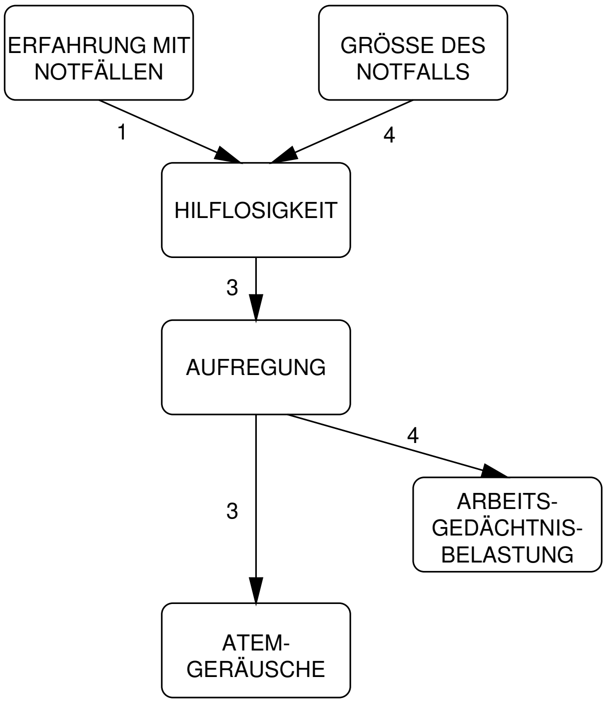

Abb. 3.1: *Rekonstruktion der Hypothesen der Versuchspersonen über Ursachen und Symptome von Aufregung und Arbeitsgedächtnisbelastung. (Ein Pfeil von A nach B bedeutet, daß A B beeinflußt. Die Zahl neben dem Pfeil spiegelt die Häufigkeit der Nennung wider).*

Aufregung. Viele Personen, die in einer Notrufzentrale anrufen, sind sehr aufgeregt. In einigen Fällen besteht Gefahr für das eigene Leben oder das von anderen. Die Versuchspersonen erwähnten dazu folgende Zusammenhänge (siehe Abb. 3.1): Eine Ursache von Arbeitsgedächtnisbelastung ist eine hohe Aufregung (Vp 2), worauf z.B. viele Atemgeräusche hinweisen (Vp 3). Ein Grund für hohe Aufregung ist, daß viele Anrufer dem Notfall hilflos gegenüber stehen. Die Größe des Notfalls und die fehlende Erfahrung mit Notfällen sind offensichtliche Ursachen dafür (Vp 1, Vp 6). Typische Äußerungen der Versuchspersonen waren:

Frage 1.1: Wie gut hat sie die Situation beschrieben?

Vp 2: *Schlecht.* – retrospektives Denken – *Sie ist aufgeregt, und sie weiß nicht was brennt.*

Frage 1.2: [3] Warum, denken Sie, beschreibt sie die Situation auf diese Weise?

Vp 1: *Ja, ich merke wohl, daß sie ängstlich ist und unter Streß steht und nicht genau mit der Situation fertig wird. Weiß nicht was sie damit anfangen soll.*

Vp 6: *Weil sie sich nicht anders weiß auszudrücken. Weil sie keine Erfahrung mit solchen Sachen hat.* – retrospektives Denken – *Sie ist mit so einer Sache*

[3] Anmerkung: Frage 1.2 bedeutet erster Notrufdialog, Frage 2.

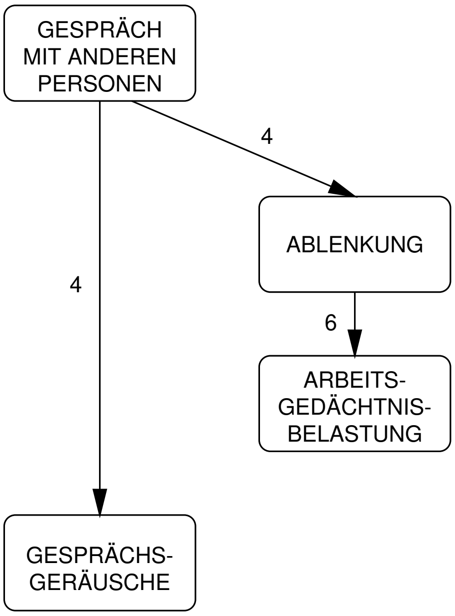

Abb. 3.2: *Rekonstruktion der Hypothesen der Versuchspersonen über Ursachen und Symptome von Ablenkung und Arbeitsgedächtnisbelastung. (Ein Pfeil von A nach B bedeutet, daß A B beeinflußt. Die Zahl neben dem Pfeil spiegelt die Häufigkeit der Nennung wider).*

> *noch nie konfrontiert worden und weiß eben nicht besser darauf zu antworten bzw. uns bessere Angaben zu geben.*
>
> Vp 3: *Weil sie arg aufgeregt ist, das hört man schon durch die Atemgeräusche.*

In dem Experiment von Calvo und Eysenck (1996) (siehe Abschnitt 2.1) zeigten ängstliche Versuchspersonen eine schlechtere Aufgabenbewältigung oder aktivierten andere Strategien, wie z.B. eine stärkere Benutzung der phonologischen Schleife (was sich durch offensichtliche Lippenbewegungen bemerkbar machte). Beides könnte auf eine höhere Arbeitsgedächtnisbelastung hindeuten.

Ablenkung. Im zweiten Notruf unterhielt sich der Anrufer neben dem Anruf noch mit anderen Personen in seiner Umgebung. Dieser Umstand wirkte sich, so die Versuchspersonen, negativ auf die Gesprächsführung aus. Die meisten Versuchspersonen brachten diese Ablenkung in Zusammenhang mit einer schlechten Situationsbeschreibung (Vp 5). Die Äußerungen der Versuchspersonen führen zu weiteren Zusammenhängen (siehe Abb. 3.2): Ablenkung ist eine Ursache von Arbeitsgedächtnisbelastung (Vp 5, Vp 6). Ein Gespräch mit anderen Personen, das durch Gesprächsgeräusche bemerkbar ist, ist eine Ursache für die Ablenkung (Vp 6). Typische Äußerungen der Versuchspersonen waren:

> Frage 2.1: Wie gut hat er die Situation beschrieben?

Vp 5: *Die war nicht so besonders gut.* – retrospektives Denken – *Er hat mit einem anderen im Hintergrund gesprochen, und ich muß ihn auf mich lenken, damit er meine Fragen beantwortet und nicht die Fragen von seinem Nachbar.*

Frage 2.2: Warum, denken Sie, hat er die Ortsbeschreibung auf diese Weise gegeben?

Vp 6: *Ich nehme an, daß er so nervös war bzw. durch irgendwas abgelenkt wurde, daß er das nicht genauer beschreiben konnte.*

Frage 2.11: Wurde die Gesprächsführung beeinträchtigt, wenn der Mann sich noch mit anderen unterhält?

Vp 6: *Ja.* – retrospektives Denken – *Ich muß den Mann auffordern bzw. versuchen, daß er die Gespräche mit anderen Leuten beendet und sich darauf konzentriert, mir klare Angaben zu machen.*

Das gleichzeitige Führen eines weiteren Gesprächs könnte man aus kognitionspsychologischer Sicht als Nebenaufgabe ansehen, die die Ausführung der Hauptaufgabe, ein Notrufgespräch zu führen, beeinträchtigen kann (siehe Kapitel 2.1, Doppelaufgabenparadigma). Der Einfluß einer Nebenaufgabe auf die Arbeitsgedächtnisbelastung wurde z.B. von Power (1985) und Roßnagel (1995a) untersucht.

Alkoholeinfluß. Im zweiten Notrufdialog (siehe Tabelle 3.2) rief der Anrufer aus einer Gaststätte an. Aufgrund seines Verhaltens und der Dauer, die notwendigen Informationen über die Unfallstelle durchzugeben, kamen viele Versuchspersonen zu dem Schluß, daß er möglicherweise betrunken war (Vp 2). Die Versuchspersonen erwähnten dazu folgende Zusammenhänge (siehe Abb. 3.3): Durch den Alkoholeinfluß werden Qualität und Geschwindigkeit der Antwort beeinflußt (Vp 3). Aus dieser Sicht kann man den Alkoholeinfluß als Ursache von Arbeitsgedächtnisbelastung ansehen (siehe „Folgen hoher Arbeitsgedächtnisbelastung“). Ruft der Anrufer aus einer Gaststätte an, worauf typische Hintergrundgeräusche hinweisen (Vp 5), so verstärkt sich die Annahme des Alkoholeinflusses des Anrufers. Eine undeutlichliche Aussprache ist eine Folge davon. Die Versuchspersonen äußerten sich dazu so:

Frage 2.4: Wie schnell wird er die Frage beantworten?

Vp 2: *Er wird lange brauchen.* – retrospektives Denken – *Das ganze Gespräch ist schon keine besondere Qualität, für einen Notruf abzusetzen. Ich weiß nicht, entweder hat er keine Erfahrung oder ist alkoholisiert.*

Frage 2.7: Wieviele Ziffern hätte wohl die längste Nummer, die der Mann jetzt richtig durchgeben könnte?

Vp 3: *Maximal fünf.* – retrospektives Denken – *Erstens mal, ich würde immernoch davon ausgehen, daß er leicht betrunken ist, also angeschwippst, und er ist jetzt auch noch in einer Gaststätte. Und von der Verfassung her wird das wohl nicht so hinhauen.*

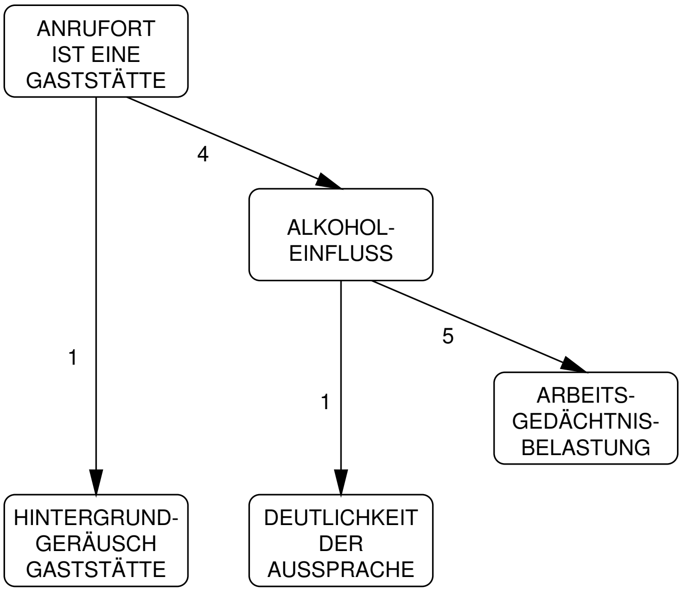

Abb. 3.3: *Rekonstruktion der Hypothesen der Versuchspersonen über den Zusammenhang zwischen Alkoholeinfluß und Arbeitsgedächtnisbelastung. (Ein Pfeil von A nach B bedeutet, daß A B beeinflußt. Die Zahl neben dem Pfeil spiegelt die Häufigkeit der Nennung wider).*

> Frage 2.9: Wie gut hat er die Telefonnummer durchgegeben?
>
> Vp 5: *schlecht.* – retrospektives Denken – *Also, im Hintergrund sind Gespräche. Er dürfte in der Gaststätte sein. Ich hake nach.*

Als nächstes werden die Folgen von hoher Arbeitsgedächtnisbelastung beschrieben, die von den Versuchspersonen genannt wurden.

3.3.3 Folgen hoher Arbeitsgedächtnisbelastung

Die Mitarbeiter der FNZ nannten im wesentlichen Symptome von Arbeitsgedächtnisbelastung, die sich auf die sprachliche Äußerungen der Anrufer beziehen. Die Zusammenhänge (siehe Abb. 3.4) werden wieder beispielhaft anhand von Aussagen der Mitarbeiter belegt.

Qualität der Situationsbeschreibung. Ein Notruf beginnt meistens damit, daß der Anrufer beschreibt, was passiert ist. Aufgrund dieser Informationen kann der Mitarbeiter der FNZ sich einen Überblick über das Ausmaß des Notfalls machen, um adäquat darauf reagieren zu können. Eine schlechte Situationsbeschreibung wurde von den Versuchspersonen als Folge hoher Arbeitsgedächtnisbelastung gesehen (siehe Abb. 3.4):

> Frage 2.1: Wie gut hat er die Lage beschrieben?
>
> Vp 3: *Er war also ziemlich unkonzentriert und die Aussage war ziemlich oberflächlich. Wie ich das so mitbekommen habe „Inselstraße, im Keller brennt es“.*

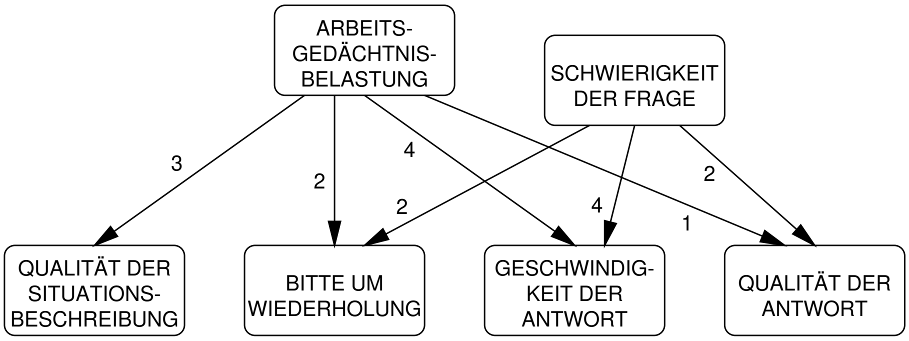

Abb. 3.4: *Rekonstruktion der Hypothesen der Versuchspersonen über Folgen von Arbeitsgedächtnisbelastung. (Ein Pfeil von A nach B bedeutet, daß A B beeinflußt. Die Zahl neben dem Pfeil spiegelt die Häufigkeit der Nennung wider).*

Durch zusätzliche Fragen versucht der Mitarbeiter der FNZ nähere Informationen über den Notfall zu erfahren. Abhängig von der Schwierigkeit der Fragen werden weitere Folgen von Arbeitsgedächtnisbelastung von den Versuchspersonen genannt.

Geschwindigkeit und Qualität der Antwort. Die Versuchspersonen meinten, daß von der Arbeitsgedächtnisbelastung und der Schwierigkeit der Frage die Geschwindigkeit und Qualität der Antwort abhängen (siehe Abb. 3.4). Schwierige Fragen bei hoher Arbeitsgedächtnisbelastung führen dann zu einer schlechteren Antwort (Vp 8) oder zu einer längeren Antwortzeit (Vp 2). Ähnliche Zusammenhänge wurden in Experimenten von Power (1985) und Roßnagel (1995a) beschrieben (siehe Abschnitt 2.1). Von den Mitarbeitern der FNZ wurde dieser Zusammenhang wie folgt beschrieben:

Frage 1.3: Wie schnell und deutlich wird sie die Frage beantworten können?

Vp 2: *schnell.* – retrospektives Denken – *Also sie weiß noch, wo sie wohnt. Weil sie hat ihre Hausnummer gewußt, und die Straße hat sie gewußt. Das kann sie schnell beantworten.*

Frage 1.9: Die Antwort war ein bißchen undeutlich. Wie verstehen Sie das, was sie gesagt hat?

Vp 8: *Sie hat mir die Adresse angegeben. Und nach der Frage „welche Etage" hat sie gesagt hoch.* – retrospektives Denken – *Sie war doch etwas aufgeregt und konnte mir aus diesem Grund mir keine genaueren Angaben machen in welcher Etage das ist.*

Bitte um Wiederholung. Die Anruferin im ersten Notruf bat nach der Frage „In welchem Stadtteil?" um eine Wiederholung der Frage (siehe Tabelle 3.1, Zeile 7). Einige Versuchspersonen meinten daraufhin, daß die Ursache dafür eine hohe Arbeitsgedächtnisbelastung der Anruferin (Vp 9) oder die Schwierigkeit der Frage sei (Vp 7):

Frage 1.5: Wie schnell und deutlich wird sie die Frage beantworten können?

Vp 7: *Vielleicht hat sie die Frage nicht verstanden. Wie er gesagt hat „in welchem Stadtteil“.* – retrospektives Denken – *Er hat gefragt in welchem Stadtteil und sie hat es nicht verstanden.*

Vp 9: *Vielleicht ist das beim ersten Mal untergegangen.* – retrospektives Denken – *Vielleicht deutlicher sprechen. Deutlicher und langsamer sprechen.*

Faktoren, die solche Fragen, wie die nach dem Stadtteil, erschweren, werden im nächsten Abschnitt beschrieben.

3.3.4 Weitere Beobachtungen

In der Untersuchung bekamen die Versuchspersonen die Aufgabe gestellt, bestimmte Aufgabenschwierigkeiten einzuschätzen. Dabei ging es um die Einschätzung der Schwierigkeit der Stadtteilfrage (siehe Notrufdialog 1 in Tabelle 3.1) und die Einschätzung der Schwierigkeit, eine Telefonnummer durchzugeben (siehe Notrufdialog 2 in Tabelle 3.2). Beide Aufgaben lieferten weitere Ergebnisse, die im folgenden näher erläutert werden.

Schwierigkeit der Stadtteilfrage. Die Reaktion der Anruferin auf die Stadtteilfrage läßt den Schluß zu, daß die Frage nicht so einfach ist, wie es im ersten Augenblick scheint. Die Versuchspersonen sollten diesen Aspekt beurteilen. Nach „In welchem Stadtteil?“ (siehe Tabelle 3.1, Zeile 6) und nach „Bitte?“ (Zeile 7) wurde das Band gestoppt. Danach wurde gefragt, wie schnell die Anruferin die Frage beantworten kann (Fragen 3 und 5). Die Tabelle 3.3 fast einige typische Aussagen der Versuchspersonen zusammen und zeigt wie die Einschätzung der Schwierigkeit der Frage sich nach der Bitte um eine Wiederholung der Frage geändert hat.

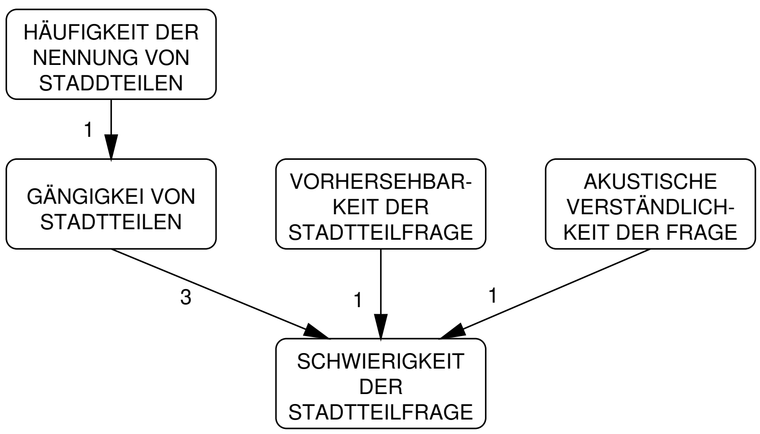

Abb. 3.5: *Schwierigkeit der Stadtteilfrage. (Ein Pfeil von A nach B bedeutet, daß A B beeinflußt. Die Zahl neben dem Pfeil spiegelt die Häufigkeit der Nennung wider).*

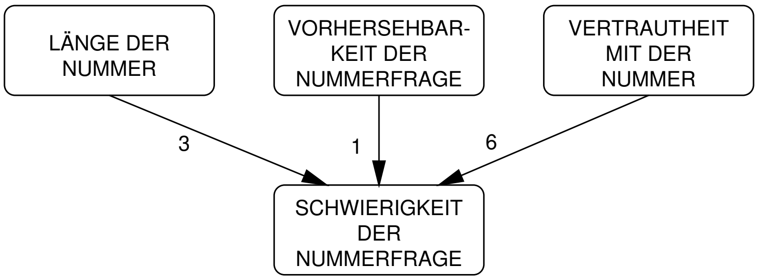

Abb. 3.6: *Schwierigkeit der Nummerfrage. (Ein Pfeil von A nach B bedeutet, daß A B beeinflußt. Die Zahl neben dem Pfeil spiegelt die Häufigkeit der Nennung wider).*

Die Aussagen der Versuchspersonen lassen zusätzliche Zusammenhänge erkennen (siehe Abb. 3.5). Die Schwierigkeit der Stadtteilfrage, die wie gesehen die Qualität und die Geschwindigkeit der Antwort beeinflußt, hängt ab von drei verschiedenen Ursachen: die Vorhersehbarkeit der Stadtteilfrage, die Gängigkeit von Stadtteilen und die akustische Verständlichkeit der Stadtteilfrage. Die Gängigkeit von Stadtteilen wird wiederum beeinflußt von der Häufigkeit der Nennung von Stadtteilen.

Schwierigkeit der Nummerfrage. Ähnlich wie bei der Stadtteilfrage sollten die Versuchspersonen einschätzen, wie schwierig es ist, sich eine fremde Telefonnummer zu merken und danach durchzugeben. Im ersten Notrufdialog wurde gefragt, wieviele Ziffern die längste Nummer hätte, die die Anruferin jetzt richtig durchgeben könnte. Im zweiten Notrufdialog sollten die Versuchspersonen beurteilen, wie gut der Anrufende die Nummer der Gaststätte durchgegeben hat. Die Tabelle 3.4 fast wieder einige typische Aussagen der Versuchspersonen zusammen. Daraus lassen sich wiederum zusätzliche Zusammenhänge erkennen (siehe Abb. 3.6). Die Schwierigkeit der Nummerfrage wird maßgeblich von drei Ursachen beeinflußt: die Vorhersehbarkeit der Nummerfrage, die Vertrautheit und die Länge der Nummer.

Instruktionen an die Anrufer. Am Schluß des ersten Notrufs wurden die Versuchspersonen gefragt, wie kompliziert die Anweisungen sein könnten, die die Anruferin noch gut verstehen und behalten könnte. Es stellte sich dabei heraus, daß die Mitarbeiter der FNZ mit Rücksicht auf die Belastung und Überforderung der Anrufer keine komplizierten Anweisungen geben. Einige Versuchspersonen äußerten sich dazu so:

> Frage 1.11: Stellen Sie sich vor, Sie müßten nach der Alarmierung der Einsatzkräfte einige Anweisungen geben. Wie kompliziert könnten die Anweisungen sein, die sie jetzt noch gut verstehen und behalten könnte?
>
> Vp 1: *Komplizierte Anweisungen würden wahrscheinlich gar nichts bringen. Da würde ich also nur die einfachsten Anweisungen geben, daß ihr nichts zustößt. Einfache Anweisungen, wie der Kollege schon gesagt hat, Tür zu und am besten raus aus der Wohnung.*

Vp 10: *Es müssen kurze, präzise Anweisungen sein und es sollte sich höchstens auf eine oder zwei beschränken, weil mehr wird wahrscheinlich doch nicht hängenbleiben.* – retrospektives Denken – *Daß man in einer solchen Aufregungssituation vielleicht doch nicht mehr so alles mitbekommt, oder die Frau nicht alles mitbekommt, was ich ihr sage.*

3.4 Diskussion

Die Analyse der Äußerungen der Versuchspersonen lieferte ein Bild von qualitativen Kausalzusammenhängen. Quantitative Zusammenhänge, also wie stark eine Variable eine andere beeinflußt, waren aus der explorativen Studie nicht zu entnehmen. Diese müssen für die folgende Implementation so definiert werden, daß sie den gemeinten Abhängigkeiten Rechnung tragen können. Die Ergebnisse lassen sich wie folgt zusammenfassen:

(1) Die Versuchspersonen machten sich kein differenziertes Bild des Arbeitsgedächtnisses, sondern sprachen global von *Konzentriertheit*. Die Versuchspersonen nannten auch keine direkten Hinweise auf Arbeitsgedächtnisbelastung, so daß hier lediglich ein einfaches und undifferenziertes Modell des Arbeitsgedächtnisses zugrunde gelegt wird.

(2) Vier von sieben der von den Versuchspersonen genannten Zusammenhänge zwischen Ursachen und Folgen von Arbeitsgedächtnisbelastung stimmen mit den Ergebnissen der in Abschnitt 2.1 beschriebenen psychologischen Untersuchungen überein. (Das sind: Aufregung, Ablenkung, Geschwindigkeit und Qualität der Antwort)

(3) Ein wichtiger Aspekt der ressourcenadaptiven Dialogführung ist die richtige Auswahl von Äußerungen und damit verbunden die richtige Einschätzung der Anforderungen, die die Äußerung an das Arbeitsgedächtnis des Dialogpartners stellt. Aber oft ist es nicht möglich, sich ein vollständiges Bild über die Situation zu machen, in der sich der Dialogpartner befindet. Neue Informationen, die von der getroffenen Einschätzung abweichen, führen zu einer Veränderung des Bildes über die Arbeitsgedächtnisbelastung des Dialogpartners. Z.B. erkannten die meisten Versuchspersonen zuerst nicht, daß eine scheinbar einfache Frage nach dem Stadtteil der Anruferin für die Hörerin schwer einzuordnen und zu beantworten sein kann.

(4) Die Mitarbeiter der FNZ berücksichtigen bei ihrer Dialogführung die Arbeitsgedächtnisbelastung ihrer Dialogpartner. Z.B. verzichten sie oft aus Sorge über die Arbeitsgedächtnisüberlastung beim Anrufenden bewußt auf das Anbieten schwieriger Anweisungen.

Im nächsten Kapitel werden die Ergebnisse der explorativen empirischen Studie mit einem Bayesschen Netz modelliert und implementiert.

Tabelle 3.3: *Aufgabenschwierigkeit: Stadtteilfrage*

Nach „... in welchem Stadtteil?“: Frage 1.1: **Wie schnell und deutlich wird sie die Frage beantworten können?** Vp 2: schnell. – RLD – Also sie weiß noch wo sie wohnt. Weil sie hat ihre Hausnummer gewußt, und die Straße hat sie gewußt. Das kann sie schnell beantworten. Vp 5: Ich würde sagen sofort. – RLD – Ja, sie ist ja vor Ort, sie weiß ja wo sie wohnt. Deshalb kann sie ja auf meine Frage hin sofort antworten.	Die meisten Versuchspersonen schätzen die Stadtteilfrage zuerst als einfach ein und rechnen mit einer schnellen Beantwortung. (9 Versuchspersonen erwarten eine schnelle Antwort, 1 Versuchsperson eine langsame).
Nach „Bitte?“: Frage 1.3: **Wie schnell und deutlich wird sie die Frage beantworten können?** Vp 1: Denke, eher, daß es länger dauert. – RLD – Ich merke schon, die Frau ist sehr aufgeregt. Das merkt man schon an ihrer Atmung. Und unter dieser Voraussetzung ist es oft so, daß die Leute selbst nicht wissen, wo sie wohnen. Und manchmal auch nicht, wie sie heißen, geschweige dann eine Telefonnummer. Vp 2: Na gut, weiß man nicht so genau. Normal. – RLD – Ich habe überlegt: hat sie ihn nicht verstanden oder konnte sie mit dem Wort Stadtteil nichts anfangen. [...] Für die meisten Leute ist der Begriff Stadtteil sowieso schon altmodisch.	Nach der Bitte um Wiederholung der Frage ändern die meisten Versuchspersonen ihre Meinung. Nur noch 5 Versuchspersonen erwarten eine schnelle Antwort wohingegen 5 eine langsame Antwort vorhersagen. Als Gründe dafür wurden die hohe Arbeitsgedächtnisbelastung und die fehlende Gängigkeit des Begriffs „Stadtteil“ angegeben.
Nach „... LVA“: Frage 1.6: **Wie gut hat sie die Frage beantwortet?** Vp 2: Das war gut. – RLD – Sie hat gezögert bei Stadtteil. Sie hat selbst nicht genau gewußt, ist es jetzt St. Johann. Ist es St. Johann? wohne ich jetzt dort? Weil es wahrscheinlich so wenig benutzt wird. Es steht ja auf keiner Postleitzahl mehr drauf. Nur als Beispiel jetzt. Dann kennen die Leute das nicht mehr so. Frage 1.7: **Warum denken Sie, hat sie auf diese Weise geantwortet?** Vp 3: Weil ansich die Ortsteile hier in Saarbrücken nicht so bekannt sind und manche sagen dann noch Saarbrücken 6, Saarbrücken 5 und so. Weil das mit den Ortsteilen ist nicht so gängig. Vp 4: Sie mußte vermutlich selbst noch einmal überlegen, weil kaum jemand fragt ja nach dem Stadtteil.	Nach der Wiederholung der Frage, gibt die Anruferin eine gute Antwort. Ihr erstes Zögern wird mit der fehlenden Gängigkeit von Stadtteilen und der fehlenden Vertrautheit der Stadtteilfrage erklärt („[...] es steht auf keiner Postleitzahl mehr drauf.“ und „kaum jemand fragt ja nach dem Stadtteil.“).

Anmerkung: RLD = Retrospektives lautes Denken nach der Frage „Was haben Sie dabei gedacht?“.

Tabelle 3.4: *Aufgabenschwierigkeit: Nummerfrage*

Dialog 2 – Der Mann in der Gaststätte: Nach „... Geben Sie mir die Nummer der Gaststätte.“: Frage 2.7: **Wieviele Ziffern hätte wohl die längste Nummer, die der Mann jetzt richtig durchgeben könnte?** Vp 9: Höchstens 3 Nummern. – RLD – Weil er die Nummer nicht kennt, weil er aus einer Gaststätte aus anruft und weil er die Nummer erst erfragen muß. Oder er hat die Nummer im Kopf, bringt sie aber nicht genau rüber.	Von den meisten Versuchspersonen wird die Arbeitsgedächtnisbelastung des Anrufers als hoch eingeschätzt. Sie erwarten nicht, daß der Anrufer alle Ziffern korrekt durchgeben kann.
Nach Durchgabe der Nummer: Frage 2.9: **Wie gut hat er die Nummer durchgegeben?** Vp 9: Zuerst muß er sie erfragen. Und wie ich das vorhin schon sagte, mehr als 3 Nummern kann er nicht durchgeben. – RLD – Daß er die Nummer nicht kennt und daß er also in einer gewissen Streßsituation ist und daß er das bzgl. der Telefonnummer nicht ganz mitbekommt und halt nur bruchstückhaft weitergibt. Frage 2.10: **Warum hat er, nach Ihrer Meinung, Probleme, die Telefonnummer richtig durchzugeben?** Vp 5: Weil er sie nicht kennt. Er ist in der Gaststätte Gast. Er fragt nach, im Hintergrund wird auch eine Nummer gesagt, die aber von mir nicht verständlich war. Vp 8: Ich gehe davon aus, er ist in einer Gaststätte, hat vielleicht etwas getrunken, daß es auf Grund dessen schwer gefallen ist, die Nummer durchzugeben.	Die Einschätzung stellte sich als richtig heraus. Als Grund für die schlechte Durchgabe der Nummer wird die fehlende Vertrautheit mit der Telefonnummer der Gaststätte angegeben. Einige Versuchspersonen führen die Probleme bei der Durchgabe der Nummer auf den Alkoholeinfluß des Anrufers zurück.
Dialog 1 – Die Frau in St. Johann: Frage 1.12: **[...] Wieviele Ziffern hätte wohl die längste Nummer, die sie jetzt richtig durchgeben könnte?** Vp 3: Drei, maximal vier. – RLD – Weil es bei größeren Telefonnummern kritisch wird, die also aus dem Kopf heraus zu sagen. Da gibt es normalerweise Probleme. Eine drei- oder vierstellige kann man im Kopf also noch gut sagen. Vp 1: Ich schätze drei bis vier. – RLD – Eigentlich nur, daß die Frau sehr unter Streß ist, sehr ängstlich ist und sich wahrscheinlich auf eine Telefonnummer nicht konzentrieren kann. Das weiß man aus Erfahrung, daß selbst, wenn man die Telefonnummer des Teilnehmers oder des Anrufers erfragt, das es da schon Probleme gibt, auch Unverständnis von den Leuten.	Wie im obigen Fall wird auch hier die fehlende Vertrautheit der Nummer als Grund für eine schlechtere Durchgabe der Nummer angegeben. Zusätzlich meinen einige Versuchspersonen, daß auch die Länge der Nummer dabei eine Rolle spielen könnte.

Anmerkung: RLD = Retrospektives lautes Denken nach der Frage „Was haben Sie dabei gedacht?“.

Kapitel 4

Modellierung der Notrufdialoge mit dynamischen Bayesschen Netzen

Basierend auf den Ursachen und Folgen hoher Arbeitsgedächtnisbelastung, die aufgrund der explorativen empirischen Studie in Kapitel 3 aus den Aussagen der Versuchspersonen rekonstruiert werden konnten, wird im folgenden ein Benutzermodell erstellt. Um die zeitveränderlichen Aspekte adäquat zu behandeln, wird für das Benutzermodell ein dynamisches Bayessches Netz (DBN) verwendet. Im folgenden werden die Entwicklungsphasen des Benutzermodells näher beschrieben:

(1) *Explorative empirische Studie und Literaturrecherche* Vor der eigentlichen Modellierung der Zusammenhänge stand die Überlegung, welche Faktoren die Arbeitsgedächtnisbelastung eines Dialogpartners beeinflussen und welche Folgen eine erhöhte Arbeitsgedächtnisbelastung hat. Die Literaturrecherche (Kapitel 2) und die explorative empirische Studie (Kapitel 3) haben wichtige Anhaltspunkte in Form von Kausalzusammenhängen gegeben, die in der Modellierung berücksichtigt werden.

(2) *Qualitative Modellierung der Zusammenhänge* Im Abschnitt 4.1 werden die einzelnen Zusammenhänge der empirischen Studie miteinander in Verbindung gesetzt und durch die zeitlichen Abhängigkeiten ergänzt. Das Ergebnis ist ein qualitatives Kausalmodell für die Feuerwehrnotrufe. Für die Beschreibung der zeitlichen Zusammenhänge werden Modellierungsschemata benutzt, die in Abschnitt 4.2 beschrieben werden.

(3) *Quantitative Modellierung der Zusammenhänge* Im Abschnitt 4.3 werden die Abhängigkeiten zwischen den Kausalzusammenhängen quantitativ beschrieben. Dabei wird jedem Wurzelknoten a-priori Wahrscheinlichkeiten für das Zutreffen seiner Zustände zugeordnet. Jedem Kindknoten wird in Form einer *conditional propability table*[1] (kurz *CPT*) die bedingten Wahrscheinlichkeiten über alle Kombinationen der Zustände der Knoten, von denen der Kindknoten abhängt, definiert.

[1]Der deutsche Begriff für die *conditional probability table* (CPT) ist *Verbindungsmatrix*. Da die Abkürzung *CPT* sich in der Literatur eingebürgert hat (siehe z.B. Russell & Norvig, 1995), wird sie hier übernommen.

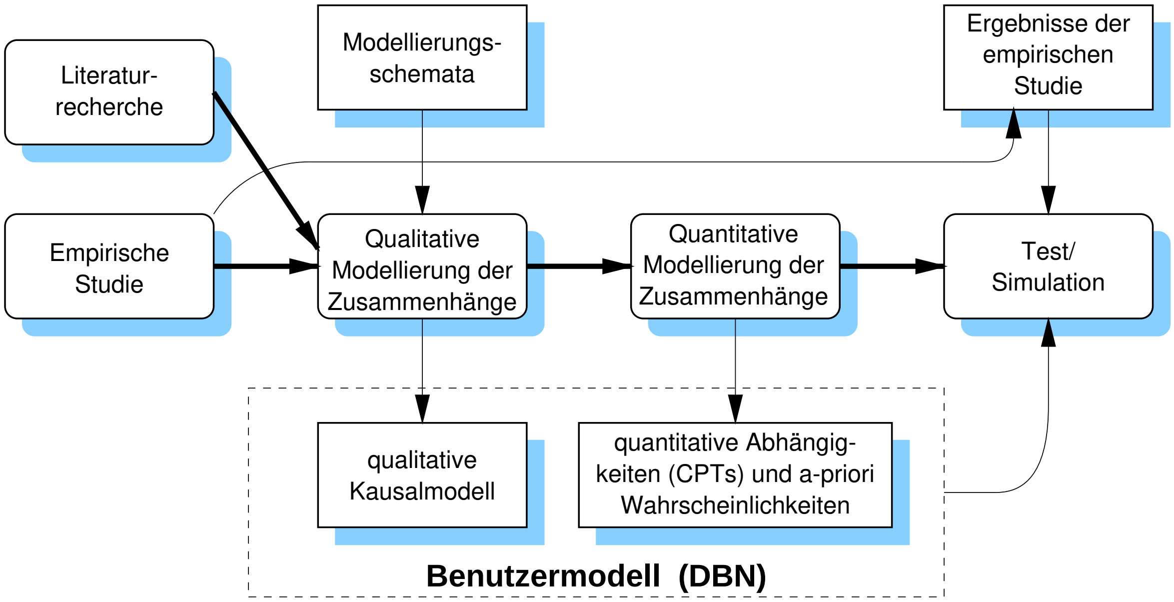

Abb. 4.1: *Workflow-Diagramm für die Entwicklung des Benutzermodells der Feuerwehrdomäne. (Erklärung im Text).*

(4) *Testen der Modellierung anhand von Fallbeispielen* Im Abschnitt 4.4 wird das DBN anhand der beiden Notrufdialoge aus der empirischen Untersuchung, die als Fallbeispiele dienen, getestet. Der Test erfolgt durch eine Simulation der beiden Notrufdialoge. Dabei dienen die Aussagen der Versuchspersonen aus der empirischen Studie als Vergleich, um festzustellen, ob das Benutzermodell plausible Einschätzungen der Arbeitsgedächtnisbelastung des Dialogpartners zieht.

Abbildung 4.1 zeigt eine grafische Beschreibung der Entwicklung. Die gerundeten Rechtecke symbolisieren die einzelnen Arbeitsphasen, die mit Ausnahme der Literaturrecherche und der empirischen Studie in den nächsten Abschnitten beschrieben werden. Die dicken Pfeile zwischen den Phasen spiegeln den zeitlichen Verlauf der Entwicklung wider. Die normalen Rechtecke symbolisieren die Eingabe- bzw. Ausgabedaten. Die dünnen Pfeile symbolisieren die Datenflüsse. Das Benutzermodell, bestehend aus dem qualitativen Kausalmodell und den quantitativen Abhängigkeiten zwischen den einzelnen Kausalzusammenhängen, wird durch ein DBN repräsentiert.

4.1 Qualitative Modellierung der Zusammenhänge

Die Ergebnisse der explorativen empirischen Untersuchung ergaben eine Reihe von Ursache-Wirkungs-Beziehungen bezüglich der Arbeitsgedächtnisbelastung von Dialogteilnehmern in Notrufdialogen. Mit Blick auf die beiden Notrufdialoge aus der empirischen Studie (siehe Tabellen 3.1 und 3.2), die später als Fallbeispiele dienen, werden zwei getrennte Modelle ($\mathcal{N}1$ und $\mathcal{N}2$) entwickelt. Das Modell $\mathcal{N}1$ (siehe Abb. 4.2) beschreibt eine Situation, wie sie im Dialog „Die Frau in St. Johann“ vorzufinden ist und $\mathcal{N}2$ (siehe Abb. 4.3) eine Situation wie in „Der Mann in der Gaststätte“.

4.1.1 Modellierung des Arbeitsgedächtnisses

Die in der kognitiven Psychologie entwickelten Arbeitsgedächtnismodelle versuchen das Arbeitsgedächtnis eines Menschen sehr differenziert zu beschreiben. Diese Modelle unterscheiden verschiedene Subsysteme, wie z.B. eine zentrale Exekutive, eine phonologische Schleife und einen räumlich-visuellen Notizblock (siehe Baddeleys Modell, Abschnitt 2.1).

Für eine On-line Modellierung des Benutzers sind solche Modelle des Arbeitsgedächtnisses zu feinkörnig, da sie voraussetzen, daß das Benutzermodell entsprechende Informationen besitzt, die es erlauben, auf die verschiedenen Aspekte der Belastung Rückschlüsse zu ziehen.[2] Die Informationen über die Arbeitsgedächtnisbelastung des Benutzers, die dem zu entwickelten Benutzermodell als beobachtbare Symptome vorliegen, sind dazu nicht detailliert genug.

Die beiden Modelle $\mathcal{N}1$ (Abb. 4.2) und $\mathcal{N}2$ (Abb. 4.3) verwenden daher lediglich ein vereinfachtes Bild des Arbeitsgedächtnisses als einen homogenen Speicher mit einer gewissen Kapazität. $\mathcal{P}$s Arbeitsgedächtnisbelastung wird als *verfügbare Arbeitsgedächtniskapazität* modelliert. Eine hohe Arbeitsgedächtnisbelastung soll einer geringen verfügbaren Arbeitsgedächtniskapazität entsprechen und umgekehrt.

4.1.2 Notrufmodelle $\mathcal{N}1$ und $\mathcal{N}2$

In beiden Modellen steht die VERFÜGBARE ARBEITSGEDÄCHTNISKAPAZITÄT[3] (VAK) im Fokus der Modellierung. Die übrigen Variablen, die in den beiden Modellen vorkommen, können bezüglich der VAK in drei Gruppen eingeteilt werden:

- *Externe Ursachen* Damit sind Ursachen von VAK gemeint, die nicht direkt, sondern über andere Variablen auf die Arbeitsgedächtniskapazität eines Dialogpartners einwirken. Eine externe Ursache ist z.B. $\mathcal{P}$s HILFLOSIGKEIT, die $\mathcal{P}$s AUFREGUNG beeinflußt.
- *Interne Ursachen* Interne Ursachen von Arbeitsgedächtnisbelastung haben einen direkten Einfluß auf die VAK. Ein Beispiel dafür ist $\mathcal{P}$s AUFREGUNG. Eine interne Ursache ist direkt durch einen Pfeil mit der VAK verbunden.
- *Beobachtbare Symptome* Symptome beschreiben spezielle Sachverhalte oder Eigenschaften, die sich in einem Dialog direkt beobachten lassen. Solche Symptome basieren entweder auf direkten Wahrnehmungen, wie z.B. ATEMGERÄUSCHE, oder auf Merkmale der sprachlichen Äußerungen von $\mathcal{P}$, wie z.B. die QUALITÄT DER ANTWORT.

[2] Jameson (1997a) bespricht diese Problematik näher.

[3] Die Namen der Knoten in einem Bayesschen Netz werden mit Großbuchstaben geschrieben.

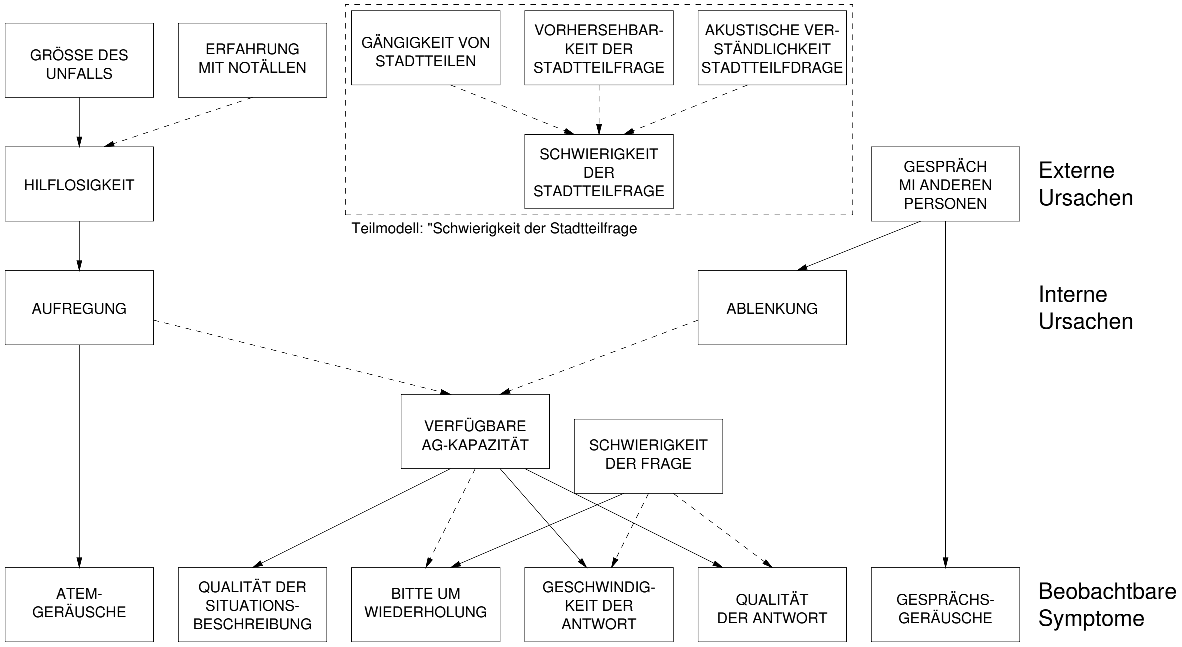

Abb. 4.2: *Modell $\mathcal{N}$1: Zusammenhänge für den Notrufdialog „Die Frau in St. Johann“ ohne Berücksichtigung der Zeitveränderlichkeit.*
Ein durchgezogener bzw. gestrichelter Pfeil von A nach B bedeutet: A beeinflußt B positiv bzw. negativ.

Zusätzlich sind in den beiden Abbildungen 4.2 und 4.3 die Teilmodelle „Schwierigkeit der Stadtteilfrage“ und „Schwierigkeit der Nummerfrage“ zu sehen, die die allgemeine Variable SCHWIERIGKEIT DER FRAGE ersetzen, wenn sie zusätzliche Informationen für die Einschätzung der Schwierigkeit der Frage beitragen können.

Die meisten Variablen kommen in beiden Modellen vor. Ein Beispiel dafür ist $\mathcal{P}$s VAK und Symptome wie QUALITÄT DER SITUATIONSBESCHREIBUNG oder GESCHWINDIGKEIT DER ANTWORT. Andere Variablen kommen aber nur in einem Modell vor. Der Unterschied der beiden Modelle ist auf die spezifische Situation zurückzuführen, die in dem jeweiligen Notrufdialog vorgefunden wurde. So rief z.B. der Mann im Notrufdialog 2 (siehe Tabelle 3.2) aus einer Gaststätte an. Die meisten Versuchspersonen vermuteten daraufhin, daß der Anrufer unter Alkoholeinfluß steht, und berücksichtigten diesen Umstand offensichtlich für ihre Einschätzung seiner Arbeitsgedächtnisbelastung. Auf der anderen Seite wurde von den Versuchspersonen die Tatsache „kein Alkoholeinfluß von $\mathcal{P}$“ nicht für ihre Einschätzungen der Arbeitsgedächtnisbelastung benutzt. Die Aussagen der Versuchspersonen bzgl. des Notrufdialogs 1 („Die Frau in St. Johann“) lassen diesen Schluß zu. Offensichtlich werden nicht alle *möglichen* Zusammenhänge in einem Inferenzprozeß berücksichtigt.

Andererseits können einige Variablen, auch wenn sie nicht beobachtet werden, die Einschätzung bzgl. $\mathcal{P}$s VAK beeinflussen. Ein Beispiel dafür ist $\mathcal{P}$s ABLENKUNG. Die Tatsache, daß kein GESPRÄCH MIT ANDEREN PERSONEN wahrgenommen wird, kann sich positiv auf die Einschätzung der ABLENKUNG von $\mathcal{P}$ auswirken. Aus diesem Grund wurde diese Variable in beiden Modellen aufgenommen.

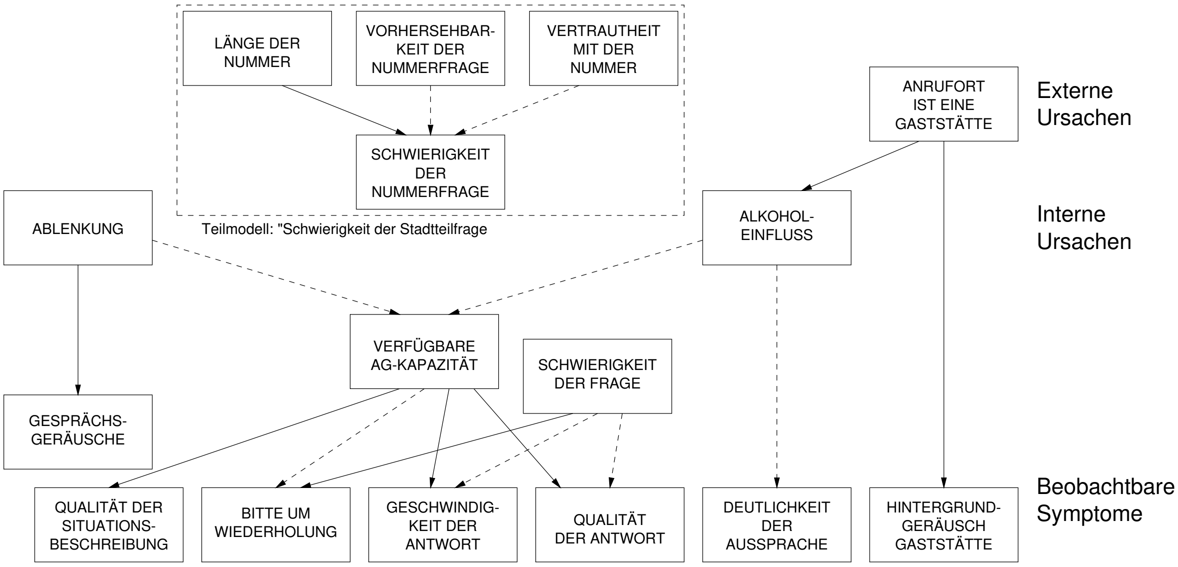

Abb. 4.3: *Modell $\mathcal{N}2$: Zusammenhänge für den Notrufdialog „Der Mann in der Gaststätte" ohne Berücksichtigung der Zeitveränderlichkeit.*
Ein durchgezogener bzw. gestrichelter Pfeil von A nach B bedeutet: A beeinflußt B positiv bzw. negativ.

4.1.3 Modellierung der Zeitveränderlichkeit

Wie in Abschnitt 2.3.5.2 beschrieben wurde, werden die Variablen eines DBN verschiedenen Knotentypen (temporär, statisch oder dynamisch) zugeordnet.

Die beobachtbaren Symptome in $\mathcal{N}1$ und $\mathcal{N}2$ sind temporär, d.h. sie werden nur in den Zeitscheiben berücksichtigt, in denen sie zusätzliche Informationen über den Benutzer liefern. ALKOHOLEINFLUSS und ANRUFORT IST EINE GASTSTÄTTE sind statische Variablen. Es wird davon ausgegangen, daß ihre Änderung während der Interaktion unerheblich ist. Statische Variablen sind in allen Zeitscheiben vorhanden, werden aber nur einmal definiert. AUFREGUNG, ABLENKUNG und VAK sind zeitveränderlich. Diese Variablen werden durch dynamische Knoten repräsentiert, die über Zeitscheiben hinweg verbunden sind.

Wie in den beiden Modellen $\mathcal{N}1$ und $\mathcal{N}2$ zu erkennen ist, gibt es drei Fälle von Zusammenhängen, in denen dynamische Knoten vorkommen:

- Zusammenhänge zwischen dynamischen und temporären Knoten
  Beispiel: ABLENKUNG $\longrightarrow$ ATEMGERÄUSCHE
- Zusammenhänge zwischen dynamischen und dynamischen Knoten
  Beispiel: ABLENKUNG $\longrightarrow$ VAK
- Zusammenhänge zwischen statischen und dynamischen Knoten
  Beispiel: ALKOHOLEINFLUSS $\longrightarrow$ VAK

Wie solche Zusammenhänge modelliert werden, wird im nächsten Abschnitt beschrieben.

4.2 Schemata für die Modellierung zeitveränderlicher Variablen

Das Thema dieses Abschnittes sind generelle Modellierungsschemata, die helfen, DBNs einfacher zu entwickeln. Je nach Art des Zusammenhangs zwischen den Knoten (dynamisch – temporär, dynamisch – dynamisch oder statisch – dynamisch) können unterschiedliche Schemata für die Modellierung benutzt werden. Wann und zu welchem Zweck ein bestimmtes Schema verwendet wird und welche Vor- und Nachteile das Schema hat, wird anhand von Beispielen diskutiert. Eine kürzere Beschreibung wurde von Schäfer und Weyrath (1997) veröffentlicht. Schäfer (1998) bietet eine ausführliche und erweiterte Definition der Schemata.

4.2.1 Zusammenhänge zwischen dynamischen und temporären Knoten

Im einfachsten Fall ist der Kindknoten eines dynamischen Knotens D ein temporärer Knoten T, z.B. ein Evidenzknoten. Ein solcher Zusammenhang liegt z.B. zwischen AUFREGUNG (D) und ATEMGERÄUSCHE (T) vor. Der Zusammenhang kann adäquat durch das Modell $\mathcal{M}1$, gezeigt in Abbildung 4.4, modelliert werden. Dieses Modell wird auch in Forbes et al. (1995) benutzt, um die zeitliche Abhängigkeiten zwischen den Zustandsknoten und Sensorknoten zu modellieren. Das Modell wird nun durch Beispiel 1 illustriert.

Beispiel 1: Zu Anfang eines Telefongesprächs sind viele Atemgeräusche zu hören. Im Laufe der Zeit nehmen diese jedoch ab, was auf eine Beruhigung von $\mathcal{P}$ hinweist.

Abbildung 4.5 zeigt das dazugehörige DBN. Um das Beispiel einfacher zu gestalten, wird vorübergehend angenommen, daß AUFREGUNG keine Elternknoten besitzt.

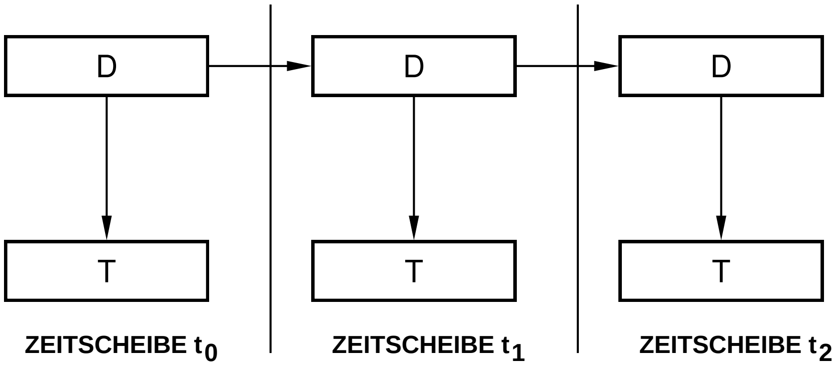

Abb. 4.4: *$\mathcal{M}1$: Modellierung des Zusammenhang zwischen einem dynamischen Knoten* D *und einem temporären Knoten* T.

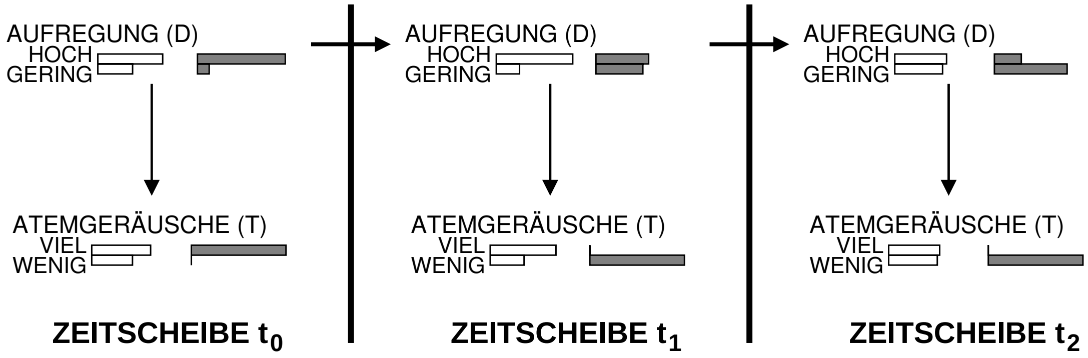

Abb. 4.5: *DBN von Beispiel 1 mit Modell $\mathcal{M}1$:* AUFREGUNG *ist dynamisch und* ATEMGERÄUSCHE *ist temporär.*

Das obige Beispiel deutet auf folgende Evidenzen hin: In Zeitscheibe t_0 gibt es viele Atemgeräusche, in t_1 und t_2 wenig. Die linken, hellen Histogramme in jeder Zeitscheibe zeigen die Wahrscheinlichkeitsverteilung des Knotens vor der Beobachtung einer Evidenz (d.h. die Vorhersage). Die rechten, dunklen Histogramme zeigen die Wahrscheinlichkeitsverteilung des Knotens nach der Beobachtung der Evidenz (d.h. die Interpretation).

Im Beispiel 1 werden in t_0 viele ATEMGERÄUSCHE beobachtet. Aus diesem Grund wird die Wahrscheinlichkeit für eine hohe AUFREGUNG in t_0 höher eingeschätzt. Der Übergang von einer Zeitscheibe zur anderen ist so modelliert, daß sich $\mathcal{P}$ beruhigen kann, d.h. $\mathcal{P}$s Aufregung kann sinken (vgl. dunkle Histogramme von AUFREGUNG in t_0 mit den hellen in t_1). Aufgrund der hohen Wahrscheinlichkeit für eine hohe AUFREGUNG in t_1 wird eine eher hohe Wahrscheinlichkeit für ATEMGERÄUSCHE vorhergesagt. Doch im Gegensatz dazu werden nur wenige ATEMGERÄUSCHE in t_1 beobachtet. Diese Evidenz führt zur Revidierung der Einschätzung von $\mathcal{P}$s AUFREGUNG. Der neue Wert von AUFREGUNG deutet verglichen mit der vorherigen Einschätzung auf eine Beruhigung von $\mathcal{P}$ hin. In Beispiel 1 wurde diese Einschätzung auch getroffen. Die Einschätzungen in der Zeitscheibe t_2 verlaufen analog.

4.2.2 Zusammenhänge zwischen dynamischen Knoten

Komplizierter wird es, wenn ein dynamischer Knoten D1 einen anderen dynamischen Knoten D2 beeinflußt. Ein solcher Zusammenhang liegt zwischen den Knoten AUFREGUNG (D1) und VAK (D2) vor. Beide Eigenschaften verändern sich im Laufe der Zeit.

In einer einfachen Modellierung kann man Modell $\mathcal{M}1$ adaptieren. D2, der Kindknoten des dynamischen Knotens D1, wird in diesem Fall nicht als dynamischer Knoten repräsentiert, sondern als temporärer Knoten. Abbildung 4.6 (linke Seite) zeigt dieses vereinfachte Modell $\mathcal{M}2.1$. In jeder Zeitscheibe wird eine neue Instanz von D ALS T erzeugt. Es gibt keine Verbindung zwischen D ALS T in der Zeitscheibe $t-1$ und D ALS T in der Zeitscheibe t.

Der Knoten D ALS T in $t-1$ beeinflußt D ALS T in t einzig über ihre gemeinsamen Elternknoten (D1 in $t-1$ und D1 in t). Die Vorhersage von D ALS T in t, die auf D1 in t basiert,

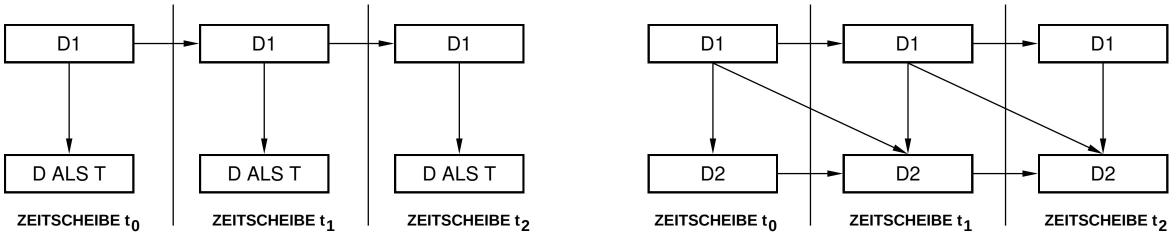

Abb. 4.6: *$\mathcal{M}2.1$ (links) und $\mathcal{M}2.2$ (rechts): Modellierung des Zusammenhangs zwischen zwei dynamischen Knoten* D1 *und* D2*. Im Modell $\mathcal{M}2.1$ wird der dynamische Knoten* D2 *als temporären Knoten* T *(*D ALS T*) vereinfacht.*

wird weniger genau sein, als wenn sie zusätzlich von der Einschätzung von D ALS T in der Zeitscheibe zuvor abhängt. Beispiel 2 illustriert das Modell $\mathcal{M}2.1$:

Beispiel 2: Zu Anfang eines Telefongesprächs sind viele Atemgeräusche zu hören. Im Laufe der Zeit nehmen diese jedoch ab, was auf eine Beruhigung von $\mathcal{P}$ hinweist. Zusätzlich werden drei einfache Fragen gestellt, die sehr schlecht beantwortet werden.

Wieder werden die Abhängigkeiten etwas vereinfacht: AUFREGUNG wird als Wurzelknoten, also ohne Elternknoten, modelliert und VAK soll in diesem Beispiel nur AUFREGUNG als Elternknoten besitzen. Abbildung 4.7 zeigt das Ergebnis von Beispiel 2 als DBN, das auf Modell $\mathcal{M}2.1$ basiert. Aus Gründen der Übersicht wird in der Abbildung 4.7 und den folgenden Abbildungen der Knoten SCHWIERIGKEIT DER FRAGE weggelassen.

Viele ATEMGERÄUSCHE in t_0 deuten darauf hin, daß $\mathcal{P}$ mit hoher Wahrscheinlichkeit eine hohe AUFREGUNG besitzt. Aus diesem Grund und weil $\mathcal{P}$ eine schlechte Antwort gibt, wird $\mathcal{P}$s VAK eher gering eingeschätzt. Diese Einschätzung kann man an den dunklen Histogrammen in t_0 erkennen. In t_1 muß der Knoten VAK neu vorhergesagt werden, da er im Modell als temporärer Knoten modelliert wurde (siehe helle Histogramme in t_1).

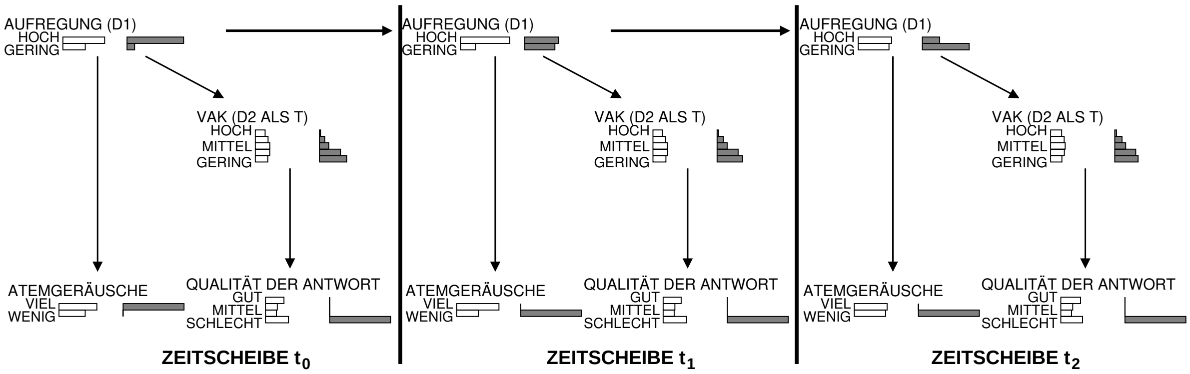

Abb. 4.7: *DBN von Beispiel 2 mit Modell $\mathcal{M}2.1$:* AUFREGUNG *ist dynamisch und* VAK *ist temporär.*

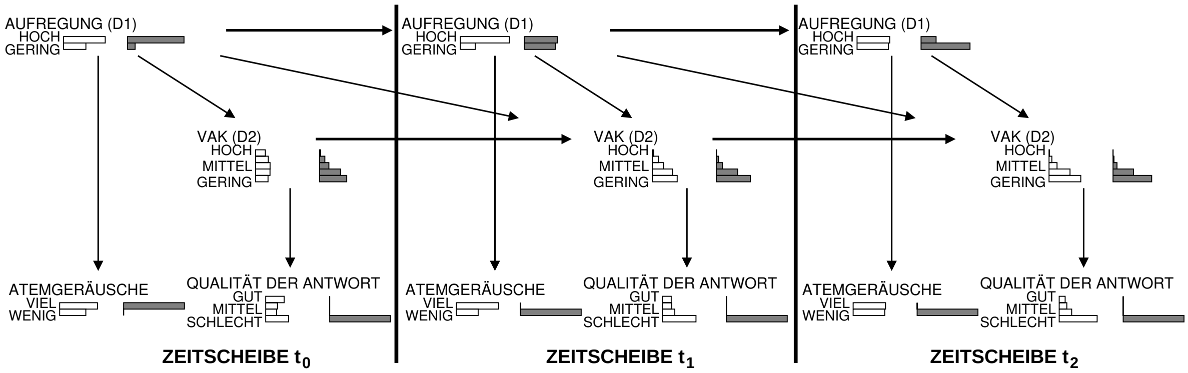

Abb. 4.8: *DBN von Beispiel 2 mit Modell $\mathcal{M}2.2$:* AUFREGUNG *und* VAK *sind dynamisch.*

Das Ergebnis ist weniger genau, als die Einschätzung in t_0. In t_1 weisen wenige ATEMGERÄUSCHE darauf hin, daß sich $\mathcal{P}$ beruhigt. Die AUFREGUNG in t_1 wird geringer eingeschätzt als in t_0 (siehe dunkle Histogramme). Obwohl weiterhin $\mathcal{P}$ eine schlechte Antwort gibt, erhöht sich leicht die VAK von $\mathcal{P}$. Diese Erhöhung ist zwar nur sehr gering, aber die Tendenz ist ein sehr fragliches Ergebnis (vgl. dunkle Histogramme von VAK). Der Grund für die geringe Erhöhung der VAK trotz schlechter QUALITÄT DER ANTWORT ist der, daß die VAK einzig von der AUFREGUNG abhängt und AUFREGUNG aufgrund der wenigen ATEMGERÄUSCHE als gering eingeschätzt wird. In t_2 setzt sich diese fragliche Tendenz weiter fort.

Im Modell $\mathcal{M}2.2$ (siehe rechte Seite der Abb. 4.6) gibt es keinen derartigen Informationsverlust, da D2 als dynamischer Knoten modelliert wird. Mit anderen Worten: D2 in t hängt zusätzlich von D2 in $t-1$ ab. Es bleibt zu erklären, warum D2 in t auch von D1 in $t-1$ abhängt. Der Einfluß von D1 in $t-1$ auf D2 in $t-1$ wurde bereits berücksichtigt. Der Knoten D1 bleibt auch in t die Ursache dieses Effekts. Es ist darum wichtig, daß dieser Einfluß von D1 auf D2 nicht zweifach gezählt wird. Deshalb ist D2 in t mit dem Knoten D1 in $t-1$ und mit D1 in t verbunden. Beim Definieren der CPT dieses Zusammenhangs muß D1 in $t-1$ und in t berücksichtigt werden. In der Notrufdomäne, ist die Einschätzung der AUFREGUNG in $t-1$ gleich der in t, wenn keine Beruhigung oder Verstärkung der Aufregung stattfindet. Wenn sich $\mathcal{P}$ beruhigt, wird die VAK höher sein und umgekehrt. Zusätzlich soll die Einschätzung der VAK nicht zu hoch sein, wenn $\mathcal{P}$ sehr aufgeregt ist. Abbildung 4.8 zeigt, wie das Modell $\mathcal{M}2.2$ Beispiel 2 modelliert. In diesem Fall tritt kein Informationsverlust auf. $\mathcal{P}$s VAK wird konstant als wahrscheinlich gering eingeschätzt. Dieses Ergebnis ist besser als das von Abbildung 4.7, modelliert mit $\mathcal{M}2.1$.

Die Interpretation in Modell $\mathcal{M}2.2$ ist kostspieliger als in $\mathcal{M}2.1$. Außerdem ist $\mathcal{M}2.2$ schwieriger zu modellieren: Die CPT, die die Abhängigkeiten zwischen D2 in t und seine Eltern D1 in t, D1 in $t-1$ und D2 in $t-1$ definiert, benötigt mehr Werte als die CPT, die den Zusammenhang zwischen D und T in $\mathcal{M}2.1$ beschreibt. Wenn man nur an der Einschätzung von $\mathcal{P}$s AUFREGUNG interessiert ist, kann man Provans (1993) Vorschlag, Effektivität vor Genauigkeit, und Modell $\mathcal{M}2.1$ vorziehen. Wenn $\mathcal{P}$s VAK im Vordergrund steht – wie in der Notrufdomäne –, dann ist $\mathcal{M}2.2$ das adäquate Modell.

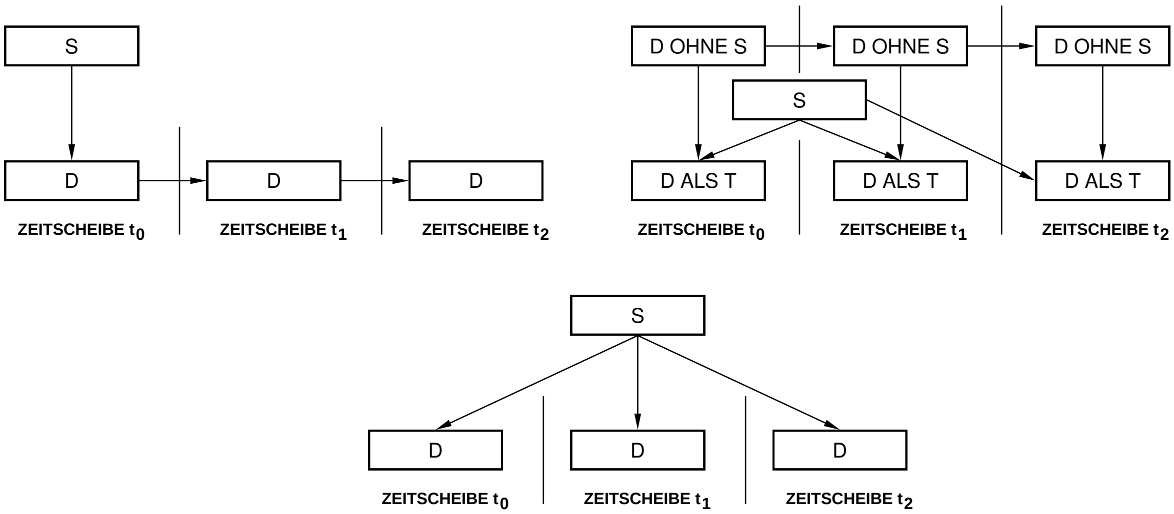

Abb. 4.9: *$\mathcal{M}3.1$ (links), $\mathcal{M}3.2$ (rechts) und $\mathcal{M}3.3$ (unten): Modellierung des Zusammenhangs zwischen einem statischen Knoten* S *und einem dynamischen Knoten* D. *Der zusätzliche dynamische Knoten* D OHNE S *im Modell $\mathcal{M}3.2$ erlaubt eine Vereinfachung von* D *als temporären Knoten* D ALS T.

4.2.3 Zusammenhänge zwischen statischen und dynamischen Knoten

Im letzten Fall wird der Einfluß eines statischen Elternknotens auf einen dynamischen Kindknoten untersucht.[4] Zum Beispiel beeinflußt ALKOHOLEINFLUSS die VAK von $\mathcal{P}$. Die Abbildung 4.9 zeigt drei unterschiedliche Schemata für die Modellierung eines solchen Zusammenhangs. Die Modelle werden anhand von Beispiel 3 illustriert.

Beispiel 3: Die Hintergrundgeräusche eines Anrufes weisen darauf hin, das der Anrufer aus einer Kneipe anruft. Eine undeutliche Aussprache signalisiert, daß $\mathcal{P}$ betrunken ist. Dennoch werden alle drei Fragen von $\mathcal{P}$ gut beantwortet.

Aus Vereinfachungsgründen wird wieder angenommen, daß der Knoten VAK in diesem Beispiel keine weiteren Elternknoten besitzen. Damit die Abbildungen etwas kleiner und übersichtlicher werden, wird ALKOHOLEINLFUSS als Wurzelknoten modelliert. Es wird angenommen, daß $\mathcal{P}$ in einer Gaststätte und möglicherweise betrunken ist.

Jensen (1995) schlägt vor, den Einfluß des statischen Knotens S in den dynamischen Knoten D zu vereinigen. So repräsentiert der dynamische Knoten D alle Information von S und das resultierende Netzwerk wird dadurch vereinfacht. Diese Idee wird in Modell $\mathcal{M}3.1$ übernommen (siehe linke Seite der Abb. 4.9). Dabei entstehen zwei Probleme:

(1) Der Knoten D muß den Einfluß von S repräsentieren können. Z.B, in Jensens (1995) Weizendomäne kann der Einfluß der AUSSÄHZEIT (S) in dem Knoten ENTWICKLUNGSSTADIUM (D) repräsentiert werden. Das ENTWICKLUNGSSTADIUM hängt nur von seinem Vorgänger in

[4] Der umgekehrte Fall dürfte wohl nie auftreten und wird daher nicht weiter berücksichtigt.

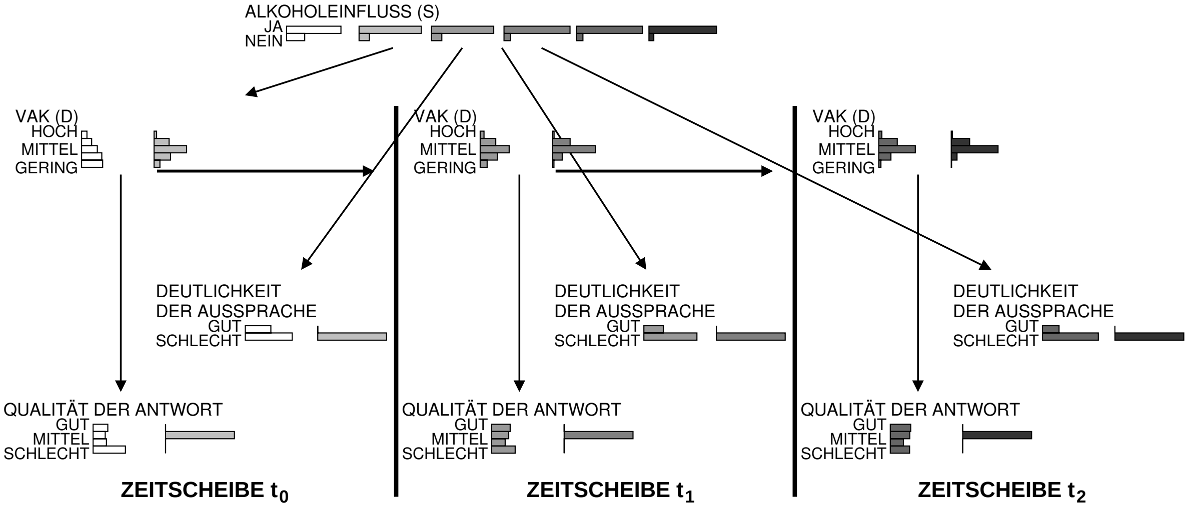

Abb. 4.10: *DBN von Beispiel 3 mit Modell $\mathcal{M}3.1$:* ALKOHOLEINFLUSS *ist statisch und* VAK *ist dynamisch.*

der vorherigen Zeitscheibe ab. In der Notrufdomäne würde diese Modellierung nicht so geeignet sein, da $\mathcal{P}$s VAK in t von $\mathcal{P}$s VAK in $t-1$ und $\mathcal{P}$s ALKOHOLEINFLUSS in t abhängt. Wenn $\mathcal{P}$ betrunken ist, sollte man vorsichtiger sein, $\mathcal{P}$ eine hohe VAK zuzuschreiben.

(2) Oft ist es nicht möglich einen statischen Knoten S mit Gewißheit zu initialisieren, sondern es gibt nur eine vage Einschätzung darüber. In der Weizendomäne von Jensen (1995) ist eine genaue Initialisierung möglich, da die Aussähzeit bekannt ist. Anders verhält es sich, wenn die Einschätzung im Laufe der Zeit verfeinert wird: Z.B. kann es sein, daß am Anfang eines Dialogs nicht klar ist, ob $\mathcal{P}$ betrunken ist. Wird erst im späteren Verlauf des Dialogs festgestellt das $\mathcal{P}$ betrunken ist, wird sich diese Evidenz nicht mehr stark genug auf die VAK der aktuellen Zeitscheibe auswirken.

Abbildung 4.10 zeigt ein DBN, das Beispiel 3 mit dem Modell $\mathcal{M}3.1$ beschreibt. Das Ergebnis ist fragwürdig. Obwohl $\mathcal{P}$ als betrunken eingeschätzt wird, wird $\mathcal{P}$s VAK als mittel bis eher hoch angenommen (in t_3): Das Netz *vergißt*, daß $\mathcal{P}$ betrunken ist und die gute QUALITÄT DER ANTWORT von $\mathcal{P}$ wird dadurch zu optimistisch interpretiert.

Das Modell $\mathcal{M}3.2$ (siehe rechte obere Darstellung in Abb. 4.9) erlaubt, daß der statische Knoten in jeder Zeitscheibe neu eingeschätzt werden kann und immer einen direkten Einfluß auf $\mathcal{P}$s VAK hat. Um das Netz zu vereinfachen, verwendet Modell $\mathcal{M}3.2$ Provans (1993) Vorschlag von Effizienz vor Genauigkeit: D wird als temporärer Knoten D ALS T repräsentiert, der in jeder Zeitscheibe neu erzeugt wird (vgl. Modell $\mathcal{M}2.1$).

D OHNE S repräsentiert die Eigenschaft von D *ohne* Einfluß von S. D ALS T wird erzeugt auf der Basis von S und D OHNE S. In dem Beispiel wird VAK (D ALS T) in jeder Zeitscheibe abhängig von VAK BEI NÜCHTERNHEIT (D OHNE S) und ALKOHOLEINFLUSS (S) erzeugt.

Abbildung 4.11 zeigt das Netz für Beispiel 3, das mit Modell $\mathcal{M}3.2$ erstellt wurde. In diesem Fall entsteht das gleiche Problem wie in Modell $\mathcal{M}2.1$ (vgl. Abb. 4.7). Nach jeder

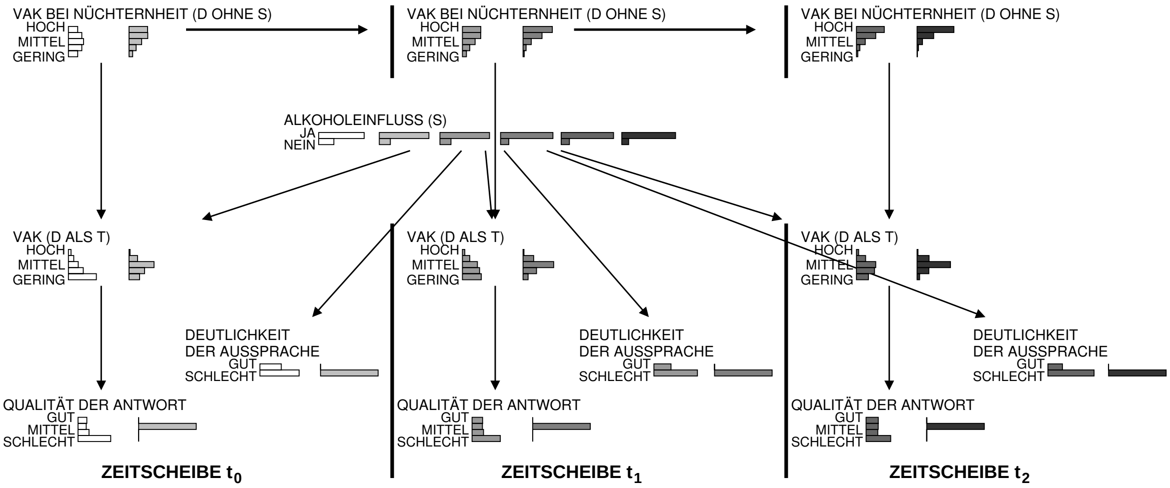

Abb. 4.11: *DBN von Beispiel 3 mit Modell $\mathcal{M}3.2$:* ALKOHOLEINFLUSS *ist statisch und* VAK *ist dynamisch.*

Neuerzeugung der VAK gibt es einen Informationsverlust. Die Einschätzung des Knotens VAK in jeder neuen Zeitscheibe ist weniger genau als in der Zeitscheibe zuvor (vgl. z.B. dunkle Histogramme in t_0 mit den weißen Histogramme in t_1). Wenn der Informationsverlust einen weniger wichtigen Knoten betrifft, ist dieses Modell eine sinnvolle Vereinfachung. In dem vorliegenden Fall interessiert aber gerade $\mathcal{P}$s VAK und deswegen ist dieses Modell hier nicht anwendbar.

Das Modell $\mathcal{M}3.3$ (siehe Abb. 4.9 unten) modelliert S als statischen und D als dynamischen Knoten. Eine derartige Struktur wird von Jensen (1995) als erstes Modell verwendet.

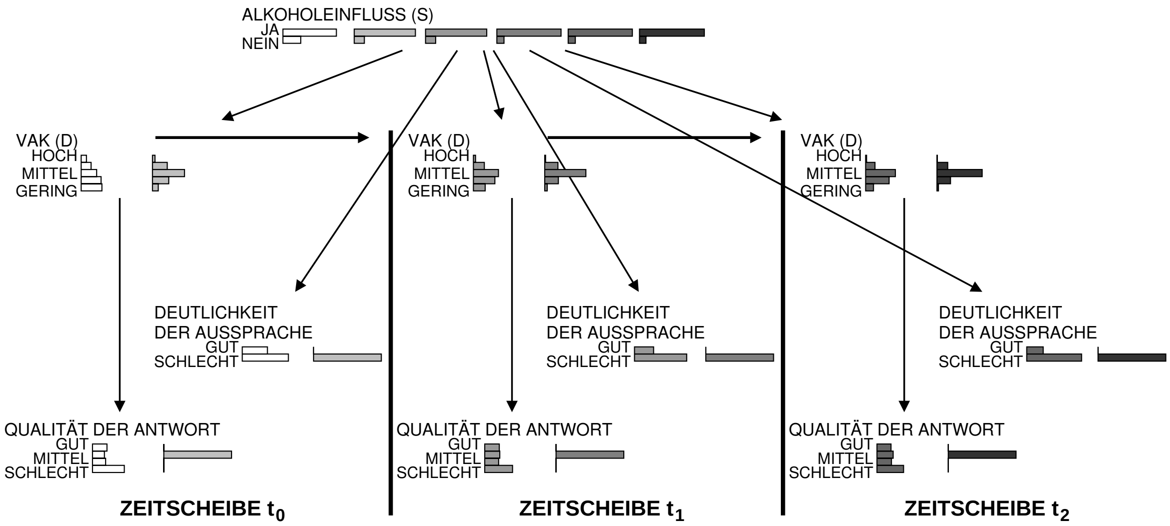

Abb. 4.12: *DBN von Beispiel 3 mit Modell $\mathcal{M}3.3$:* ALKOHOLEINFLUSS *ist statisch und* VAK *ist dynamisch.*

Jensen (1995) verwirft dieses Modell aber aus Effizienzgründen und vereinfacht laut Modell $\mathcal{M}3.1$. Wie beschrieben wurde, ist diese Vereinfachung in seinem Fall berechtigt.

In Modell $\mathcal{M}3.3$ gibt es Knoten, die durch zwei oder mehr Pfade miteinander verbunden sind. D.h. das resultierende Netz ist mehrfach verbunden und benötigt daher sehr kostspielige Inferenzalgorithmen. Abbildung 4.12 zeigt das DBN, das das o.g. Beispiel 3 mit $\mathcal{M}3.3$ beschreibt.

Im Beispiel 3 führt $\mathcal{M}3.2$ zu einem ähnlichen Ergebnis nach der Interpretation der beobachteten Evidenz einer eher guten Antwort. Gäbe es nicht diese Evidenz, würde $\mathcal{P}$s VAK inadäquat gering eingeschätzt werden (vgl. helle Histogrammen der VAK in Zeitscheibe t_1 nach t_2 in den Abbildungen 4.11 und 4.12). Modell $\mathcal{M}3.3$ berücksichtigt in t_1 die Einschätzung der VAK für die neue Einschätzung in t_2. Aus diesem Grund wird das Modell $\mathcal{M}3.3$ für die Modellierung der Notrufdomäne bevorzugt.

Alternativ könnte man auch den statischen Knoten (S) durch einen dynamischen Knoten (D) ersetzen. Diese Lösung wird von Heckerman (1993) vorgeschlagen. Das resultierende Modell wäre dann $\mathcal{M}2.2$.

4.3 Quantitative Modellierung der Zusammenhänge

Die Zusammenhänge, die in den vorherigen Abschnitten beschrieben wurden, werden im folgenden quantitativ beschrieben. Für jeden Knoten werden die folgenden Attribute spezifiziert:

- *Knoten* eindeutiger Name des Knotens
- *Status* statisch, dynamisch oder temporär
- *Eltern* Liste der Elternknoten
- *Zustände* Menge von sich ausschließenden und die Variable erschöpfend beschreibenden Zustände (Hypothesen)
- *A-priori* a-priori Wahrscheinlichkeiten für das Zutreffen der Zustände, wenn der Knoten ein Wurzelknoten ist
- *CPT* bedingte Wahrscheinlichkeiten über alle möglichen Kombinationen der Zustände der verbundenen Knoten, wenn der Knoten ein Kindknoten ist
- *Beschreibung* natürlichsprachliche Beschreibung der CPT bzw. der a-priori Wahrscheinlichkeiten

Die CPT eines Kindknotens (bzw. a-priori Wahrscheinlichkeit eines Elternknotens) kann auf verschiedene Arten definiert werden. Z.B. können die bedingten Wahrscheinlichkeiten direkt in Form einer Verbindungsmatrix angeben werden. Diese Vorgehensweise bietet sich an, wenn die Verbindungsmatrix nicht zu groß ist.

Erhöht sich die Menge der Elternknoten oder die Anzahl der Zustände eines Knotens, ist eine Beschreibung der CPT durch direkte Angabe der bedingten Wahrscheinlichkeiten der

Tabelle 4.1: *Beschreibung der Knoten* GESPRÄCH MIT ANDEREN PERSONEN *und* ABLENKUNG

Knoten: GESPRÄCH MIT ANDEREN PERSONEN (G)	Knoten: ABLENKUNG$_t$ (A_t)
Status: temporär	Status: dynamisch
Eltern: –	Eltern: ABLENKUNG$_{t-1}$ (A_{t-1}), GESPRÄCH MIT ANDEREN PERSONEN (G)
Zustände: (2) nein, ja	Zustände: (2) gering, hoch
A-priori: $\begin{bmatrix} 0,80 & 0,20 \end{bmatrix}$	CPT: $P(A_t = a_t \mid A_{t-1} = a_{t-1}, G = g) = M_{A_t \mid A_{t-1}, G=g}(a_t, a_{t-1})$ $$\begin{array}{cc} M_{A_t\mid A_{t-1},G=nein} & M_{A_t\mid A_{t-1},G=ja} \\ \begin{bmatrix} 0,95 & 0,05 \\ 0,30 & 0,70 \end{bmatrix} & \begin{bmatrix} 0,15 & 0,85 \\ 0,10 & 0,90 \end{bmatrix} \end{array}$$
Beschreibung: Ein Gespräch mit anderen Personen während des Notrufs ist nicht wahrscheinlich.	Beschreibung: Wenn $\mathcal{P}$ sich mit anderen unterhält und in $t-1$ abgelenkt war, dann ist $\mathcal{P}$ sehr wahrscheinlich in t abgelenkt.

Anmerkung: $P(X = x \mid Y = y)$ = Wahrscheinlichkeit für $X = x$ bedingt $Y = y$, $M_A(x, y)$ = Matrixelement in Zeile x und Spalte y

Verbindungsmatrix schwierig. Die Zahl der zu definierenden bedingten Wahrscheinlichkeiten bei n zusammenhängenden Knoten (einschließlich des Kindknotens) mit z_i Zuständen $(i = 1 \cdots n)$ beträgt: $\prod_{i=1}^{n} z_i$. Wenn die Zahl der bedingten Wahrscheinlichkeiten zu groß ist, ist es besser, den Zusammenhang mit einer geeigneten Funktion anzugeben.

Beide Fälle werden in den folgenden zwei Abschnitten exemplarisch beschrieben. Die übrigen Knotendefinitionen werden im Anhang A.4 angegeben.

4.3.1 Angabe der CPT durch eine Verbindungsmatrix

Die Definition einer CPT durch die direkte Angabe der bedingten Wahrscheinlichkeiten wird anhand des Zusammenhangs GESPRÄCH MIT ANDEREN PERSONEN, ABLENKUNG$_{t-1}$ und ABLENKUNG$_t$ illustriert (siehe Tabelle 4.1). Es handelt sich hierbei um den einfachen Zusammenhang, bei dem ein temporärer Knoten (GESPRÄCH MIT ANDEREN PERSONEN) einen dynamischen Knoten (AUFREGUNG) beeinflußt.

Der temporäre Knoten GESPRÄCH MIT ANDEREN PERSONEN (G) ist ein Wurzelknoten mit zwei Zuständen, *nein* und *ja*. Für das Zutreffen der Zustände werden a-priori Wahrscheinlichkeiten angegeben: mit 80% Wahrscheinlichkeit führt $\mathcal{P}$ kein GESPRÄCH MIT ANDEREN PERSONEN. ABLENKUNG ist eine zeitveränderliche Variable, die als dynamischer Knoten mit zwei Zuständen, *gering* und *hoch*, modelliert wird. ABLENKUNG$_t$ (A_t) in Zeitscheibe t hat zwei Elternknoten,

Tabelle 4.2: *Beschreibung der Knoten* ABLENKUNG *und* VAK

Knoten: AUFREGUNG (A_t)

Status: dynamisch

Eltern: AUFREGUNG$_{t-1}$ (A_{t-1})

Zustände: gering, hoch

CPT: $P(A_t = a_t \mid A_{t-1} = a_{t-1}) = M_{A_{t-1}|A_{t-1}}(a_t, a_{t-1})$

$$M_{A_{t-1}|A_{t-1}} = \begin{bmatrix} 0,85 & 0,15 \\ 0,15 & 0,85 \end{bmatrix}$$

Beschreibung: $\mathcal{P}$s Aufregung kann mit der Zeit nachlassen, wenn $\mathcal{P}$ z.B. von $\mathcal{S}$ beruhigt wird.

Knoten: VAK (V) (Notruf $\mathcal{N}1$)

Status: dynamisch

Eltern: VAK$_{t-1}$ (V_{t-1}), AUFREGUNG$_{t-1}$ (A_{t-1}), AUFREGUNG$_t$ (A_t), ABLENKUNG$_{t-1}$ (B_{t-1}), ABLENKUNG$_t$ (B_t)

Zustände: sehr gering, gering, mittel, hoch, sehr hoch

CPT: $P(V_t = v_t \mid V_{t-1} = v_{t-1}, A_t = a_t, A_{t-1} = a_{t-1}, B_t = b_t, B_{t-1} = b_{t-1}) = N_{\mu,\sigma}(v_t)$

$$dA = \begin{cases} 0,0 & \text{, wenn} \quad a_{t-1} = a_t \\ 0,3 & \text{, wenn} \quad a_{t-1} > a_t \\ -0,3 & \text{, wenn} \quad a_{t-1} < a_t \end{cases}$$

$$dB = \begin{cases} 0,0 & \text{, wenn} \quad b_{t-1} = b_t \\ 0,3 & \text{, wenn} \quad b_{t-1} > b_t \\ -0,3 & \text{, wenn} \quad b_{t-1} < b_t \end{cases}$$

$$\mu = v_{t-1} + v_{t-1} * (dA + dB), \quad \sigma = 0,15$$

Beschreibung: Hat sich die Aufregung oder Ablenkung von $\mathcal{P}$ erhöht, dann erhöht sich die Wahrscheinlichkeit für eine geringe VAK.

Anmerkung: $P(X = x \mid Y = y)$ = Wahrscheinlichkeit für $X = x$ bedingt $Y = y$, $M_A(x, y)$ = Matrixelement in Zeile x und Spalte y, $N_{\mu,\sigma}(x)$ = Normalverteilung mit Erwartungswert μ und Standardabweichung σ.

GESPRÄCH MIT ANDEREN PERSONEN (G) in t und ABLENKUNG$_{t-1}$ (A_{t-1}) in $t-1$. Die CPT wird durch die Verbindungsmatrix $M_{A_t|A_{t-1},G}$ definiert: die Wahrscheinlichkeit für eine geringe Ablenkung in t ist 95%, wenn $\mathcal{P}$ in $t-1$ nur gering abgelenkt war und $\mathcal{P}$ in t kein Gespräch mit anderen Personen führt.

4.3.2 Berechnung der bedingten Wahrscheinlichkeiten durch eine Funktion

Komplizierter wird es im Fall der Variablen VAK. Diese Variable ist zeitveränderlich und wird daher als dynamischen Knoten modelliert. Da zwei zeitveränderliche Variablen, nämlich ABLENKUNG und AUFREGUNG, die VAK beeinflussen, wird für diesen Zusammenhang das Modellierungsschema $\mathcal{M}3.3$ benutzt. Das bedeutet: VAK$_t$ in t ist abhängig von VAK$_{t-1}$, ABLENKUNG$_{t-1}$, AUFREGUNG$_{t-1}$ in $t-1$ sowie von ABLENKUNG$_t$ und AUFREGUNG$_t$ in t (siehe Tabelle 4.2).

Die Verbindungsmatrix dieses Zusammenhangs enthält 400 bedingte Wahrscheinlichkeiten. Für die Angabe der CPT eignet sich deshalb besser eine Funktion, die die Werte berechnet. Die passende Funktion dieses Zusammenhangs orientiert sich an folgender sprachlichen Beschreibung: „ $\mathcal{P}$s VAK in t wird im Mittel ähnlich der VAK in $t-1$ sein (mit einer gewissen Abweichung). Erhöht sich aber $\mathcal{P}$s AUFREGUNG oder ABLENKUNG von $t-1$ nach t, wird $\mathcal{P}$s VAK in t etwas geringer sein als in $t-1$. Sinkt $\mathcal{P}$s AUFREGUNG oder ABLENKUNG von $t-1$ nach t, wird $\mathcal{P}$s VAK in t etwas höher sein als in $t-1$.“ Dieser Zusammenhang läßt sich gut anhand einer Normalverteilung definieren (siehe Tabelle 4.2).

4.3.2.1 Technische Realisierung der Normalverteilungsfunktion

Eine Normalverteilung wird durch einen Erwartungswert μ und eine Standardabweichung σ definiert. Vorher ist ein entsprechendes Intervall zu wählen, z.B [0,1]. Da nur das Intervall betrachtet wird, muß die Fläche außerhalb des Intervalls prozentual innerhalb des Intervalls verteilt werden (siehe Abb. 4.13 oben). Danach wird das Intervall in diskrete Intervallsegmente gemäß der Anzahl der Zustände des Knotens eingeteilt. Jedem Zustand wird eines dieser Intervallsegmente zugeordnet. Die Wahrscheinlichkeit für das Zutreffen des Zustandes wird dann mit der Fläche des entsprechenden Segments gleichgesetzt (siehe Abb. 4.13 unten).

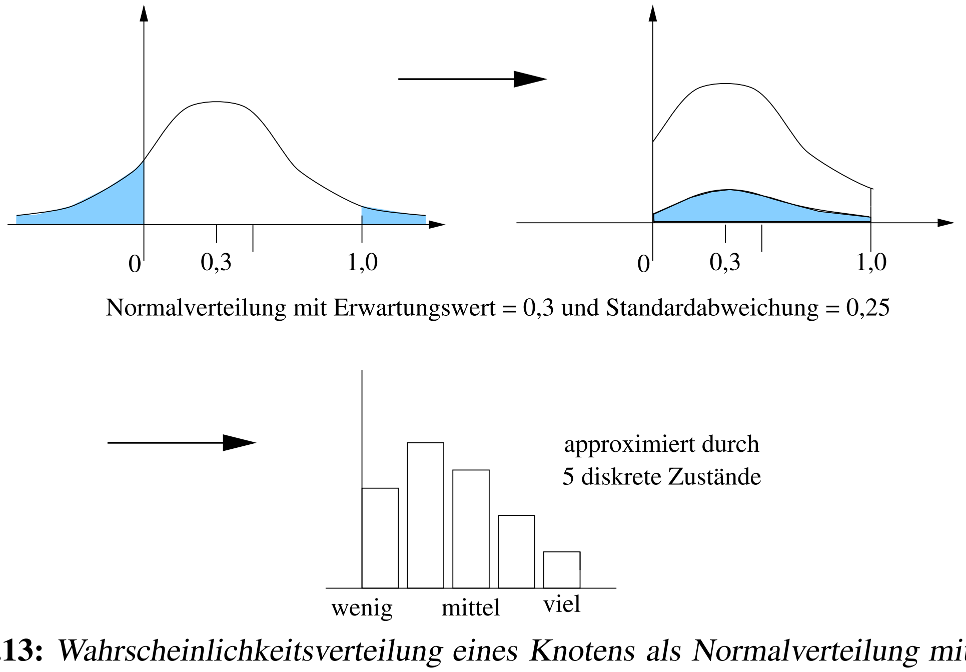

Abb. 4.13: *Wahrscheinlichkeitsverteilung eines Knotens als Normalverteilung mit Erwartungswert μ und Standardabweichung σ*

4.4 Tests und Simulation

Anhand der beiden Notrufdialoge aus der empirischen Untersuchung wird das beschriebene Benutzermodell getestet. Der Test erfolgt durch eine Simulation der beiden Notrufdialoge. Dabei dienen die Aussagen der Versuchspersonen aus der empirischen Studie als Vergleich, um festzustellen, ob das Benutzermodell plausible Einschätzungen der Arbeitsgedächtnisbelastung des Dialogpartners zieht.

4.4.1 Notrufdialog: Die Frau in St. Johann

Im ersten Fallbeispiel wird ein Teil des Notrufdialogs „Die Frau in St. Johann“ (siehe Tabelle 3.1) simuliert. Das Beispiel beschränkt sich auf die ersten fünf Dialogäußerungen. Im ersten Schritt der Modellierung wird der Dialog in Zeitscheiben eingeteilt. Die beiden ersten Zeitscheiben t_0 und t_1 (siehe Abb. 4.14, oben) zeigen den Anfang des Notrufdialogs mit der Situationsbeschreibung. In den beiden folgenden zwei Zeitscheiben t_2 und t_3 (siehe Abb. 4.14, unten) fragt der Mitarbeiter der FNZ nach dem Stadtteil, in dem die Anruferin wohnt. Da die Frau beim ersten Mal um eine Wiederholung der Frage bittet (in t_2), wird die Frage ein zweites Mal gestellt (in t_3). Die Tabelle 4.3 zeigt die Zuordnung der Dialogäußerungen zu den einzelnen Zeitscheiben und die in den Zeitscheiben beobachteten Symptome.

In der Zeitscheibe t_0 werden viele ATEMGERÄUSCHE festgestellt, so daß die Wahrscheinlichkeit für eine hohe AUFREGUNG höher eingeschätzt wird. Aufgrund dieser relativ hohen AUFREGUNG wird der Grad der HILFLOSIGKEIT entsprechend höher bewertet. Die neue Einschätzung wird dadurch erklärt, daß entweder der NOTFALL größer ist oder die ERFAHRUNG MIT NOTFÄLLEN der Anruferin geringer ist, als a-priori angenommen wurde. Außerdem werden keine GESPRÄCHSGERÄUSCHE beobachtet. Die ABLENKUNG wird daher etwas geringer eingeschätzt als zuvor. Da die AUFREGUNG ziemlich hoch angenommen wird, wird die VAK als mittel bis gering eingeschätzt.

In Zeitscheibe t_1 erfolgt die Situationsbeschreibung. Der Übergang von einer Zeitscheibe zur anderen ist so modelliert, daß sich $\mathcal{P}$ im Laufe der Zeit beruhigen kann (vgl. Abb. 4.5). Die Vorhersage für $\mathcal{P}$s ABLENKUNG und $\mathcal{P}$s VAK berücksichtigt zusätzlich zu den statischen Abhängigkeiten von den Elternknoten die jeweilige Einschätzung der letzten Zeitscheibe. In t_1 werden wie in t_0 viele ATEMGERÄUSCHE wahrgenommen, so daß wiederum die AUFREGUNG höher eingeschätzt wird als vorhergesagt. Da die AUFREGUNG bereits relativ hoch eingeschätzt wurde, ist es jetzt sehr wahrscheinlich, daß die Anruferin sehr aufgeregt ist. Wieder sind keine GESPRÄCHSGERÄUSCHE wahrzunehmen. Als Folge davon wird wie in t_0 eine geringe ABLENKUNG wahrscheinlicher. Die sehr hohe Wahrscheinlichkeit für eine hohe AUFREGUNG ist auch der Grund dafür, daß die VAK doch eher gering bewertet wird. Eine schlechte QUALITÄT DER SITUATIONSBESCHREIBUNG trägt dazu bei.

Die Einschätzungen der Versuchspersonen waren ähnlich: Sechs von elf Versuchspersonen meinten, daß die Anruferin eine schlechte Situationsbeschreibung gegeben hat und führten diese Tatsache auf eine eher hohe Arbeitsgedächtnisbelastung zurück.

Tabelle 4.3: *Beobachtbare Symptome im Notrufdialog „Die Frau in St. Johann“*

Zeitscheibe 0:	
$\mathcal{P}$: <#Stöhnen>.	ATEMGERÄUSCHE = viel GESPRÄCHSGERÄUSCHE = nein
Zeitscheibe 1:	
$\mathcal{P}$: Ja, hier ist Frau Müller, Schopenhauerstraße 10. Bei mir brennt der Badeautomat. Da war heute morgen noch ein Installateur da. <lautes Atmen>. Jetzt brennt das Ding. <#Atmen>.	ATEMGERÄUSCHE = viel GESPRÄCHSGERÄUSCHE = nein QUALITÄT DER SITUATIONSBESCHREIBUNG = schlecht
Zeitscheibe 2:	
$\mathcal{S}$: **In welchem Stadtteil?** $\mathcal{P}$: Bitte?	ATEMGERÄUSCHE = wenig GESPRÄCHSGERÄUSCHE = nein BITTE UM WIEDERHOLUNG = ja
Zeitscheibe 3:	
$\mathcal{S}$: **In welchem Stadtteil?** $\mathcal{P}$: Ei, das is' hier in Saarbrücken <Z> in St. Johann, in der Nähe der LVA.	ATEMGERÄUSCHE = wenig GESPRÄCHSGERÄUSCHE = nein QUALITÄT DER ANTWORT = sehr gut GESCHWINDIGKEIT DER ANTWORT = schnell

Anmerkung: links: Äußerungen, rechts: beobachtbare Symptome

In der Zeitscheibe t_2 fragt der Mitarbeiter der FNZ nach dem Stadtteil, in dem die Anruferin wohnt. Die SCHWIERIGKEIT DER FRAGE wird als mittel bis gering angenommen. Der Grund dafür ist eine eher hohe VORHERSEHBARKEIT DER FRAGE, daß Stadtteile eher häufig genannt werden, sie also im sprachlichen Gebrauch relativ gängig sind, und schließlich, daß die Frage akustisch eher gut verständlich ist. Aber aufgrund der geringen VAK wird mit einer mittel bis schlechten QUALITÄT DER ANTWORT gerechnet, und daß die Antwort eher langsam gegeben wird (siehe helle Histogramme). Die Anruferin bittet aber um eine WIEDERHOLUNG der Frage. Die BITTE UM WIEDERHOLUNG führt zu einer Neueinschätzung der SCHWIERIGKEIT DER FRAGE. Die Frage wird jetzt als eher mittel bis schwer eingeschätzt, was dadurch begründet wird, daß die Stadtteilfrage doch nicht so vorhersehbar war wie vorher angenommen wurde. Auch die Einschätzung der GÄNGIGKEIT VON STADTTEILEN ändert sich. Die eher optimistische Einschätzung wird nun etwas relativiert.

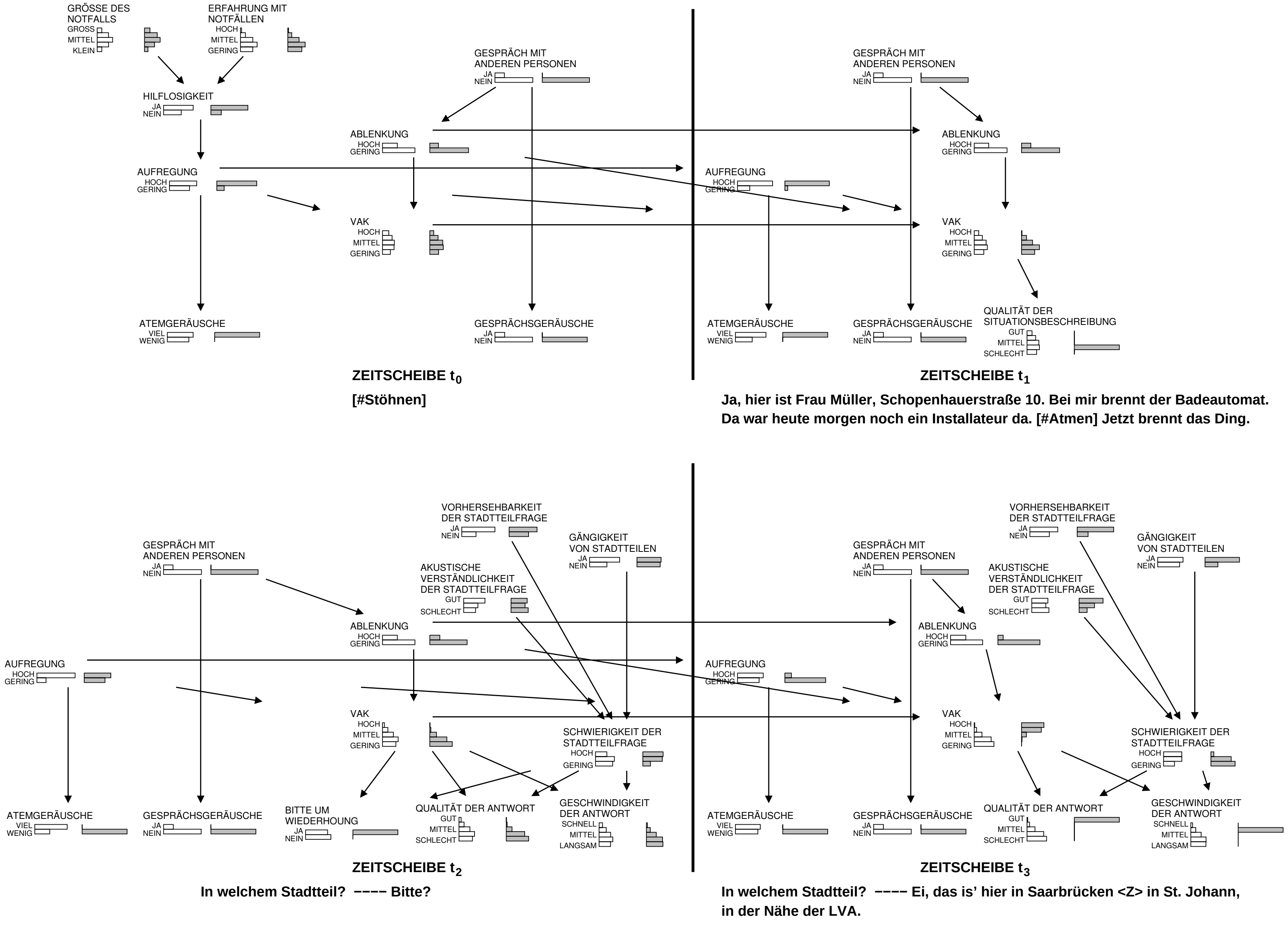

Abb. 4.14: *Notrufdialog: Die Frau in St. Johann.*

Diese Revision der Einschätzung der Schwierigkeit der Stadtteilfrage ist auch in den Aussagen der Versuchspersonen zu finden (vgl. Abschnitt 3, Tabelle 3.3).

Die BITTE UM WIEDERHOLUNG führt auch dazu, daß die VAK noch geringer eingeschätzt wird als zuvor. Die Beruhigung der Anruferin, die sich dadurch zeigt, daß die AUFREGUNG weniger hoch eingeschätzt wird als in der Zeitscheibe zuvor (aufgrund weniger ATEMGERÄUSCHE), kann an der pessimistischen Einschätzung der VAK wenig ändern.

In Zeitscheibe t_3 wiederholt der Mitarbeiter der FNZ die Frage nach dem Stadtteil. In Folge der geringen Wahrscheinlichkeit der VAK in t_2 wird auch eine geringe VAK in t_3 vorhergesagt. Eine geringe VAK führt wahrscheinlich zu einer schlechten QUALITÄT DER ANTWORT und einer eher langsamen GESCHWINDIGKEIT DER ANTWORT. Die Vorhersage in t_3 ist noch pessimistischer als die in t_2. Diese Einschätzung ist vergleichbar mit den Aussagen der meisten Versuchspersonen: Nach der Bitte um Wiederholung der Stadtteilfrage erwarten nur noch fünf Versuchspersonen eine schnelle Antwort, wohingegen vor der Bitte um Wiederholung neun Versuchspersonen eine schnelle Antwort erwartet haben.

Dieses Mal antwortet die Anruferin schnell und mit einer sehr guten Qualität. Wenig ATEMGERÄUSCHE signalisieren zusätzlich, daß die Anruferin sich etwas beruhigt hat. Aus diesem Grund und wegen der schnellen und sehr guten QUALITÄT DER ANTWORT wird die VAK jetzt eher hoch eingeschätzt (siehe schwarze Histogramme in t_3). Auch die Versuchspersonen urteilten dementsprechend (vgl. Tabelle 3.3).

4.4.2 Notrufdialog: Der Mann in der Gaststätte

Im zweiten Fallbeispiel wird ein Teil des Notrufdialogs „Der Mann in der Gaststätte“ (siehe Tabelle 3.2) simuliert. Das Beispiel beschränkt sich wieder auf bestimmte Ausschnitte des Originaldialogs. Die Dialogteile 5-8 des Originals wurden weggelassen, da sie keine neuen Erkenntnisse liefern. Der Dialog wird wieder in Zeitscheiben eingeteilt, die bis auf t_0 (die Situationsbeschreibung) je eine Äußerung des Mitarbeiters der FNZ und eine Äußerung des Anrufers enthalten. Die Tabelle 4.4 zeigt die Zuordnung der Dialogäußerungen zu den einzelnen Zeitscheiben und die in den Zeitscheiben beobachteten Symptome.

In Zeitscheibe t_0 (siehe Abb. 4.15) hört man, daß der Anrufer mit anderen Personen ein Gespräch führt. Als Folge davon erhöht sich die Wahrscheinlichkeit, daß er abgelenkt ist. Die Ablenkung und eine mittelhohe Wahrscheinlichkeit, daß der Anrufer unter ALKOHOLEINFLUSS steht, worauf eine schlechte DEUTLICHKEIT DER AUSSPRACHE hindeutet, führt zu einer geringen Einschätzung der VAK. Die schlechte QUALITÄT DER SITUATIONSBESCHREIBUNG verstärkt diese pessimistische Einschätzung.

In Zeitscheibe t_1 fragt der Mitarbeiter der FNZ nach dem Namen des Anrufers. Die hohe Wahrscheinlichkeit für eine hohe ABLENKUNG in t_0 wird nach t_1 übernommen. Aus diesem Grund und weil die VAK in t_0 sehr gering eingeschätzt wurde, wird die VAK in t_1 wiederum als sehr gering vorhergesagt. Als Folge davon ist auch mit einer eher schlechten QUALITÄT DER ANTWORT zu rechnen, obwohl die SCHWIERIGKEIT DER FRAGE selbst als eher gering angenommen wird (siehe helle Histogramme). Aber entgegen der Vorhersage beantwortet der Anrufer die

Tabelle 4.4: *Beobachtbare Symptome im Notrufdialog „Der Mann in der Gaststätte“*

Zeitscheibe 0:	
$\mathcal{P}$: Dieter Müller is mein Name. Un zwar <P> in de <#> <;redet mit jemanden anderes> direkt do vorne. nä, nä, do hinne, <P> Trifelsstroß. <;redet jetzt mit Feuerwehrmann> Äh, Trifelsstroß, owwe im Dachgeschoß is e Feuer.	GESPRÄCHSGERÄUSCHE = ja QUALITÄT DER SITUATIONSBESCHREIBUNG = sehr schlecht DEUTLICHKEIT DER AUSSPRACHE = schlecht
Zeitscheibe 1:	
$\mathcal{S}$: **Wie war noch ihr Name?**	GESPRÄCHSGERÄUSCHE = nein
$\mathcal{P}$: Dieter Müller, 47081511 DRK-Nummer.	QUALITÄT DER ANTWORT = gut
Zeitscheibe 2:	
$\mathcal{S}$: **Geben's mir noch ihre Telefonnummer!**	GESPRÄCHSGERÄUSCHE = nein
$\mathcal{P}$: Äh, die Telefonnummer von hier? <%><P> Ich ruf aus einer Gaststätte an.	QUALITÄT DER ANTWORT = sehr schlecht HINTERGRUNDGERÄUSCH EINER GASTSTÄTTE = ja
Zeitscheibe 3:	
$\mathcal{S}$: **Aus einer Gaststätte. Genn 'se mir die Nummer von der Gaststätte!**	GESPRÄCHSGERÄUSCHE = ja
$\mathcal{P}$: <;Zur Wirtin> Könn 'se mir gerade die Gaststättennummer äh äh die Nummer? <P> Bitte? <P> 25 <P> <;Im Hintergrund: 25 - 27> — 27 — 25 <P> — 27	QUALITÄT DER ANTWORT = sehr schlecht GESCHWINDIGKEIT DER ANTWORT = langsam HINTERGRUNDGERÄUSCH EINER GASTSTÄTTE = ja

Anmerkung: links: Äußerungen, rechts: beobachtbare Symptome

Frage nach seinem Namen gut. Auch ist in t_1 kein GESPRÄCH MIT ANDEREN PERSONEN mehr zu hören, was eine geringere Wahrscheinlichkeit für die ABLENKUNG zur Folge hat. Aus diesen Gründen wird auch die VAK jetzt eher mittel eingeschätzt.

In Zeitscheibe t_2 fragt der Mitarbeiter nach der Telefonnummer des Anrufers, der daraufhin zu erkennen gibt, daß er aus einer Gaststätte anruft. Da die Frage nach der Telefonnummer als wahrscheinlich vorhersehbar bewertet wird, die LÄNGE DER NUMMER wahrscheinlich zwischen 4 und 7 Ziffern liegt und der Anrufer mit der eigenen Nummer vertraut sein müßte (bei der Vorhersage ist noch nicht bekannt, daß der Anrufer aus einer Gaststätte anruft), wird die SCHWIERIGKEIT DER FRAGE als eher mittel bis gering angenommen.

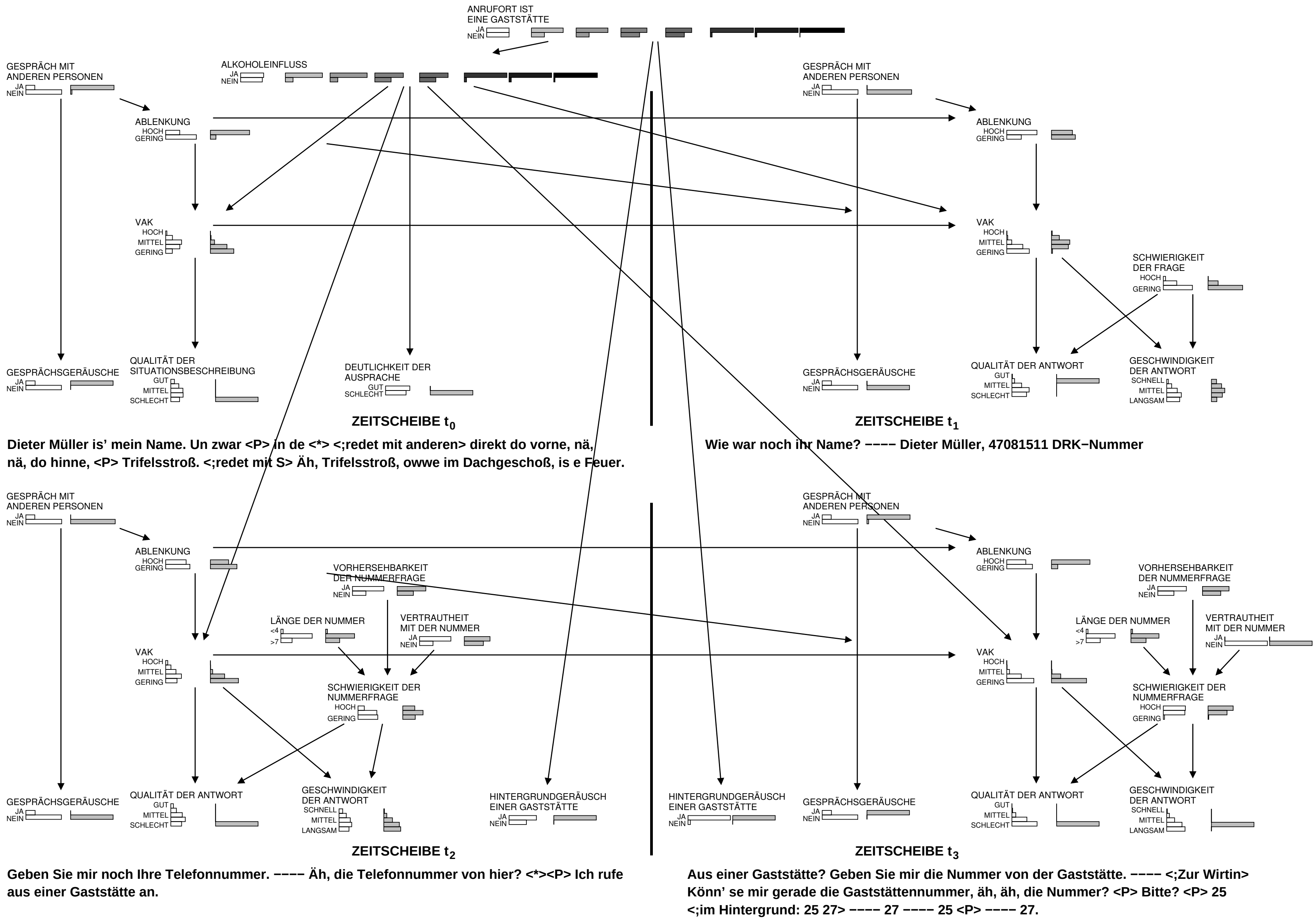

Abb. 4.15: *Notrufdialog: Der Mann in der Gaststätte.*

Aus diesem Grund und weil die VAK eher mittel bis gering eingeschätzt wurde, wird mit einer mittleren QUALITÄT und einer mittleren GESCHWINDIGKEIT der Antwort gerechnet (siehe helle Histogramme).

Der Anrufer beantwortet aber die Frage nicht, was hier als schlechte QUALITÄT DER ANTWORT modelliert wird. Zusätzlich erkennt man typische HINTERGRUNDGERÄUSCHE EINER GASTSTÄTTE, woraufhin die Wahrscheinlichkeit dafür, daß der ANRUFORT EINE GASTSTÄTTE ist, sehr hoch ist. Die Folge davon ist eine sehr hohe Wahrscheinlichkeit für den ALKOHOLEINFLUSS des Anrufers. Die sehr hohe Wahrscheinlichkeit des ALKOHOLEINFLUSSES und die schlechte Antwort bewirken eine noch geringere Einschätzung der VAK des Anrufers (siehe dunkle Histogramme).

In Zeitscheibe t_3 fordert der Mitarbeiter der FNZ den Anrufer auf, die Telefonnummer der Gaststätte durchzugeben. Da nicht davon ausgegangen werden kann, daß die Nummer der Gaststätte dem Anrufer bekannt ist, wird die VERTRAUTHEIT MIT DER NUMMER auf nein gesetzt. Damit wird die SCHWIERIGKEIT DER FRAGE jetzt höher beurteilt. Dementsprechend wird auch mit einer eher schlechten QUALITÄT DER ANTWORT gerechnet (siehe helle Histogramme). Wie erwartet wurde, hat der Anrufer Probleme, die Telefonnummer durchzugeben (siehe Text unter Zeitscheibe t_3). Aus diesem Grund und wegen der hohen Wahrscheinlichkeit für eine hohe ABLENKUNG wird die VAK des Anrufers als sehr gering eingeschätzt. Eine ähnliche Einschätzung haben auch die Versuchspersonen getroffen.

4.5 Diskussion

Es wurde gezeigt, wie man DBNs einsetzen kann, um die Arbeitsgedächtnisbelastung des Benutzers in einem Notrufdialog einzuschätzen.

Die beschriebenen Netze machten vergleichbare Einschätzungen der Arbeitsgedächtnisbelastung eines Anrufers/Benutzers wie die Mitarbeiter der FNZ. Mit Hilfe des Benutzermodells ist es also möglich, plausible Inferenzen über den Grad der Arbeitsgedächtnisbelastung des Benutzers in einem Notrufdialog zu ziehen.

Das Ziel dieser Modellierung war es, die Inferenzen der Mitarbeiter der FNZ in den zwei Notrufdialogen post hoc zu rekonstruieren, damit die oben genannten Vergleiche gezogen werden konnten. Die hier verwendeten DBNs können aber jeweils nur ihren Norufdialog beschreiben. Für ein System, das verschiedene Dialoge mit einem Benutzer führen soll, sind solche starre Netze nicht brauchbar. Im System READY werden die Bayesschen Netze daher dynamisch, d.h. abhängig von der jeweiligen Dialogsituation, konstruiert (siehe Abschnitt 6.2).

Kapitel 5

Empirie: Einschätzung von Zeitdruck und Arbeitsgedächtnisbelastung

Mit Hilfe der Mitarbeiter der FNZ, die viel Erfahrung im Umgang mit Arbeitsgedächtnisbelastung haben, konnte eine erste empirische Grundlage für READY geschaffen werden. Der bisherige Stand läßt aber noch Fragen offen:

(1) In der Feuerwehrdomäne treten Beschränkungen des Arbeitsgedächtnisses besonders stark auf. Sind die Ergebnisse der Feuerwehrstudie auch auf weniger extreme Situationen übertragbar?

(2) Ein Notrufdialog wird fast immer mit Zeitdruck geführt. Die erste Studie ergab in dieser Hinsicht keine konkreten Kausalzusammenhänge. Welche Hinweise auf Zeitdruck treten in Dialogen auf?

(3) Welche Ursache-Wirkungs-Beziehungen können im READY-Szenario zusätzlich berücksichtigt werden, die für die Aussagen über die Arbeitsgedächtnisbelastung und den Zeitdruck des Benutzers relevant sind?

Eine zweite Studie mit ähnlicher Methodologie wie die der Feuerwehrstudie soll Antworten auf diese Fragen liefern. Eine erste Auswertung wurde von Weyrath (1997) vorgenommen. Aufbauend auf diesen Ergebnissen wird im nächsten Kapitel das Benutzermodell von READY entwickelt.

5.1 Fragestellung

In dem READY-Szenario (siehe Kapitel 1) spielen die Beschränkungen des Arbeitsgedächtnisses und der Zeit von $\mathcal{P}$ eine wichtige Rolle. Einerseits muß sich $\mathcal{P}$ auf die Instruktionen und Fragen von $\mathcal{S}$ konzentrieren und andererseits die notwendigen Handlungen ausführen. $\mathcal{S}$ hat das Ziel seine Instruktionen und Frage so zu stellen, daß $\mathcal{P}$ so wenig wie möglich belastet wird. Zusätzlich wird $\mathcal{S}$ auch $\mathcal{P}$s Zeitdruck berücksichtigen wollen. Dabei tritt folgende Frage unwillkürlich auf: „Wie macht $\mathcal{S}$ das?“ Oder genauer:

Tabelle 5.1: *Transliteration eines Pannenhilfedialogs*

1	$\mathcal{S}$:	**Hodab.**
2	$\mathcal{P}$:	Ja ich habe vier Minuten Zeit, das Auto wieder flott zu kriegen.
3	$\mathcal{S}$:	**Aha.**
4	$\mathcal{P}$:	OK, also <Z>, ich bin gefahr' und dann hat's <P> angefang' äh zu ruckeln <P> und <Z> ich kann soviel Gas geb'n wie ich will, es wird nicht schneller, es Auto. Und jetzt?
5	$\mathcal{S}$:	**Aber er läuft noch?**
6	$\mathcal{P}$:	Ach so ja %und %es %geht ab und zu <%> aus. Ne?
7	$\mathcal{S}$:	**Ab und zu geht er aus? <P>**
8	$\mathcal{P}$:	Ja.
9	$\mathcal{S}$:	**Motorhaube schon geöffnet?**
10	$\mathcal{P}$:	Nee, Motorhaube hab' ich noch nicht geöffnet.
11	$\mathcal{S}$:	**Äh <Z>. Im <Z> Fußraum, daß heißt <P> unterm Lenkrad <P> links <P> is'n Hebel. <P> daran ziehen.**
12	$\mathcal{P}$:	Moment %grad %bitte <#Knistern> <P> OK <P> jetzt.
13	$\mathcal{S}$:	**Jetzt Motorhaube öffnen?**
14	$\mathcal{P}$:	<P> Wie rum <Ächz> %geht'n der auf? So rum? [...]

Anmerkung: $\mathcal{P}$ = Anrufer, $\mathcal{S}$ = Mechaniker, <P> = Pause, <Z> = Zögern, <#> = nichtartikulartorisches Geräusch, % = Schwerverständliches, <%> = Unverständliches.

Wie erkennt man, daß das Arbeitsgedächtnis eines Dialogpartners stärker als normal belastet ist bzw. daß er unter Zeitdruck steht?

Die Untersuchung wurde mit Studenten durchgeführt, die über kein psychologisches Fachwissen verfügen. Da solche Ressourcenbeschränkungen in alltäglichen Dialogen sehr häufig vorkommen, wurde angenommen, daß auch dieser Personenkreis gute Hypothesen beitragen können.

5.1.1 Vorstudie

Ausgangspunkt für die eigene empirische Studie bildet eine Vorstudie, die die verwendeten Daten erhoben hat. In dieser Vorstudie, die von Roger Knopp und Dr. Anthony Jameson vorbereitet und ausgeführt wurde, wurde das READY-Beispielszenario so genau wie möglich abends an einem Straßenrand nachgestellt. An zwei Autos wurden abwechselnd echte Störungen verursacht, die von den neun vorher geworbenen Autofahrer, die als Versuchspersonen dienten, behoben werden sollten. Die Rolle von READY wurde von einem Kraftfahrzeugmechaniker ($\mathcal{S}$) übernommen, der über Mobiltelefon den Versuchspersonen ($\mathcal{P}$) Pannenhilfe erteilte. Das Ergebnis der Vorstudie waren neun Autopannendialoge. Einer davon ist in der Tabelle 5.1 abgebildet. Bei einem der neun Dialogen ist die Aufnahme aus technischen Gründen nicht brauchbar. Die Versuchspersonen wurden mit 20 DM für ihre Teilnahme an der Untersuchung belohnt.

In vier der neun Dialogen wurde zusätzlich durch das Angebot einer Extrabelohnung (20 DM) für besonders schnelle Behebung der Störung Zeitdruck erzeugt.

Ausschnitte dieser Dialoge wurden in der vorliegenden Studie einer anderen Gruppe von Versuchspersonen vorgespielt. Diese hatten die Aufgabe einzuschätzen, inwieweit die Dialogteilnehmer den Beschränkungen von Zeitdruck und Arbeitsgedächtnisbelastung unterliegen. Diese Studie wird nun im folgenden näher beschrieben.

5.2 Methode

5.2.1 Versuchspersonen

An der empirischen Untersuchung nahmen 12 Studenten der Universität des Saarlandes teil.[1]

5.2.2 Material

Aus fünf der aufgezeichneten Dialoge der Vorstudie wurden Teile entnommen und zusammengeschnitten.[2] Damit die Versuchspersonen nicht zu monotonen Einschätzungen verleitet werden, wurde das Band so geschnitten, daß abwechselnd Dialoge mit vielen Ressourcenbeschränkungen und solche mit wenigen Ressourcenbeschränkungen abgespielt werden konnten. Die Gesamtlänge der Ausschnitte betrug 20 Minuten. Für die Studie wurden zwei Tonbandgeräte benutzt, eines zum Abspielen der Pannendialoge und ein weiteres zum Aufnehmen des Gesprächs.

5.2.3 Durchführung

Jede Versuchsperson nahm an einem Gespräch von ca. 50 Minuten teil. Die Durchführung ist ähnlich die der Feuerwehrstudie. Zu Beginn des Gesprächs bekam jede Versuchsperson eine Instruktionsanleitung vorgelegt, in der das Szenario und das Ziel der Untersuchung beschrieben wurde. Da davon ausgegangen wurde, daß den Versuchspersonen der Begriff Arbeitsgedächtnisbelastung nicht geläufig ist, wurde er in der Instruktionsanleitung mit *zeitweiser Überforderung* umschrieben und anhand eines Beispiels illustriert. So wurde sichergestellt, daß die Versuchspersonen, auch wenn sie den Begriff nicht kannten, sich eine möglichst genaue Vorstellung davon machen konnten.

Dann bekam jede Versuchsperson das Band mit den o.g. Dialogausschnitten vorgespielt. An bestimmten Stellen wurde das Band gestoppt, und die Versuchsperson wurde um eine kurze und spontane Einschätzung gebeten. Danach wurde jede Versuchsperson aufgefordert, durch retrospektives lautes Denken ihre Gedächtnisinhalte zu verbalisieren.

Zwei Typen von Einschätzungen standen in der Untersuchung im Vordergrund:

[1] Ich bedanke mich bei den 12 Studenten, die an der empirischen Untersuchung teilgenommen haben.

[2] Der transliterierte Dialog mit den Fragen ist in Anhang A.3 abgedruckt.

(1) Fragen nach dem Ausmaß des Zeitdrucks, in dem der Autopannendialog stattfindet: „Wie hoch ist der Grad der Eile des Anrufers/Mechanikers?"
(2) Fragen nach der Einschätzung der Arbeitsgedächtnisbelastung: „Inwieweit wird der Anrufer durch die Situation/Frage/Äußerung des Mechanikers beansprucht?".

Die Bänder mit den Gesprächen wurden transliteriert und daraufhin untersucht, wie die Versuchspersonen die Ursache-Wirkungs-Beziehungen bzgl. Arbeitsgedächtnisbelastung und Zeitdruck gesehen haben.

5.3 Ergebnisse

5.3.1 Erste Beobachtung

Ähnlich wie in der Feuerwehrstudie erkannten die Versuchspersonen Hinweise (Ursachen und Symptome) auf Arbeitsgedächtnisbelastung, machten sich aber kein differenziertes Bild vom Arbeitsgedächtnis der Dialogteilnehmer. Des weiteren drückten fast alle Aussagen Unsicherheit und Vagheit aus. Typische Äußerungen waren z.B. die folgenden:

„Es kann *vielleicht* daran liegen"
„Er hat es *wahrscheinlich* nicht eilig".
„Sie ist *ziemlich* überfordert"
„Also *relativ* beansprucht"

Die Hinweise im Verhalten der Dialogteilnehmer erlauben daher kaum sichere Schlußfolgerungen über die die Arbeitsgedächtnisbelastung oder den Zeitdruck der Dialogteilnehmer. Aus diesem Grund werden, wie schon in der Feuerwehrstudie, die betreffenden Kausalzusammenhänge nicht als deterministisch sondern eher als probabilistisch betrachtet.

Im Unterschied zur Feuerwehrstudie nannten die Versuchspersonen auch viele Zusammenhänge, die sich auf die sprachlichen Merkmale der Äußerungen der Dialogteilnehmer bezogen. In der Feuerwehrstudie dominierten überwiegend situationsabhängige Merkmale wie Aufregung oder Alkoholeinfluß. Diese Tatsache ist durch die extreme Situation zu erklären, in der ein Notrufdialog stattfindet. Im folgenden werden wieder die Zusammenhänge getrennt nach Ursachen und Folgen beschrieben (siehe Abb. 5.1).

5.3.2 Ursachen hoher Arbeitsgedächtnisbelastung

Zu den Ursachen von Arbeitsgedächtnisbelastung seitens des Hörers gehören laut Versuchspersonen syntaktische und semantische Merkmale der Äußerungen des Sprechers.

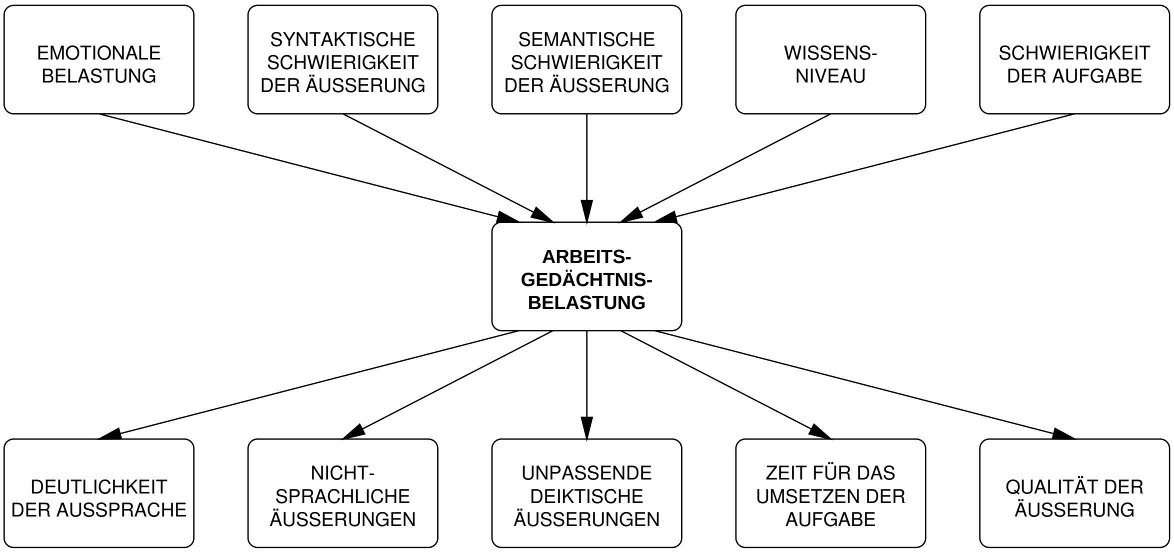

Abb. 5.1: *Rekonstruktion der Hypothesen der Versuchspersonen über Ursachen und Folgen hoher Arbeitsgedächtnisbelastung. (Ein Pfeil von A nach B bedeutet, daß A B beeinflußt).*

Syntaktische Schwierigkeit. Ein syntaktisches Merkmal ist z.B. die *Syntaktische Schwierigkeit der Äußerung*. Diese Variable beschreibt die Bewertung der Schwierigkeit, die Äußerung syntaktisch zu verstehen. Darunter fallen z.B. Aspekte wie Äußerungslänge, Informationsdichte und die Benutzung tiefverschachtelter Nebensätze. Der Satz „Jetzt die Motorhaube öffnen!“ ist syntaktisch einfacher als z.B. „Sie können jetzt ohne Probleme durch das Ziehen am vorderen Hebel im Fußraum ihres Wagens die Motorhaube öffnen“.

Der Hörer muß bei syntaktisch schwierigen Äußerungen mehr Inferenzen ziehen und mehr Informationen im Arbeitsgedächtnis behalten (siehe Abschnitt 2.1.3.2). Einige Versuchspersonen empfanden manche Instruktionen des Mechanikers $\mathcal{S}$ syntaktisch zu schwer und nannten diese Schwierigkeit als Ursache von Arbeitsgedächtnisbelastung bei $\mathcal{P}$.

Die Abbildung 5.2 zeigt, wie häufig der Zusammenhang genannt wurde: 6 Versuchspersonen nannten diesen Zusammenhang, davon 3 Versuchspersonen einmal, 2 Versuchspersonen zweimal und eine Versuchsperson äußerte sich sogar 7 mal über diesen Zusammenhang. Beispiele von Äußerungen der Versuchspersonen sind:

Frage 5.4: [3] Wie hoch ist der Grad der Eile des Mechanikers?

Vp 6: *Anscheinend ziemlich hoch.* – retrospektives Denken – *Ich habe halt gemerkt, daß er doch ziemlich überfordernd geredet hat. Er hat halt eine Information nach der anderen gebracht [...].*

Frage 3.9: Inwieweit wird die Anruferin durch die Äußerung beansprucht?

Vp 2: *Jetzt stark.* – retrospektives Denken – *Ungenauigkeit des Mechanikers. Er hat sich eben öfters korrigiert [...]*

[3] Anmerkung: Frage 5.4 bedeutet 4. Frage im 5. Dialog.

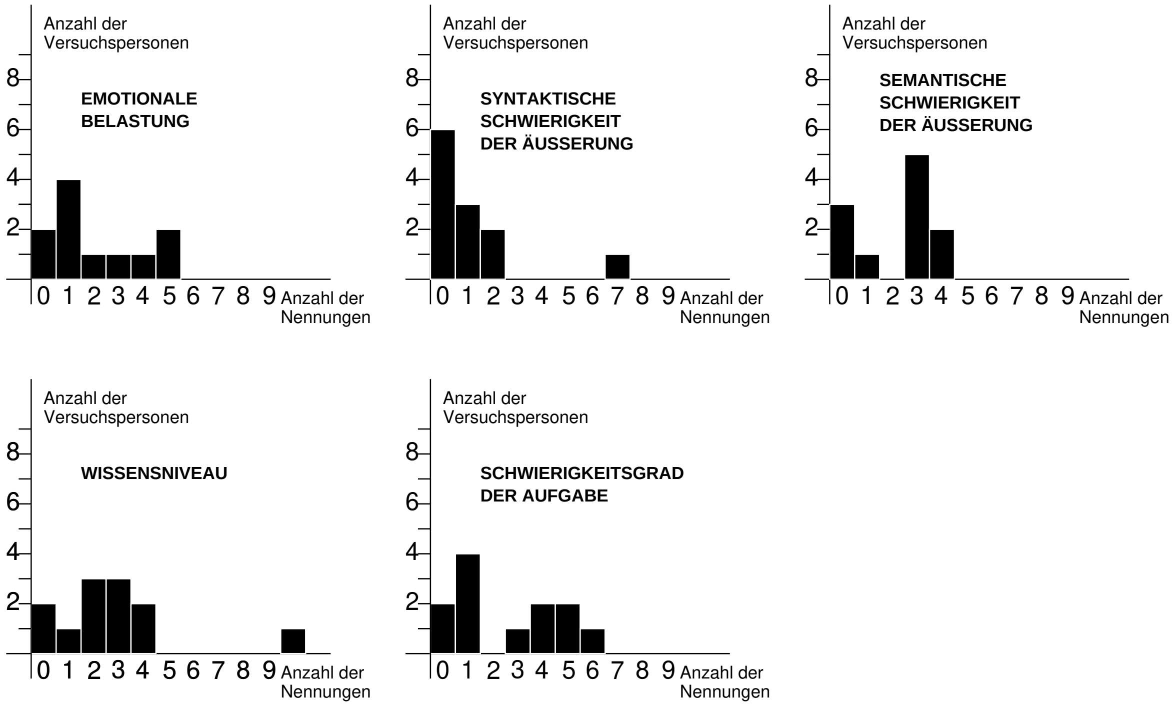

Abb. 5.2: *Häufigkeiten der Nennungen der Zusammenhänge bzgl. Ursachen hoher Arbeitsgedächtnisbelastung. Die Diagramme geben an, wie oft ein Hinweis auf einen Zusammenhang geben wurde, der durch die Variablen beschrieben wird.*

Semantische Schwierigkeit. Häufiger als die syntaktische Schwierigkeit wurde von den Versuchspersonen die *semantische Schwierigkeit der Äußerung* genannt. Dieser Einfluß auf die Arbeitsgedächtnisbelastung beschreibt die Schwierigkeit, eine Äußerung in ihrer Bedeutung zu verstehen. Die semantische Schwierigkeit der Äußerung erhöht sich z.B. durch die Verwendung von Fachbegriffen, die der Hörer nicht kennt oder von denen er nur ein vages Bild hat. Kintsch und Vipond (1979) erklären diesen Zusammenhang durch eine höhere propositionale Dichte neuer oder fremder Konzepte. Die semantische Schwierigkeit der Äußerung wurde von neun Versuchspersonen als Ursache hoher Arbeitsgedächtnisbelastung gesehen. Die meisten von ihnen nannten den Zusammenhang sogar drei- oder viermal (siehe Abb. 5.2). Die Häufigkeit der Nennung deutet auf ein starkes Indiz für Arbeitsgedächtnisbelastung hin.

> Frage 5.3: Inwieweit wird der Anrufer durch die Äußerung beansprucht?
>
> Vp 2: *Er wird schon etwas mehr beansprucht, weil so Begriffe wie Lichtmaschine, da weiß nicht unbedingt jeder, was er damit anfangen soll.* – retrospektives Denken – *Ich habe mir versucht vorzustellen, ja, eben vorne die Kühlerhaube und die ganzen Instrumente.*
>
> Frage 3.7: Inwieweit wird die Anruferin durch die Äußerung beansprucht?
>
> Vp 1: *Es ist anscheinend schwierig für die Anruferin, Fachausdrücke wie Kotflügel. Wenn jemand keine Ahnung vom Auto hat, wird er es nicht verstehen.*

Schwierigkeitsgrad der Aufgabe und Wissensniveau. Als weitere Ursachen von Arbeitsgedächtnisbelastung wurde von den Versuchspersonen der *Schwierigkeitsgrad der Aufgabe* im Verhältnis zum *Wissensniveau* genannt. Die Instruktionen von $\mathcal{S}$ müssen von $\mathcal{P}$ ausgeführt werden. Hat $\mathcal{P}$ ein zu geringes Wissensniveau bzgl. der Reparatur von Autos oder ist der Schwierigkeitsgrad der Aufgabe zu hoch, ist laut Versuchspersonen mit einer hohen Arbeitsgedächtnisbelastung zu rechnen. Die beiden dazugehörigen Diagramme in Abbildung 5.2 zeigen, daß diese Variablen von fast allen Versuchspersonen als Ursache hoher Arbeitsgedächtnisbelastung empfunden wurde.

Frage 3.5: Inwieweit wird die Anruferin durch die Situation beansprucht?

Vp 4: *Man hat schon gemerkt, daß sie von Autos und so technischen Sachen keine Ahnung hat, weil sie nicht wußte, wie man eine Motorhaube öffnet, was man wahrscheinlich eher einem Mann zutrauen würde. Von daher, ein bißchen beansprucht halt.*

Frage 5.5: Inwieweit wird der Anrufer durch die Äußerung beansprucht?

Vp 10: *Er wird doch einigermaßen beansprucht.* – retrospektives Denken – *Weil er hat zurückgefragt, wo die Lichtmaschine ist und insofern scheint er sich nicht so auszukennen, wo was zu finden ist. Er muß das jetzt erst suchen. Also er ist höher beansprucht.*

Emotionale Belastung. Seitens des Sprechers wurde von den Versuchspersonen die *Emotionale Belastung* von $\mathcal{P}$ als Ursache von Arbeitsgedächtnisbelastung gesehen. Diese Variable steht für äußere Einflüsse, die $\mathcal{P}$s Arbeitsgedächtnis unabhängig vom Dialog belasten können. Darunter fallen z.B. Aspekte wie Angst oder Aufregung (siehe z.B. Calvo & Eysenck, 1996). Auch dieser Einfluß wurde sehr häufig (von 10 Versuchspersonen) genannt (siehe Abb. 5.2).

Frage 3.2: Wie schätzt du den Grad der Eile der Anruferin ein?

Vp 1: *Klingt etwas aufgeregt und anscheinend doch ein bißchen in Panik.* – retrospektives Denken – *Es kann eine gefährliche Situation ergeben, wenn der Wagen immer schneller wird.*

Frage 2.2: Inwieweit war der Anrufer durch die Beschreibung der Situation beansprucht?

Vp 7: *Er war relativ stark beansprucht, denn er scheint sehr unruhig zu sein.*

5.3.3 Folgen hoher Arbeitsgedächtnisbelastung

Zu den Folgen von Arbeitsgedächtnisbelastung nannten die Versuchspersonen einerseits Symptome in bezug auf den Inhalt sprachlicher Äußerungen und andererseits Symptome, die sich auf die sprachlichen Merkmale der Äußerungen beziehen.

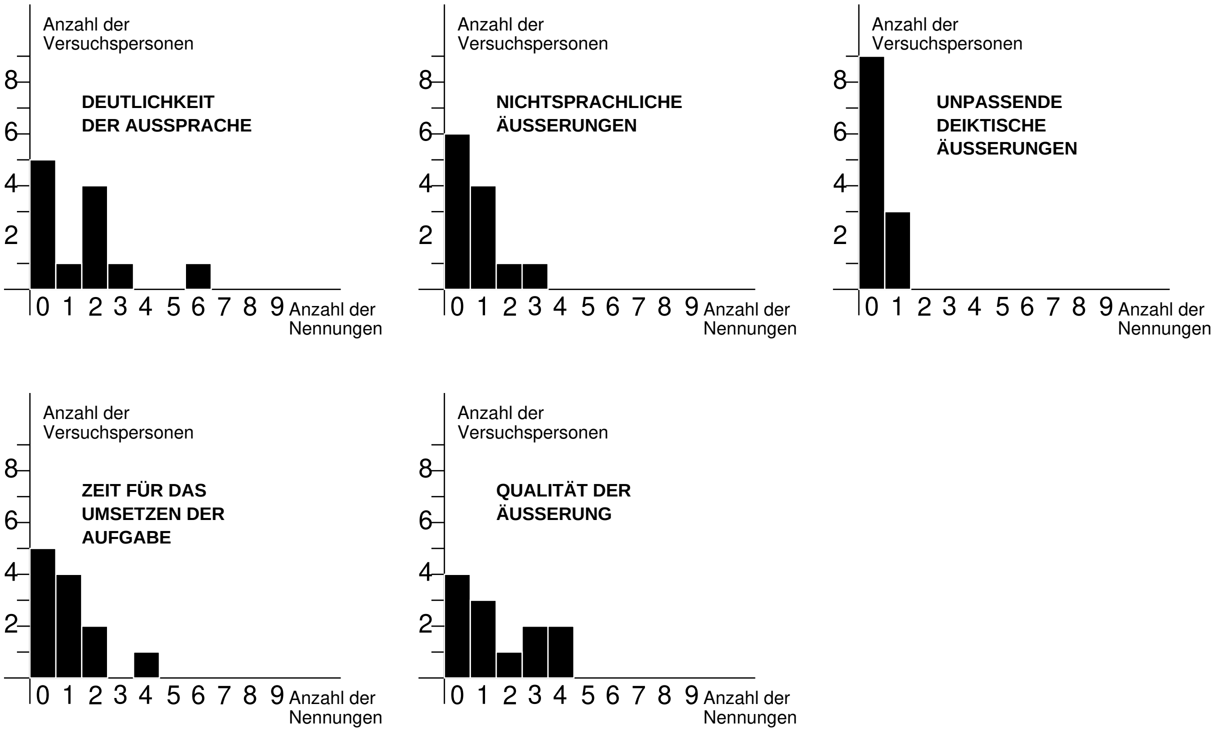

Abb. 5.3: *Häufigkeiten der Nennungen der Zusammenhänge bzgl. Folgen hoher Arbeitsgedächtnisbelastung. Die Diagramme geben an, wie oft ein Hinweis auf einen Zusammenhang geben wurde, der durch die Variablen beschrieben wird.*

Unpassende deiktische Äußerungen. Zu der ersten Kategorie gehören z.B. die Verwendung *unpassender deiktischer Äußerungen*. Damit sind Äußerungen gemeint, die räumliche Beschreibungen aus der Sicht des Sprechers darstellen und im Dialogkontext schwer oder gar nicht verstanden werden können. Drei Versuchspersonen (siehe Abb. 5.3) fiel die Verwendung einer unpassenden deiktischen Äußerung im Dialog 3 (siehe Tabelle 5.1, Zeile 14) auf. Die Frau fragte den Mechaniker, ob die Motorhaube „so oder so 'rum“ aufgeht, was der Mechaniker am Telefon natürlich nicht sehen konnte. Daß dieser Zusammenhang nicht von mehr Versuchspersonen genannt wurde, hängt vielleicht daran, daß die Äußerung nur sehr schwer im Dialog zu hören war.

> Frage 3.5: Inwieweit wird die Anruferin durch die Situation beansprucht?
>
> Vp 4: *Das hat man halt daran gemerkt, daß sie dann sagt, so rum oder so. Ich meine das kann er nicht sehen, am Telefon. Schon ein bißchen durcheinander. Also relativ beansprucht.*
>
> Vp 6: *Ich glaube ziemlich* – retrospektives Denken – *Sie fragt „so rum“, was eigentlich unlogisch ist am Handy [...] sie ist ziemlich überfordert.*

Qualität der Äußerung. Als weiteres Symptom für hohe Arbeitsgedächtnisbelastung bzgl. des Inhaltes der sprachlichen Äußerungen wurde von den Versuchspersonen die *Qua-*

lität der Äußerung genannt. Mit anderen Worten: eine schlechte Qualität der Äußerung wurde von den Versuchspersonen als Folge von Arbeitsgedächtnisbelastung gesehen. Die Versuchspersonen verbindeten eine schlechte Qualität mit unpräzisen oder schlecht formulierten Äußerungen. Die Abbildung 5.3 zeigt in der Hinsicht eine hohe Übereinstimmung der Versuchspersonen in dieser Sache deutlich.

> Frage 2.2: Inwieweit war der Anrufer durch die Beschreibung der Situation beansprucht?
>
> Vp 8: *Klack, klack, klack zu sagen, und irgendwas und irgendwie sind Umschreibungen, die so ungenau sind, daß die Beanspruchung wohl so hoch ist, um die genauen Worte zu finden.*

Nichtsprachliche Äußerungen. Zu den Merkmalen sprachlicher Äußerungen, die als Folgen hoher Arbeitsgedächtnisbelastung von den Versuchspersonen gesehen wurden, gehören *nichtsprachliche Äußerungen*, wie z.B. Lachen, Stöhnen oder Äh's. Füllphrasen, wie z.B. „äh", geben dem Sprecher mehr Zeit, seine sprachlichen Äußerungen dem Planungsaufwand anzupassen (vgl. Abschnitt 2.1.3). Lachen wurde von den Versuchspersonen so aufgefaßt, daß $\mathcal{P}$ nicht mit den gestellten Anweisungen zurechtkommt, und Stöhnen wird i.a. als Zeichen hoher Belastung gewertet.

> Frage 3.9: Inwieweit wird die Anruferin durch die Situation beansprucht?
>
> Vp 1: *[...] weil sie lacht, ein Zeichen, daß sie das nicht richtig macht.*
>
> Frage 3.7: Inwieweit wird die Anruferin durch die Äußerung beansprucht?
>
> Vp 1: *Ziemlich.* – retrospektives Denken – *Sie sucht die ganze Zeit und wird auch nervös und macht auch solche mißmutige Laute und deshalb ist sie ziemlich überfordert in dem Moment.*

Deutlichkeit der Äußerung. Neben der Qualität der Äußerung berücksichtigten die Versuchspersonen auch die *Deutlichkeit der Aussprache.* Eine schlechte Aussprache wird oft als Reduzierung der sprachlichen Verarbeitung gesehen, was auf eine hohe Arbeitsgedächtnisbelastung hindeutet. Mehr als die Hälfte der Versuchspersonen bemerkten einen derartigen Zusammenhang (siehe Abb. 5.3).

> Frage 4.1: Inwieweit war der Anrufer durch die Beschreibung der Situation beansprucht?
>
> Vp 4: *Ich denke, der redet vielleicht auch sonst normal. Eher ziemlich stark beansprucht.* – retrospektives Denken – *Daß er auch deshalb undeutlich gesprochen hat [...]*

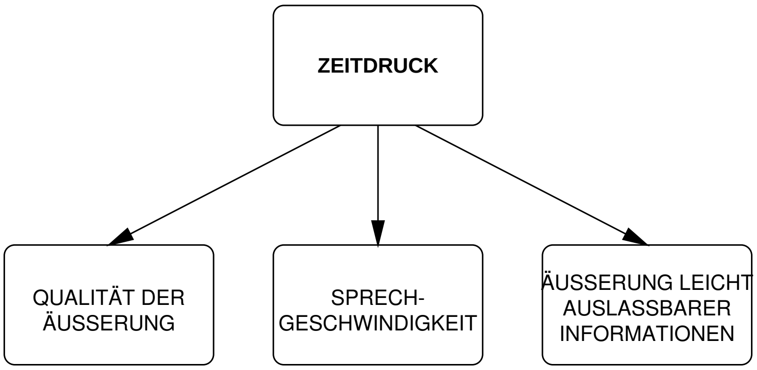

Abb. 5.4: *Rekonstruktion der Hypothesen der Versuchspersonen über Folgen von Zeitdruck. (Ein Pfeil von A nach B bedeutet, daß A B beeinflußt).*

Zeit für das Umsetzen der Aufgabe. Neben den sprachlichen Merkmalen der Äußerungen wurden auch situative Aspekte von den Versuchspersonen berücksichtigt. Ein Beispiel dafür ist die Aufrechthaltung der Kommunikation auch während der Reparatur. Da sich der Mechaniker $\mathcal{S}$ und der Anrufer $\mathcal{P}$ nur über Mobiltelefon verständigen, wird eine lange *Zeit für das Umsetzen der Aufgabe* zusammen mit dem Fehlen von Rückmeldungen des Dialogpartners von einigen Versuchspersonen als nachteilig empfunden und der hohen Arbeitsgedächtnisbelastung zugeschoben.

> Frage 5.3: Inwieweit wird der Anrufer durch die Äußerung beansprucht?
>
> Vp 6: *Ziemlich.* – retrospektives Denken – *Das er auch lange nichts sagt. Also wenn sich das auf die Frage davor bezieht. Daß er die ganze Zeit sucht und überlegt.*

Der nächste Abschnitt beschreibt die von den Versuchspersonen genannten Folgen von Zeitdruck.

5.3.4 Folgen von Zeitdruck

Während der Instruktion bekamen die Versuchspersonen zusätzlich die Information, daß sich einige Anrufer aufgrund eines bevorstehenden Termins in Zeitdruck befinden können. Durch die Frage „Wie hoch ist der Grad der Eile des Anrufers/Mechanikers?“ sollten von den Versuchspersonen Einschätzungen bzgl. des Zeitdrucks herausgelockt werden. Die Versuchspersonen nannten daraufhin Symptome in bezug auf den Inhalt sprachlicher Äußerungen und Merkmale des Sprechens (siehe Abb. 5.4)

Qualität der Äußerung. Die *Qualität der Äußerung* wurde von den Versuchspersonen nicht nur als Folge hoher Arbeitsgedächtnisbelastung sondern auch als Folge von Zeitdruck gesehen. Eine ausführliche Beschreibung kann somit als Symptom betrachtet werden, das $\mathcal{P}$ keine Eile hat. Acht Versuchspersonen nannten diesen Zusammenhang (siehe Abb. 5.5).

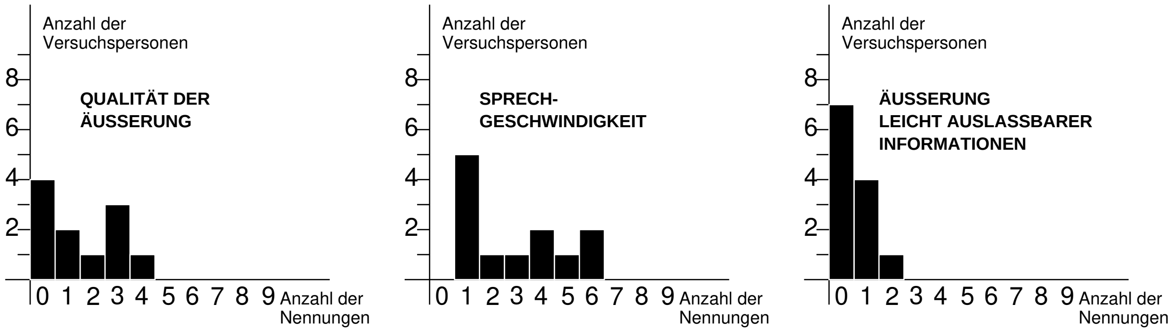

Abb. 5.5: *Häufigkeiten der Nennungen der Zusammenhänge bzgl. Folgen von Zeitdruck. Die Diagramme geben an, wie oft ein Hinweis auf einen Zusammenhang geben wurde, der durch die Variablen beschrieben wird.*

Frage 3.8: Wie hoch ist der Grad der Eile des Mechanikers?

Vp 2: *Gering.* – retrospektives Denken – *Der läßt sich Zeit. Er beschreibt das alles auch ausführlich.*

Äußerung leicht auslaßbarer Informationen. Ähnlich wie die Qualität der Äußerung wurde von den Versuchspersonen die *Äußerung leicht auslaßbarer Informationen* gewertet. Die Beschreibung vieler Einzelheiten, die nicht direkt mit der Autopanne zu tun haben, wird mit wenig Zeitdruck assoziiert.

Frage 5.1: Wie hoch ist der Grad der Eile des Anrufers?

Vp 8: *Mäßig.* – retrospektives Denken – *Er beschreibt es relativ oberflächlich. „Das Radio ging aus" also läßt er sich Zeit für so Kleinigkeiten. Eile nicht zu groß.*

Sprechgeschwindigkeit. Als starkes Indiz für eine Folge von Zeitdruck wird die *Sprechgeschwindigkeit* gewertet. Obwohl sie bei verschiedenen Personen unterschiedlich ist, wurde von den Versuchspersonen eine hohe Sprechgeschwindigkeit als Folge von Zeitdruck gewertet. Dieses Symptom wurde von allen Versuchspersonen genannt (siehe Abb. 5.5).

Frage 4.2: Wie hoch ist der Grad der Eile des Anrufers?

Vp 3: *Ziemlich.* – retrospektives Denken – *Er hat ziemlich schnell gesprochen und daraus schließe ich, daß er eben in Eile war.*

Frage 1.2: Wie hoch ist der Grad der Eile des Anrufers?

Vp 10: *Er hat Zeit, er spricht langsam.* – retrospektives Denken – *Das er halt langsam spricht und genau beschreibt, was er für ein Problem hat.*

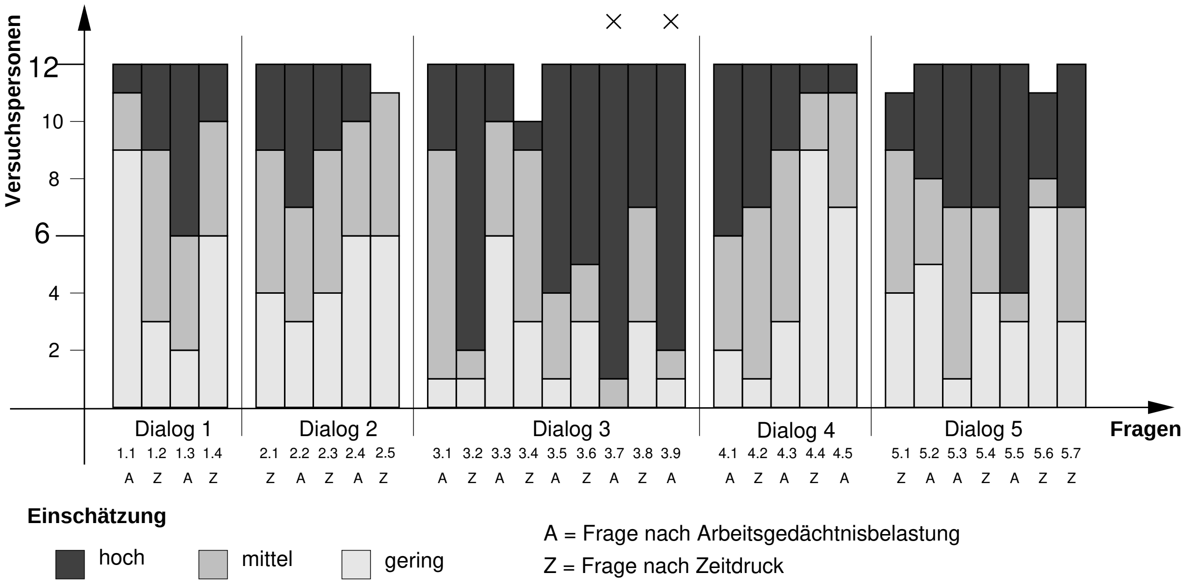

Abb. 5.6: *Häufigkeiten der Einschätzungen pro Frage. Auf der horizontale Achse sind die Fragen bzgl. der einzelnen Dialoge angeordnet, die vertikale Achse die Anzahl der Versuchspersonen. Die einzelnen Balkenhöhen spiegeln die Häufigkeiten der Einschätzungen wieder.*

5.3.5 Übereinstimmungen der Versuchspersonen

Neben den qualitativen Zusammenhängen, die die Versuchspersonen nannten, wurde auch untersucht, inwieweit die Versuchspersonen mit ihren Einschätzungen übereinstimmten.

Wie bereits beschrieben wurde, gab es nur zwei Fragetypen in der empirischen Untersuchung: Fragen nach der Arbeitsgedächtnisbelastung und nach dem Zeitdruck des Anrufers. Die beiden Fragen wurden so gestellt, daß sie von den Versuchspersonen eine Einschätzung über den Grad der jeweiligen Ressourcenbeschränkung entlocken sollten. Die Antworten der Versuchspersonen wurden dann eine der Kategorien *gering*, *mittel* und *hoch* zugeordnet. Die Unsicherheit und Vagheit in den Antworten der Versuchspersonen (z.B. durch *vielleicht* oder *relativ*) wurden nicht berücksichtigt.

Die Abbildung 5.6 zeigt, wie die Versuchspersonen mit ihren Einschätzungen übereinstimmten. Man kann sehen, daß die Versuchspersonen sehr oft unterschiedlicher Meinung über den Grad der Belastung des Anrufers waren. Diese Tatsache unterstreicht den unsicheren Charakter solcher Einschätzungen. Dennoch fallen einige Stellen auf, in denen die Versuchspersonen im hohen Maße in ihrer Antwort übereinstimmten. Z.B. im Dialog 3, Fragen 3.7 und 3.9 (siehe Abb. 5.6, Spalten mit Kreuz), meinten fast alle Versuchspersonen das die Arbeitsgedächtnisbelastung der Anruferin hoch ist. Die Tabelle 5.2 gibt die zugehörige Szene und einige typische Aussagen der Versuchspersonen wieder, die aufgrund dieser Fragen gegeben wurden.

Tabelle 5.2: *Übereinstimmung der Versuchspersonen*

Dialog 3:

$\mathcal{S}$**: Ai, in der Mitte der Haube <oah> <!Stöhnen> zwischen Haube und Kühlergrill durchgreifen.** $\mathcal{P}$: Ja ich hab's. $\mathcal{S}$**: Da ... Da ... ja?** $\mathcal{P}$: <%> <P> wie geht'n ... wie hält denn das Ding <Ächz>? $\mathcal{S}$**: Ach so, äh, <Z> <P> äh auf der rechten Seite liegt <Z> glaube ich <Z> ein Stab, auf der rechten Seite. Am ... äh, äh auf'm Kotflügel, so zu sagen.**

Frage 3.7: *Inwieweit war die Anruferin durch die Äußerung beansprucht?*

Vp 3: *mittel bis stark.* – retrospektives Denken – *Also ich denke, daß sie sich schon ziemlich darauf konzentrieren muß, um das zu finden und zu verstehen, was der Mechaniker meint.*

Vp 8: *Immer höher.* – retrospektives Denken – *Weil es ihr so langsam unangenehm wird, weil sie überhaupt nicht weiß, was zu machen ist.*

Vp 10: *Die Anruferin ist doch recht beansprucht, weil sie scheint sich mit Autos ganz und gar nicht auszukennen und kann noch nicht einmal eine Motorhaube öffnen.*

Vp 11: *Sehr arg.* – retrospektives Denken – *Weil das gar nicht auf dem Kotflügel liegt, die Stange, sondern die liegt ja vorne drauf. Kotflügel ist ja seitlich. Das war ein Fehler von dem.*

$\mathcal{P}$: <%> <;T> $\mathcal{S}$**: Bitte? <P> Ja?** $\mathcal{P}$: <;T> <%> Ah, oh Scheiße. Oh ihr Leut' <Lachen>. $\mathcal{S}$**: <lachen>** $\mathcal{P}$: <Lachen> <%> Ai, wo denn <Lachen>, ja ok, ja jetzt. $\mathcal{S}$**: Ja?** $\mathcal{P}$: Ja <P> und jetzt? $\mathcal{S}$**: Äh, <Z> <P> in der Mitte <Z> ist <Z> ein großes rundes Ding, <P> so'n Luftfilter <P>. Ah, nee, Moment, <P> Stop. <P> Äh, weiter rechts ist eine orangene Kappe.**

Frage 3.9: *Inwieweit war die Anruferin durch die Äußerung beansprucht?*

Vp 2: *Jetzt stark.* – retrospektives Denken – *Ungenauigkeit des Mechaniker. Er hat sich eben öfters korrigiert und wenn man ohnehin nicht so viel Ahnung hat von einer Sache, dann ist es schlecht für das Verständnis, wenn man jemanden vor sich hat der einem zum Teil etwas falsches am Anfang erzählt.*

Vp 3: *Ja, sie wird beansprucht.* – retrospektives Denken – *Daß also die Beanspruchung von daher kommt, weil sie sich erstmal in diesem ungewohnten Motorraum zurechtfinden muß und dann dadurch, daß er sich nochmal verbessert hat, daß er dadurch ein bißchen unsicher war.*

Vp 7: *Sie wird relativ stark beansprucht, das zeigt sich an ihrem nervösen lachen.* – retrospektives Denken – *Sie scheint mit dieser Situation nicht ganz fertig zu werden, weicht also auf dieses Lachen aus.*

Vp 8: *Auch wieder höher als zuvor.* – retrospektives Denken – *Das nervöse Lachen und das Stottern und all das zeigt, daß die Nervosität so langsam gestiegen ist, weil einfach die Kommunikation nicht so funktioniert.*

5.4 Diskussion

Die vorliegende Studie sollte untersuchen, wie Menschen erkennen, daß das Arbeitsgedächtnis eines Dialogpartners stärker als normal belastet ist bzw. er unter Zeitdruck steht. Die Ergebnisse lassen sich wie folgt zusammenfassen:

(1) Wie schon in der Feuerwehrstudie, machten sich die Versuchspersonen kein differenziertes Bild des Arbeitsgedächtnisses der Dialogteilnehmer.

(2) Fast alle Versuchspersonen waren sich über die Ressourcenbeschränkungen der Dialogteilnehmer unsicher und machten daher nur vage Aussagen.

(3) Einige der Zusammenhänge zwischen Arbeitsgedächtnis bzw. Zeitdruck und Sprachverarbeitung wurden von mehreren Versuchspersonen übereinstimmend genannt. Einige Zusammenhänge stimmen mit Ergebnissen aus früheren Untersuchungen (siehe Abschnitt 2.1) überein.

(4) Die genannten Kausalzusammenhängen bzgl. der Arbeitsgedächtnisbelastung und des Zeitdrucks eines Dialogpartners sind zum Teil mit denen der Notrufstudie vergleichbar.

Im nächsten Kapitel wird ein Benutzermodell entwickelt, das auf den beschriebenen Kausalzusammenhängen basiert.

Kapitel 6

Benutzermodellierung in dem System READY

In diesem Kapitel wird beschrieben, wie READY $\mathcal{P}$s Arbeitsgedächtnisbelastung und Zeitdruck erkennen und einschätzen kann. Anhand dieser Einschätzung kann READY seine eigenen Dialogäußerungen der aktuelle Ressourcenlage von $\mathcal{P}$ anpassen.[1] Für den Aufbau des Benutzermodells werden wie schon in der Feuerwehrdomäne (siehe Kapitel 4) dynamische Bayessche Netze verwendet, die mit Hilfe von Modellierungsschemata (siehe Abschnitt 4.2) aufgebaut werden. Die Modellierung baut auf den Ergebnissen der empirischen Studie auf, die im Abschnitt 5 beschrieben wurde.

Am Schluß dieses Kapitels wird anhand eines Beispieldialogs illustriert, wie READY die Ressourcenbeschränkungen seines Dialogpartners einschätzen kann.

6.1 Modellierung der Ressourcenbeschränkungen

Wie bereits in der Notrufdomäne wird auch in READY das Arbeitsgedächtnis (AG) als ein einziger Speicher mit einer gewissen verfügbaren Kapazität aufgefaßt. Feinkörnigere Modelle, wie sie in der kognitiven Psychologie, z.B. von Baddeley (1986), ausgearbeitet wurden (siehe Kapitel 2.1), werden in der ersten Entwicklungsphase von READY nicht verwendet. Der Grund dafür liegt darin, daß im Dialog nur wenige Informationen über das Arbeitsgedächtnis des Dialogpartners verfügbar sind. Daher gibt es nur wenige Anhaltspunkte für Inferenzen über den aktuellen Zustand eines Subsystems des Arbeitsgedächtnisses (z.B. der phonologischen Schleife) eines Dialogpartners. Es ist klar, daß diese eher undifferenzierte Betrachtung des Arbeitsgedächtnisses zu ungenauen Inferenzen führen kann.

[1] Die Dialogplanung im System READY wird in Weis (1999) beschrieben.

6.1.1 Beschränkungen des Arbeitsgedächtnisses

6.1.1.1 Anforderung und Verwendung von Arbeitsgedächtniskapazität

Zusätzlich zur Modellierung der Arbeitsgedächtnisbelastung in der Notrufdomäne wird im Benutzermodell von READY angenommen, daß $\mathcal{P}$ gleichzeitig mit der Ausführung von zwei Aufgaben beschäftigt sein kann. Dadurch kann es gemäß des Doppelaufgabenparadigmas (siehe Abschnitt 2.1.2) zur Überforderung des Arbeitsgedächtnisses von $\mathcal{P}$ kommen. Der Zusammenhang zwischen dem Ausführen von zwei Aufgaben und der verfügbaren Arbeitsgedächtniskapazität wird wie folgt berücksichtigt:

Zu jedem Zeitpunkt hat der Dialogpartner $\mathcal{P}$ eine gewisse *verfügbare Arbeitsgedächtniskapazität*, die für die Bewältigung der Aufgaben im Szenario eingesetzt werden kann. Diese verfügbare Kapazität kann zeitweise geringer sein als die gesamte Arbeitsgedächtniskapazität von $\mathcal{P}$, wenn $\mathcal{P}$ emotional belastet ist oder durch Ereignisse in seiner Umgebung abgelenkt wird.

Soll $\mathcal{P}$ eine bestimmte Aufgabe ausführen, entsteht eine *Anforderung an das Arbeitsgedächtnis* von $\mathcal{P}$ entsprechend der Höhe der durch die Aufgabe verursachten Belastung. Die Anforderung beschreibt die Menge an Arbeitsgedächtniskapazität, die für die problemlose Ausführung der Aufgabe notwendig ist. Abhängig von der aktuellen Kapazität des Arbeitsgedächtnisses wird entschieden, wieviel Arbeitsgedächtniskapazität für die Ausführung der Aufgabe verwendet wird. Die Kapazität kann maximal die Höhe der gestellten Anforderung bzw. der verfügbaren Arbeitsgedächtniskapazität erreichen.

Eine solche Aufgabe könnte z.B. das Verstehen einer Frage sein, die $\mathcal{S}$ stellt. Abbildung 6.1 zeigt die eben beschriebenen Zusammenhänge am Beispiel eines Sprachverstehensprozesses. Die Frage ist verbunden mit einer ihrer Schwierigkeit angemessenen ANFORDERUNG DER SPRACHANALYSE AN AG, die von der VERFÜGBAREN ARBEITSGEDÄCHTNISKAPAZITÄT in Form einer VERWENDUNG VON AG FÜR DIE SPRACHANALYSE, wie oben beschrieben, erfüllt wird.

6.1.1.2 Gleichzeitige Ausführen von zwei Aufgaben

Müssen zwei Aufgaben gleichzeitig ausgeführt werden, kann es sein das die Gesamtkapazität des Arbeitsgedächtnisses zu gering für die (problemlose) Ausführung beider Aufgaben ist. Folge von dieser Überbelastung des Arbeitsgedächtnisses von $\mathcal{P}$ ist ein Leistungsabfall bzgl. der Hauptaufgabe (gemäß des Doppelaufgabenparadigmas). Welche Aufgabe $\mathcal{P}$ als Hauptaufgabe ansieht, hängt davon ab, welche Prioritäten $\mathcal{P}$ den beiden auszuführenden Aufgaben gibt. Die Zuteilung von Kapazität an die beiden Aufgaben hängt dann zusätzlich von dieser Prioritätsverteilung ab. Der Aufgabe mit Priorität wird soviel Kapazität zugeteilt, wie sie anfordert (maximal die verfügbare Arbeitsgedächtniskapazität). Der Aufgabe ohne Priorität wird die restliche verfügbare Kapazität (maximal in Höhe ihrer eigenen Anforderung).

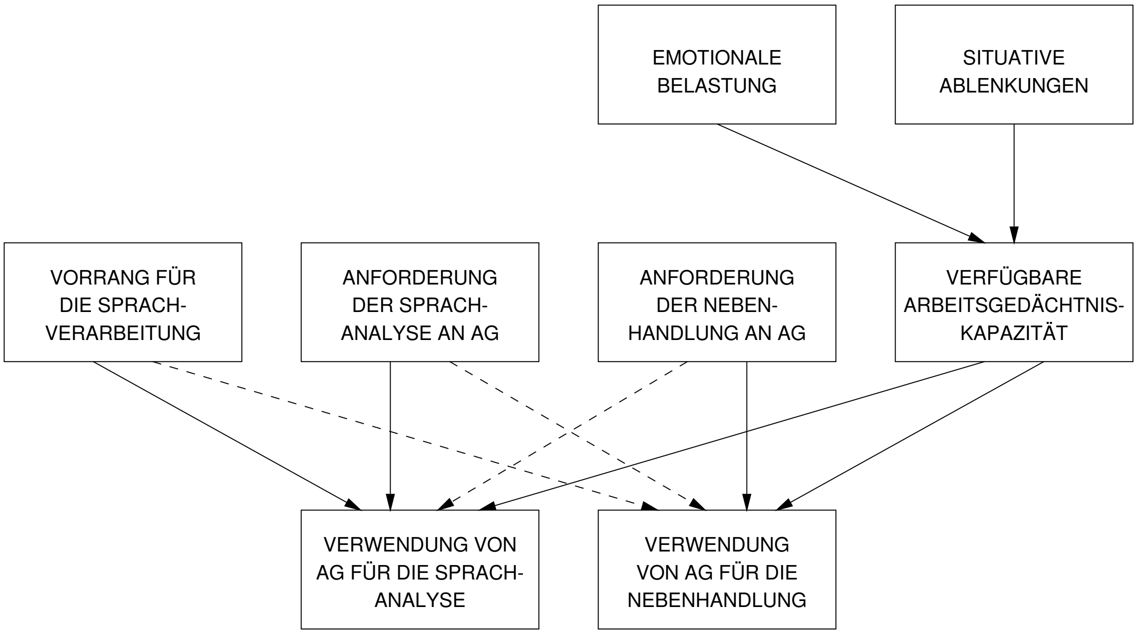

Abb. 6.1: *Modell der Zusammenhänge zwischen Anforderung und Verwendung von verfügbarer Arbeitsgedächtniskapazität bei zwei Aufgaben.*
Ein durchgezogener bzw. gestrichelter Pfeil von A *nach* B *bedeutet:* A *beeinflußt* B *positiv bzw. negativ.*

6.1.2 Beschränkungen der Zeit

Zeitbeschränkungen werden in READY mithilfe der Variablen ZEITDRUCK modelliert. Diese Variable drückt aus, inwieweit $\mathcal{P}$ sich subjektiv zeitlich unter Druck fühlt (siehe auch Abschnitt 6.2.1).

6.1.3 Auswirkungen von Ressourcenbeschränkungen

6.1.3.1 Erfolg bei der Ausführung einer Aufgabe

Die Einschätzung, wie $\mathcal{P}$ seine Arbeitsgedächtniskapazität an die einzelnen Aufgaben verteilt, ist deshalb wichtig, weil das Verhältnis zwischen Anforderung und Verwendung von Arbeitsgedächtniskapazität einer Aufgabe den *Erfolg für die Ausführung der Aufgabe* bestimmt. Abbildung 6.2 veranschaulicht diesen Zusammenhang wiederum am Beispiel der Sprachanalyse. Die SPRACHLICHEN EIGENSCHAFTEN DER FRAGE, die in der Abbildung als eine Gruppe von Knoten zusammengefaßt werden (gestricheltes Kästchen), umfassen Knoten, die in Hinblick auf die Anforderung der Frage von Bedeutung sind, wie z.B. die Äußerungslänge, die syntaktische Schwierigkeit der Äußerung oder Verwendung von Fachbegriffen.

Stellt $\mathcal{S}$ eine Frage, so hängt $\mathcal{P}$s ANFORDERUNG DER SPRACHANALYSE AN AG von diesen SPRACHLICHEN EIGENSCHAFTEN DER FRAGE ab. Abhängig von der ANFORDERUNG DER SPRACHANALYSE, der

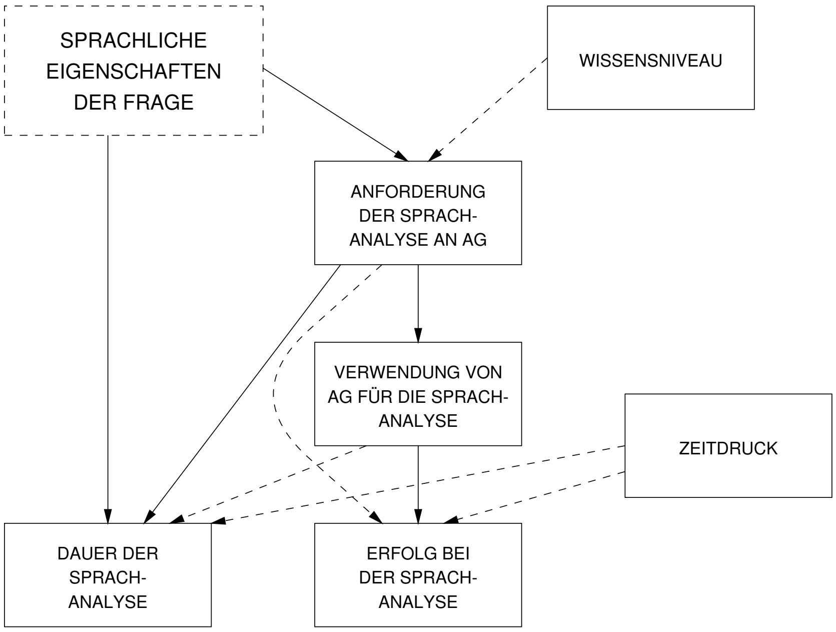

Abb. 6.2: *Modellierung der Auswirkungen der sprachlichen Eigenschaften der Frage auf die Dauer und den Erfolg der Sprachanalyse.*
Ein durchgezogener bzw. gestrichelter Pfeil von A *nach* B *bedeutet:* A *beeinflußt* B *positiv bzw. negativ.*

ANFORDERUNG DER HANDLUNG, des VORRANGS FÜR DIE SPRACHANALYSE und der VERFÜGBAREN ARBEITSGEDÄCHTNISKAPAZITÄT (nicht in Abb. 6.2) wird, wie im letzten Abschnitt gesehen, die VERWENDUNG VON AG FÜR DIE SPRACHANALYSE ermittelt. Wenn die Verwendung im Vergleich zur Anforderung gering ist, bedeutet dies, daß die Sprachanalyse nur unzureichend ausgeführt werden kann. In diesem Fall und bei zu hohem ZEITDRUCK wird der ERFOLG BEI DER SPRACHANALYSE wahrscheinlich leiden. Umgekehrt, wird genügend Kapazität für die Sprachanalyse verwendet und ist der ZEITDRUCK gering, so wird der ERFOLG BEI DER SPRACHANALYSE wahrscheinlich.

6.1.4 Dauer der Aufgabe

Ähnlich wie der Erfolg wird auch die DAUER DER SPRACHANALYSE[2] eingeschätzt (siehe Abb. 6.2). Auch hier führt eine Abweichung der Verwendung von Arbeitsgedächtniskapazität von ihrer Anforderung zu einer längeren Dauer der Ausführung der Sprachanalyse. Des weiteren führt ein erhöhter ZEITDRUCK tendenziell zu einer kürzeren *Dauer der Ausführung der Aufgabe*, da $\mathcal{P}$ sich beeilt. Auch die SPRACHLICHEN EIGENSCHAFTEN DER FRAGE beeinflußt diesen Knoten, in der Form, daß eine schwer zu verstehende Frage mehr Zeit für die Sprachanalyse benötigt.

[2] READY berücksichtigt bei seiner Dialogplanung u.a. auch die Zeit, die $\mathcal{P}$ zur Ausführung der Aufgabe benötigen wird (z.B. anhand von der DAUER DER SPRACHANALYSE). Für die Erkennung von $\mathcal{P}$s Arbeitsgedächtnisbelastung und Zeitdruck wird diese Variable z.Z. nicht berücksichtigt, da das System keine Evidenzen über Ausführungszeiten von Aufgaben behandeln kann.

6.2 Allgemeines Schema für die Erkennung und Einschätzung von Arbeitsgedächtnisbelastung und Zeitdruck

Abbildung 6.3 zeigt anhand von zwei Zeitscheiben, wie die bisher beschriebenen Zusammenhänge miteinander in Verbindung gebracht werden. Ein gestrichelter Kasten steht wieder für eine Gruppe von Knoten. Gestrichelte Pfeile symbolisieren hier Abhängigkeiten, die nur in einigen konkreten Situationen vorkommen. zum Beispiel hängt der ERFOLG BEI DER AUSFÜHRUNG DER HANDLUNG in $t+1$ vom ERFOLG BEI DER SPRACHANALYSE in t ab. Erfolgt nach der Ausführung der Handlung direkt eine weitere Instruktion von $\mathcal{S}$ in $t+2$ so hängt aber nicht der ERFOLG DIESER SPRACHANALYSE in $t+2$ vom ERFOLG BEI DER AUSFÜHRUNG DER (LETZTEN) HANDLUNG in $t+1$ ab.

Das allgemeine Schema in Abbildung 6.3 wird für die Modellierung der beiden Dialogsequenzen Frage–Antwort und Instruktion–Handlung herangezogen, wobei AUFGABE 1 und AUFGABE 2 wie folgt zu ersetzen sind:

(1) *Frage (von $\mathcal{S}$) – Antwort (von $\mathcal{P}$)*: In Zeitscheibe t nimmt $\mathcal{P}$ eine Sprachanalyse vor (AUFGABE 1), dabei kann $\mathcal{P}$ vielleicht mit einer Nebenhandlung beschäftigt sein (AUFGABE 2). In Zeitscheibe $t+1$ erfolgt dann die Sprachgenerierung von $\mathcal{P}$ für die Antwort (AUFGABE 1), wiederum vielleicht gestört durch eine Nebenhandlung (AUFGABE 2).

(2) *Instruktion (von $\mathcal{S}$) – Ausführung der Handlung (von $\mathcal{P}$)*: In Zeitscheibe t nimmt $\mathcal{P}$ eine Sprachanalyse vor (AUFGABE 1), dabei kann $\mathcal{P}$ vielleicht mit einer Nebenhandlung beschäftigt sein (AUFGABE 2). In Zeitscheibe $t+1$ erfolgt dann die Ausführung der Handlung durch $\mathcal{P}$ (AUFGABE 1), wiederum vielleicht gestört durch eine Nebenhandlung (AUFGABE 2).

6.2.1 Zeitveränderliche Eigenschaften

Wie bereits in der Notrufdomäne sind auch im READY-Szenario einige Benutzereigenschaften zeitveränderlich. Diese Variablen werden wie in der Notrufdomäne als dynamische Knoten modelliert.

- SITUATIVE ABLENKUNGEN: Durch Einflüsse aus der Umgebung kann $\mathcal{P}$ abgelenkt werden. Solche SITUATIVE ABLENKUNGEN können z.B. durch vorbeifahrende Autos verursacht werden.
- EMOTIONALE BELASTUNG: Entstand durch den Defekt des Autos beim Fahren eine gefährliche Situation, kann es sein, daß $\mathcal{P}$ am Anfang des Dialogs emotional belastet ist. Es wird aber angenommen, daß die EMOTIONALE BELASTUNG von $\mathcal{P}$ kann im Laufe des Dialogs sinkt, wenn z.B. $\mathcal{P}$ durch $\mathcal{S}$ beruhigt wird.
- VERFÜGBARE ARBEITSGEDÄCHTNISKAPAZITÄT (VAK): Im Fall der Variablen VAK sind die Zusammenhänge komplizierter, da die VAK zusätzlich von den dynamischen Knoten EMOTIONALE BELASTUNG und SITUATIVE ABLENKUNGEN beeinflußt wird. Die VAK in $t+1$ ist so hoch,

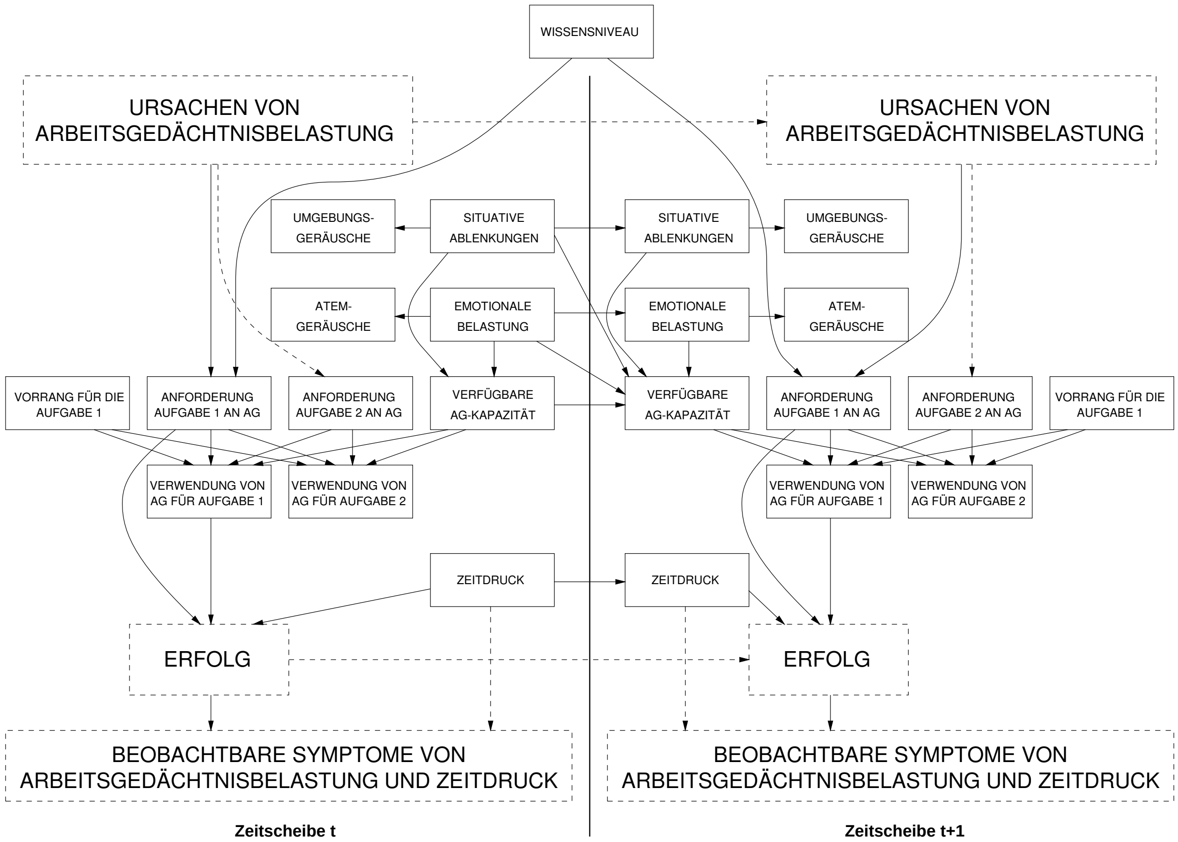

Abb. 6.3: *Schema zur Modellierung der Szenario-Dialoge für zwei Zeitscheiben.*
Ein gestrichelter Kasten steht für eine Menge von Knoten, gestrichelte Pfeile bezeichnet Zusammenhänge, die nur in bestimmten Situationen vorkommen können.

wie in t. Steigt $\mathcal{P}$s EMOTIONALE BELASTUNG oder SITUATIVE ABLENKUNGEN von t nach $t+1$, wird die VAK wahrscheinlich sinken.

- ZEITDRUCK: In bezug auf ZEITDRUCK wird folgende Annahme getroffen: Das subjektive Gefühl von Zeitdruck hat die Tendenz zuzunehmen, da man sich immer stärker des Zeitdrucks bewußt wird, je weniger Zeit verbleibt. Wenn das System in t eine gewisse Einschätzung bzgl. $\mathcal{P}$s ZEITDRUCK hat, wird es in der folgenden Zeitscheibe $t+1$ etwas mehr ZEITDRUCK vermuten, wenn es keine andere Evidenz erhalten hat.

6.2.2 Zeitlich unveränderliche Eigenschaften

In dem Benutzermodell von READY wird das WISSENSNIVEAU des Benutzers als zeitlich unveränderlich angenommen. Das gleiche gilt auch für die NORMALE ARTIKULATIONSGESCHWINDIGKEIT von $\mathcal{P}$. Diese Variablen werden als statische Knoten dargestellt.

6.3 Modellierung der kognitiven Aufgaben

In READY werden z.Z. drei verschiedene Aufgaben von $\mathcal{P}$ modelliert. Diese sind: Verstehen einer Frage von $\mathcal{S}$, Beantwortung einer Frage und Ausführung einer Handlung nach einer Instruktion von $\mathcal{S}$. Alle Tätigkeiten werden einer der entsprechenden Kategorien *Sprachanalyse*, *Sprachgenerierung* bzw. *Ausführung einer Handlung* zugeordnet. Zusätzlich kann $\mathcal{P}$ mit zwei Aufgaben gleichzeitig beschäftigt sein, z.B. mit der Ausführung einer Handlung und mit einer Sprachanalyse.

Das Ziel ist es, zu erkennen wie stark $\mathcal{P}$s Arbeitsgedächtnis belastet ist und wie hoch $\mathcal{P}$s Zeitdruck ist, damit READY seine eigenen Dialogäußerungen (Fragen und Instruktionen) daran anpassen kann. Im folgenden wird für jede Aufgabe anhand eines Beispiels beschrieben, wie READY die o.g. Ressourcenbeschränkungen von $\mathcal{P}$ erkennen und einschätzen kann. Die Knoten des Benutzermodells werden im Anhang A.5 definiert.

6.3.1 Arbeitsgedächtnisbelastung und Zeitdruck bei der Sprachanalyse

Ein Sprachanalyseprozeß seitens $\mathcal{P}$ tritt dann auf, wenn $\mathcal{S}$ eine Frage stellt oder eine Instruktion gibt. Probleme bei der Sprachanalyse werden mit Arbeitsgedächtnisbelastung und Zeitdruck assoziiert (vgl. Abschnitt 2.1.3.2). Da der Vorgang der Sprachanalyse sich naturgemäß nicht genau beobachten läßt, können daher nur wenige Symptome für die Erkennung von $\mathcal{P}$s Arbeitsgedächtnisbelastung und Zeitdruck berücksichtigt werden.

Ein Symptom für den Erfolg der Sprachanalyse ist eine kurze *Rückmeldung* seitens $\mathcal{P}$. Wenn $\mathcal{P}$ eine positive Rückmeldung gibt, wie z.B. „Ok", kann READY diese Rückmeldung als eine Evidenz für eine spontane Erfolgsmeldung bzgl. der Sprachanalyse von $\mathcal{P}$ betrachten. Eine negative Rückmeldung, wie z.B. „Wie bitte?", signalisiert den umgekehrten Fall. Das Ausbleiben einer Rückmeldung wird neutral bewertet, da Menschen nicht immer das Verstehen einer Frage in Form einer spontanen Erfolgsmeldung zu erkennen geben.

Andererseits läßt sich der Grad der Arbeitsgedächtnisbelastung beim Sprachverstehen anhand der sprachlichen Eigenschaften der Äußerung gut vorhersagen. Dazu werden Ursachen von Arbeitsgedächtnisbelastung beim Hörer (siehe Abschnitt 2.1.3.2 und 5.3) herangezogen: Der Einfluß der sprachlichen Eigenschaften einer Äußerung auf die ANFORDERUNG DER SPRACHANALYSE AN AG erfolgt einerseits durch die SCHWIERIGKEIT DER FORMULIERUNG, d.h. die Schwierigkeit der Interpretation einer Äußerung aufgrund syntaktisch-struktureller Aspekte, und andererseits durch die KOMPLEXITÄT DER KONZEPTUALISIERUNG, d.h. die Schwierigkeit der Interpretation einer Äußerung aufgrund semantisch-pragmatischer Aspekte. Die SCHWIERIGKEIT DER FORMULIERUNG wird wiederum durch die SYNTAKTISCHE SCHWIERIGKEIT DER ÄUSSERUNG, die VERWENDUNG VON FACHBEGRIFFEN und ELLIPSEN UND ANAPHERN bestimmt. Bei der KOMPLEXITÄT DER KONZEPTUALISIERUNG muß man zwischen einer Frage und einer Instruktion unterscheiden. Stellt $\mathcal{S}$ eine Frage, so wird die KOMPLEXITÄT DER KONZEPTUALISIERUNG von dem GRAD DER VORSTRUKTURIERUNG (Ja/Nein- und Entscheidungsfragen bieten eine hohe Vorstrukturierung im

Vergleich zu W-Fragen) und der INTRINSISCHEN SCHWIERIGKEIT DER FRAGE (die Schwierigkeit, die von Struktur der Äußerung unabhängig ist) beeinflußt. Gibt $\mathcal{S}$ eine Instruktion, so muß $\mathcal{P}$ einen Plan für die Ausführung der Handlung erzeugen. Die KOMPLEXITÄT DER KONZEPTUALISIERUNG hängt dann von der EXPLIZITHEIT DER INSTRUKTION und der KOMPLEXITÄT DES PLANS ab.[3]

Wie READY die verfügbare Arbeitsgedächtniskapazität und den Zeitdruck bei der Sprachanalyse einschätzt, wird anhand von Beispiel 1 illustriert.

Beispiel 1: $\mathcal{S}$ stellt eine lange und eher schwerverständliche Frage. Viele Umgebungsgeräusche weisen darauf hin, daß $\mathcal{P}$ abgelenkt ist. Außerdem sind viele Atemgeräusche von $\mathcal{P}$ zu hören. $\mathcal{P}$ gibt eine negative Rückmeldung.

Abbildung 6.4 zeigt das Bayessche Netz, das Beispiel 1 beschreibt. Die hellen Histogramme zeigen die Wahrscheinlichkeitsverteilungen der Knoten vor der Beobachtung der Evidenzen (d.h. die Vorhersage) und die dunklen Histogramme die Interpretation der beobachteten Evidenzen.

Die Zeitscheibe t_0 (linke Seite der Abb. 6.4) beschreibt den Zeitpunkt kurz vor der Frage von $\mathcal{S}$. Aus Platzgründen werden nur die dynamischen Variablen gezeigt, die einen Einfluß auf die Zeitscheibe t_1 haben, in der die Sprachanalyse von $\mathcal{P}$ stattfindet.

Aufgrund der sprachlichen Eigenschaften der Frage (hohe SYNTAKTISCHE SCHWIERIGKEIT DER ÄUSSERUNG und geringer GRAD DER VORSTRUKTURIERUNG) wird eine eher mittel bis hohe ANFORDERUNG DER SPRACHANALYSE AN AG gestellt. Außerdem wird angenommen, daß die ANFORDERUNG DER NEBENHANDLUNG AN AG eher mittel bis gering ist. Da es wahrscheinlicher ist, daß $\mathcal{P}$ mehr Aufmerksamkeit der Frage von $\mathcal{S}$ geben wird als der Ausführung der Nebenhandlung, wird die Sprachanalyse höchstwahrscheinlich den Vorrang erhalten. Die SITUATIVE ABLENKUNGEN und die EMOTIONALE BELASTUNG sind eher mittel bis gering, was sich in einer eher hohen VAK auswirkt.

Abhängig von der verfügbaren Arbeitsgedächtniskapazität und den Anforderungen wird die Arbeitsgedächtniskapazität den einzelnen Aufgaben zugeteilt (siehe Abschnitt 6.1). Die VERWENDUNG VON AG FÜR DIE SPRACHANALYSE ist eher mittel hoch, erreicht aber nicht die Menge an Arbeitsgedächtniskapazität, die von ihr angefordert wurde. Der Unterschied kann dadurch erklärt werden, daß einerseits nicht genügend VAK vorhanden ist und andererseits der VORRANG FÜR DIE SPRACHVERARBEITUNG nicht zu hundert Prozent auf der Sprachanalyse liegt. Die VERWENDUNG VON AG FÜR DIE NEBENHANDLUNG liegt deutlich unter der angeforderten Menge. Der Grund dafür ist, daß hierauf viel weniger Vorrang gelegt wird.

Aufgrund des relativ geringen ZEITDRUCKS und dem tendenziell ähnlichen Verhältnis zwischen Anforderung und Verwendung von Arbeitsgedächtniskapazität für die Sprachanalyse, wird ein eher hoher ERFOLG BEI DER SPRACHANALYSE vorhergesagt, was sich in einer eher positiven SPONTANEN ERFOLGSMELDUNG äußern wird.

[3] Da die ANFORDERUNG DER SPRACHGENERIERUNG AN AG (bei der Antwort) bzw. die ANFORDERUNG DER HANDLUNG AN AG (bei der Ausführung der Handlung) von der INTRINSISCHEN SCHWIERIGKEIT DER FRAGE bzw. der KOMPLEXITÄT DES PLANS abhängt, werden diese Knoten in der jeweiligen Zeitscheibe für die Sprachgenerierung bzw. Handlungsausführung dargestellt.

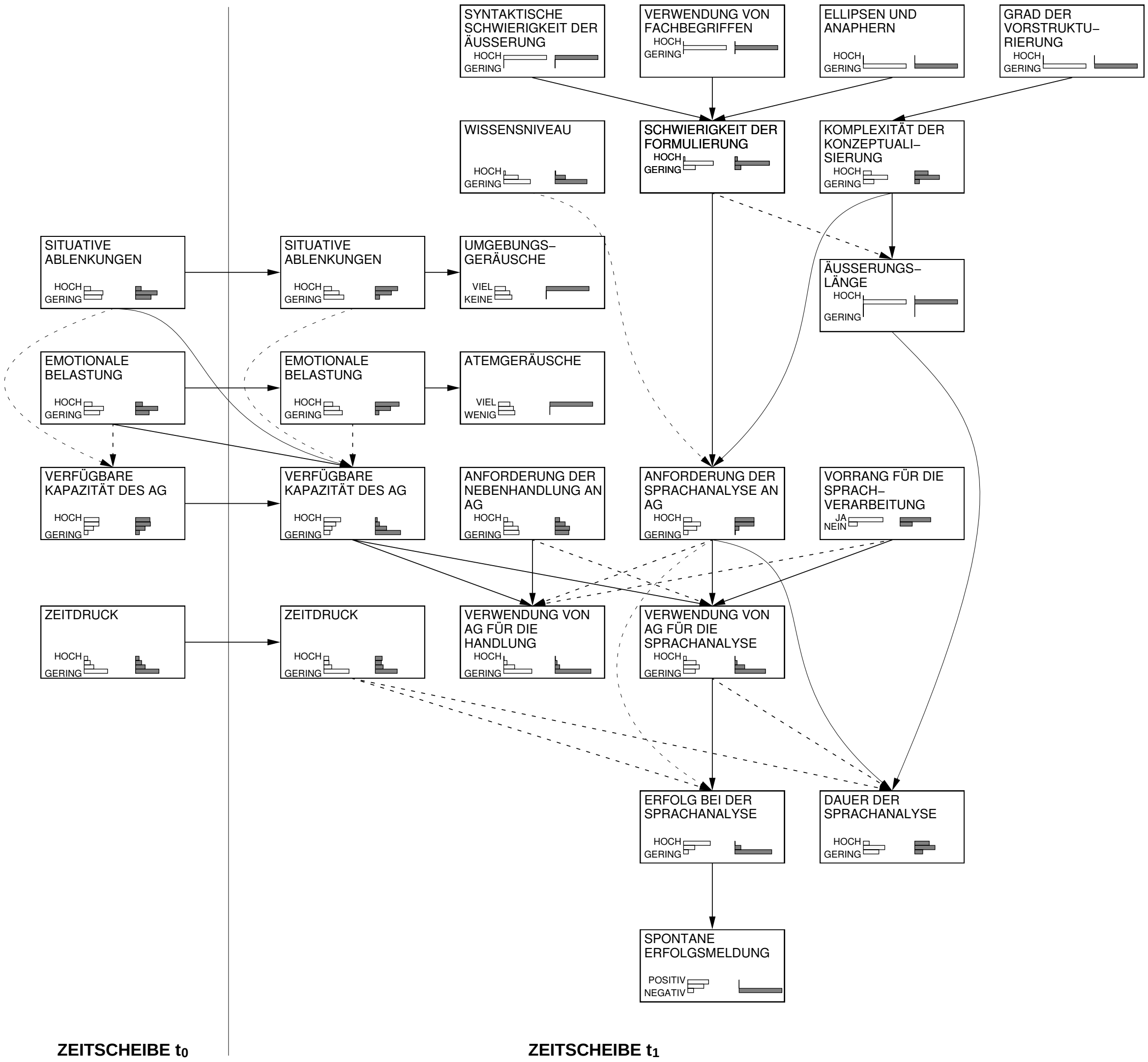

Abb. 6.4: *DBN von Beispiel 1: Die von $\mathcal{S}$ gestellte Frage wird von $\mathcal{P}$ nicht richtig verstanden. die verfügbare Arbeitsgedächtniskapazität wird daher geringer eingeschätzt.*

Im Beispiel 1 werden viele Umgebungsgeräusche und Atemgeräusche beobachtet. Zusätzlich gibt $\mathcal{P}$ eine negative Rückmeldung. Diese Evidenzen führen zu einer neuen Interpretation der Benutzervariablen (siehe dunkle Histogramme). Die negative SPONTANE ERFOLGSMELDUNG hat zur Folge, daß der ERFOLG BEI DER SPRACHANALYSE jetzt eher als sehr gering eingeschätzt wird. Als Ursache davon wird jetzt angenommen, daß $\mathcal{P}$'s VERWENDUNG VON AG FÜR DIE SPRACHANALYSE doch geringer und der ZEITDRUCK von $\mathcal{P}$ etwas höher war, als im voraus angenommen wurde. Die geringe VERWENDUNG VON AG FÜR DIE SPRACHANALYSE wird dadurch begründet, daß die Wahrscheinlichkeit für die ANFORDERUNG DER SPRACHANALYSE AN AG höher war oder die VAK geringer war oder der VORRANG FÜR DIE SPRACHVERARBEITUNG von $\mathcal{P}$ geringer war. Letzteres wird auch durch eine höhere EMOTIONALE BELASTUNG (durch viele ATEMGERÄUSCHE) und mehr SITUATIVE ABLENKUNGEN (durch viele UMGEBUNGSGERÄUSCHE) begründet.

6.3.2 Arbeitsgedächtnisbelastung und Zeitdruck bei der Sprachgenerierung

Ein Sprachgenerierungsprozeß seitens $\mathcal{P}$ tritt dann auf, wenn $\mathcal{P}$ eine Antwort auf eine Frage geben muß. Probleme bei der Sprachgenerierung werden wieder mit Arbeitsgedächtnisbelastung und Zeitdruck assoziiert (vgl. Abschnitt 2.1.3.1). Im Gegensatz zur Sprachanalyse können bei der Sprachproduktion viele Beobachtungen vorgenommen werden (z.B. die sprachlichen Merkmale der Äußerung von $\mathcal{P}$), so daß sich die Einschätzung der Belastung genauer vornehmen läßt.

In der empirische Untersuchung (siehe Kapitel 5) wurden viele Zusammenhänge identifiziert, die in das Modell einfließen können.

Symptome von ZEITDRUCK sind die MENGE LEICHT AUSLASSBARER INFORMATIONEN und die Benutzung von vielen ELLIPSEN UND ANAPHERN BEI P. Die beiden Zusammenhänge lassen sich wie folgt beschreiben: Wenn $\mathcal{P}$s ZEITDRUCK hoch ist, wird $\mathcal{P}$ wahrscheinlich nicht viele Informationen äußern, die nichts mit der Frage zu tun haben, und $\mathcal{P}$s Äußerung wird wahrscheinlich viele ELLIPSEN UND ANAPHERN enthalten.

Ein weiteres Symptom für ZEITDRUCK von $\mathcal{P}$ ist die Artikulationsgeschwindigkeit. Dieser Zusammenhang wurde in der empirischen Untersuchung von vielen Versuchspersonen genannt (siehe Abschnitt 5.3).[4] Durch einen Vergleich mit der NORMALEN ARTIKULATIONSGESCHWINDIGKEIT wird der Wert für den Knoten ABWEICHUNG (VON DER NORMALEN) DER ARTIKULATIONSGESCHWINDIGKEIT berechnet: Ist die BEOBACHTETE ARTIKULATIONSGESCHWINDIGKEIT höher als der ermittelte Normalwert, erhöht sich die Wahrscheinlichkeit für eine hohe ABWEICHUNG DER ARTIKULATIONSGESCHWINDIGKEIT, was als Indiz für einen hohen ZEITDRUCK gewertet wird. Dieser Zusammenhang wird in Berthold (1997) ausführlich besprochen.

Der Grad der Arbeitsgedächtnisbelastung hat Auswirkungen auf den Erfolg der Sprachgenerierung. Dabei wird zwischen dem ERFOLG BEI DER KONZEPTUALISIERUNG und dem ERFOLG BEI DER FORMULIERUNG unterschieden. Der ERFOLG BEI DER KONZEPTUALISIERUNG beschreibt, inwieweit $\mathcal{P}$ die Frage inhaltlich konzeptualisiert hat. Der ERFOLG BEI DER FORMULIERUNG gibt an, wie gut $\mathcal{P}$ die Antwort gegeben hat. Zusätzlich ist der Knoten ERFOLG BEI DER KONZEPTUALISIERUNG (in $t+1$) von dem ERFOLG BEI DER SPRACHANALYSE (in t) abhängig.

Ein hohes Defizit zwischen der ANFORDERUNG DER SPRACHGENERIERUNG AN AG und der VERWENDUNG VON AG FÜR DIE SPRACHGENERIERUNG macht einen geringen ERFOLG BEI DER FORMULIERUNG wahrscheinlicher, was sich dadurch zeigt, daß die DEUTLICHKEIT DER AUSSPRACHE leiden wird.

Ähnliches gilt in bezug auf den ERFOLG BEI DER KONZEPTUALISIERUNG. Einen geringen ERFOLG BEI DER KONZEPTUALISIERUNG erkennt man dadurch, daß $\mathcal{P}$ wahrscheinlich keine INHALTLICH PASSENDE ANTWORT[5] geben und möglicherweise ORTSDEIKTISCHE ÄUSSERUNGEN verwenden wird, wie die Frau in dem Beispieldialog (siehe Tabelle 5.1).

[4] In der empirischen Untersuchung wurde allgemein von der Sprechgeschwindigkeit gesprochen, die sich aus der Gesamtdauer der Äußerung (unter Einbeziehung aller Pausen) und der Äußerungslänge berechnet. In dieser Arbeit werden aber keine Pausen, die der Sprecher macht, berücksichtigt (siehe dafür Berthold, 1998).

[5] Diese Variable wurde in der empirischen Studie QUALITÄT DER ANTWORT genannt.

Wie READY die Arbeitsgedächtnisbelastung und den Zeitdruck von $\mathcal{P}$ bei einer Antwort erkennen kann, wird anhand von Beispiel 2 illustriert. Um den dynamischen Charakter der Einschätzungen besser zu verdeutlichen, setzt Beispiel 2 das vorangegangene Beispiel fort.

Beispiel 2: Die von $\mathcal{S}$ gestellte Frage (aus Beispiel 1) wurde von $\mathcal{P}$ mit hoher Wahrscheinlichkeit nicht richtig verstanden. $\mathcal{P}$ gibt erwartungsgemäß keine inhaltlich passende Antwort. Die mittel-lange Antwort enthält wenige leicht auslaßbare Informationen, des weiteren viele Ellipsen und Anaphern, was auf erhöhten Zeitdruck hinweist. Die Deutlichkeit der Aussprache und die beobachtete Artikulationsgeschwindigkeit ist mittel. Mittel-viele Umgebungs- und Atemgeräusche deuten auf eine Senkung der situativen Ablenkungen und der emotionalen Belastung hin.

Abbildung 6.5 zeigt das zu Beispiel 2 gehörende Bayessche Netz. Auf der linken Seite der Abbildung ist ein Teil des Netzes der Zeitscheibe t_1 (Sprachanalyse aus Beispiel 1) abgebildet, um die Einschätzung der zeitveränderlichen Variablen zu illustrieren. In t_2 werden die Knoten gezeigt, die für die Erkennung und Einschätzung der Arbeitsgedächtnisbelastung bzw. des Zeitdrucks während der Sprachgenerierung von Bedeutung sind.

Nach der negativen Rückmeldung wurde $\mathcal{P}$s verfügbare Arbeitsgedächtniskapazität als eher gering eingeschätzt (t_1). Diese Einschätzung wird aufgrund der mittel bis hohen EMOTIONALEN BELASTUNG und SITUATIVEN ABLENKUNG in die Zeitscheibe t_2 übernommen.

Aufgrund der hohen INTRINSISCHEN SCHWIERIGKEIT DER FRAGE und des eher geringen WISSENSNIVEAUS wird die ANFORDERUNG DER SPRACHGENERIERUNG AN AG als mittel bis hoch eingeschätzt. Die ANFORDERUNG DER NEBENHANDLUNG AN AG wird eher als mittel eingeschätzt. Da $\mathcal{P}$ eine Antwort generiert, gibt $\mathcal{P}$ wahrscheinlich der Sprachverarbeitung den Vorrang. Da die VAK als eher gering eingeschätzt wird, wird die VERWENDUNG VON AG FÜR DIE SPRACHGENERIERUNG eher gering sein. Der Unterschied zwischen der Anforderung und der Verwendung von Arbeitsgedächtniskapazität macht einen hohen ERFOLG BEI DER FORMULIERUNG somit nicht wahrscheinlich. Noch schlechter sieht es bei dem ERFOLG BEI DER KONZEPTUALISIERUNG aus. Dieser Knoten hängt zusätzlich von dem ERFOLG BEI DER SPRACHANALYSE ab, der in der vorherigen Zeitscheibe als sehr gering eingeschätzt wurde.

Nach den Beobachtungen der sprachlichen Merkmalen von $\mathcal{P}$s Äußerung werden die Evidenzen gesetzt. Die Verwendung von vielen ELLIPSEN UND ANAPHERN und wenig LEICHT AUSLASSBARE INFORMATIONEN machen einen hohen ZEITDRUCK eher wahrscheinlich. Der eher geringe ERFOLG BEI DER KONZEPTUALISIERUNG und der mittel bis geringe ERFOLG BEI DER FORMULIERUNG von $\mathcal{P}$s Äußerung tragen zu der Einschätzung des erhöhten Zeitdrucks bei. Da der ERFOLG BEI DER SPRACHANALYSE (in t_1) sehr gering eingeschätzt wurde und $\mathcal{P}$ keine INHALTLICH PASSENDE ANTWORT gibt, wird der geringe ERFOLG BEI DER KONZEPTUALISIERUNG noch ein bißchen unwahrscheinlicher.[6]

Da die Einschätzung der beiden Erfolgsknoten sich nur geringfügig ändert, fällt die Änderung der Elternknoten auch geringer aus. Man kann sehen, daß die Knoten VAK, ANFORDERUNG DER SPRACHGENERIERUNG AN AG und VERWENDUNG VON AG FÜR DIE SPRACHGENERIERUNG

[6]Für die Variable VERWENDUNG VON ORTSDEIKTISCHEN ÄUSSERUNGEN liegt keine Evidenz vor, da in der Antwort von $\mathcal{P}$ keine ortsdeiktischen Äußerungen vorkommen konnten.

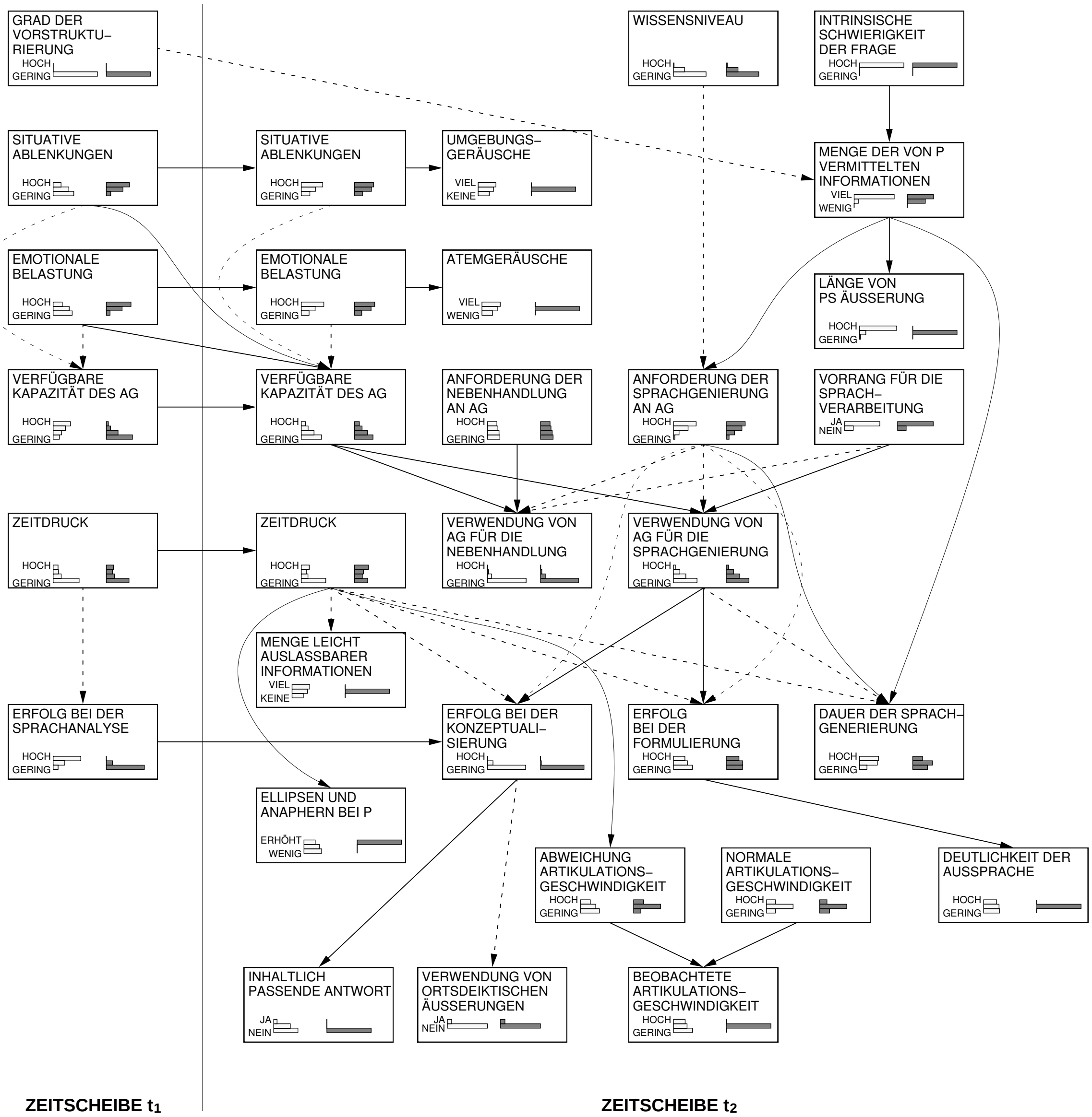

Abb. 6.5: *DBN von Beispiel 2: Die schlechte Beantwortung, der von $\mathcal{S}$ gestellten Frage, deutet auf eine mittel bis geringe verfügbare Arbeitsgedächtniskapazität hin. Wenig leicht auslaßbare Informationen und viele Ellipsen bzw. Anaphern lassen auf erhöhten Zeitdruck zurückschließen.*

in etwa so eingeschätzt wurden, wie vorhergesagt. Begründen läßt sich das dadurch, daß $\mathcal{P}$ sich gemäß den vorhergesagten Einschätzungen verhält.

6.3.3 Arbeitsgedächtnisbelastung und Zeitdruck bei der Ausführung einer Handlung

Nach einer Instruktion von $\mathcal{S}$ führt $\mathcal{P}$ eine Handlung aus. Da im READY-Szenario $\mathcal{S}$ und $\mathcal{P}$ via Mobiltelefon kommunizieren, ist READY bei der Erkennung von Problemen, die $\mathcal{P}$ bei der Ausführung der Handlung hat, auf das Vorkommen von bestimmten sprachlichen Äußerungen beschränkt. Problematische Äußerungen von $\mathcal{P}$ werden wiederum als Folgen von $\mathcal{P}$s Arbeitsgedächtnisbelastung betrachtet.

Ein Zeichen einer hohen Belastung seitens $\mathcal{P}$ ist das laute Wiederholen von Teilen der Instruktion von $\mathcal{S}$. Dadurch kann $\mathcal{P}$ sich die Instruktion besser einprägen, ähnlich dem *rehearsal-Prozeß* (vgl. Kapitel 2.1). Das Wiederholen von Instruktionsteilen wird durch den Knoten ECHO VON ÄUSSERUNGEN dargestellt. Ein weiteres Symptom bzgl. der Belastung ist – wie beim Sprachanalyseprozeß – eine Rückmeldung von $\mathcal{P}$. Bei einer erfolgreichen Ausführung einer Handlung kann $\mathcal{P}$ eine positive SPONTANE ERFOLGSMELDUNG geben und eine negative Erfolgsmeldung im umgekehrten Fall. Wie schon bei der Sprachanalyse, wird das Ausbleiben einer Rückmeldung als neutral bewertet.

Eine weitere verbale Rückmeldung von $\mathcal{P}$ nach der Ausführung einer Instruktion ist ein SPRECHERWECHSELSIGNAL. Damit signalisiert $\mathcal{P}$, daß er die Ausführung der Handlung erfolgreich beendet hat und $\mathcal{S}$ mit weiteren Instruktionen fortfahren soll. Bei einer positiven SPONTANEN ERFOLGSMELDUNG wird ein SPRECHERWECHSELSIGNAL als Symptom für Zeitdruck interpretiert. Bei negativen ERFOLGSMELDUNGEN werden keine SPRECHERWECHSELSIGNALE gesetzt, da eher der Mißerfolg im Vordergrund von $\mathcal{P}$s Äußerung steht, nicht $\mathcal{P}$s Zeitdruck.

Die Einschätzung der Arbeitsgedächtnisbelastung und des Zeitdrucks von $\mathcal{P}$ wird anhand des Beispiels 3 illustriert. Das Beispiel ist eine erweiterte Fortsetzung von Beispiel 2.

Beispiel 3: $\mathcal{S}$ gibt eine einfache Instruktion. $\mathcal{P}$ gibt eine positive Rückmeldung als Zeichen, daß $\mathcal{P}$ die Instruktion verstanden hat. Daraufhin führt $\mathcal{P}$ die Instruktion aus, wobei $\mathcal{P}$ Teile der Instruktion wiederholt. Nach der Ausführung der Handlung äußert $\mathcal{P}$ „Ok. Weiter!“, was auf eine erfolgreiche Ausführung der Handlung schließen läßt. Die Aufforderung „Weiter!“ wird als Sprecherwechselsignal gedeutet. Wenige Atemgeräusche deuten auf eine Beruhigung von $\mathcal{P}$ hin. Außerdem sind keine Umgebungsgeräusche zu hören.

Abbildung 6.6 zeigt das zu Beispiel 3 gehörende Bayessche Netz. in Zeitscheibe t_3 ist wiederum ein Teil des Netzes abgebildet, das die Einschätzungen bei der Sprachanalyse während der Instruktion von $\mathcal{S}$ wiederspiegelt. Die Darstellung des Netzes in t_3 ist, wie in den Beispielen zuvor, auf die dynamischen Knoten und den ERFOLG BEI DER SPRACHANALYSE, der einen Einfluß auf den ERFOLG BEI DER AUSFÜHRUNG DER HANDLUNG hat, beschränkt.

Die Hauptaufgabe in Zeitscheibe t_4 ist die Ausführung der Handlung, deren Anforderung einerseits vom WISSENSNIVEAU und andererseits von der ANFORDERUNG AN PS PLANUNG abhängt. Die ANFORDERUNG AN PS PLANUNG ist ein Maß für den Aufwand für die Planexpansion seitens

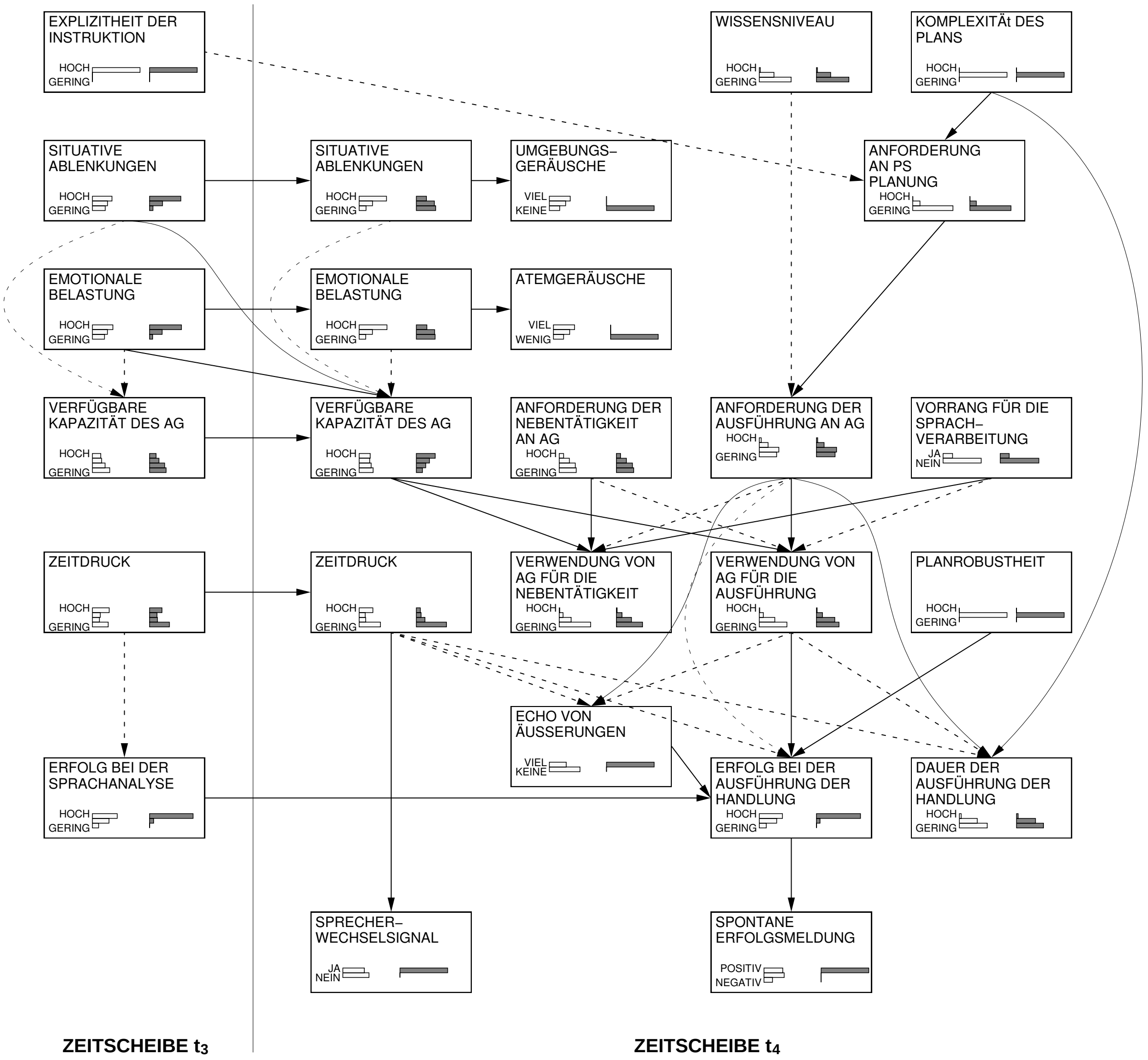

Abb. 6.6: *DBN von Beispiel 3: Die von $\mathcal{S}$ gestellte Frage wird von $\mathcal{P}$ nicht richtig verstanden. Die verfügbare Arbeitsgedächtniskapazität wird daher geringer eingeschätzt.*

$\mathcal{P}$, das einerseits von der EXPLIZITHEIT DER INSTRUKTION (z.B. elementare oder abstrakte Aktionsbeschreibung) und andererseits von der KOMPLEXITÄT DES PLANS abhängt. Die KOMPLEXITÄT DES PLANS bestimmt auch die DAUER DER AUSFÜHRUNG DER HANDLUNG. Zusätzlich beeinflußt die PLANROBUSTHEIT den ERFOLG BEI DER AUSFÜHRUNG DER HANDLUNG. Die PLANROBUSTHEIT gibt an, wie robust der Plan gegenüber dem Fehlschlagen einzelner Aktionen ist.

In Beispiel 3 gibt $\mathcal{S}$ eine einfache und explizite Instruktion. Die Plankomplexität ist mittel. Dadurch wird die ANFORDERUNG AN PS PLANUNG als wahrscheinlich gering eingeschätzt, so daß trotz eher mittlerem bis geringem WISSENSNIVEAU die ANFORDERUNG DER AUSFÜHRUNG AN AG auch als eher gering eingeschätzt wird. Da $\mathcal{P}$ mehr Aufmerksamkeit der Ausführung der Handlung gibt, fällt diesmal der VORRANG zuungunsten der Sprachverarbeitung aus.

Die Vorhersage der restlichen Knoten in t_4 erfolgt ähnlich wie in den Beispielen zuvor. Daher wird ihre Beschreibung nicht wiederholt.

Geringe UMGEBUNGSGERÄUSCHE und ATEMGERÄUSCHE deuten auf eine Verringerung der SITUATIVEN ABLENKUNGEN und der EMOTIONALEN BELASTUNG hin, was sich positiv VAK von $\mathcal{P}$ auswirkt.

Nach der Ausführung der Handlung gibt $\mathcal{P}$ eine positive SPONTANE ERFOLGSMELDUNG, die zu einer höheren Einschätzung eines hohen ERFOLGS BEI DER AUSFÜHRUNG DER HANDLUNG führt. Dieser Eindruck wird durch das ECHO VON ÄUSSERUNGEN zusätzlich verstärkt.

Viele ECHOS VON ÄUSSERUNGEN und die hohe Wahrscheinlichkeit für einen hohen ERFOLG BEI DER AUSFÜHRUNG DER HANDLUNG führen dazu, daß ein geringer ZEITDRUCK wahrscheinlicher wird. Diese Einschätzung wird allerdings durch das SPRECHERWECHSELSIGNAL abgemildert, das ein Indiz für Zeitdruck ist.

Insgesamt erkennt man anhand des Bayesschen Netzes in Abbildung 6.6, daß durch die Beobachtung einer erfolgreichen Ausführung der Instruktion der Grad der Arbeitsgedächtnisbelastung geringer eingeschätzt wird.

6.4 Einschätzung der Arbeitsgedächtnisbelastung und des Zeitdrucks anhand eines Beispieldialogs

Zum Abschluß wird das Benutzermodell von READY anhand des Beispieldialogs der empirischen Untersuchung (siehe Tabelle 5.1) getestet. Wie bereits in Kapitel 4 werden die Einschätzungen des Systems bzgl. der Ressourcenbeschränkungen mit den der Versuchspersonen verglichen. Ähnliche Einschätzungen der Ressourcenlage des Benutzers führen wieder zur Annahme, daß das Benutzermodell von READY plausible Inferenzen zieht. Schäfer (1998) beschreibt ausführlich zwei weitere Dialoge.

die Beschreibung des Beispieldialogs konzentriert sich auf die Inferenzen, die für die Einschätzung $\mathcal{P}$s VAK und ZEITDRUCK wesentlich sind. Die folgenden Abbildungen der DBNs zeigen im Unterschied zu den DBNs im vorherigen Abschnitt das ganze Zeitscheibenpaar, das die entsprechende Frage/Antwort-Situation bzw. Instruktion/Handlungs-Situation darstellt. Dabei zeigen die hellen Histogramme die Einschätzungen, nachdem das System die Äußerung verbalisiert hat (für beide Zeitscheiben). Die dunklen Histogramme zeigen die Einschätzungen, die nach der Interpretation der Evidenzen vorgenommen wurden (wiederum für beide Zeitscheiben).

Zusätzlich wird für jede Frage/Antwort- bzw. Instruktion/Handlungs-Situation der betreffende Ausschnitt der Dialoghistorie angegeben.

Abbildung 6.7 zeigt ein Beispiel der Dialoghistorie der ersten beiden Zeitscheiben. Zu den einzelnen Äußerungen werden die wichtigsten Merkmale und Werte bzgl. der Einschätzungen bestimmter Knoten angegeben. Die Zahl beschreibt den Erwartungswert bzgl. der Wahrscheinlichkeitsverteilung der Zustände des Knoten in Promille. Zur besseren Lesbarkeit werden die Erwartungswerte durch linguistische Hecken beschrieben. Auf

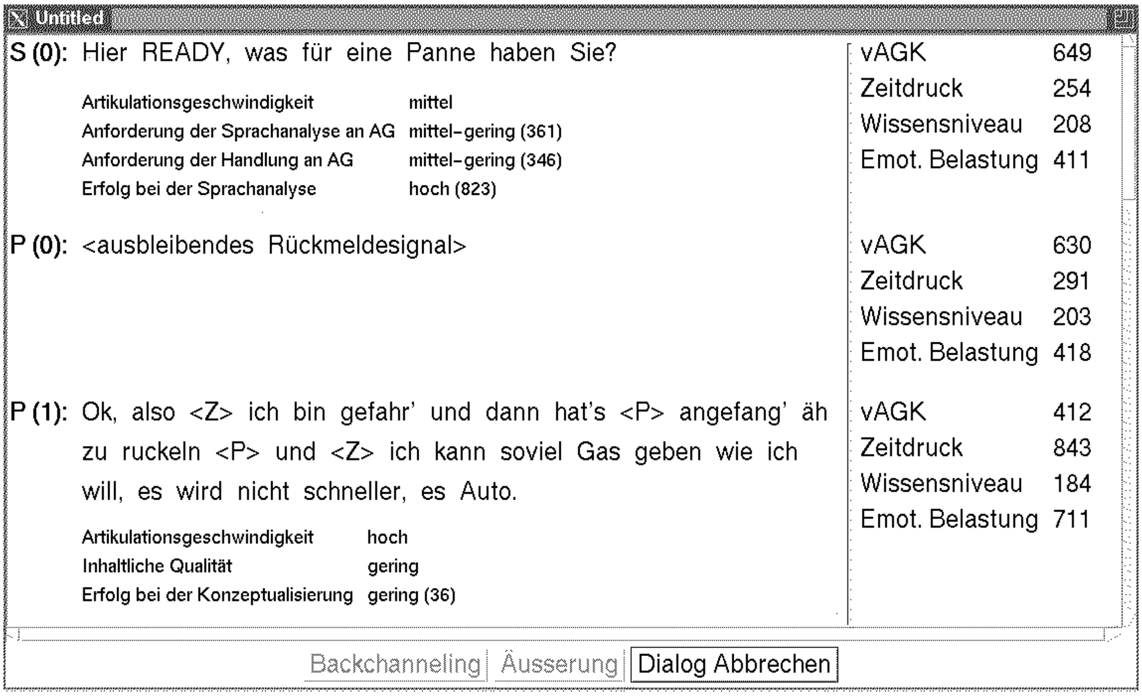

Abb. 6.7: *Ausschnitt der Dialoghistorie von* READY*: Zeitscheiben t_0 und t_1.*
Die linke Seite enthält die einzelnen Äußerungen mit ausgewählten Merkmalen. Die rechte Seite zeigt die wichtigsten Benutzereigenschaften. Die Zahlen geben die Erwartungswerte in Promille an.

der rechten Seite der Dialoghistorie werden die Einschätzungen der wichtigsten Benutzereigenschaften (VAK, ZEITDRUCK, WISSENSNIVEAU und EMOTIONALE BELASTUNG) mit ihren Erwartungswerten dargestellt.

6.4.1 Zeitscheiben t_0 und t_1

Zu Beginn des Dialogs (in t_0) meldet sich READY in Form einer Frage über die Art der Autopanne: „Hier READY, was für eine Panne haben Sie?" Daraufhin gibt $\mathcal{P}$ eine Situationsbeschreibung (in t_1). $\mathcal{P}$ muß also in t_0 eine *Sprachanalyse* und in t_1 eine *Sprachgenerierung* ausführen.

Da aufgrund der Eigenschaften der Frage die ANFORDERUNG DER SPRACHANALYSE AN AG mittel bis gering eingeschätzt wird, erwartet $\mathcal{S}$ einen eher hohen ERFOLG BEI DER SPRACHANALYSE (in t_0). Weil noch keine Informationen über $\mathcal{P}$s Arbeitsgedächtnisbelastung bzw. Zeitdruck vorliegen, wird erst einmal von einer eher mittel bis hohen VAK und einem geringen ZEITDRUCK ausgegangen (siehe Abb. 6.7).

Die Abbildung 6.11 (Seite 105) zeigt das dazugehörige DBN. $\mathcal{P}$ gibt keine Rückmeldung, was als neutrale SPONTANE ERFOLGSMELDUNG interpretiert wird.

In der Zeitscheibe t_1 gibt $\mathcal{P}$ die Antwort auf die gestellte Frage. $\mathcal{P}$ redet schnell und gibt keine LEICHT AUSLASSBARE INFORMATIONEN, was als Indiz für hohen ZEITDRUCK gedeutet wird.

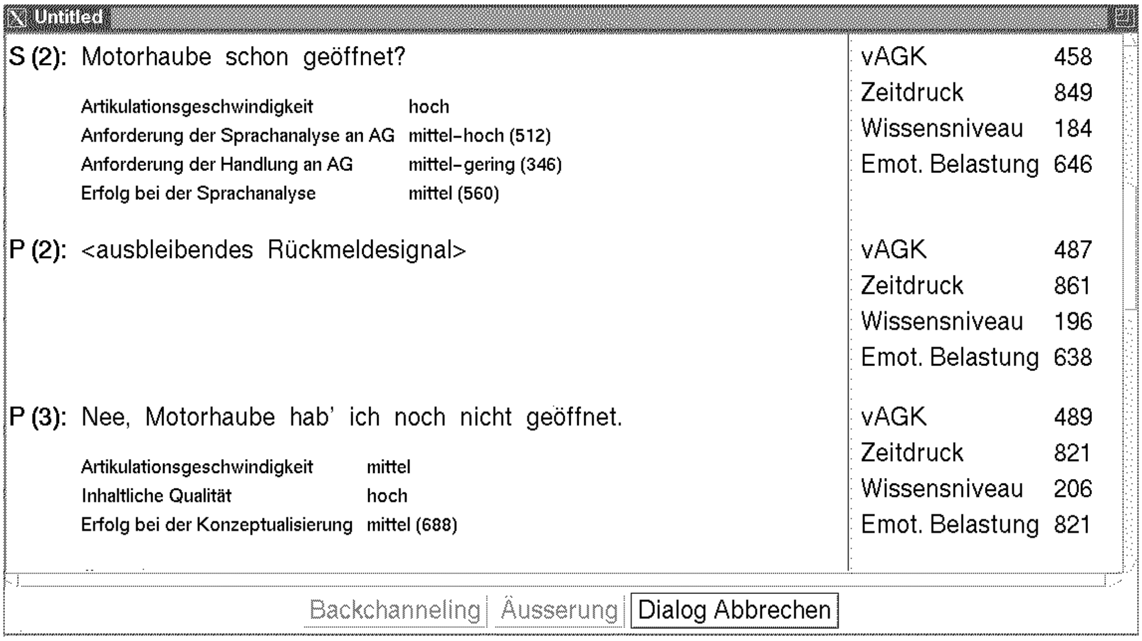

Abb. 6.8: *Ausschnitt der Dialoghistorie von* READY*: Zeitscheiben t_2 und t_3.*
Die linke Seite enthält die einzelnen Äußerungen mit ausgewählten Merkmalen. Die rechte Seite zeigt die wichtigsten Benutzereigenschaften. Die Zahlen geben die Erwartungswerte in Promille an.

Auch in der Abbildung 6.7 ist ein deutlicher Anstieg des ZEITDRUCKS von t_0 nach t_1 zu erkennen.[7] $\mathcal{P}$ äußert keine INHALTLICH PASSENDE ANTWORT und hat eine undeutliche Aussprache, so daß nun ein hoher ERFOLG BEI DER KONZEPTUALISIERUNG bzw. ERFOLG BEI DER FORMULIERUNG unwahrscheinlicher als zu Beginn angesehen wird.

In der empirischen Studie wurde an dieser Stelle das Band gestoppt und die Versuchspersonen nach ihrer Einschätzung der Eile der Anruferin gebeten. Zehn von zwölf Versuchspersonen meinten, daß die Anruferin unter Zeitdruck stünde (vgl. Frage 3.2 in Abb. 5.6, Seite 82). Diese Einschätzung ist vergleichbar mit der des Systems.

6.4.2 Zeitscheiben t_2 und t_3

In den Zeitscheiben t_2 und t_3 liegt wiederum eine Frage/Antwort-Situation vor. $\mathcal{S}$ stellt eine einfache Frage: „Motorhaube schon geöffnet?" Die Abbildung 6.8 zeigt die Dialoghistorie dieser beiden Zeitscheiben und die Abbildung 6.12 (Seite 106) das dazugehörige DBN.

Die SYNTAKTISCHE SCHWIERIGKEIT DER FRAGE und die Verwendung von ELLIPSEN UND ANAPHERN ist mittel und die VERWENDUNG VON FACHBEGRIFFEN ist gering. Da es sich um eine Entscheidungsfrage handelt, ist der GRAD DER VORSTRUKTURIERUNG hoch. Aus diesen Gründen werden

[7] Die hohe ABWEICHUNG DER ARTIKULATIONSGESCHWINDIGKEIT resultiert daher, daß die NORMALE ARTIKULATIONSGESCHWINDIGKEIT durch eine schmale a-priori Verteilung definiert wird. Aufgrund der hohen Häufigkeit der Nennung der Sprechgeschwindigkeit in der empirischen Studie entspricht diese Definition am ehesten den impliziten Annahmen der Versuchspersonen.

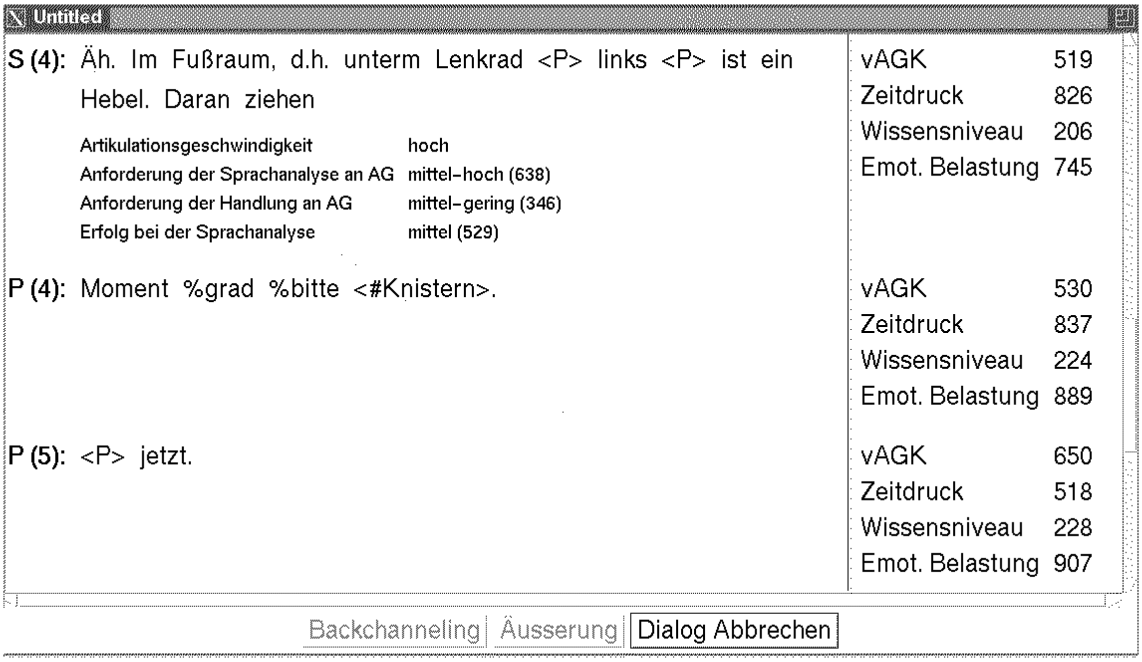

Abb. 6.9: *Ausschnitt der Dialoghistorie von* READY*: Zeitscheiben t_4 und t_5.*
Die linke Seite enthält die einzelnen Äußerungen mit ausgewählten Merkmalen. Die rechte Seite zeigt die wichtigsten Benutzereigenschaften. Die Zahlen geben die Erwartungswerte in Promille an.

die SCHWIERIGKEIT DER FORMULIERUNG und die KOMPLEXITÄT DER KONZEPTUALISIERUNG, die beide ein Maß für die Schwierigkeit der Äußerung sind, als wahrscheinlich mittel bis gering eingeschätzt. $\mathcal{P}$s WISSENSNIVEAU wird etwas geringer eingeschätzt als am Anfang des Dialogs. Die ANFORDERUNG DER SPRACHANALYSE AN AG wird daher etwas höher eingeschätzt als in der Zeitscheibe t_0. Daher wird mit einem wahrscheinlich mittleren ERFOLG BEI DER SPRACHANALYSE gerechnet.

In t_3 gibt $\mathcal{P}$ die Antwort: „Nee, Motorhaube hab' ich noch nicht geöffnet.“ In Abbildung 6.12 (Seite 106) läßt sich erkennen, daß der ERFOLG BEI DER KONZEPTUALISIERUNG aufgrund der Ressourcenlage von $\mathcal{P}$ als eher mittel bis gering vorhergesagt wird (helle Histogramme). $\mathcal{P}$ gibt aber eine INHALTLICH PASSENDE ANTWORT, so daß das System seine Einschätzung des ERFOLGS BEI DER KONZEPTUALISIERUNG anpassen muß. Die erhöhte Wahrscheinlichkeit für einen hohen Erfolg macht auch einen hohen ZEITDRUCK nicht mehr so wahrscheinlich.

Die hohe Wahrscheinlichkeit für eine hohe EMOTIONALE BELASTUNG (aufgrund vieler ATEMGERÄUSCHE) wirkt sich negativ auf die Einschätzung der VAK aus. Wegen der guten Antwort von $\mathcal{P}$ fällt die Wahrscheinlichkeit für eine geringe VAK nicht so hoch aus.

6.4.3 Zeitscheiben t_4 und t_5

In den Zeitscheiben t_4 und t_5 liegt eine Instruktion/Handlungs-Situation vor. $\mathcal{P}$ muß also in t_4 eine *Sprachanalyse* durchführen und in t_5 die Handlung ausführen. Die Dialoghistorie ist in Abbildung 6.9 zu sehen.

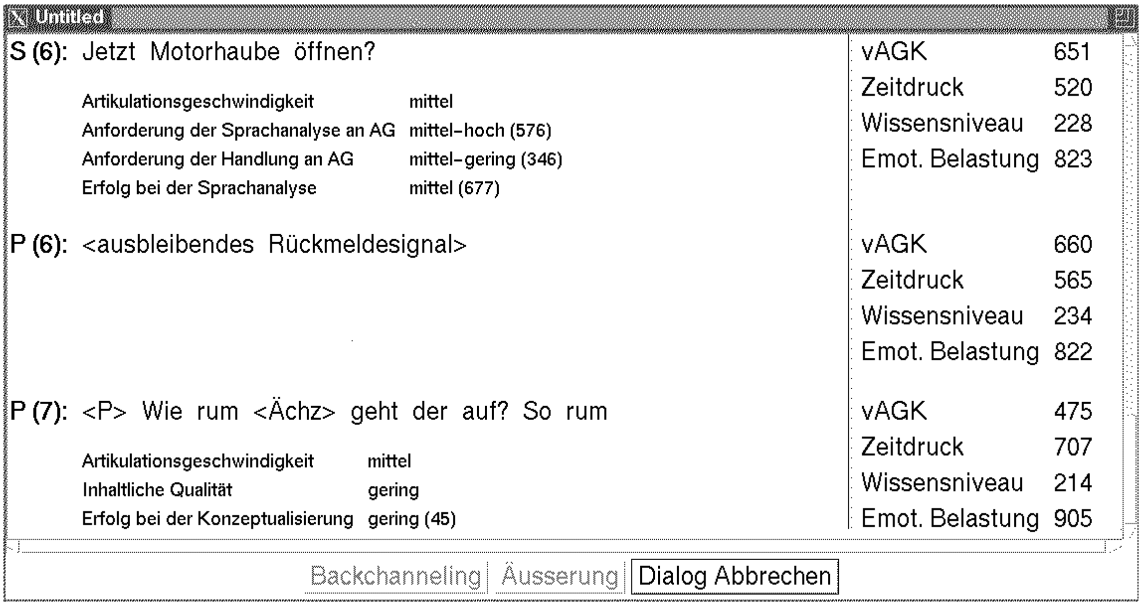

Abb. 6.10: *Ausschnitt der Dialoghistorie von* READY*: Zeitscheiben t_6 und t_7.*
Die linke Seite enthält die einzelnen Äußerungen mit ausgewählten Merkmalen. Die rechte Seite zeigt die wichtigsten Benutzereigenschaften. Die Zahlen geben die Erwartungswerte in Promille an.

$\mathcal{S}$ gibt eine lange und explizite Instruktion. Aus diesem Grund wird die ANFORDERUNG DER SPRACHANALYSE AN AG (in t_4) eher mittel bis hoch, die ANFORDERUNG DER HANDLUNG AN AG (in t_5) aber eher mittel bis gering vorhergesagt (siehe Abbildung 6.13). $\mathcal{P}$ gibt in t_4 eine Rückmeldung „Moment %grad“, die als neutral gewertet wird. In dem Dialogteil ist zu hören, daß $\mathcal{P}$ die Situation nicht bewältigen kann. Viele ATEMGERÄUSCHE deuten auf eine erhöhte EMOTIONALE BELASTUNG hin.

Aufgrund des wahrscheinlich hohen ZEITDRUCKS von $\mathcal{P}$ und einer eher mittleren VAK (in t_5) wird ein eher geringer ERFOLG BEI DER AUSFÜHRUNG DER HANDLUNG vorhergesagt. $\mathcal{P}$ gibt aber eine positive SPONTANE ERFOLGSMELDUNG in Form von „Ok <P> jetzt.“ Daraufhin wird die Einschätzung des ERFOLGS BEI DER AUSFÜHRUNG DER HANDLUNG revidiert. Die Wahrscheinlichkeit für einen hohen Erfolg ist jetzt höher. Die Einschätzung des Erfolgs und kein SPRECHERWECHSELSIGNAL machen einen hohen ZEITDRUCK unwahrscheinlicher. Aufgrund des eher hohen Erfolges ist eine hohe VAK wahrscheinlicher.

6.4.4 Zeitscheiben t_6 und t_7

In der Zeitscheibe t_6 stellt $\mathcal{S}$ wieder eine Frage: „Jetzt Motorhaube öffnen?“ $\mathcal{P}$ scheint von der Situation überfordert zu werden und gibt in t_7 eine schlechte Antwort, die in Form einer Gegenfrage geäußert wird [8]: „Wie rum <Ächz> %geht' n der auf? So rum?“ Die Abbildung

[8] Dieser Dialogausschnitt entspricht eher einer Instruktion/Gegenfrage-Situation. Da zur Zeit eine solche Situation in READY nicht modelliert werden kann, wird hier ein Frage/Antwort-Schema verwendet. Durch ansteigende Intonation der Äußerung des Mechanikers (angedeutet durch das Fragezeichen) kann die Äußerung

6.10 zeigt wieder den betreffenden Ausschnitt der Dialoghistorie.

Das dazugehörige DBN wird in Abbildung 6.14 (Seite 108) dargestellt. Die Antwort wird als nicht passend gewertet, außerdem verwendet $\mathcal{P}$ ORTSDEIKTISCHE ÄUSSERUNGEN, so daß ein geringer ERFOLG BEI DER KONZEPTUALISIERUNG sehr wahrscheinlich ist. Eine geringe DEUTLICHKEIT DER AUSSPRACHE führt dazu, daß der ERFOLG BEI DER FORMULIERUNG geringer eingeschätzt wird als zuvor.

Da ein geringer Erfolg wahrscheinlicher ist als vorhergesagt wurde, wird die VAK von $\mathcal{P}$ (auch aufgrund der eher hohen EMOTIONALEN BELASTUNG) geringer eingeschätzt. Ähnliches gilt für die Einschätzung des ZEITDRUCKS.

An dieser Stelle des Dialogs wurde wieder in der empirischen Untersuchung das Band gestoppt und die Versuchspersonen um ihre Einschätzungen bzgl. der Arbeitsgedächtnisbelastung und des Zeitdrucks der Anruferin gebeten. Acht von zwölf Versuchspersonen meinten, daß die Arbeitsgedächtnisbelastung der Anruferin hoch ist (vgl. Frage 3.5 in der Abb. 5.6, Seite 82). Sieben Versuchspersonen schätzten den Zeitdruck der Anruferin als hoch ein (vgl. Frage 3.6 in der Abb. 5.6). Diese Einschätzungen entsprechen den Inferenzen des Benutzermodells.

auch wie folgt interpretiert werden: „Können Sie jetzt die Motorhaube öffnen?“

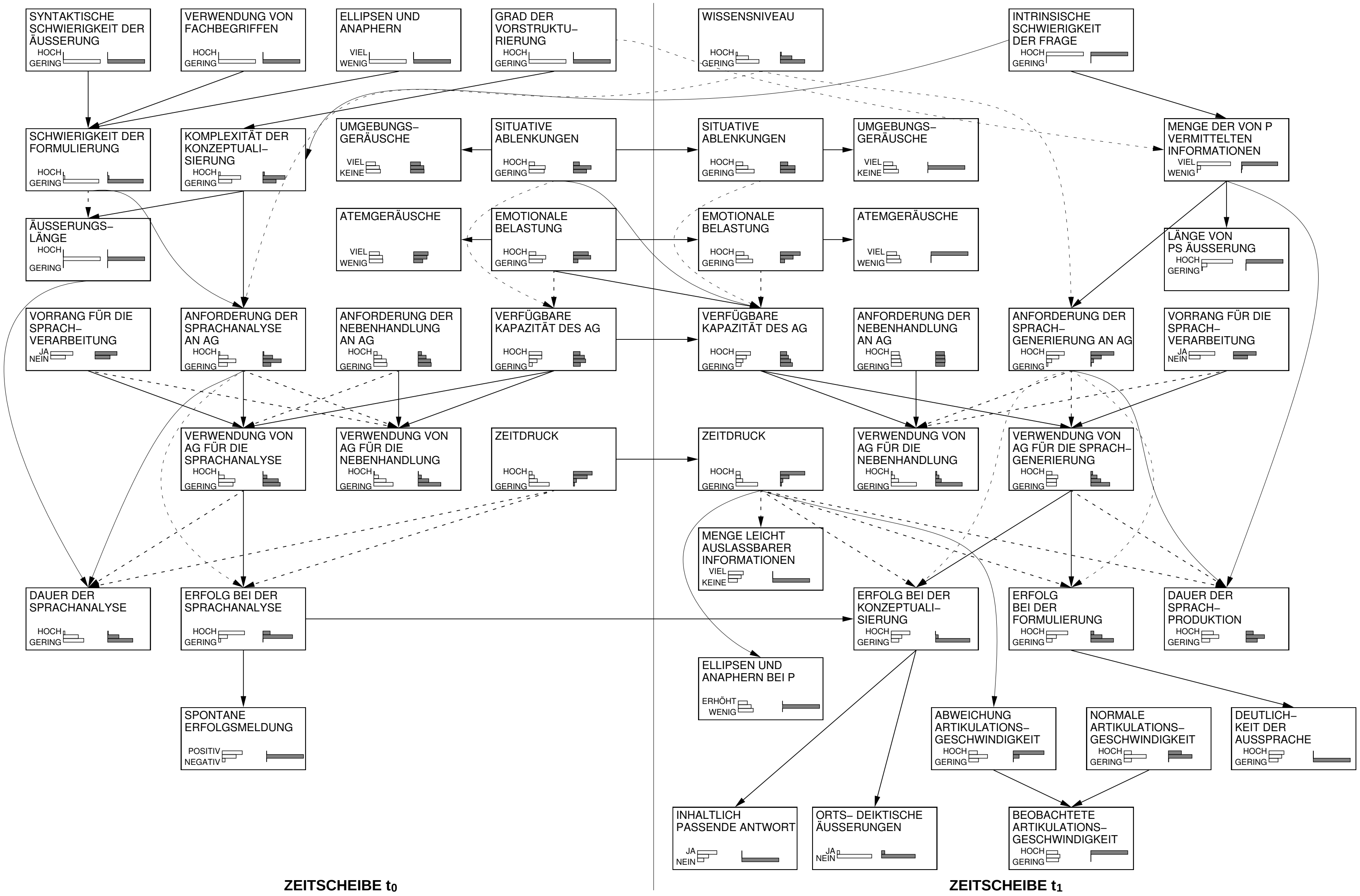

Abb. 6.11: *Zeitscheibe t_0 (Frage) $\mathcal{S}$: „Hier* READY, *was für eine Panne haben Sie?.“ Zeitscheibe t_1 (Antwort) $\mathcal{P}$: „OK, also <Z>, ich bin gefahr' und dann hat's <P> angefang' äh zu ruckeln <P> und ...“.*

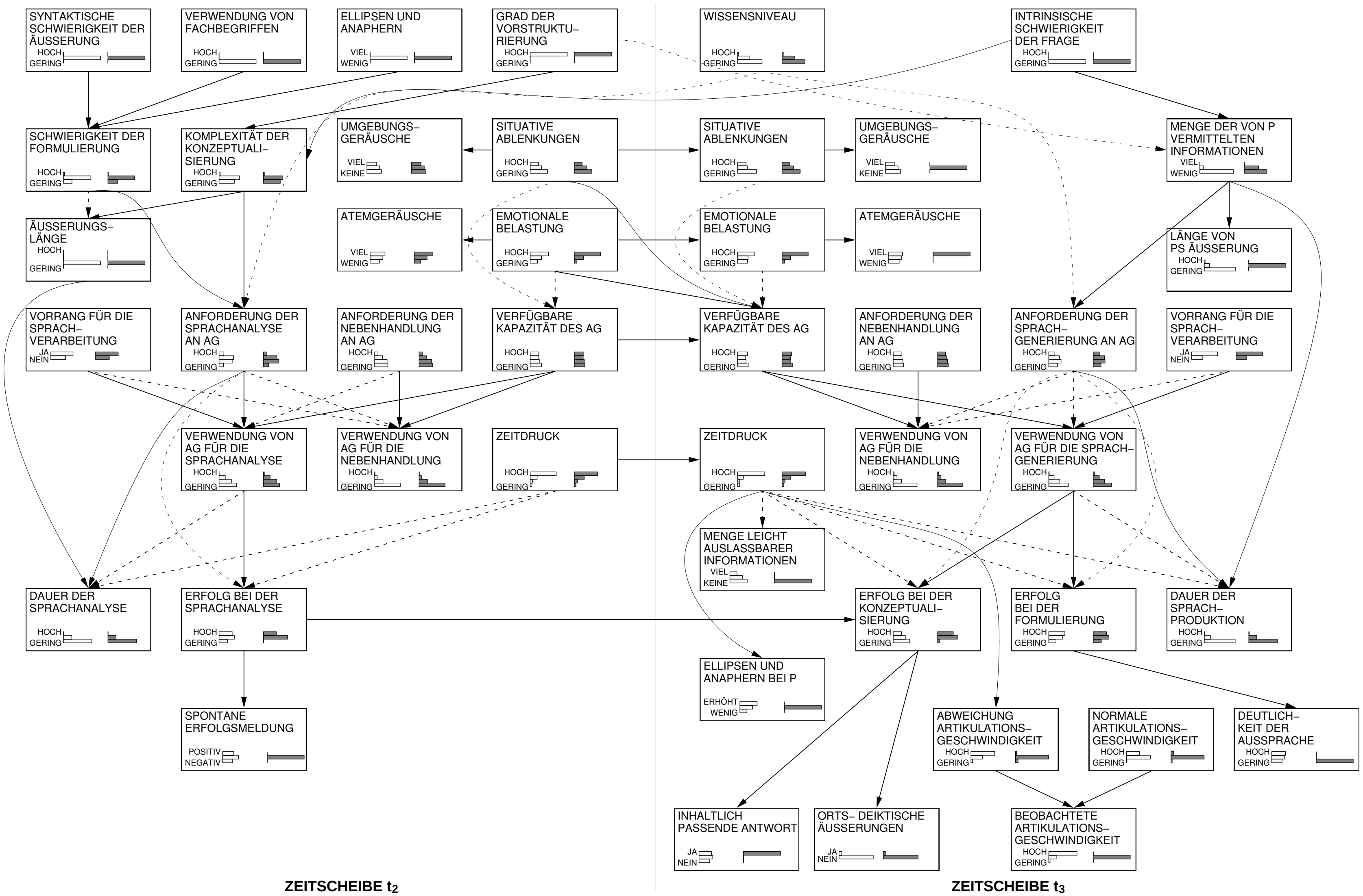

Abb. 6.12: *Zeitscheibe t_2 (Frage) $\mathcal{S}$: „Motorhaube schon geöffnet?" Zeitscheibe t_3 (Antwort) $\mathcal{P}$: „Nee, Motorhaube hab' ich noch nicht geöffnet.".*

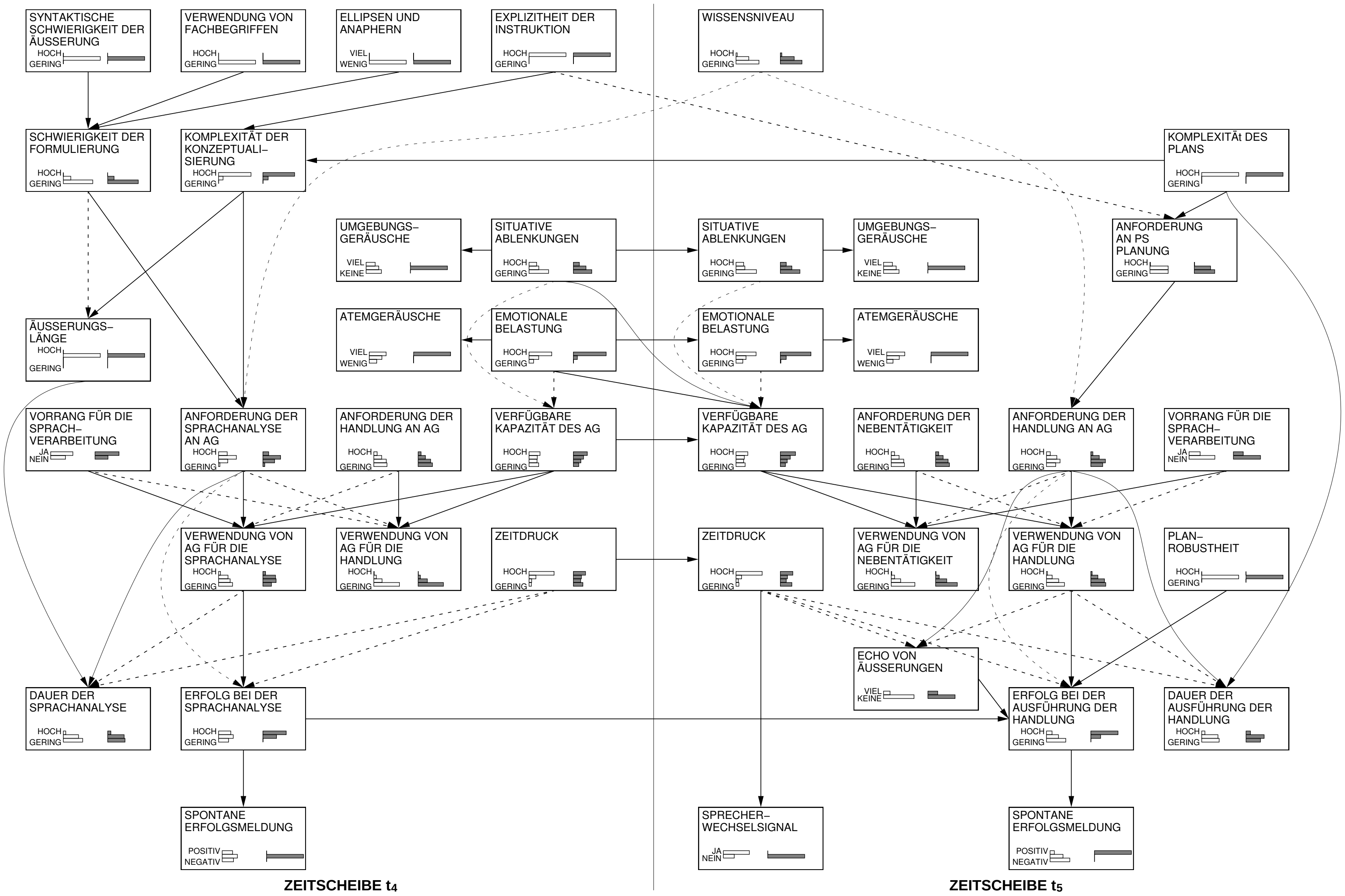

Abb. 6.13: *Zeitscheibe t_4 (Anweisung) $\mathcal{S}$: „Äh. Im Fußraum, d.h. unterm Lenkrad <P> links <P> ist ein Hebel. Daran ziehen“ Zeitscheibe t_5 (Handlung) $\mathcal{P}$: „Moment %grad %bitte <#Knistern> <P> OK <P> <P> jetzt.“.*

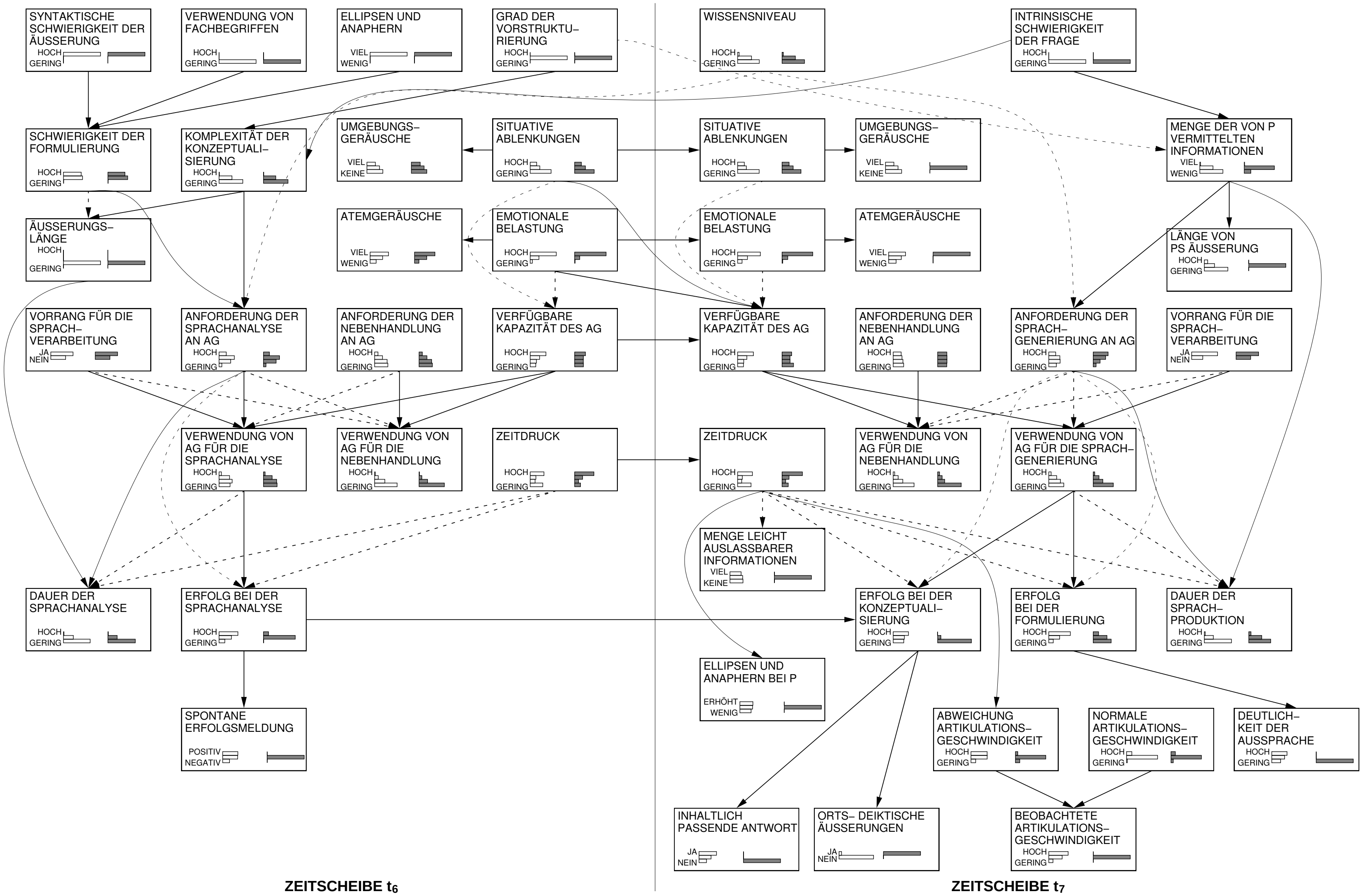

Abb. 6.14: *Zeitscheibe t_6 (Frage) $\mathcal{S}$: „Jetzt Motorhaube öffnen?“ Zeitscheibe t_7 (Antwort) $\mathcal{P}$: „<P> Wie rum <Ächz> %geht'n der auf? So rum?“.*

Kapitel 7

Zusammenfassung und Ausblick

In dieser Arbeit wurde gezeigt, wie ein Dialogsystem mit Hilfe des Inferenzmechanismus der dynamischen Bayesschen Netze (DBNs) die Beschränkungen der Zeit und des Arbeitsgedächtnisses eines Dialogpartners erkennen kann.

Die verwendeten Kausalzusammenhänge basierten einerseits auf den Ergebnissen von psychologischen Studien, in denen Einflüsse der Arbeitsgedächtnisbelastung auf die Sprachverarbeitung untersucht wurden, und andererseits auf Ergebnissen von zwei explorativen empirischen Untersuchungen, die im Rahmen dieser Arbeit durchgeführt wurden.

Die erste empirische Untersuchung fand in der Saarbrücker Feuerwehrnotrufzentrale (FNZ) statt. Anhand von Aufnahmen von Notrufdialogen schätzten elf Mitarbeiter der FNZ, die als Versuchspersonen dienten, die Arbeitsgedächtnisbelastung der Anrufer ein. Über retrospektives lautes Denken wurden ihre Annahmen über Kausalzusammenhänge ermittelt. Die Formulierungen der Versuchspersonen waren durchgehend durch Vagheit und Unsicherheit gekennzeichnet, so daß die betreffenden Kausalzusammenhänge nicht als deterministisch sondern als probabilistisch betrachtet werden. Die Kausalzusammenhänge wurden danach durch DBNs beschrieben und anhand von zwei Fallbeispielen getestet. Es wurde gezeigt, daß die Inferenzen der DBNs den Einschätzungen der Versuchspersonen entsprachen.

Die Studie in der FNZ lieferte eine Vielzahl von Ursachen und Folgen von Arbeitsgedächtnisbelastung. Aber in dieser Domäne treten Beschränkungen des Arbeitsgedächtnisses besonders stark auf. Andererseits ergab die Studie keine Hinweise auf Ursachen oder Folgen von Zeitdruck. Eine zweite Studie mit ähnlicher Methodologie hatte die Aufgabe zu untersuchen, welche Ursache-Wirkungs-Beziehungen im READY-Szenario zusätzlich berücksichtigt werden müssen. Als Versuchspersonen dienten dabei Studenten der Universität des Saarlandes.

Die Analyse der Aussagen der Versuchspersonen lieferte wieder ein Bild von Kausalzusammenhängen bzgl. der Arbeitsgedächtnisbelastung und des Zeitdrucks eines Dialogpartners, die zum Teil mit denen der Notrufstudie vergleichbar waren. Unterschiede in den Zusammenhängen entstanden dadurch, daß erstens die beiden Domänen unterschiedlich sind und zweitens die Erfahrung der Versuchspersonen teilweise anders waren.

Im Anschluß daran wurde ein Benutzermodell entwickelt, mit dem die Ressourcenlage des Benutzers erkannt und eingeschätzt werden kann. Mit dem Modell können Arbeitsgedächtnisbelastung und Zeitdruck eines Dialogpartners bei der Sprachanalyse, Sprachgenerierung und Ausführung einer Handlung eingeschätzt werden. Zusätzlich wird in dem Modell die Möglichkeit der Ausführung von Doppelaufgaben berücksichtigt. Im Anschluß daran wurde demonstriert, wie das Benutzermodell in READY eingesetzt wird. Es wurde gezeigt, daß die Einschätzungen des Modells bzgl. der Ressourcenlage des Dialogpartners mit den Einschätzungen der Versuchspersonen vergleichbar sind. Somit erfüllt das Benutzermodell die zu Beginn festgelegten Ansprüche.

7.1 Ausblick

Im Hinblick auf eine Erweiterung der in dieser Arbeit vorgestellten Thematik sind noch verschiedene Fragen offen, die in der nächsten Förderungsperiode (siehe Wahlster & Jameson, 1998) untersucht werden sollen:

(1) *Experimentelle Studien zur quantitativen Untersuchung der Zusammenhänge*
Die durchgeführten empirischen Untersuchungen lieferten hauptsächlich qualitative Ergebnisse zur Bestimmung der einzelnen Ursachen und Folgen von Arbeitsgedächtnisbelastung und Zeitdruck. Eine experimentelle Studie könnte die in dieser Arbeit beschriebenen Kausalzusammenhänge objektiv und quantitativ untersuchen und so Daten für die quantitative Modellierung liefern.

(2) *Lernen der dynamischen Bayesschen Netze*
Die quantitativen Abhängigkeiten, die in den DBNs definiert wurden, mußten grob geschätzt werden. Die empirischen Untersuchungen und die Literaturrecherche ergaben hierzu keine Daten. Es konnte zwar ein Systemverhalten erzielt werden, das nachvollziehbar und mit den Einschätzungen von Menschen vergleichbar war. Dennoch wäre es sinnvoll, wenn das Netz aus „seinen Erfahrungen" lernen könnte. Dazu müßten z.B. *Lernalgorithmen* für Bayessche Netze angewendet werden, die die bedingten Wahrscheinlichkeiten lernen.

(3) *Differenzierte Modelle des Arbeitsgedächtnisses*
Wie in Abschnitt 2.1 beschrieben wurde, ist das Bild des Arbeitsgedächtnisses als ein homogener Speicher mit einer gewissen Kapazität eine Vereinfachung der komplexen Modelle, die in der kognitiven Psychologie entwickelt wurden. Es könnte sich als sinnvoll erweisen, ein differenzierteres Modell des Arbeitsgedächtnisses zu verwenden. Ein solches Modell könnte z.B. von der phonologischen Schleife des Baddeleyschen Modells Gebrauch machen.

(4) *Andere Arten von Interaktionsmedien berücksichtigen*
In dieser Arbeit wurde vorausgesetzt, daß der Benutzer und READY in natürlicher Sprache kommunizieren. Andere Interaktionsarten sind auch denkbar: Eingabe über Tastatur oder Handschrift, Ausgabe als Text oder Grafiken. Auch hier spielen Arbeitsgedächtnisbelastung und Zeitdruck eine Rolle.

Anhang A

Daten zu den empirischen Untersuchungen und Modellierungen

A.1 Versuchsablauf der Feuerwehrstudie

A.1.1 Notruf 1 – Die Frau in St. Johann

$\mathcal{P}$: <#Stöhnen>.

$\mathcal{S}$: **Feuerwehr.**

$\mathcal{P}$: Ja, hier ist Frau Müller, Schopenhauerstraße 10. Bei mir brennt der Badeautomat. Da war heute morgen noch ein Installateur da. <lautes Atmen>. Jetzt brennt das Ding. <#Atmen>

Frage 1.1: *Wie gut hat sie die Situation beschrieben?*

Frage 1.2: *Warum denken Sie, beschreibt sie die Situation auf diese Weise?*

$\mathcal{S}$: **Schopenhauerstraße 10?**

$\mathcal{P}$: Schopenhauerstraße 10.

$\mathcal{S}$: **In welchem Stadtteil?**

Frage 1.3: *Wie schnell und deutlich wird sie die Frage beantworten können?*

Frage 1.4: *Hätten Sie sich in dieser Situation die Zeit genommen, nach dem Stadtteil zu fragen?*

$\mathcal{P}$: Bitte?

Frage 1.5: *Wie schnell und deutlich wird sie die Frage beantworten können?*

$\mathcal{S}$: **In welchem Stadtteil?**

$\mathcal{P}$: Ei, das is' hier in Saarbrücken <Z> in St. Johann, in der Nähe der LVA.

Frage 1.6: *Wie gut hat sie die Frage beantwortet?*

Frage 1.7: *Warum denken Sie, hat sie auf diese Weise geantwortet?*

$\mathcal{S}$: **Welche Etage?**

Frage 1.8: *Wie schnell wird sie die Frage beantworten können?*

$\mathcal{P}$: <%> hoch

Frage 1.9: *Die Antwort war ein bißchen undeutlich. Wie verstehen Sie das, was sie gesagt hat?*

$\mathcal{S}$: **Welche Etage is denn das?**

$\mathcal{P}$: Äh, die erste. <#Stöhnen>

Frage 1.10: *Warum hat sie nicht unmittelbar geantwortet, nachdem die Frage gestellt wurde?*

$\mathcal{S}$: **mache, mache, <P>**

$\mathcal{P}$: <#leise> oah schnell! <#lautes Atmen>

$\mathcal{S}$: **joh, mache 'se die Tür zu, wir kommen sofort, ne?**

$\mathcal{P}$: ja, es iss gut.

$\mathcal{S}$: **ja, tschüß.**

Frage 1.11: *Stellen Sie sich vor, Sie müßten nach der Alarmierung der Einsatzkräfte einige Anweisungen geben. Wie kompliziert könnten die Anweisungen sein, die sie jetzt noch gut verstehen und behalten könnte?*

Frage 1.12: *Stellen Sie sich vor, der Kollege fragt nach der Telefonnummer der Wohnung. Die Anruferin ist aber Gast in der Wohnung und muß daher die ihr nicht bekannte Nummer erst erfragen. Wieviele Ziffern hätte wohl die längste Nummer, die sie jetzt richtig durchgeben könnte?*

Frage 1.13: *Wie hätten Sie an der Stelle Ihres Kollegen das Gespräch evtl. schneller führen können?*

A.1.2 Notruf 2 – Der Mann in der Gaststätte

$\mathcal{S}$: <#>

$\mathcal{P}$: Bitte?

$\mathcal{S}$: **Feuerwehr.**

$\mathcal{P}$: Dieter Müller is mein Name. Un zwar <P> in de <#> <;redet mit jemanden anderes> direkt do vorne. nä, nä, do hinne, <P> Trifelsstroß. <;redet jetzt mit Feuerwehrmann> Äh, Trifelsstroß, owwe im Dachgeschoß is e Feuer.

Frage 2.1: *Wie gut hat er die Lage beschrieben?*

Frage 2.2: *Warum, denken Sie, hat er die Ortsbeschreibung auf diese Weise gegeben?*

Frage 2.3: *Inwieweit beeilt er sich bei der Durchgabe der notwendigen Informationen?*

$\mathcal{S}$: **Trifelsstraße?**

$\mathcal{P}$: Ja, <P>

$\mathcal{S}$: **Was für Hausnummer denn?**

Frage 2.4: *Wie schnell wird er die Frage beantworten?*

$\mathcal{P}$: 12.

$\mathcal{S}$: **Wie war noch ihr Name?**

$\mathcal{P}$: Dieter Müller, 47081511 DRK-Nummer.

Frage 2.5: *Wie gut hat er die Frage beantwortet?*

Frage 2.6: *Inwieweit beeilt er sich bei der Durchgabe der notwendigen Informationen?*

$\mathcal{S}$: **Geben's mir noch ihre Telefonnummer!**

$\mathcal{P}$: Äh, die Telefonnummer von hier? <%><P> Ich ruf aus einer Gaststätte an.

$\mathcal{S}$: **Aus einer Gaststätte. Genn 'se mir die Nummer von der Gaststätte!**

Frage 2.7: *Wieviele Ziffern hätte wohl die längste Nummer, die der Mann jetzt richtig durchgeben könnte?*

Frage 2.8: *Hätten Sie sich in dieser Situation die Zeit genommen, nach der Telefonnummer zu fragen?*

$\mathcal{P}$: <;Zur Wirtin> Könn 'se mir gerade die Gaststättennummer äh äh die Nummer? <P> Bitte? <P> 25 <P> <;Im Hintergrund werden Zahlen genannt>

$\mathcal{S}$: **27.**

$\mathcal{P}$: 25 <P>

$\mathcal{S}$: **27.**

Frage 2.9: *Wie gut hat er die Telefonnummer durchgegeben?*

Frage 2.10: *Warum hat er, nach Ihrer Meinung, Probleme, die Telefonnummer richtig durchzugeben?*

$\mathcal{S}$: **Also in der Trifelsstraße Nummer 12?**

$\mathcal{P}$: Bin vorm Haus dann.

$\mathcal{S}$: **Im Dachgeschoß?**

$\mathcal{P}$: Im Dachgeschoß owwe.

$\mathcal{S}$: **Ja, wir kommen sofort.**

$\mathcal{P}$: Alles klar.

Frage 2.11: *Wurde die Gesprächsführung beeinträchtigt, wenn der Mann sich noch mit anderen unterhält?*

Frage 2.12: *Wie hätten Sie an der Stelle Ihres Kollegen das Gespräch evtl. schneller führen können?*

A.2 Probleme des retrospektiven lauten Denkens

Notrufe werden oft unter Zeitdruck geführt. Daher kann man annehmen, daß die Mitarbeiter der FNZ nicht lange über die Belastung des Dialogpartners nachdenken können. Ein Aspekt der Untersuchung war, die Einschätzungen von den Versuchspersonen herauszulocken, die in echten Notrufsituationen vorkommen. Deshalb wurde den Versuchspersonen wenig Zeit für die Einschätzung gegeben. Die Verbalisierung der Denkprozesse sollte durch das *retro-*

spektive laute Denken[1] herbeigeführt werden (vgl. Ericsson & Simon, 1993). Im Idealfall sollte die retrospektive Verbalisierung unmittelbar nach der Beendigung der Einschätzung gegeben werden, solange sich noch viele Informationen darüber im Kurzzeitgedächtnis der Versuchsperson befinden. Aber der Prozeß des lauten Denkens kann auch zu Problemen führen, die bei der Analyse der Daten berücksichtigt werden müssen. Folgende allgemeine Schwierigkeiten können auftreten (aus Ericsson & Simon, 1993):

(1) Ericsson und Simon: Es werden auf Gedächtnisstrukturen zugegriffen, die nicht durch den gerade beendeten kognitiven Prozeß kreiert wurden.

Vorliegende Studie: Die Versuchspersonen generierten oftmals Verbalisierungen, die nicht mit den Einschätzungen übereinstimmten:

Frage 1.3: Wie schnell und deutlich wird sie die Frage beantworten können?

Vp 4: *Ich schätze mal 5 Sekunden.* – retrospektives Denken – *Ich habe überlegt, wo der Stadtteil ist und daraufhin habe ich geschlossen, sie ist aufgeregt [...]*

(2) Ericsson und Simon: Ältere oder untergeordnete Informationen, die mit dem kognitiven Prozeß zu tun haben, werden mitverbalisiert oder mit den Informationen über den Prozeß vermischt.

Vorliegende Studie: Die Versuchspersonen verglichen neue Situationen mit älteren, bekannten Situationen:

Frage 1.5: Wie schnell und deutlich wird sie die Frage beantworten können?

Vp 4: *Jetzt dauert es etwas länger.* – retrospektives Denken – *Gut, ich stelle mir das vor, weil wir desöfteren schon solche Fälle gehabt haben, und das macht die Praxis.*

(3) Ericsson und Simon: Wird die Versuchsperson aufgefordert, über Informationen zu berichten, die nicht beobachtet wurden (z.B. durch Beeinflussung des Fragestellers), so basiert ihre Antwort möglicherweise nicht auf direkte Inhalte des Arbeitsgedächtnisses. Oftmals wird auch nicht mit „Ich weiß nicht." geantwortet, sondern es wird eine Antwort auf der Basis von Informationen generiert, die durch das Langzeitgedächtnis unterstützt werden. Dieser Prozeß wird durch Warum-Fragen besonders unterstützt.

Vorliegende Studie: Die Versuchspersonen haben, insbesondere nach der Frage „Was haben Sie dabei gedacht", sich die einleitenden Instruktionen erinnert und ihre Antworten daran angepaßt:

Frage 1.3: „Wie schnell und deutlich wird sie die Frage beantworten können?"

Vp 7: *Sofort.* – retrospektives Denken – *Das ist eine Frage, Was soll ich mir dabei denken? - Ach so, jetzt versteh ich. Sie war nicht so <P> so aufgeregt war sie wiederum nicht.*

[1] Ein anderer Typ von Verbalisierung zur Reflexion kognitiver Prozesse ist das „direkte Verbalisieren" von Gedächtnisinhalten, auch „lautes Denken" genannt. Diese Technik wird häufig bei der Aquisition von Expertenwissen benutzt.

Desweiteren muß darauf geachtet werden, daß die Versuchsperson ihre spontan gegebene Einschätzung in der retrospektiven Verbalisierung nicht generalisiert. Folgende Gründe können dafür vorliegen und müssen daher berücksichtigt werden (aus Ericsson & Simon, 1993):

(1) Ericsson und Simon: Die Versuchsperson ist sich ihrer benutzten generellen Prozeduren (Programme) bewußt und beschreibt diese direkt während der Beantwortung der Frage ohne Referenz auf ihre spezifischen Gedankenabläufe.

Vorliegende Studie: Allgemeine Aspekte von Notrufen wurden mit der spezifischen Situation vermischt:

Frage 1.4: „Hätten Sie sich in dieser Situation die Zeit genommen, nach dem Stadtteil zu fragen?"

Vp 4: *Ja, daß ist richtig. Das ist wichtig, weil wir haben verschiedene Straßen, die doppelt vorkommen.* – retrospektives Denken – *Tja, die genaue Einsatzstelle muß man genau wissen, und das läuft halt so routinemäßig ab und aus dem Grund geht das automatisch.*

(2) Ericsson und Simon: Die Versuchsperson kann auf verschiedene Arten von vorhandenen Informationen (z.B. generelles Wissen) zurückgreifen, um eine Verbalisierung zu generieren, die eine generelle Prozedur oder Strategie beschreibt.

Vorliegende Studie: Die Versuchspersonen machten häufig allgemeine Aussagen über Notrufe, anstatt die Gedankenabläufe bei der Einschätzung zu beschreiben:

Frage 1.1: „Wie gut hat sie die Situation beschrieben"

Vp 5: *Also ich würde sagen, ersteinmal hat sie eine schlechte Aussprache und die Situation ist schlecht beschrieben.* – retrospektives Denken – *Dann muß ich also nachfragen. Vor allen Dingen der Name, der Wohnort, die Straße, weil bei uns im Einzugsgebiet gibt es Straßen, die öfter in verschiedenen Orten vorkommen.*

Im Idealfall sollten nur Verbalisierungen für die Analyse benutzt werden, die auf den Gedankenabläufen bei der Beantwortung der Fragen beruhen. Die oben genannten Problemfälle müssen daher berücksichtigt werden.

A.3 Versuchsablauf der READY-Studie

Dialog 1

$\mathcal{P}$: Waschbüsch, hallo. Ich bin der Jens und ich steh hier unten in der Saarbrücker Straße und in meinem Auto da riecht der Innenraum seit einer Viertelstunde nach Benzin während der Fahrt.

Frage 1.1: *Inwieweit war der Anrufer durch die Beschreibung der Situation beansprucht?*

Frage 1.2: *Wie hoch ist der Grad der Eile des Anrufers?*

$\mathcal{S}$**: <Äh> auf der linken Seite ist <äh> ein Benzinfilter. So'n durchsichtiges Teil.** $\mathcal{P}$: Ja. $\mathcal{S}$**: Siehst du das?** $\mathcal{P}$: Ja. $\mathcal{S}$**: Ist da, da ist 'ne Leitung dran.** $\mathcal{P}$: Ja. $\mathcal{S}$**: Guck mal, ob die überall noch <äh> befestigt ist oder ob die irgendwo abgerissen ist oder ein Leck hat. Hallo?** $\mathcal{P}$: Ja vorne, ganz vorne ist es ein bißchen, riecht's so ein bißchen nach Benzin. $\mathcal{S}$**: <Ah>.** $\mathcal{P}$: Das wackelt da dran. Das wackelt da. Ist nicht so ganz fest, <eh>. $\mathcal{S}$**: <Äh> ist da eine Schelle drum? Oder irgendwas zum anziehen, zum befestigen?**

Frage 1.3: *Inwieweit war der Anrufer durch die Frage beansprucht?*

Frage 1.4: *Wie hoch ist der Grad der Eile des Anrufers?*

Dialog 2

$\mathcal{P}$: <Mhm> an der Saarbrücker Straße. Da ist mir beim Fahren aufgefallen, daß am linken Vorderreifen, wenn ich fahr klingt plötzlich hat's angefangen zu klackern. So klack klack klack klack und ist mir aufgefallen, wenn ich Gas gegeben hab' oder wenn ich bremse, dann ist das klackern nicht mehr.

Frage 2.1: *Wie hoch ist der Grad der Eile des Anrufers?*

Frage 2.2: *Inwieweit war der Anrufer durch die Beschreibung der Situation beansprucht?*

$\mathcal{S}$**: Beim Gasgeben wird's Klackern weniger.** $\mathcal{P}$: Beim Bremsen. $\mathcal{S}$**: Und beim Bremsen.** $\mathcal{P}$: Ja. $\mathcal{S}$**: Hast du Radkappen drauf?** $\mathcal{P}$: Nein die hab' ich schon abgemacht.

Frage 2.3: *Wie hoch ist der Grad der Eile des Anrufers?*

$\mathcal{S}$**: Sind die Radschrauben richtig angezogen?** $\mathcal{P}$: Müßte ich mal gucken. Ich hol mal gerade ... $\mathcal{S}$**: Ja.** $\mathcal{P}$: Ich hab nur mal es Kreuz gerade geholt. <%> $\mathcal{S}$**: Ja.** $\mathcal{P}$: Die Radschrauben sind also ... $\mathcal{S}$**: Ja.** $\mathcal{P}$: ... locker. Ich zieh' sie jetzt mal fest. $\mathcal{S}$**: Äh, sind ganz locker?** $\mathcal{P}$: Die waren total locker. Ich bin sie gerade am festdrehen <%>. $\mathcal{S}$**: Bitte?** $\mathcal{P}$: Also drei von den Rad als drei von den Muttern waren total locker. $\mathcal{S}$**: Äh, guck' mal an der Lauffläche vom Reifen, falls, ich weiß es nicht <%> falls es das nicht gewesen ist, an der Lauffläche mit 'm Reifen mit 'm Handschuh mal drübergehen, ob da irgendwas steckt, oder ob da ein Stein drin ist, oder ein Nagel oder sonstwas.**

Frage 2.4: *Inwieweit war der Anrufer durch die Äußerung beansprucht?*

Frage 2.5: *Wie hoch ist der Grad der Eile des Mechanikers?*

Dialog 3

$\mathcal{S}$**: Hodab.** $\mathcal{P}$: Ok, also <Z>, ich bin gefahr' und dann hat's <P> angefang' äh zu ruckeln <P> und <Z> ich kann soviel Gas geb'n wie ich will, es wird nicht schneller, es Auto. Und jetzt?

Frage 3.1: *Inwieweit war die Anruferin durch die Beschreibung der Situation beansprucht?*

Frage 3.2: *Wie hoch ist der Grad der Eile der Anrufering?*

$\mathcal{S}$: **Aber er läuft noch?** $\mathcal{P}$: Ach so ja %und %es %geht ab und zu <%> aus. Ne? bf $\mathcal{S}$: Ab und zu geht er aus? <P>. $\mathcal{P}$: Ja. $\mathcal{S}$: **Motorhaube schon geöffnet?**

Frage 3.3: *Inwieweit war die Anruferin durch die Frage beansprucht?*

Frage 3.4: *Wie hoch ist der Grad der Eile des Mechanikers?*

$\mathcal{P}$: Nee, Motorhaube hab' ich noch nicht geöffnet. $\mathcal{S}$: **Äh <Z>. Im <Z> Fußraum, daß heißt <P> unterm Lenkrad <P> links <P> is'n Hebel. <P> daran ziehen.** $\mathcal{P}$: Moment %grad %bitte <#Knistern> <P> OK <P> jetzt. $\mathcal{S}$: **Jetzt Motorhaube öffnen?** $\mathcal{P}$: <P> Wie rum <Ächz> %geht' n der auf? So rum?

Frage 3.5: *Inwieweit war die Anruferin durch die Situation beansprucht?*

Frage 3.6: *Wie hoch ist der Grad der Eile der Anruferin?*

$\mathcal{S}$: **Ai, in der Mitte der Haube <oah> <!Stöhnen> zwischen Haube und Kühlergrill durchgreifen.** $\mathcal{P}$: Ja ich hab's. $\mathcal{S}$: **Da ... Da ... ja?** $\mathcal{P}$: <%> <P> wie geht' n ... wie hält denn das Ding <Ächz>? $\mathcal{S}$: **Ach so, äh, <Z> <P> äh auf der rechten Seite liegt <Z> glaube ich <Z> ein Stab, auf der rechten Seite. Am ... äh, äh auf'm Kotflügel, so zu sagen.**

Frage 3.7: *Inwieweit war die Anruferin durch die Äußerung beansprucht?*

Frage 3.8: *Wie hoch ist der Grad der Eile des Mechanikers?*

$\mathcal{P}$: <%> <;T> $\mathcal{S}$: **Bitte? <P> Ja?** $\mathcal{P}$: <;T> <%> Ah, oh Scheiße. Oh ihr Leut' <Lachen>. $\mathcal{S}$: **<lachen>** $\mathcal{P}$: <Lachen> <%> Ai, wo denn <Lachen>, ja ok, ja jetzt. $\mathcal{S}$: **Ja?** $\mathcal{P}$: Ja <P> und jetzt? $\mathcal{S}$: **Äh, <Z> <P> in der Mitte <Z> ist <Z> ein großes rundes Ding, <P> so'n Luftfilter <P>. Ah, nee, Moment, <P> Stop. <P> Äh, weiter rechts ist eine orangene Kappe.**

Frage 3.9: *Inwieweit war die Anruferin durch die Äußerung beansprucht?*

Dialog 4

$\mathcal{P}$: [...] Und <äh> %Motorhaub' und da steigt auf einmal so ein komischer Wasserdampf auf, ja. $\mathcal{S}$: **Ja.** $\mathcal{P}$: Und dann trau' ich mich nicht, weiterzu <unverständlich> und zu <unverständlich> und auf einmal ist da ein so ein. $\mathcal{S}$: **Was?** $\mathcal{P}$: Ist so ein rotes Lämpchen, das leuchtet unten links im Cockpit bei mir.

Frage 4.1: *Inwieweit war der Anrufer durch die Beschreibung der Situation beansprucht?*

Frage 4.2: *Wie hoch ist der Grad der Eile des Anrufers?*

$\mathcal{S}$: **<Äh> Motorhaube offen?** $\mathcal{P}$: Ja, Moment, mach' ich. Ah, so, ist offen. $\mathcal{S}$: **Ganz vorne, direkt da wo ungefähr wo der Schnapper ist für die Motorhaube ist ein Deckel.** $\mathcal{P}$: Aha. $\mathcal{S}$: **Gefunden?** $\mathcal{P}$: Ja, hab' ich. $\mathcal{S}$: **So.** $\mathcal{P}$: Hoff' ich mal. $\mathcal{S}$: **<Äh> schwarzer Deckel.** $\mathcal{P}$: Schwarzer Deckel. $\mathcal{S}$: **Ja?** $\mathcal{P}$: Ja. $\mathcal{S}$: **<Äh> mal ganz vorsichtig oder mal probieren, ob der offen ist. Also mal zudrehen.** $\mathcal{P}$: <Äh> nein, der, der ist zu, ja. $\mathcal{S}$: **Der ist zu?** $\mathcal{P}$: Der ist zu. $\mathcal{S}$: **<Äh> mal vorsichtig aufdrehen. Nicht ganz, nur mal lösen.** $\mathcal{P}$: Ja, er ist gelöst. $\mathcal{S}$: **Kommt Wasserdampf raus?**

Frage 4.3: *Inwieweit wird der Anrufer durch die Frage beansprucht?*

$\mathcal{P}$: <Öh>, nein. $\mathcal{S}$**: Gut. Ganz abdrehen.** $\mathcal{P}$: Ach, so, es zischt. $\mathcal{S}$**: Es zischt, es zischt. Ah. Okey, <äh> ja, mal zart weiter aufdrehen.** $\mathcal{P}$: Aha, ich drehe, ich drehe. $\mathcal{S}$**: Langsam.** $\mathcal{P}$: So. $\mathcal{S}$**: Zischt's immer noch?** $\mathcal{P}$: Ich dreh' immer noch. Zischt immer noch, ja. $\mathcal{S}$**: Zischt's viel?** $\mathcal{P}$: Bitte? $\mathcal{S}$**: Ob es viel zischt.** $\mathcal{P}$: Ja, es zischt viel und qualmt hier raus soundso viel.

Frage 4.4: *Wie hoch ist der Grad der Eile des Mechanikers?*

Frage 4.5: *Inwieweit wird der Anrufer durch die Äußerung beansprucht?*

Dialog 5

$\mathcal{S}$**: Hodab.** $\mathcal{P}$: [...] und zwar <äh> es es Auto es Licht wird schwächer, es Radio wird geht langsam aus und ich bin stehngeblieben.

Frage 5.1: *Wie hoch ist der Grad der Eile des Anrufers?*

Frage 5.2: *Inwieweit war der Anrufer durch die Beschreibung der Situation beansprucht?*

$\mathcal{S}$**: Auf der linken Seite ist 'ne Lichtmaschine. Das heißt, da ist ein Rad dran <äh>, da sollt ein Keilriemen rundrum gehen. Das Rad läßt sich wahrscheinlich jetzt von Hand drehen. Gefunden?**

Frage 5.3: *Inwieweit wird der Anrufer durch die Äußerung beansprucht?*

Frage 5.4: *Wie hoch ist der Grad der Eile des Mechanikers?*

$\mathcal{P}$: Einen Moment, ja. Keilriemen gefunden, ja. $\mathcal{S}$**: Ist der Keilriemen noch drauf?** $\mathcal{P}$: hä? $\mathcal{S}$**: Hallo?** $\mathcal{P}$: Ja, er ist drauf. $\mathcal{S}$**: Ist drauf. Ist er locker?** $\mathcal{P}$: Nein, fest. $\mathcal{S}$**: Ist fest. <Äh> an <äh> an der Lichtmaschine hinten gehen Stecker dran.** $\mathcal{P}$: Wo ist die Lichtmaschine?

Frage 5.5: *Inwieweit wird der Anrufer durch die Äußerung beansprucht?*

$\mathcal{S}$**: <Ei äh> da, wo das Rad dran ist. Wo der Keilriemen <%rundum>.** $\mathcal{P}$: Ja, hab' ich, ja. $\mathcal{S}$**: Da hintendran gehen Stecker dran. Sind alle Stecker noch dran?** $\mathcal{P}$: Nein. $\mathcal{S}$**: Draufstecken.**

Frage 5.6: *Wie hoch ist der Grad der Eile des Anrufers?*

Frage 5.7: *Wie hoch ist der Grad der Eile des Mechanikers?*

A.4 Definition der Knoten der Modelle $\mathcal{N}1$ und $\mathcal{N}2$

Im folgenden werden die Definitionen der einzelnen Knoten angegeben, die in den beiden Modellen $\mathcal{N}1$ und $\mathcal{N}2$ verwendet werden. Die bedingten Wahrscheinlichkeiten in der CPT eines Kindknotens werden entweder direkt in Form einer Verbindungsmatrix angegeben oder durch eine Funktion berechnet. Für die a-priori Wahrscheinlichkeiten eines Wurzelknotens gilt ähnliches. Dabei werden folgende Abkürzungen benutzt: $P(X = x \mid Y = y)$ ist die Wahrscheinlichkeit für $X = x$ bedingt $Y = y$. $M_A(x, y)$ bezeichnet das Element der Matrix M_A in Zeile x und Spalte y. $N_{\mu,\sigma}(x)$ beschreibt die Normalverteilung mit Erwartungswert μ und Standardabweichung σ.

GRÖSSE DES UNFALLS (G) *statisch*

Eltern: –

Zustände: (5) klein, ..., groß

A-priori: $P(G = g) = N_{\mu,\sigma}(g) \quad \text{mit} \quad \mu = 0,5, \quad \sigma = 0,25$

Beschreibung: Die Größe des Unfalls ist eher mittel.

ERFAHRUNG MIT NOTFÄLLEN (E) *statisch*

Eltern: –

Zustände: (5) wenig, ..., viel

A-priori: $P(E = e) = N_{\mu,\sigma}(e) \quad \text{mit} \quad \mu = 0,3, \quad \sigma = 0,25$

Beschreibung: $\mathcal{P}$ hat eher mittel bis wenig Erfahrung mit Notfällen.

HILFLOSIGKEIT (H) *temporär*

Eltern: GRÖSSE DES UNFALLS (G), ERFAHRUNG MIT NOTFÄLLEN (E)

Zustände: (2) nein, ja

CPT: $P(H = h \mid G = g, E = e)$

$$= \begin{cases} \frac{exp(g-e)}{1+exp(g-e)} & \text{, wenn h=ja} \\ 1 - \frac{exp(g-e)}{1+exp(g-e)} & \text{, wenn h=nein} \end{cases}$$

Beschreibung: Je größer der Unfall ist und je weniger $\mathcal{P}$ Erfahrung mit Notfällen hat, desto wahrscheinlicher ist $\mathcal{P}$s Hilflosigkeit.

GESPRÄCH MIT ANDEREN PERSONEN (G) *temporär*

Eltern: –

Zustände: (2) nein, ja

A-priori: $\begin{bmatrix} 0,80 & 0,20 \end{bmatrix}$

Beschreibung: Ein Gespräch mit anderen Personen während des Notrufes ist nicht wahrscheinlich.

GESPRÄCHSGERÄUSCHE (H) *temporär*

Eltern: GESPRÄCH MIT ANDEREN PERSONEN (G)

Zustände: (2) nein, ja

CPT: $P(H = h \mid G = g) = M_{H|G}(h, g)$

$$M_{H|G} = \begin{bmatrix} 0,99 & 0,01 \\ 0,01 & 0,99 \end{bmatrix}$$

Beschreibung: Wenn $\mathcal{P}$ ein Gespräch mit anderen Personen führt, dann hört man dies sehr wahrscheinlich.

ANRUFORT IST EINE GASTSTÄTTE (G) *statisch*

Eltern: –

Zustände: (2) nein, ja

A-priori: $\begin{bmatrix} 0,50 & 0,50 \end{bmatrix}$

Beschreibung: Es gibt kein explizites Wissen darüber, ob der Anrufort ist eine Gaststätte ist.

HINTERGRUNDGERÄUSCHE EINER GASTSTÄTTE (H) *temporär*

Eltern: ANRUFORT IST EINE GASTSTÄTTE (G)

Zustände: (2) nein, ja

CPT: $P(H = h \mid G = g) = M_{H|G}(h, g)$

$$M_{H|G} = \begin{bmatrix} 0,90 & 0,10 \\ 0,02 & 0,98 \end{bmatrix}$$

Beschreibung: Wenn der Anrufort ist eine Gaststätte ist, dann wird man mit hoher Wahrscheinlichkeit (98%) typische Hintergrundgeräusche hören.

ALKOHOLEINFLUSS (A) *statisch*

Eltern: ANRUFORT IST EINE GASTSTÄTTE (G)

Zustände: (2) nein, ja

CPT: $P(A = a \mid G = g) = M_{A|G}(h, g)$

$$M_{A|G} = \begin{bmatrix} 0,82 & 0,18 \\ 0,15 & 0,85 \end{bmatrix}$$

Beschreibung: Ist der Anrufort ist eine Gaststätte, ist es wahrscheinlich, daß $\mathcal{P}$ betrunken ist.

AUFREGUNG (A_t) *dynamisch*

Eltern: AUFREGUNG$_{t-1}$ (A_{t-1})

Zustände: (2) gering, hoch

CPT: $P(A_t = a1 \mid A_{t-1} = a2) = M_{A_{t-1}|A_{t-1}}(a1, a2)$

$$M_{A_{t-1}|A_{t-1}} = \begin{bmatrix} 0,85 & 0,15 \\ 0,15 & 0,85 \end{bmatrix}$$

Beschreibung: $\mathcal{P}$s Aufregung kann mit der Zeit nachlassen, d.h. wenn $\mathcal{P}$ in $t-1$ aufgeregt war, ist $\mathcal{P}$ in t nur noch mit 85% aufgeregt.

in Zeitscheibe t_0:

Eltern: HILFLOSIGKEIT (H)

CPT: $P(A_0 = a \mid H = h) = M_{A_0|H}(a, h)$

$$M_{A_0|H} = \begin{bmatrix} 0,80 & 0,20 \\ 0,20 & 0,80 \end{bmatrix}$$

Beschreibung: Wenn $\mathcal{P}$ hilflos ist, dann ist $\mathcal{P}$s Aufregung wahrscheinlich hoch (80%, 2. Zeile, 2. Spalte). Dieser Knoten ersetzt in Zeitscheibe t_0 den dynamischen Knoten AUFREGUNG.

ABLENKUNG (A_t) *dynamisch*

Eltern: ABLENKUNG$_{t-1}$ (A_{t-1}), GESPRÄCH MIT ANDEREN PERSONEN (G)

Zustände: (2) gering, hoch

CPT: $P(A_t = a1 \mid A_{t-1} = a2, G = g) = M_{A_t|A_{t-1},G=g}(a1, a2)$

$$\begin{array}{cc} M_{A_t|A_{t-1},G=nein} & M_{A_t|A_{t-1},G=ja} \\ \begin{bmatrix} 0,95 & 0,05 \\ 0,30 & 0,70 \end{bmatrix} & \begin{bmatrix} 0,15 & 0,85 \\ 0,10 & 0,90 \end{bmatrix} \end{array}$$

Beschreibung: Wenn $\mathcal{P}$ sich mit anderen unterhält und in $t-1$ abgelenkt war, dann ist $\mathcal{P}$ sehr wahrscheinlich in t abgelenkt.

in Zeitscheibe t_0:

Eltern: GESPRÄCH MIT ANDEREN PERSONEN (G)

CPT: $P(A_0 = a \mid G = g) = M_{A_0|G}(a, g)$

$$M_{A_0|G} = \begin{bmatrix} 0,82 & 0,18 \\ 0,15 & 0,85 \end{bmatrix}$$

Beschreibung: Wenn $\mathcal{P}$ ein Gespräch mit anderen Personen führt, dann ist $\mathcal{P}$ wahrscheinlich abgelenkt (85%). Dieser Knoten ersetzt in Zeitscheibe t_0 den dynamischen Knoten ABLENKUNG.

VAK$_t$ (V_t) (Modell $\mathcal{N}1$) *dynamisch*

Eltern: VAK$_{t-1}$ (V_{t-1}), AUFREGUNG$_{t-1}$ (A_{t-1}), AUFREGUNG$_t$ (A_t), ABLENKUNG$_{t-1}$ (B_{t-1}), ABLENKUNG$_t$ (B_t)

Zustände: (5) gering, ..., hoch

CPT: $P(V_t = v1 \mid V_{t-1} = v2, A_t = a1,$
$A_{t-1} = a2, B_= - b1, B_{t-1} = b2) = N_{\mu,\sigma}(v1)$

$$dA = \begin{cases} 0,0 & \text{, wenn} \quad a2 = a1 \\ 0,3 & \text{, wenn} \quad a2 > a1 \\ -0,3 & \text{, wenn} \quad a2 < a1 \end{cases}$$

$$dB = \begin{cases} 0,0 & \text{, wenn} \quad b2 = b1 \\ 0,3 & \text{, wenn} \quad b2 > b1 \\ -0,3 & \text{, wenn} \quad b2 < b1 \end{cases}$$

$$\mu = v2 + v2 * (dA + dB), \quad \sigma = 0,15$$

Beschreibung: Hat sich die Aufregung oder Ablenkung von $\mathcal{P}$ erhöht, dann verringert sich $\mathcal{P}$s VAK.

in Zeitscheibe t_0:

Eltern: AUFREGUNG$_0$ (A_0), ABLENKUNG$_0$ (B_0)

CPT: $P(V_0 = v \mid A_0 = a, B_0 = b) = N_{\mu,\sigma}(v)$
$\mu = M_{A_0,B_0}(a, b), \quad \sigma = 0,25$

$$M_{A_0,B_0} = \begin{bmatrix} 0,80 & 0,40 \\ 0,40 & 0,15 \end{bmatrix}$$

Beschreibung: Je mehr $\mathcal{P}$ aufgeregt und abgelenkt ist, desto geringer ist $\mathcal{P}$s VAK. Dieser Knoten ersetzt in Zeitscheibe t_0 den dynamischen Knoten VAK.

VAK$_t$ (V_t) (Modell $\mathcal{N}2$) *dynamisch*

Eltern: VAK$_{t-1}$ (V_{t-1}), ALKOHOLEINFLUSS (A), ABLENKUNG$_{t-1}$ (B_{t-1}), ABLENKUNG$_t$ (B_t)

Zustände: (5) gering, ..., hoch

CPT: $P(V_t = v1 \mid V_{t-1} = v2, A = a,$
$B_t = b1, B_{t-1} = b2) = N_{\mu,\sigma}(v1)$

$X = v2 - (b1 - b2),$
Intervallgrenzen von V berücksichtigt

$$\mu = \begin{cases} 1,0 * X & \text{, wenn} \quad a = nein \\ 0,6 * X & \text{, wenn} \quad a = ja \end{cases}$$

$\sigma = 0.15$

Beschreibung: Die Verrringerung von $\mathcal{P}$s VAK resultiert entweder aus einem vorhandenen Alkoholeinfluß oder, wenn die Ablenkung in t höher ist als in $t-1$.

in Zeitscheibe t_0:

Eltern: ALKOHOLEINFLUSS (A), ABLENKUNG$_0$ (B_0)

CPT: $P(V_0 = v \mid A = a, B_0 = b) = N_{\mu,\sigma}(v)$
$\mu = M_{A,B_0}, \quad \sigma = 0,25$

$$M_{A,B_0} = \begin{bmatrix} 0,50 & 0,40 \\ 0,35 & 0,20 \end{bmatrix}$$

Beschreibung: Ist $\mathcal{P}$ betrunken und je höher $\mathcal{P}$ abgelenkt ist, desto geringer ist $\mathcal{P}$s VAK.

ATEMGERÄUSCHE (A) *temporär*

Eltern: AUFREGUNG (B)

Zustände: (2) wenig, viel

CPT: $P(A = a \mid B = b) = M_{A|B}(a, b)$

$$M_{A|B} = \begin{bmatrix} 0,80 & 0,20 \\ 0,20 & 0,80 \end{bmatrix}$$

Beschreibung: Ist $\mathcal{P}$s Aufregung hoch, dann sind wahrscheinlich viele Atemgeräusche zu hören (80%, 2. Zeile, 2. Spalte)

DEUTLICHKEIT DER AUSSPRACHE (D) *temporär*

Eltern: ALKOHOLEINFLUSS (A)

Zustände: (2) schlecht, gut

CPT: $P(D = d \mid A = a) = M_{D|A}(d, a)$

$$M_{D|A} = \begin{bmatrix} 0,20 & 0,80 \\ 0,70 & 0,30 \end{bmatrix}$$

Beschreibung: Ist $\mathcal{P}$ betrunken, ist mit hoher Wahrscheinlichkeit eine schlechte Aussprache zu erwarten.

SCHWIERIGKEIT DER FRAGE (S) *temporär*

Eltern: –

Zustände: (3) gering, mittel, hoch

A-priori: $\begin{bmatrix} 0,65 & 0,30 & 0,05 \end{bmatrix}$

Beschreibung: Die Schwierigkeit der Frage ist eher mittel bis gering.

QUALITÄT DER SITUATIONSBESCHREIBUNG (Q) *temporär*

Eltern: VAK (V)

Zustände: (5) schlecht, ..., gut

CPT: $P(Q = q \mid V = v) = N_{\mu,\sigma}(q)$
$\mu = M_V(v), \quad \sigma = 0,15$
$M_V = \begin{bmatrix} 0,10 & 0,30 & 0,50 & 0,60 & 0,90 \end{bmatrix}$

Beschreibung: Bei geringer VAK ist mit einer schlechten Qualität der Situationsbeschreibung zu rechnen.

BITTE UM WIEDERHOLUNG (W) *temporär*

Eltern: VAK (V), SCHWIERIGKEIT DER FRAGE (S)

Zustände: (5) nein, ja

CPT: $P(W = w \mid V = v, S = s) = N_{\mu,\sigma}(w)$
$\mu = M_{V,S}(v, s), \quad \sigma = 0,15$

$$M_{V,S} = \begin{bmatrix} 0,55 & 0,70 & 0,80 \\ 0,30 & 0,55 & 0,70 \\ 0,20 & 0,30 & 0,65 \\ 0,05 & 0,15 & 0,40 \\ 0,01 & 0,05 & 0,30 \end{bmatrix}$$

Beschreibung: $\mathcal{P}$ bittet um Wiederholung der Frage, wenn $\mathcal{S}$ eine schwere Frage stellt oder $\mathcal{P}$ eine geringe VAK besitzt.

GESCHWINDIGKEIT DER ANTWORT (G) *temporär*

Eltern: VAK (V), SCHWIERIGKEIT DER FRAGE (S)

Zustände: (5) langsam, ..., schnell

CPT: $P(G = g \mid V = v, S = s) = N_{\mu,\sigma}(g)$
$\mu = M_{V,S}(v, s), \quad \sigma = 0,15$

$$M_{V,S} = \begin{bmatrix} 0,20 & 0,10 & 0,05 \\ 0,40 & 0,30 & 0,15 \\ 0,50 & 0,50 & 0,25 \\ 0,75 & 0,60 & 0,35 \\ 0,95 & 0,90 & 0,65 \end{bmatrix}$$

Beschreibung: Bei geringer VAK oder schwerer Frage ist mit einer langsamen Antwort zu rechnen.

QUALITÄT DER ANTWORT (Q) *temporär*

Eltern: VAK (V), SCHWIERIGKEIT DER FRAGE (S)

Zustände: (5) schlecht, ..., gut

CPT: $P(Q = q \mid V = v, S = s) = N_{\mu,\sigma}(q)$
$\mu = M_{V,S}(v, s), \quad \sigma = 0,15$

$$M_{V,S} = \begin{bmatrix} 0,20 & 0,10 & 0,05 \\ 0,40 & 0,30 & 0,15 \\ 0,50 & 0,50 & 0,25 \\ 0,75 & 0,60 & 0,35 \\ 0,95 & 0,90 & 0,65 \end{bmatrix}$$

Beschreibung: Bei geringer VAK oder schwerer Frage ist mit einer schlechten Antwort zu rechnen.

GÄNGIGKEIT VON STADTTEILEN (G) *temporär*

Eltern: –

Zustände: (2) nein, ja

A-priori: $\begin{bmatrix} 0,36 & 0,64 \end{bmatrix}$

Beschreibung: Die Stadtteile sind eher gängig.

AKUSTISCHE VERSTÄNDLICHKEIT DER STADTTEILFRAGE (A) *temporär*

Eltern: –

Zustände: (3) schlecht, mittel, gut

A-priori: $\begin{bmatrix} 0,25 & 0,30 & 0,45 \end{bmatrix}$

Beschreibung: Die akkustische Verständlichkeit ist mittel bis gut.

VORHERSEHBARKEIT DER STADTTEILFRAGE (V) *temporär*

Eltern: –

Zustände: (2) nein, ja

A-priori: $\begin{bmatrix} 0,30 & 0,70 \end{bmatrix}$

Beschreibung: Mitarbeiter der FNZ fragen oft nach den Stadtteil, in dem der Anrufer wohnt.

SCHWIERIGKEIT DER STADTTEILFRAGE (S) *temporär*

Eltern: GÄNGIGKEIT VON STADTTEILEN (G) VORHERSEHBARKEIT DER STADTTEILFRAGE (V), AKUSTISCHE VERSTÄNDLICHKEIT DER FRAGE (A)

Zustände: (3) gering, mittel, hoch

CPT: $P(S = s \mid G = g, V = v, A = a) = N_{\mu,\sigma}(s)$

$$\mu = M_{V=v,G,A}(g, a), \quad \sigma = 0,1$$

$$M_{V=nein,G,A} \begin{bmatrix} 0,95 & 0,90 \\ 0,85 & 0,60 \\ 0,70 & 0,40 \end{bmatrix} \qquad M_{V=ja,G,A} \begin{bmatrix} 0,90 & 0,50 \\ 0,60 & 0,25 \\ 0,40 & 0,05 \end{bmatrix}$$

Beschreibung: Wenn die Stadtteilfrage nicht vorhersehbar ist, Stadtteile nicht gängig sind und Akkustik schlecht ist, dann ist die Frage eher schwer.

LÄNGE DER NUMMER (L) *temporär*

Eltern: –

Zustände: (3) >7, 4-7, <4

A-priori: $\begin{bmatrix} 0,25 & 0,70 & 0,05 \end{bmatrix}$

Beschreibung: Die meisten Telefonnummern haben 4-7 Ziffern.

VORHERSEHBARKEIT DER NUMMERFRAGE (V) *temporär*

Eltern: –

Zustände: (2) nein, ja

A-priori: $\begin{bmatrix} 0,30 & 0,70 \end{bmatrix}$

Beschreibung: Die Mitarbeiter der FNZ fragen oft nach der Nummer des Anrufers.

VERTRAUTHEIT MIT DER NUMMER (T) *temporär*

Eltern: –

Zustände: (2) nein, ja

A-priori: $\begin{bmatrix} 0,30 & 0,70 \end{bmatrix}$

Beschreibung: Die meisten Menschen kennen ihre Telefonnummer.

SCHWIERIGKEIT DER NUMMERFRAGE (S) *temporär*

Eltern: LÄNGE DER NUMMER (L), VORHERSEHBARKEIT DER NUMMERFRAGE (V), VERTRAUTHEIT MIT DER NUMMER (T)

Zustände: (3) gering, mittel, hoch

CPT: $P(S = s \mid L = l, V = v, T = t) = N_{\mu,\sigma}(s)$

$$\mu = M_{V=v,T,L}(t,l), \quad \sigma = 0,1$$

$$M_{V=nein,T,L} \begin{bmatrix} 0,95 & 0,60 \\ 0,85 & 0,30 \\ 0,70 & 0,10 \end{bmatrix} \qquad M_{V=ja,T,L} \begin{bmatrix} 0,70 & 0,40 \\ 0,50 & 0,25 \\ 0,40 & 0,05 \end{bmatrix}$$

Beschreibung: Wenn die Telefonnummer kurz ist, die Frage danach vorhersehbar und man mit der Nummer vertraut ist, dann ist die Frage nach der Nummer einfach.

A.5 Beschreibung der Knoten des Benutzermodells von READY

Im folgenden werden die Definitionen der einzelnen Knoten angegeben, die in READY verwendet werden. Wie im Abschnitt zuvor werden folgende Abkürzungen benutzt: $P(X = x \mid Y = y)$ ist die Wahrscheinlichkeit für $X = x$ bedingt $Y = y$. $M_A(x, y)$ bezeichnet das Element der Matrix M_A in Zeile x und Spalte y. $N_{\mu,\sigma}(x)$ beschreibt die Normalverteilung mit Erwartungswert μ und Standardabweichung σ.

A.5.1 Modell der Zusammenhänge zwischen Anforderung und Verwendung von verfügbarer Arbeitsgedächtniskapazität bei zwei Aufgaben

EMOTIONALE BELASTUNG$_t$ (E_t) *dynamisch*

Eltern: EMOTIONALE BELASTUNG$_{t-1}$ (E_{t-1})

Zustände: (3) gering, mittel, hoch

CPT: $P(E_t = e_t \mid E_{t-1} = e_{t-1}) = N_{\mu,\sigma}(e_t)$
$\mu = 0,9 * e_{t-1}$ und $\sigma = 0.2$

Beschreibung: $\mathcal{P}$s emotionale Belastung nimmt im Laufe der Zeit ab.

in Zeitscheibe t_0:

Eltern: –

A-priori: $P(E_0 = e) = N_{\mu,\sigma}(e)$ $\mu = 0,4$ und $\sigma = 0.3$

Beschreibung: $\mathcal{P}$s emotionale Belastung ist a-priori eher mittel bis gering.

ATEMGERÄUSCHE (A) *temporär*

Eltern: EMOTIONALE BELASTUNG (E)

Zustände: (3) wenig, mittel, viel

CPT: $P(A = a \mid E = e) = N_{\mu,\sigma}(a)$
$\mu = e$ und $\sigma = 0.4$

Beschreibung: Hat $\mathcal{P}$ eine hohe emotionale Belastung, werden wahrscheinlich viele Atemgeräusche zu hören sein.

SITUATIVE ABLENKUNGEN$_t$ (A_t) *dynamisch*

Eltern: SITUATIVE ABLENKUNGEN$_{t-1}$ (A_{t-1})

Zustände: (3) gering, mittel, hoch

CPT: $P(A_t = a_t \mid A_{t-1} = a_{t-1}) = N_{\mu,\sigma}(a_t)$
$\mu = 0,9 * a_{t-1}$ und $\sigma = 0.2$

Beschreibung: $\mathcal{P}$ kann zeitweise abgelenkt sein. Die Ablenkung nimmt aber im Laufe der Zeit ab.

in Zeitscheibe t_0:

Eltern: –

A-priori: $P(A_0 = a) = N_{\mu,\sigma}(a)$ $\mu = 0,35$ und $\sigma = 0.3$

Beschreibung: Die situativen Ablenkungen sind a-priori eher mittel bis gering.

UMGEBUNGSGERÄUSCHE (U) *temporär*

Eltern: SITUATIVE ABLENKUNGEN (A)

Zustände: (3) keine, mittel, viel

CPT: $P(U = u \mid A = a) = N_{\mu,\sigma}(u)$
$\mu = a$ und $\sigma = 0.4$

Beschreibung: Situative Ablenkungen werden wahrscheinlich anhand von vielen Umgebungsgeräuschen zu hören sein.

VAK$_t$ (V_t) *dynamisch*

Eltern: VAK$_{t-1}$ (V_{t-1}), EMOTIONALE BELASTUNG$_t$ (E_t), EMOTIONALE BELASTUNG$_{t-1}$ (E_{t-1}), SITUATIVE ABLENKUNGEN$_t$ (A_t), SITUATIVE ABLENKUNGEN$_{t-1}$ (A_{t-1})

Zustände: (4) gering, ..., hoch

CPT: $P(V_t = v_t \mid V_{t-1} = v_{t-1}, E_t = e_t, E_{t-1} = e_{t-1}, A_t = a_t, A_{t-1} = a_{t-1}) = N_{\mu,\sigma}(v_t)$
$\mu = max(1, v_{t-1} + 1/2 * ((e_{t-1} - e_t) + (a_{t-1} - a_t)))$
$\sigma = 0.2$

Beschreibung: $\mathcal{P}$s VAK in t ist so hoch wie in $t-1$. Steigt aber $\mathcal{P}$s emotionale Belastung oder die situativen Ablenkungen von $t-1$ nach t, wird die VAK wahrscheinlich sinken.

in Zeitscheibe t_0:

Eltern: EMOTIONALE BELASTUNG$_0$ (E_0), SITUATIVE ABLENKUNGEN$_0$ (A_0)

CPT: $P(V_0 = v \mid E_0 = e, A_0 = a) = N_{\mu,\sigma}(v)$
$\mu = 1 - 3/8 * (e + a)$ und $\sigma = 0.3$

Beschreibung: Je höher die Ablenkung und emotionale Belastung ist, desto geringer ist die VAK.

ANFORDERUNG DER SPRACHANALYSE (A) *temporär*

Eltern: SCHWIERIGKEIT FORMULIERUNG (S), KOMPLEXITÄT DER ÄUSSERUNG (K), WISSENSNIVEAU (W)

Zustände: (4) gering, ..., hoch

CPT: $P(A = a \mid S = s, K = k, W = w) = N_{\mu,\sigma}(a)$
$\mu = 1/3 * (s + k + (1 - w))$ und $\sigma = 0.15$

Beschreibung: Eine hohe Anforderung der Sprachanalyse ist umso wahrscheinlicher, je höher die Schwierigkeit der Formulierung und Komplexität der Äußerung von $\mathcal{S}$ und je geringer das Wissensniveau von $\mathcal{P}$ ist.

ANFORDERUNG DER SPRACHGENERIERUNG (A) *temporär*

Eltern: MENGE DER VON P VERMITTELTEN INFORMATIONEN (M), WISSENSNIVEAU (W)

Zustände: (4) gering, ..., hoch

CPT: $P(A = a \mid M = m, W = w) = N_{\mu,\sigma}(a)$
$\mu = 1/2(m + (1 - w))$ und $\sigma = 0.3$

Beschreibung: Eine hohe Anforderung der Sprachgenerierung ist umso wahrscheinlicher, je mehr Informationen $\mathcal{P}$ vermitteln muß und je geringer das Wissensniveau von $\mathcal{P}$ ist.

ANFORDERUNG DER AUSFÜHRUNG DER HANDLUNG (A) *temporär*

Eltern: ANFORDERUNG AN PS PLANUNG (P), WISSENSNIVEAU (W)

Zustände: (4) gering, ..., hoch

CPT: $P(A = a \mid S = s, P = p, W = w) = N_{\mu,\sigma}(a)$
$\mu = 1/5(s + 3 * p + (1 - w))$ und $\sigma = 0.2$

Beschreibung: Eine hohe Anforderung der Ausführung der Handlung ist umso wahrscheinlicher, je höher die Anforderung an $\mathcal{P}$s Planung und je geringer $\mathcal{P}$s Wissensniveau ist.

ANFORDERUNG DER NEBENHANDLUNG (A) *temporär*

Eltern: NEBENTÄTIGKEIT (N)

Zustände: (4) gering, ..., hoch

CPT: $P(A = a \mid N = n) = N_{\mu,\sigma}(a)$ $\mu = n$ und $\sigma = 0.3$

Beschreibung: Eine hohe Anforderung der Nebenhandlung ist umso wahrscheinlicher, wenn $\mathcal{P}$ eine Nebentätigkeit ausführen muß.

VORRANG FÜR DIE SPRACHVERARBEITUNG (V) *temporär*

Eltern: –

Zustände: (2) nein, ja

A-priori: $\begin{bmatrix} 0,40 & 0,60 \end{bmatrix}$

Beschreibung: Der Vorrang liegt wahrscheinlich auf der Sprachverarbeitung.

VERWENDUNG VON AG FÜR DIE AUFGABE 1 (Z_1) *temporär*

Eltern: VAK (K), ANFORDERUNG DER AUFGABE 1 AN AG (A_1), ANFORDERUNG DER AUFGABE 2 AN AG (A_2), VORRANG FÜR AUFGABE 1 (V)

Zustände: (4) gering, ..., hoch

CPT: $P(Z_1 = z \mid K = k, A_1 = a_1, A_2 = a_2, V = v) =$

$$= \begin{cases} 1, & \text{wenn} \quad v = \text{ja} \quad \wedge \quad z = min(a_1, k) \\ 1, & \text{wenn} \quad v = \text{nein} \quad \wedge \quad z = min(a_1, max(0, k - a_2)) \\ 0, & \text{sonst} \end{cases}$$

$\sigma = 0.2$

Beschreibung: Die Aufgabe, die eine höhere Priorität hat, wird maximal Arbeitsgedächtniskapazität zugeteilt. Die verbleibende Kapazität wird dann für die Aufgabe mit geringerer Priorität verwendet.

A.5.2 Modellierung der Dauer der Aufgabe und des Erfolgs bei der Ausführung einer Aufgabe

ZEITDRUCK$_t$ (Z_t) *dynamisch*

Eltern: ZEITDRUCK$_{t-1}$ (Z_{t-1})

Zustände: (4) gering, ..., hoch

CPT: $P(Z_t = z_t \mid Z_{t-1} = z_{t-1}) = N_{\mu,\sigma}(z_t)$
$\mu = max(1, c + z_{t-1})$ c = Konstante für Zunahme des Zeitdrucks (hier c=0)
$\sigma = 0.1$

Beschreibung: Das subjektive Gefühl von Zeitdruck hat die Tendenz abzunehmen.

in Zeitscheibe t_0:

Eltern: –

A-priori: $P(Z_0 = a) = N_{\mu,\sigma}(a) \quad \mu = 0,2$ und $\sigma = 0.4$

Beschreibung: Der Zeitdruck ist a-priori eher gering.

WISSENSNIVEAU (W) *statisch*

Eltern: –

Zustände: (3) gering, mittel, hoch

A-priori: $P(W = w) = N_{\mu,\sigma}(w), \quad \mu = 0,2$ und $\sigma = 0.25$

Beschreibung: Es wird angenommen, daß $\mathcal{P}$ ein eher mittel bis geringes Wissensniveau besitzt.

DAUER DER AUFGABE (D) *temporär*

Eltern: ANFORDERUNG DER AUFGABE AN AG (A), VERWENDUNG VON AG FÜR DIE AUFGABE (V), ZEITDRUCK (Z), EIGENSCHAFTEN DER AUFGABE (E)

Zustände: (4) gering, ..., hoch

CPT: $P(D = d \mid A = a, V = v, Z = z, E = e) = N_{\mu,\sigma}(d)$
$\mu = 1/2 * (1 + (a - min(a, v))) * (1 - cz) * (0,2 + 0,8 * e) \qquad \sigma = 0.2$
c Faktor, wie stark Zeitdruck die Dauer verkürzt (hier $c = 0,5$)

Beschreibung: EIGENSCHAFTEN DER AUFGABE ist bei der Sprachanalyse ÄUSSERUNGSLÄNGE, bei Sprachgenierung MENGE DER VON P VERMITTELTEN INFORMATIONEN und bei der Ausführung der Handlung KOMPLEXITÄT DES PLANS.

ERFOLG BEI DER SPRACHANALYSE (E) *temporär*

Eltern: ANFORDERUNG DER SPRACHANALYSE AN AG (A), VERWENDUNG VON AG FÜR DIE SPRACHANALYSE (V), ZEITDRUCK (Z)

Zustände: (4) gering, ..., hoch

CPT: $P(E = e \mid A = a, V = v, Z = z) = N_{\mu,\sigma}(e)$
$\mu = (1 - (a - min(a, v))) * (1 - cz)$ und $\sigma = 0.2$
c = Faktor, wie stark Zeitdruck den Erfolg verringert (hier c=1)

Beschreibung: Ein hoher Erfolg bei der Sprachanalyse ist umso wahrscheinlicher, je höher die Verwendung von AG für die Sprachanalyse ist im Vergleich zu der Anforderung der Sprachanalyse an AG und je geringer der Zeitdruck ist. (Der Zusammenhang für die Variable ERFOLG BEI DER FORMULIERUNG wird genauso definiert.)

ERFOLG BEI DER KONZEPTUALISIERUNG (E) *temporär*

Eltern: ANFORDERUNG DER SPRACHGENERIERUNG AN AG (A), VERWENDUNG VON AG FÜR DIE SPRACHGENERIERUNG (V), ERFOLG BEI DER SPRACHANALYSE$_{t-1}$ (E_{t-1}), ZEITDRUCK (Z)

Zustände: (4) gering, ..., hoch

CPT: $P(E = e \mid A = a, V = v, Z = z, E_{t-1} = e_{t-1}) = N_{\mu,\sigma}(e)$
$\mu = e_{t-1} * (1 - (a - min(a, v))) * (1 - cz)$ und $\sigma = 0.2$
c = Faktor, wie stark Zeitdruck den Erfolg verringert (hier c=1)

Beschreibung: Ein hoher Erfolg bei der Konzeptualisierung ist umso wahrscheinlicher, je höher die Verwendung von AG für die Sprachgenerierung ist im Vergleich zu der Anforderung der Sprachgenerierung an AG, je höher der Erfolg der Sprachanalyse in $t-1$ war und je geringer der Zeitdruck ist.

ERFOLG BEI DER AUSFÜHRUNG DER HANDLUNG (E) *temporär*

Eltern: ANFORDERUNG DER AUSFÜHRUNG DER HNADLUNG AN AG (A), VERWENDUNG VON AG FÜR DIE AUSFÜHRUNG DER HANDLUNG (V), ZEITDRUCK (Z), PLANROBUSTHEIT (P), ECHO VON ÄUSSERUNGEN (C), ERFOLG DER SPRACHANALYSE$_{t-1}$ (E_{t-1})

Zustände: (4) gering, ..., hoch

CPT: $P(E = e \mid A = a, V = v, Z = z, P = p, C = c, E_{t-1} = e_{t-1}) = N_{\mu,\sigma}(e)$
$\mu = e_{t-1} * (1 - (a - min(a, v))) * (1 - cz) * (0,8 + 0,2 * r)$
$*(0,9 + 0,1 * c) \qquad \sigma = 0.2$
c = Faktor, wie stark Zeitdruck den Erfolg verringert (hier c=1)

Beschreibung: Ähnlich wie ERFOLG BEI DER KONZEPTUALISIERUNG. Zusätzlich führt eine höhere Planrobustheit und viele Echos von Äußerungen zu einem wahrscheinlich höherem Erfolg.

A.5.3 Sprachlichen Eigenschaften einer Äußerung von S

SYNTAKTISCHE SCHWIERIGKEIT DER ÄUSSERUNG (S) *temporär*

Eltern: –

Zustände: (3) gering, mittel, hoch

A-priori: wird vom System gesetzt.

Beschreibung: Hörer-unabhängige Bewertung der Schwierigkeit, die Äußerung syntaktisch zu verstehen. Darunter fallen Aspekte wie Einbettung von Nebensätzen, adverbiale Beschreibungen usw.

VERWENDUNG VON FACHBEGRIFFEN (V) *temporär*

Eltern: –

Zustände: (3) gering, mittel, hoch

A-priori: wird vom System gesetzt.

Beschreibung: Häufigkeit des Auftretens von Fachbegriffen in der Äußerung.

ELLIPSEN UND ANAPHERN (BEI $\mathcal{S}$) (E) *temporär*

Eltern: –

Zustände: (3) wenig, mittel, viel

A-priori: wird vom System gesetzt.

Beschreibung: Häufigkeit des Auftretens von Ellipsen und Anaphern in der Äußerung.

GRAD DER VORSTRUKTURIERUNG (V) *temporär*

Eltern: –

Zustände: (3) wenig, mittel, viel

A-priori: wird vom System gesetzt.

Beschreibung: Ja/Nein- und Entscheidungsfragen bieten eine hohe Vorstrukturierung im Vergleich zu W-Fragen.

INTRINSISCHE SCHWIERIGKEIT DER FRAGE (S) *temporär*

Eltern: –

Zustände: (3) gering, mittel, hoch

A-priori: wird vom System gesetzt.

Beschreibung: Die intrinsische Schwierigkeit der Frage ist die Schwierigkeit, die von der Struktur der Äußerung unabhängig ist.

EXPLIZITHEIT DER INSTRUKTION (E) *temporär*

Eltern: –

Zustände: (3) gering, mittel, hoch

A-priori: wird vom System gesetzt.

Beschreibung: Die Explizitheit der Instruktion gibt den Grad an, zu dem die Instruktion aus elementaren, d.h. ohne eigene Planung von $\mathcal{P}$ ausführbaren, Anweisungen besteht.

SCHWIERIGKEIT DER FORMULIERUNG (F) *temporär*

Eltern: SYNTAKTISCHE SCHWIERIGKEIT DER ÄUSSERUNG (S), VERWENDUNG VON FACHBEGRIFFEN (V), ELLIPSEN UND ANAPHERN (BEI $\mathcal{S}$) (E)

Zustände: (3) gering, mittel, hoch

CPT: $P(F = f \mid S = s, V = v, E = e) = N_{\mu,\sigma}(f)$
$\mu = 1/3 * (s + v + e))$ und $\sigma = 0.15$

Beschreibung: Je höher die syntaktische Schwierigkeit der Äußerung ist und je mehr Fachbegriffe und Ellipsen bzw. Anaphern verwendet werden, desto wahrscheinlicher ist eine hohe Schwierigkeit der Formulierung.

KOMPLEXITÄT DER KONZEPTUALISIERUNG (K) (bei einer Frage) *temporär*

Eltern: GRAD DER VORSTRUKTURIERUNG (V), INTRINSISCHE SCHWIERIGKEIT DER FRAGE (I)

Zustände: (3) gering, mittel, hoch

CPT: $P(K = k \mid V = v, I = i) = N_{\mu,\sigma}(k)$
$\mu = 1/2 * (v + i)$ und $\sigma = 0.2$

Beschreibung: Je höher die Vorstrukturierung und die intrinsische Schwierigkeit der Frage ist, desto höher ist die Komplexität der Konzeptualisierung.

KOMPLEXITÄT DER KONZEPTUALISIERUNG (K) (bei einer Instruktion) *temporär*

Eltern: EXPLIZITHEIT DER INSTRUKTION (E), KOMPLEXITÄT DES PLANS (K)

Zustände: (3) gering, mittel, hoch

CPT: $P(K = k \mid E = e, I = i) = N_{\mu,\sigma}(k)$
$\mu = 1/2 * (e + i)$ und $\sigma = 0.2$

Beschreibung: Je höher die Explizitheit der Instruktion und die Komplexität des Plans ist, desto höher ist die Komplexität der Konzeptualisierung.

MENGE DER VON P VERMITTELTEN INFORMATIONEN (M) *temporär*

Eltern: GRAD DER VORSTRUKTURIERUNG (V), INTRINSISCHE SCHWIERIGKEIT DER FRAGE (S)

Zustände: (3) wenig, mittel, viel

CPT: $P(M = m \mid V = v, S - s) = N_{\mu,\sigma}(m)$
$\mu = 1/2 * ((1 - v) + s)$ und $\sigma = 0.2$

Beschreibung: Je höher die Vorstrukturierung der Frage ist und je geringer die intrinsische Schwierigkeit der Frage ist, desto wahrscheinlicher ist es, daß $\mathcal{P}$ weniger Informationen vermitteln muß.

LÄNGE VON PS ÄUSSERUNG (L) *temporär*

Eltern: MENGE DER VON P VERMITTELTEN INFORMATIONEN (M)

Zustände: (3) gering, mittel, hoch

CPT: $P(L = l \mid M = m) = N_{\mu,\sigma}(l) \qquad \mu = m \quad \text{und} \quad \sigma = 0.2$

Beschreibung: Je mehr Informationen $\mathcal{P}$ vermitteln muß, desto länger ist $\mathcal{P}$s Äußerung.

KOMPLEXITÄT DES PLANS (K) *temporär*

Eltern: –

Zustände: (3) gering, mittel, hoch

A-priori: wird vom System gesetzt.

Beschreibung: Die Komplexität des Plans gibt an, wie lang der Plan ist, der durch die Instruktion erzeugt wird.

ANFORDERUNG AN PS PLANUNG (A) *temporär*

Eltern: EXPLIZITHEIT DER INSTRUKTION (E), KOMPLEXITÄT DES PLANS (K)

Zustände: (3) gering, mittel, hoch

CPT: $P(A = a \mid E = e, K = k) = N_{\mu,\sigma}(a)$
$\mu = 1/3 * (k + 2 * (1 - e)) \quad \text{und} \quad \sigma = 0.15$

Beschreibung: Die Explizitheit der Instruktion gibt an, wie groß der Aufwand für die Planexpansion seitens $\mathcal{P}$ ist, falls nicht ausschließlich elementare Aktionen instruiert werden.

A.5.4 Symptome von Arbeitsgedächtnisbelastung und Zeitdruck bei der Sprachanalyse

---

SPONTANE ERFOLGSMELDUNG (S) *temporär*

Eltern: ERFOLG DER AUFGABE (E)

Zustände: (3) negativ, neutral, positiv

CPT: $P(S = s \mid E = e) = N_{\mu,\sigma}(s)$

$$\mu = \begin{cases} 1/2 * (1 + (e - 1/2)), & \text{wenn} \quad e > 1/2 \\ e, & \text{sonst} \end{cases}$$

$\sigma = 0.125$

Beschreibung: Ein hoher Erfolg bei der Aufgabe macht einen positive spontane Erfolgsmeldung wahrscheinlich.

A.5.5 Symptome von Arbeitsgedächtnisbelastung bei der Sprachgenerierung

INHALTLICH PASSENDE ANTWORT (I) *temporär*

Eltern: ERFOLG BEI DER KONZEPTUALISIERUNG (E)

Zustände: (3) nicht passend, mäßig passen, passend

CPT: $P(I = i \mid E = e) = N_{\mu,\sigma}(i)$
$\mu = max(1, 1/4 * +3/4 * e)$ und $\sigma = 0.5$

Beschreibung: Ist der Erfolg bei der Konzeptualisierung hoch, wird $\mathcal{P}$ wahrscheinlich eine inhaltlich passende Antwort auf die Frage geben.

VERWENDUNG VON ORTSDEIKTISCHEN ÄUSSERUNGEN (O) *temporär*

Eltern: ERFOLG BEI DER KONZEPTUALISIERUNG (E)

Zustände: (2) nein, ja

CPT: $P(O = o \mid E = e) =$

$$= \begin{cases} X, & \text{wenn} \quad o = \text{nein} \\ 1 - X, & \text{wenn} \quad o = \text{ja} \end{cases}$$

wobei $X = -0,05 * (1 - e) + 0,95$

Beschreibung: Die Wahrscheinlichkeit, daß $\mathcal{P}$ ortsdeiktische Äußerungen verwendet, ist sehr klein (5% wenn $\mathcal{P}$ einen hohen Erfolg bei der Konzeptualisierung hat, 10% bei geringem Erfolg).

DEUTLICHKEIT DER AUSSPRACHE (D) *temporär*

Eltern: ERFOLG BEI DER FORMULIERUNG (E)

Zustände: (3) gering, mittel, hoch

CPT: $P(D = d \mid E = e) = N_{\mu,\sigma}(d)$ $\mu = e$ und $\sigma = 0.5$

Beschreibung: Hat $\mathcal{P}$ einen geringen Erfolg bei der Formulierung, erhöht sich die Wahrscheinlichkeit für eine geringe Deutlichkeit der Aussprache.

ABWEICHUNG VON DER ARTIKULATIONSGESCHWINDIGKEIT (A) *temporär*

Eltern: ZEITDRUCK (Z)

Zustände: (3) langsamer, normal, schneller

CPT: $P(A = a \mid Z = z) = N_{\mu,\sigma}(a) \qquad \mu = z$ und $\sigma = 0.3$

Beschreibung: Je höher der Zeidruck von $\mathcal{P}$ ist, desto höher ist die Wahrscheinlichkeit, daß $\mathcal{P}$ von seiner normalen Artikulationsgeschwindigkeit abweicht.

NORMALE ARTIKULATIONSGESCHWINDIGKEIT (N) *statisch*

Eltern: –

Zustände: (3) gering, mittel, hoch

A-priori: $P(N = n) = N_{\mu,\sigma}(n) \qquad \mu = 0,5$ und $\sigma = 0.2$

Beschreibung: Die meisten Menschen haben eine mittlere Artikulationsgeschwindigkeit.

BEOBACHTETE ARTIKULATIONSGESCHWINDIGKEIT (B) *temporär*

Eltern: ABWEICHUNG VON DER ARTIKULATIONSGESCHWINDIGKEIT (A), NORMALE ARTIKULATIONSGESCHWINDIGKEIT (N)

Zustände: (3) gering, mittel, hoch

CPT: $P(B = b \mid A = a, N = n) = N_{\mu,\sigma}(b)$
$\mu = min(0, max(1, n + a - 0,5))$ und $\sigma = 0.1$

Beschreibung: Durch eine Abweichung der Artikulationsgeschwindigkeit nach oben (unten) erhöht (verringert) sich die beobachtete Artikulationsgeschwindigkeit gegenüber der normale Artikulationsgeschwindigkeit.

MENGE LEICHT AUSLASSBAERER INFORMATIONEN (I) *temporär*

Eltern: ZEITDRUCK (Z)

Zustände: (3) keine, mittel, viel

CPT: $P(I = i \mid Z = z) = N_{\mu,\sigma}(i)$
$\mu = (1 - z)$ und $\sigma = 0.5$

Beschreibung: Ist $\mathcal{P}$s Zeitdruck hoch, wird $\mathcal{P}$ wahrscheinlich weniger Informationen geben, die nichts mit der Autopanne zu tun haben.

ELLIPSEN UND ANAPHERN BEI P (E) *temporär*

Eltern: ZEITDRUCK (Z)

Zustände: (3) wenig, mittel, erhöht

CPT: $P(E = e \mid Z = z) = N_{\mu,\sigma}(e) \qquad \mu = z$ und $\sigma = 0.5$

Beschreibung: Ist $\mathcal{P}$s Zeitdruck hoch, wird $\mathcal{P}$ mehr Ellipsen und Anaphern in seinen Äußerungen verwenden.

A.5.6 Symptome von Zeitdruck und Arbeitsgedächtnisbelastung bei der Ausführung einer Handlung

SPONTANE ERFOLGSMELDUNG (S) *temporär*

Eltern: ERFOLG DER AUFGABE (E)

Zustände: (3) negativ, neutral, positiv

CPT: $P(S = s \mid E = e) = N_{\mu,\sigma}(s)$

$$\mu = \begin{cases} 1/2 * (1 + (e - 1/2)), & \text{wenn} \quad e > 1/2 \\ e, & \text{sonst} \end{cases}$$

$\sigma = 0.125$

Beschreibung: Ein hoher Erfolg bei der Aufgabe macht einen positive spontane Erfolgsmeldung wahrscheinlich.

SPRECHERWECHSELSIGNAL (S) *temporär*

Eltern: ZEITDRUCK (Z)

Zustände: (2) nein, ja

CPT: $P(S = s \mid Z = z) = N_{\mu,\sigma}(s) \qquad \mu = z \quad \text{und} \quad \sigma = 0.4$

Beschreibung: Je höher der Zeitdruck ist, desto höher ist die Wahrscheinlichkeit für ein Sprecherwechselsignal seitens $\mathcal{P}$ (nur bei positiver spontanen Erfolgsmeldung).

Literaturverzeichnis

Albrecht, D. W., Zukerman, I., Nicholson, A. E. & Bud, A. (1997). Towards a Bayesian model for keyhole plan recognition in large domains. In A. Jameson, C. Paris & C. Tasso (Hrsg.), *User modeling: Proceedings of the Sixth International Conference, UM97* (S. 365-376). Wien, New York: Springer Wien New York.

Anderson, A. H. & Boyle, E. A. (1994). Forms of introduction in dialogues: Their discourse contexts and communicative consequences. *Language and Cognitive Processes*, *9*, 101-122.

Atkinson, R. C. & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence (Hrsg.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, S. 89-195). San Diego, CA: Academic Press.

Baddeley, A. D. (1986). *Working memory.* Oxford: Oxford University Press.

Baecker, R. M., Grudin, J., Buxton, W. A. S. & Greenberg, S. (Hrsg.). (1995). *Readings in human-computer interaction: Toward the year 2000.* San Franzisco, California: Morgan Kaufmann.

Bauer, M. (1996). A Dempster-Shafer approach to modeling agent preferences for plan recognition. *User Modeling and User-Adapted Interaction*, *5*, 317-348.

Berthold, A. (1997). „Ich – ähm – denke, daß …“: Zu sprachlichen Indikatoren für Ressourcenbeschränkungen und ihrer Verarbeitung in natürlichsprachlichen Dialogsystemen. In R. Schäfer & M. Bauer (Hrsg.), *ABIS-97, 5. GI-Workshop: Adaptivität und Benutzermodellierung in interaktiven Systemen: Workshop-Beiträge* (S. 139-147). Saarbrücken: Universität des Saarlandes.

Berthold, A. (1998). *Repräsentation und Verarbeitung sprachlicher Indikatoren von Ressourcenbeschränkungen.* Diplomarbeit, Fachbereich Informatik, Universität des Saarlandes. (In Vorbereitung)

Brajnik, G. & Tasso, C. (1992). A flexible tool for developing user modeling applications with nonmonotonic reasoning capabilities. In *Proc. of the third international workshop on user modeling, Dagstuhl* (S. 42-66).

Brajnik, G. & Tasso, C. (1994). A shell for developing non-monotonic user modeling systems. In *International journal of human-computer studies, 40* (S. 31-62).

Britton, B. K., Glynn, S. M. & Smith, J. W. (1985). Cognitive demands of processing expository text: A cognitive workbench model. In B. K. Britton & J. B. Black (Hrsg.), *Understanding expository text: A theoretical and practical manual for analyzing explanatory text* (S. 227-248). Hillsdale, NJ: Erlbaum.

Calvo, M. G. & Eysenck, M. W. (1996). Phonological working memory and reading in test anxiety. *Memory*, *4*, 289-305.

Carberry, S. (1990). Incorporating default inferences into plan recognition. In *Proceedings of the Eighth National Conference on Artificial Intelligence* (S. 471-478). Boston, MA.

Charniak, E. (1991). Bayesian networks without tears. *AI Magazine*, *12*(4), 50-63.

Chin, D.-N. (1989). Knome: Modelling what the user knows in uc. In A. K. und Wolfgang Wahlster (Hrsg.), *User models in dialog systems* (S. 74-107). Berlin, Heidelberg: Springer.

Cooper, G. F. (1990). The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence*, *4*(4), 393–405.

Dagum, P., Galper, A. & Horvitz, E. (1992). Dynamic network models for forecasting. In *Proceedings of the Eighth Conference on Uncertainty in Artificial Intelligence* (S. 41–48). San Mateo: Morgan Kaufmann.

Ericsson, K. A. & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Verb. ed.). Cambridge, MA: MIT Press.

Eysenck, M. W. (Hrsg.). (1994). *The blackwell dictionary of cognitive psychology.* Oxford: Basil Blackwell Ltd.

Forbes, J., Huang, T., Kanazawa, K. & Russell, S. (1995). The BATmobile: Towards a Bayesian Automated Taxi. In C. S. Mellish (Hrsg.), *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence* (S. 1878-1885). San Mateo, CA: Morgan Kaufmann.

Giangrandi, P. & Tasso, C. (1997). Managing temporal knowledge in student modeling. In A. Jameson, C. Paris & C. Tasso (Hrsg.), *User modeling: Proceedings of the Sixth International Conference, UM97* (S. 415-426). Wien, New York: Springer Wien New York.

Gordon, J. & Shortliffe, E. H. (1984). The Dempster-Shafer theory of evidence. In B. G. Buchanan & E. H. Shortliffe (Hrsg.), *Rule-based expert systems: The MYCIN experiments of the Stanford Heuristic Programming Project.* Reading: MA: Addison-Wesley.

Grabowski, J. (1991). *Der propositionale Ansatz der Textverständlichkeit: Kohärenz, Interessantheit und Behalten.* Münster: Aschendorff.

Greene, J. O., Lindsey, A. E. & Hawn, J. J. (1990). Social goals and speech production: Effects of multiple goals on pausal phenomena. *Journal of Language and Social Psychology*, *9*, 119-134.

Heckerman, D. (1993). Causal independence for knowledge acquisition and inference. In D. Heckerman & A. Mamdani (Hrsg.), *Proceedings of the 9th conference on uncertainty in artificial intelligence* (S. 122-127). Morgan Kaufmann.

Herrmann, T. & Grabowski, J. (1994). *Sprechen: Psychologie der Sprachproduktion.* Heidelberg: Spektrum Akademischer Verlag.

Horvitz, E. & Barry, M. (1995). Display of information for time-critical decision making. In P. Besnard & S. Hanks (Hrsg.), *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence* (S. 296-305). San Francisco: Morgan Kaufmann.

Horvitz, E., Breese, J., Heckerman, D., Hovel, D. & Rommelse, K. (1998). The lumiere project: Bayesian user modeling for inferring the goals and needs of software users. In *Proceedings of the fourteenth conference on uncertainty in artificial intelligence* (S. 256-265). Wisconsin, USA.

Jameson, A. (1996). Numerical uncertainty management in user and student modeling: An overview of systems and issues. *User Modeling and User-Adapted Interaction*, *5*, 193-251.

Jameson, A. (1997a). Modeling the user's processing resources: Pragmatic simplicity meets psychological complexity. In R. Schäfer & M. Bauer (Hrsg.), *ABIS-97, 5. GI-Workshop: Adaptivität und Benutzermodellierung in interaktiven Systemen: Workshop-Beiträge* (S. 149-160). Saarbrücken: Universität des Saarlandes. (Verfügbar über http://www.cs.uni-sb.de/users/jameson/)

Jameson, A. (1997b). Wie gehen wir mit dem Arbeitsgedächtnis unserer Dialogpartner um? Eine Integration von Ergebnissen aus vier Forschungsrichtungen. In H. Mandl (Hrsg.), *Bericht über den 40. Kongreß der Deutschen Gesellschaft für Psychologie in München 1996* (S. 258-263). Göttingen: Hogrefe.

Jameson, A., Paris, C. & Tasso, C. (Hrsg.). (1997). *User modeling: Proceedings of the Sixth International Conference, UM97.* Wien, New York: Springer Wien New York. (Elektronisch verfügbar über http://um.org)

Jameson, A., Schäfer, R., Simons, J. & Weis, T. (1995). Adaptive provision of evaluation-oriented information: Tasks and techniques. In C. S. Mellish (Hrsg.), *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence* (S. 1886-1893). San Mateo, CA: Morgan Kaufmann.

Jensen, A. L. (1995). Quantification experience of a DSS for mildew management in winter wheat. In M. J. Druzdzel, L. C. van der Gaag, M. Henrion & F. V. Jensen (Hrsg.), *IJCAI-95 workshop on building probabilistic networks: Where do the numbers come from* (S. 22–31). Montreal, Quebec, Canada.

Kashihara, A., Hirashima, T. & Toyoda, J. (1995). A cognitive load application in tutoring. *User Modeling and User-Adapted Interaction*, *4*, 279-303.

Kemper, S., Jackson, J. D., Cheung, H. & Anagnopoulos, C. A. (1993). Enhancing older adults' reading comprehension. *Discourse Processes*, *16*, 405-428.

Kintsch, W. & Vipond, D. (1979). Reading comprehension and readability in educational practice and psychological theory. In L.-G. Nilsson (Hrsg.), *Perspectives on memory research: Essays in honor of Uppsala University's 500th anniversary* (S. 329-365). Hillsdale, NJ: Erlbaum.

Kjærulff, U. (1992). A computational scheme for reasoning in dynamic probabilistic networks. In *Proceedings of the Eighth Conference on Uncertainty in Artificial Intelligence* (S. 121–129). San Mateo: Morgan Kaufmann.

Kobsa, A. & Pohl, W. (1995). The user modeling shell system BGP-MS. *User Modeling and User Adapted Interaction*, *4*, 59-106.

Maybury, M. T. & Wahlster, W. (Hrsg.). (1998). *Readings in intelligent user interfaces.* San Franzisco, California: Morgan Kaufmann.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81-97.

Murphy, M. & McTear, M. (1997). Learner modelling for intelligent CALL. In A. Jameson, C. Paris & C. Tasso (Hrsg.), *User modeling: Proceedings of the Sixth International Conference, UM97.* Wien, New York: Springer Wien New York.

Netica API programmer's library reference manual. (1997). Vancouver, BC, Canada.

Nicholson, A. E. & Brady, J. M. (1992). The data association problem when monitoring robot vehicles using dynamic belief networks. In *Proceedings of the Tenth European Conference on Artificial Intelligence* (S. 689–693). Wien.

Nicholson, A. E. & Brady, J. M. (1994). Dynamic belief networks for discrete monitoring. *IEEE Transactions on Systems, Man, and Cybernetics*, *24*, 1593-1610.

Oviatt, S. (1995). Predicting spoken disfluencies during human-computer interaction. *Computer Speech and Language*, *9*, 19-35.

Oviatt, S. (1996). Multimodal interfaces for dynamic interactive maps. In M. J. Tauber (Hrsg.), *Proceedings of the conference on human factors in computing systems, CHI'96* (S. 95-102). Amsterdam, The Netherlands: Addison Wesley / ACM Press.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference.* San Mateo, CA: Morgan Kaufmann.

Pohl, W. & Höhle, J. (1997). Mechanismus for flexible representation and use of knowledge in user modeling shell systems. In A. Jameson, C. Paris & C. Tasso (Hrsg.), *User modeling: Proceedings of the Sixth International Conference, UM97* (S. 403-414). Wien, New York: Springer Wien New York.

Power, M. J. (1985). Sentence production and working memory. *The Quarterly Journal of Experimental Psychology*, *37A*, 367-385.

Provan, G. M. (1993). Tradeoffs in constructing and evaluating temporal influence diagrams. In D. Heckerman & A. Mamdani (Hrsg.), *Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence* (S. 40–47). San Mateo: Morgan Kaufmann.

Rich, E. (1979). User modeling via stereotypes. *Cognitive Science*, *3*, 329-354.

Rich, E. (1983). Users are individuals: Individualizing user models. In *International journal of man-machine studies 18* (S. 199-214).

Rich, E. (1989). Stereotypes and user modeling. In A. Kobsa & W. Wahlster (Hrsg.), *User models in dialog systems* (S. 35-51). Berlin: Springer.

Roßnagel, C. (1995a). Kognitive Belastung und Hörerorientierung beim monologischen Instruieren. *Zeitschrift für Experimentelle Psychologie*, *42*, 94-110.

Roßnagel, C. (1995b). Übung und Hörerorientierung beim monologischen Instruieren: Zur Differenzierung einer Grundannahme. *Sprache & Kognition*, *14*, 16-26.

Rummer, R. (1997). Verarbeitungsbeschränkungen bei der Sprachproduktion: Ein experimenteller Ansatz zur Erforschung sprachlicher Makroplanungsprozesse. In R. H. Kluwe (Hrsg.), *Strukturen und Prozesse intelligenter Systeme* (S. 41-63). Wiesbaden: Deutscher Universitäts-Verlag.

Russell, S. J. & Norvig, P. (1995). *Artificial intelligence: A modern approach.* Englewood Cliffs, NJ: Prentice-Hall.

Schäfer, R. (1998). *Benutzermodellierung mit dynamischen Bayes'schen Netzen als Grundlage adaptiver Dialogsysteme.* Dissertation, Universität des Saarlandes. (In Vorbereitung)

Schäfer, R., Weis, T., Weyrath, T. & Jameson, A. (1997). Wie können Ressourcenbeschränkungen eines Dialogpartners erkannt und berücksichtigt werden?. *Kognitionswissenschaft*, *6*, 151-164.

Schäfer, R. & Weyrath, T. (1996). Einschätzung von verfügbarer Arbeitsgedächtniskapazität mit temporalen Bayesschen Netzen. In H.-G. Lindner (Hrsg.), *4. GI-Workshop: Adaptivität und Benutzermodellierung in interaktiven Systemen: Workshop-Beiträge* (S. 5.1-5.9). Dortmund: VEW AG.

Schäfer, R. & Weyrath, T. (1997). Assessing temporally variable user properties with dynamic Bayesian networks. In A. Jameson, C. Paris & C. Tasso (Hrsg.), *User modeling: Proceedings of the Sixth International Conference, UM97* (S. 377-388). Wien, New York: Springer Wien New York.

Scott, A. C., Clayton, J. E. & Gibson, E. L. (1991). *A practical guide to knowledge acquisition.* Reading, MA: Addison-Wesley Publishing Company.

Shiffrin, R. M. & Nosofsky, R. M. (1994). Seven plus or minus two: A commentary on capacity limitations. *Psychological Review*, *101*, 357-361.

S.K. Card, T. M. & Newell, A. (1980). *The psychology of human-computer interaction.* Hillsdale, NJ: Lawrence Erlbaum Associates Ltd., Publishers.

Spies, M. (1993). *Unsicheres Wissen. Wahrscheinlichkeit, Fuzzy-Logik, neuronale Netze und menschliches Denken.* Heidelberg: Spektrum Akademischer Verlag.

Srinivas, S. & Breese, J. (1992). *IDEAL: Influence diagram evaluation and analysis in LISP - Documentation and users guide.*

Tawfik, A. Y. & Neufeld, E. (1994). Temporal Bayesian networks. In S. D. Goodwin & H. J. Hamilton (Hrsg.), *Proceedings of the time-94 – international workshop on temporal representation and reasoning.* Pensacola,FL.

van Mulken, S. (1996). Reasoning about the user's decoding of presentations in an intelligent multimedia presentation system. In S. Carberry & I. Zukerman (Hrsg.), *Proceedings of the Fifth International Conference on User Modeling* (S. 67-74). Boston, MA: User Modeling, Inc.

Wahlster, W. & Jameson, A. (1998). READY: Benutzerorientierte Ressourcenadaption. In W. H. Tack (Hrsg.), *Ressourcenadaptive kognitive Prozesse: Fortsetzungsantrag Januar 1999 – Dezember 2001.* Saarbrücken: Universität des Saarlandes.

Wahlster, W. & Kobsa, A. (1989). User models in dialog systems. In A. Kobsa & W. Wahlster (Hrsg.), *User models in dialog systems* (S. 4-34). Berlin: Springer.

Weis, T. (1999). *Ressourcenadaptive Dialogverarbeitung in natürlichsprachlichen Dialogsystemen.* Dissertation, Universität des Saarlandes. (In Vorbereitung)

Weyrath, T. (1997). Einschätzung von Zeitdruck und Arbeitsgedächtnisbelastung in ressourcenbeschränkten Dialogen: Eine explorative empirische Untersuchung. In W. Krause, U. Kotkamp & R. Goertz (Hrsg.), *Proceedings der 3. Fachtagung der Gesellschaft für Kognitionswissenschaft* (S. 236-238). Jena: Friedrich-Schiller-Universität Jena.