In M. T. Maybury (Ed.), *IUI'99: International Conference on Intelligent User Interfaces.* New York: ACM.
(Also available from http://www.cs.uni-sb.de/users/jameson/).

Making Systems Sensitive to the User's Time and Working Memory Constraints

Anthony Jameson Ralph Schäfer Thomas Weis André Berthold Thomas Weyrath
Department of Computer Science, University of Saarbrücken
P.O. Box 15 11 50, 66041 Saarbrücken, Germany
{jameson | ralph | tweis | berthold | tom}@cs.uni-sb.de
Fax: +49 681 302-4136, Phone: +49 681 302-2363, WWW: http://w5.cs.uni-sb.de/˜ready/

ABSTRACT

Recent advances in user modeling technology have brought within reach the goal of having systems adapt to temporary limitations of the user's available time and working memory capacity. We first summarize empirical research by ourselves and others that sheds light on the causes and consequences of these (continually changing) *resource limitations*. We then present a decision-theoretic approach that allows a system to *assess* a user's resource limitations and to *adapt* its behavior accordingly. This approach is illustrated with reference to the performance of the prototype assistance system READY.

Keywords

Adaptive systems, User modeling, Bayesian networks, Time pressure, Working memory, Natural language

INTRODUCTION

A user's available time and working memory can vary considerably, even within a given interaction with a system. For example, a user may become increasingly worried about whether she will finish her memo in time for an impending appointment, or she may be temporarily distracted by events in her environment. The situational variability of these user "resources" seems to be increasing: The more portable computing devices become, the more frequently they are used amid the hustle and bustle of everyday action and social encounter.

In human-human interaction people routinely take into account the temporary resource constraints of other persons, for example, by selectively minimizing or simplifying what they say. But such adaptation has rarely been attempted in the design of interactive systems.

Yet automatic adaptation would be especially appropriate here: When a user is time-pressured or distracted, she is in a poor position to cooperate in solving the problem by either (a) specifying her resource limitations to the system or (b) adapting her own behavior appropriately (e.g., by selecting an especially fast or simple method that she wouldn't normally use).

In the next section, we summarize research on adaptive systems that has some bearing on this problem. We then briefly summarize some previous experimental research that sheds light on causes and consequences of these resource limitations, as well as a study of our own that addressed this question more directly. We then motivate and present a decision-theoretic framework for handling the problem. The prototype dialog system READY shows concretely how this framework can be realized and the type of adaptation that can be achieved.

RELEVANT PREVIOUS WORK ON ADAPTIVE SYSTEMS

Time Limitations

An especially explicit type of adaptation to situational time constraints is exhibited by the text generation system PAULINE (see, e.g., [9]), which also takes into account a number of other rhetorical goals. The approach exemplified by READY builds on PAULINE's approach in two ways: (a) Instead of simplifying its own processing when time limitations are perceived, READY focuses on predicting and controlling the processing time of the user—which seems likely to become a more important factor in the long term. (b) Instead of employing heuristic rules to adapt to time constraints, it uses explicit causal models of the relationships among the relevant variables.

A method for explicitly anticipating the user's processing of a presentation in a time-critical situation was presented in [7] (see also [6]). This decision-theoretic approach takes into account, for example, the facts that (a) presenting additional information to an equipment operator in an emergency may lead to a better decision by the operator, but (b) the utility of that decision may be lower because of the additional time required for the system to transmit and the operator to process the additional information. This work illustrates how a decision-theoretic framework supports the formalization of the tradeoffs that have to be dealt with when scarce resources have to be allocated.

The problem of how a system ($\mathcal{S}$) can *recognize* the time constraints of a user ($\mathcal{U}$) in the first place has not to our knowledge been addressed yet.

Working Memory Limitations

Several adaptive systems have modeled aspects of the user's cognition that are closely related to working memory (WM). For example, the tutoring system of [16] includes a method for computing the *cognitive load* imposed on the student by a specific type of task, as well as procedures for imposing the optimal level of load in each individual case. One of the user models employed in the LUMIÈRE prototype, (see [8]) included an unobservable hypothesis USER DISTRACTED as well as assumptions about its relationship to the difficulty of $\mathcal{U}$'s current task and to $\mathcal{U}$'s observable behavior.

These systems deal with links between WM-related variables and specific other variables. The examples suggest that it would be worthwhile to develop a more general conceptualization that can be applied in a wider variety of situations.

RELEVANT PREVIOUS EMPIRICAL RESULTS

To implement a system that can reason about its user's resource limitations within a given type of interaction, we require knowledge about the relevant causal relationships, i.e., the causes and consequences of the resource limitations in that type of interaction. In general, a lot of relevant results can be found in the published experimental literature on the type of interaction in question, even though the experiments were in almost all cases conducted for different purposes. Consider, for example, Oviatt's research on the prospects and problems of spoken language as a medium for human-computer communication (see, e.g., [19]). From our perspective, it reveals some *causes* of high WM load in users (i.e., the need to supply lengthy and/or unstructured speech input) as well as some observable *consequences* of high WM load (i.e., filled pauses, self-corrections).

Other results relevant to spoken language—which is the interaction medium of the READY system discussed below—come largely from psycholinguistic experiments. For example, Roßnagel ([21]) created high WM load in one group of speakers by forcing them to retrieve a large amount of information from long-term memory while speaking; observable consequences included a large proportion of pauses and a diminished quality of the content of the utterances produced.

Experimental results concerning the consequences of time limitations for speech production are less numerous and less conclusive. For example, some studies have shown that speakers under time pressure sometimes speak faster and make more self-corrections ([17]) and produce less task-unrelated information ([22]); but these results applied only under certain conditions.

Caller: <groan>
Fireman: Fire Department.
Caller: Yes, this is Frau Schmidt. Schopenhauer Street 10. My water heater is on fire. Just this morning this repairman was here <loud breathing>. Now the thing's on fire. <loud breathing>.
Fireman: Schopenhauer Street 10?
Caller: Schopenhauer Street 10.
Fireman: In what part of town?
Caller: Pardon?
Fireman: In what part of town?
Caller: Well, that's here in Saarbrücken <pause>, in Sankt Johann, near the LVA.
Fireman: What floor?
Caller: <incomprehensible>
Fireman: So what floor is that?
Caller: Uh, the first. <groan>
Fireman: Close, close, ... <pause>
Caller: <softly:> Oh hurry! <loud breathing>
Fireman: Yeah, close the door, we're coming right away, OK?
Caller: Yes, that's OK.
Fireman: Yes, bye.

Figure 1. Translated transcript of an emergency call to a fire department.

Our distillation and integration of relevant results of previous empirical research ([3], [11]) has proven useful as a foundation for the development of the READY prototype, even though a good deal of extrapolation is in general required.

An Exploratory Empirical Study

In connection with spoken dialog as a communication medium, there exist people who have some expertise in recognizing and adapting to other people's time and WM constraints: those who regularly handle emergency telephone calls. Ten firemen from the Saarbrücken Fire Department served as subjects in a study ([25]) that made use of the method of *retrospective thinking aloud* (see, e.g., [4]). Each subject listened to three previously recorded phone calls from persons reporting fires. The tape was stopped at predetermined points. At each such point, the subject was asked to answer spontaneously a question about the caller that presumably corresponded to a type of assessment that firemen make while handling such calls (e.g., "How quickly will she be able to provide the information that she was just asked for?"). The subject was then asked to verbalize the thoughts that occurred to him while answering the question.

An analysis of the answers yielded a picture of the (largely shared) causal relationships that the subjects implicitly perceived. Some of these relationships were consistent with general results from previous experimental research, while others were fairly domain-specific (e.g., the link between audible breathing and agitation, which in turn is seen as a cause of "lack of concentration").

In addition to specific qualitative causal relationships, the analysis yielded the following general conclusions:

1. Firemen handling emergency calls perceive varying degrees of "concentration" in the callers, and they believe that these variations should be adapted to. For example, several

subjects noted that it would have been pointless to give the caller in Figure 1 further instructions after dispatching the fire engine, since the agitated caller would probably have had difficulty remembering and executing them anyway.

2. Most of the subjects' formulations included expressions of uncertainty and vagueness (e.g., "…seems to be pretty distracted …"). The firemen seem to recognize implicitly that the relevant causal relationships are probabilistic in nature and do not permit reliable, precise assessments or predictions.

3. An important source of error is a failure to assess accurately the *demands* that a specific dialog contribution will place on the listener. For example, most of the subjects were initially surprised at the difficulty that the caller in Figure 1 had in stating "what part of town" her street was in (though they were able to offer post hoc explanations).

This last result shows that, even where there is converging evidence that yields a fairly clear assessment of $\mathcal{U}$'s resource limitations, this assessment is only one step toward successful adaptation. Methods for accurately predicting the resource demands of particular dialog contributions are equally important.

CHOICE OF A SUITABLE MODELING FRAMEWORK

How can we design an intelligent interactive system so that it can recognize and adapt to the user's changing resource limitations? Research on user and student modeling has yielded a variety of techniques for assessing and adapting to properties of users, including logic- and stereotype-based techniques, machine learning methods, and a host of qualitative and quantitative application-specific procedures.[1]

For the problem at hand here, what seems most suitable is a decision-theoretic framework that includes *dynamic Bayesian networks* and closely related *influence diagrams* for modeling the user's resource limitations and making decisions about the system's behavior.[2]

For the *recognition problem*, these methods are especially well suited for

1. integration of unreliable evidence from a diverse set of observations (in particular, concerning causes and symptoms of resource limitations);
2. incremental use of sparse evidence (as opposed, e.g., to the processing of large amounts of evidence with machine learning methods); and
3. explicit reasoning about the ways in which the user's resource limitations change during the interaction.

For the *adaptation problem*, these methods allow

1. exploitation of most parts of the same probabilistic model that is used for the recognition task;
2. comparative evaluation of possible system behaviors using multiple evaluation criteria whose weights depend on $\mathcal{U}$'s resource limitations; and
3. consideration in the evaluation process of predicted user responses to system behaviors.

Before demonstrating these properties of the framework with reference to the prototype system READY, we will introduce the context and overall architecture of this system.

SCENARIO AND ARCHITECTURE OF THE READY PROTOTYPE

Example Scenario

The initial example scenario of READY is illustrated in Figure 2: Users are drivers whose cars need minor repairs; they request assistance from the system in natural language via mobile phone.

We conducted a field study in which this scenario was realistically realized with the help of a professional auto repairman and 9 naive subjects, who dealt with intentionally created auto repair problems. The transliterated dialogs confirmed that in this scenario drivers sometimes show signs of WM overload (for example when trying to perform an unfamiliar task with the car while talking to the repairman on the mobile phone)—though on the whole to a smaller degree than in the emergency call situation. In this study, time pressure was manipulated with differing deadlines and monetary rewards; natural causes of time pressure could range from general discomfort with the situation to the need to repair the car quickly so as to be able to reach an appointment on time.

The input that we give to READY when testing the system—and the dialog contributions that it is designed to be able to produce—are modeled closely on the transliterations of these dialogs. In this way we hope to ensure that READY provides realistic solutions to realistic problems, even though the system is still far from being mature and complete enough for practical application.

Overall Architecture

Figure 3 shows the current system architecture. The USER INTERFACE is designed so as to make it unnecessary for READY to deal with the challenging problems of speech processing raised by the scenario. Input is done via a natural language menu interface with which the "user" can compose utterances and specify a number of aspects of their form, such as the position and length of pauses (see Figure 4).

What is sent to the DIALOG MANAGEMENT component is a representation of $\mathcal{U}$'s utterance that contains the information that READY needs in order to update its user model and determine an appropriate response, i.e.: the meaning of $\mathcal{U}$'s utterance in the current dialog context plus any evidence that can be extracted from the input (including noises from the environment) that bears on $\mathcal{U}$'s current resource limitations.

The DIALOG MANAGEMENT component uses its knowledge

---

[1] The volume [13] includes a representative sample of these techniques, and its Reader's Guide provides a classification.

[2] The classic reference for this family of techniques is [20]. A survey of their use for user and student modeling is given in [10], while a briefer more recent discussion is offered in [8].

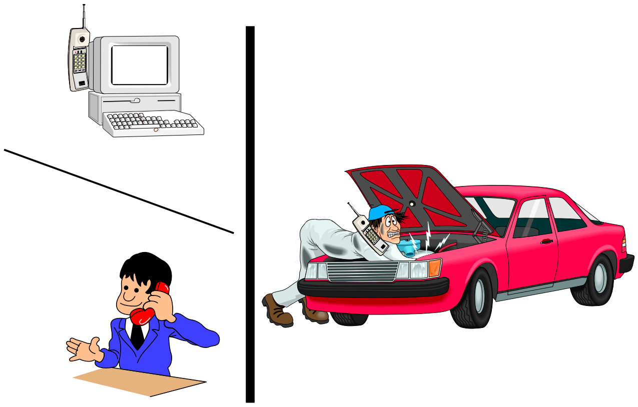


Figure 2. Illustration of READY's initial example scenario, showing the similarity of READY's role to that of an auto repairman who offers advice by phone.

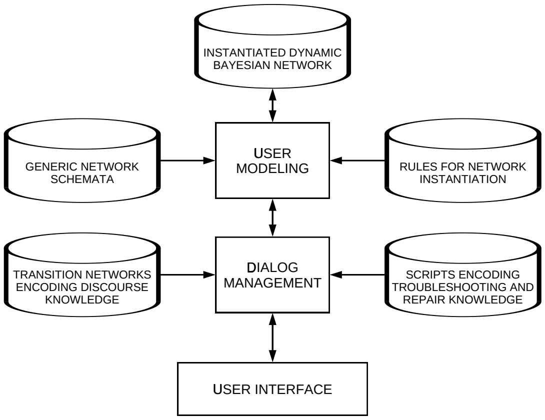
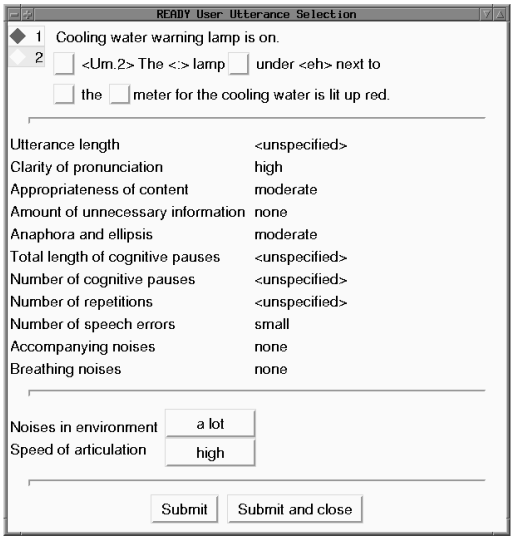


Figure 3. Architecture of the READY prototype.

(Boxes denote processing components, cylinders denote knowledge bases, and arrows show the flow of information.)

about the current dialog state and about possible diagnosis and repair actions to determine a set of possible dialog contributions (e.g., instructions) that might make sense in the current situation—not yet taking into account $\mathcal{U}$'s resource limitations.[3] These contributions may differ in their basic content (e.g., prescribing a simple or a complex action) and/or in their form (e.g., using simple, redundant formulations or concise, technical ones). The DIALOG MANAGEMENT component then sends the possible contributions to the USER MODELING component, which decides which one seems best in the light of $\mathcal{U}$'s current resource limitations.



● 1 Cooling water warning lamp is on.

○ 2 <Um.2> The <:> lamp under <eh> next to the meter for the cooling water is lit up red.

Utterance length	<unspecified>
Clarity of pronunciation	high
Appropriateness of content	moderate
Amount of unnecessary information	none
Anaphora and ellipsis	moderate
Total length of cognitive pauses	<unspecified>
Number of cognitive pauses	<unspecified>
Number of repetitions	<unspecified>
Number of speech errors	small
Accompanying noises	none
Breathing noises	none

Noises in environment: a lot
Speed of articulation: high

Figure 4. Translated screen shot of READY's menu interface for simulating speech input.

(Either of the two utterances shown at the top can be selected by the user for input; for the second, longer one, a number of variants can be specified via the pull-down menus attached to parts of the utterance. The table in the middle shows the characteristics of the currently selected utterance—here, the first one—that the system will treat as significant.)

MODELING OF RESOURCE LIMITATIONS

Example of a Network Schema

All of the USER MODELING component's assumptions about relevant causal relationships are represented in *network schemata* (cf. the knowledge source in the upper left of Figure 3). Each time the USER MODELING component is asked by the DIALOG MANAGEMENT component to evaluate a possible utterance (e.g., a question), it uses the appropriate schema in order to extend its Bayesian network model of $\mathcal{U}$. Each such schema corresponds to a typical sequence of two dialog moves. For example, Figure 5 shows, in simplified form, the schema for the sequence "Question (by system) → Answer (by user)". This schema corresponds to a pair of *time slices* of a *dynamic Bayesian network*.

Each box with a solid border represents a node that corresponds to a single variable; the arrows correspond to causal relationships among variables. Each box with a dashed border denotes a group of variables that play a similar role in the network.[4]

The variables in the left-hand and right-hand time slices are related to $\mathcal{U}$'s success in understanding and answering a question, respectively. In each time slice, there is a group of observable variables (at the bottom in the figure) that serve, among other things, as symptoms of $\mathcal{U}$'s resource limitations. The only other variables whose values are ever known to $\mathcal{S}$ with certainty are the ones in the group in the upper left; these represent properties of a possible utterance by $\mathcal{S}$ that will influence the demands that the utterance will place on $\mathcal{U}$'s time and WM.

---

[3] As with the processing of input utterances, $\mathcal{S}$'s knowledge about dialog structures and auto repair problems has been kept as simple as possible, so that we can focus on the key problem of adapting to $\mathcal{U}$'s resource limitations.

[4] A complete graphical depiction of this schema shows all of the causal relationships among variables within a group as well as their relationships to variables outside the group. By way of summary, in Figure 5 each group is depicted as if it were a single node in the network.

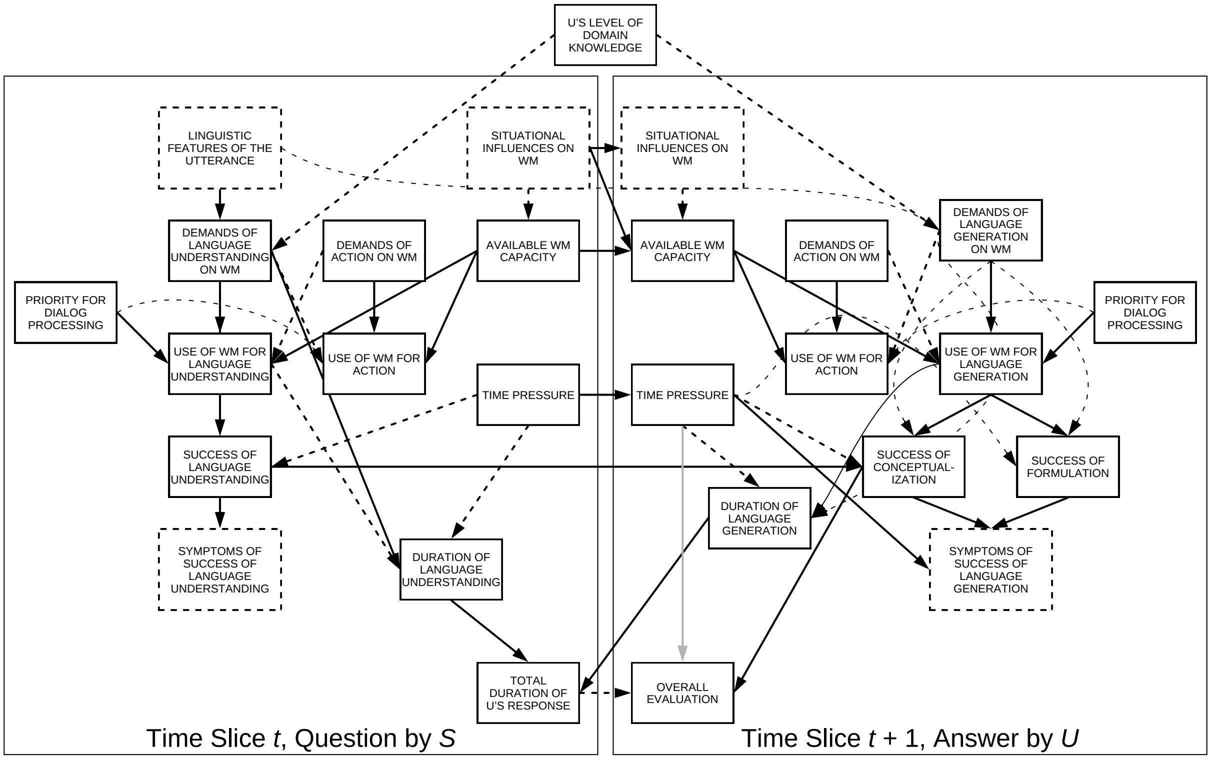
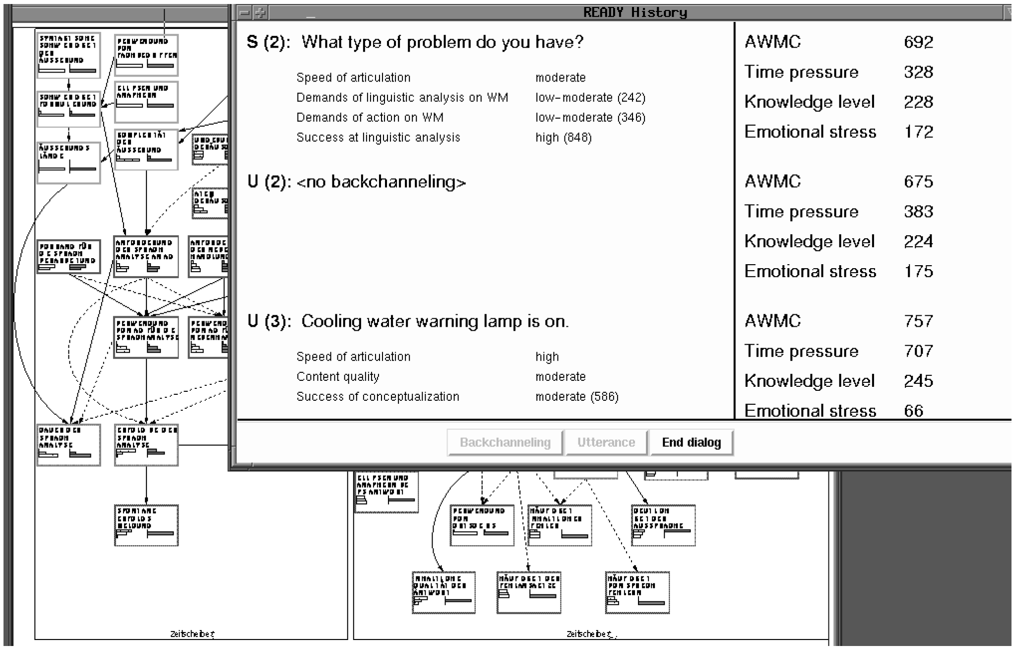


Figure 5. A schema for the construction of two time slices of a dynamic Bayesian network in READY.

(Solid and dashed arrows denote positive and negative cause influences, respectively. See the text for further explanation.)



S (2): What type of problem do you have?

Speed of articulation	moderate
Demands of linguistic analysis on WM	low–moderate (242)
Demands of action on WM	low–moderate (346)
Success at linguistic analysis	high (848)

AWMC	692
Time pressure	328
Knowledge level	228
Emotional stress	172

U (2): <no backchanneling>

AWMC	675
Time pressure	383
Knowledge level	224
Emotional stress	175

U (3): Cooling water warning lamp is on.

Speed of articulation	high
Content quality	moderate
Success of conceptualization	moderate (586)

AWMC	757
Time pressure	707
Knowledge level	245
Emotional stress	66

Figure 6. Translated version of READY's main interaction screen.

(The numbers on the right are the expected values of the system's current assessments of key variables, including available WM capacity ("AWMC"). In this case, the user has just input the problem description "Cooling water warning lamp is on". Because of its unusual terseness and the absence of any symptoms of high WM load, READY's estimate of $\mathcal{U}$'s time pressure increases quite drastically, and the estimate of $\mathcal{U}$'s available WM capacity increases somewhat as well. The reasons for these and other changes can be examined at any time in the fully instantiated Bayesian network, shown here in the background.)

Conceptualization of Resource Limitations

The variables in the middle level of Figure 5 correspond to theoretical constructs. Since their definitions are not tightly constrained by empirical evidence, they require some theoretical assumptions to be made; note that alternative assumptions could be introduced without changing the basic approach.

Time Pressure

The variable TIME PRESSURE is viewed as the subjective cost of time to $\mathcal{U}$. It is assumed that an increase in time pressure increases $\mathcal{U}$'s desire to reduce the duration of an action, even when such a reduction exacts a price in terms of other criteria (such as the likelihood of success of the action). For example, in the first time slice TIME PRESSURE is assumed to have a negative influence on both the time that $\mathcal{U}$ spends trying to understand $\mathcal{S}$'s utterance and the likelihood that $\mathcal{U}$ will succeed. In the second time slice, similar influences are shown for the tasks of conceptualizing and formulating an utterance.

Working Memory

In research in cognitive psychology, some highly differentiated models of working memory and its use have been developed which include several subsystems such as a *central executive*, a *phonological loop*, a *visuo-spatial scratchpad*, and sometimes further subsystems (see, e.g., [2]). For an initial effort at on-line modeling of a user's interaction with a interactive system, these models are too fine-grained. In the present version of READY's model, therefore, WM is treated as if it were a homogeneous store with a particular *capacity*. (A comparable simplification is employed, e.g., in [15].) In the near future we will systematically investigate more differentiated modeling schemas for WM to see if they contribute added value, at least for particular types of tasks ([12]).

The simplified conception is based on the following assumptions: At each point in the interaction, $\mathcal{U}$ has some *available WM capacity* that can be used to handle his or her tasks. This available capacity may be less than $\mathcal{U}$'s total WM capacity, for example if $\mathcal{U}$ is agitated or distracted by events in the environment (cf. the group of variables SITUATIONAL INFLUENCES ON WM in Figure 5).

The WM capacity that $\mathcal{U}$ can devote to interaction with $\mathcal{S}$ may be reduced further if $\mathcal{U}$ tries to perform another task simultaneously. For example, in the first time slice in Figure 5, $\mathcal{U}$ might be performing an action while listening to a question asked by $\mathcal{S}$. Each task that $\mathcal{U}$ performs is assumed to create a particular *demand* on $\mathcal{U}$'s WM. When two tasks have to be performed simultaneously, the available WM capacity may be too low to permit both to be performed without problems. In this case, the way $\mathcal{U}$ uses his or her WM for the two tasks is assumed to depend on the relative priority of the system-related task for $\mathcal{U}$. This relative priority is represented in each half of Figure 5 by a variable PRIORITY FOR DIALOG PROCESSING.

Temporal Relationships Among Variables

As was mentioned earlier, one of the challenges involved in the modeling of time and WM constraints is the need to deal with different types of change in the variables being modeled.

Most of the variables in Figure 5 are *temporary*: Because they refer to a brief event or state, they are defined only within a single time slice. Other variables—for example, U'S LEVEL OF DOMAIN KNOWLEDGE, are *static*: They are defined in all time slices, and any changes in their value during the course of an interaction are assumed to be negligible. But the critical variables TIME PRESSURE and AVAILABLE WM CAPACITY are *dynamic*: They are defined in all time slices, but their values can change significantly during an interaction, for example because of situational influences. The schema in Figure 5 illustrates a frequently applied method for handling a dynamic variable in a Bayesian network: A separate node for the variable is included in each time slice and is linked to the corresponding node in the previous time slice. A detailed discussion and justification of the way this method is applied in READY is given in [23], and a different application of this basic method to a user modeling problem is presented in [1].[5]

Selection of the System's Dialog Contributions

When the USER MODELING component is asked by the DIALOG MANAGEMENT component to evaluate a possible utterance of $\mathcal{S}$, it proceeds as follows: A network schema is chosen that will allow $\mathcal{S}$ to anticipate and evaluate the consequences of the utterance (e.g., if the utterance is a question, the schema in Figure 5). The Bayesian network that has been constructed so far is extended with two new time slices. The variables in the group LINGUISTIC FEATURES OF THE UTTERANCE are instantiated as dictated by the properties of the candidate utterance. The other root nodes are instantiated on the basis of information that $\mathcal{S}$ has about the current situation—for example, concerning the likelihood that $\mathcal{U}$ is (still) occupied with some physical action. The extended network is then evaluated, giving rise to new probabilistic expectations about the variables in both time slices. That is, $\mathcal{S}$ anticipates both the immediate processing of its utterance by $\mathcal{U}$ and $\mathcal{U}$'s subsequent responses to the utterance (e.g., $\mathcal{U}$'s answering of the question or $\mathcal{U}$'s carrying out of an instruction).

The purpose of this anticipation is to generate an OVERALL EVALUATION of the candidate utterance (cf. the bottom nodes in Figure 5) using evaluation criteria that are presumably in line with $\mathcal{U}$'s own interests: the total time required for $\mathcal{U}$ to respond (weighted negatively) and the success of $\mathcal{U}$'s re-

---

[5] In [8], an alternative approach is discussed that eliminates the need to create multiple time slices, while leaving implicit some of the relationships that are expressed explicitly with time slices. In a dialog system, it is useful to introduce at least one time slice for each dialog contribution, because of the qualitatively different events that are involved in each contribution.

sponse. The relative weight that $\mathcal{U}$ would attach to these two criteria is assumed to depend on $\mathcal{U}$'s time pressure, as is indicated in the figure.

The overall pattern is that $\mathcal{S}$'s perception of $\mathcal{U}$'s resource limitations can affect $\mathcal{S}$'s choice of utterances in two ways:

- Most directly: High time pressure biases the overall evaluation in favor of utterances that $\mathcal{U}$ can deal with quickly.
- Less directly: $\mathcal{U}$'s time pressure and WM availability can determine the speed and success of $\mathcal{U}$'s response in more or less complex ways, thus indirectly influencing the overall evaluation.

Use of Influence Diagrams to Narrow the Search

This procedure of systematically evaluating each possible system utterance can be quite time-consuming if there are a lot of possible utterances. We have therefore tested an alternative approach, which is now being integrated into the prototype: Instead of creating a separate possible extension of its basic network for each possible utterance, $\mathcal{S}$ creates a single *influence diagram* (see, e.g., [24], [20]), in which several aspects of an utterance (e.g., syntactic complexity and use of technical terms) are represented by *decision nodes* and the OVERALL EVALUATION is treated as a *value node*. The procedure used for processing influence diagrams (see [14]) gives values for the variables in the decision nodes that would lead to the best possible overall evaluation—i.e., desirable properties of $\mathcal{S}$'s next utterance. It is relatively easy to find the candidate utterances that come closest to having these properties.

Updating the User Model

The utility of the procedures described in the previous subsection of course depends largely on the accuracy of the user model that has been built up in the course of the dialog. This model is updated on the basis of information about $\mathcal{U}$'s behavior that is sent to the USER MODELING component by the DIALOG MANAGEMENT component. This information may concern either $\mathcal{U}$'s immediate feedback to $\mathcal{S}$'s utterance (e.g., one of the possible SYMPTOMS OF SUCCESS OF LANGUAGE UNDERSTANDING in the first time slice) or $\mathcal{U}$'s response to the utterance (e.g., one or more SYMPTOMS OF SUCCESS OF LANGUAGE GENERATION in the second time slice). The network is reevaluated, leading to new assessments of the variables in both time slices. Where static or dynamic (as opposed to temporary) variables are affected, these reassessments will influence $\mathcal{S}$'s choice of subsequent utterances.

TECHNICAL FEASIBILITY

The explicit representation of many types of causal relationships makes READY's adaptation relatively easy to understand, criticize, and improve. But it does lead to a fairly high degree of complexity. For example, the two time slices of a network constructed on the basis of the schema in Figure 5 currently comprise about 60 nodes; and the complete network that is built up during the dialog includes one time slice for each dialog contribution.

The main implementation language for READY is Lucid COMMON LISP, but the Bayesian networks and influence diagrams are processed with the tool NETICA ([18]). On a SUN Ultra 1 with 147 MHz and 256 MByte of RAM, the time required for interpretation of the evidence in a single user utterance varies from about .5 sec to about 3 sec, depending on how many time slices have been added to the network; a maximum of 12 time slices are retained, any earlier ones being eliminated through a *rolling-up* procedure. Evaluation of a single possible system utterance is slower, taking up to 12 sec under the same conditions. Fortunately, the processing of a related influence diagram, which (as mentioned above) can replace the evaluation of a number of similar utterances, takes roughly the same amount of time as the evaluation of a single utterance.

In sum, the techniques as currently implemented are one or two orders of magnitude too slow for practical application.

FURTHER RESEARCH

We are beginning to explore the application of techniques for learning Bayesian networks (see, e.g., [5]) as a way of improving the accuracy of the models. A planned adaptation of the prototype to a different domain (information provision in an airport) and different interaction techniques (e.g., gestural input and graphical output) will provide evidence concerning the generalizability of the approach ([12]).

All of the sections of this paper have shown, in different ways, that the problem of adapting to a user's changing time and WM limitations is a challenging one. The conceptual framework and the technical methods presented here are not being proposed as a final solution to the problem. In fact, one reason why we chose an elaborate and explicit modeling approach is that this approach will ultimately support principled and justifiable simplifications of the sort that will be necessary before a practical resource-adaptive system can be deployed.

ACKNOWLEDGEMENTS

This research is being supported by the German Science Foundation (DFG) in its Collaborative Research Center on Resource-Adaptive Cognitive Processes, SFB 378, Project B2, READY. We thank the Saarbrücken Fire Department for its cooperation, Wolfgang Wahlster for his continuing support, and the four anonymous reviewers for their thoughtful comments and suggestions.

REFERENCES

1. D. W. Albrecht, I. Zukerman, and A. E. Nicholson. Bayesian models for keyhole plan recognition in an adventure game. *User Modeling and User-Adapted Interaction*, 8:5–47, 1998.

2. A. D. Baddeley. *Working Memory*. Oxford University Press, Oxford, 1986.

3. A. Berthold. Repräsentation und Verarbeitung sprachlicher Indikatoren für kognitive Ressourcenbeschränkungen [Representation and processing of linguistic indicators of cognitive resource limitations]. Master's thesis, Department of Computer Science, University of Saarbrücken, 1998.
4. K. A. Ericsson and H. A. Simon. *Protocol Analysis: Verbal Reports as Data*. MIT Press, Cambridge, MA, Rev. edition, 1993.
5. D. Heckerman. A tutorial on learning with Bayesian networks. Technical Report MSR-TR-95-06, Microsoft Research, March 1995. Revised November 1996.
6. E. Horvitz. Transmission and display of information for time-critical decisions. Technical Report MSR-TR-95-13, Microsoft Research, February 1995.
7. E. Horvitz and M. Barry. Display of information for time-critical decision making. In P. Besnard and S. Hanks, editors, *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, pages 296–305. Morgan Kaufmann, San Francisco, 1995.
8. E. Horvitz, J. Breese, D. Heckerman, D. Hovel, and K. Rommelse. The Lumière project: Bayesian user modeling for inferring the goals and needs of software users. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, pages 256–265, Madison, WI, 1998.
9. E. H. Hovy. *Generating Natural Language Under Pragmatic Constraints*. Erlbaum, Hillsdale, NJ, 1988.
10. A. Jameson. Numerical uncertainty management in user and student modeling: An overview of systems and issues. *User Modeling and User-Adapted Interaction*, 5:193–251, 1996. Available from http://www.cs.uni-sb.de/users/jameson/.
11. A. Jameson. Wie gehen wir mit dem Arbeitsgedächtnis unserer Dialogpartner um? Eine Integration von Ergebnissen aus vier Forschungsrichtungen [How do we deal with the working memory of our dialog partners? An integration of results from four areas of research]. In H. Mandl, editor, *Bericht über den 40. Kongreß der Deutschen Gesellschaft für Psychologie in München 1996*, pages 258–263. Hogrefe, Göttingen, Germany, 1997.
12. A. Jameson. Adapting to the user's time and working memory limitations: New directions of research. In U. J. Timm and M. Rössel, editors, *ABIS-98, Adaptivität und Benutzermodellierung in interaktiven Softwaresystemen*. FORWISS, Erlangen, Germany, 1998.
13. A. Jameson, C. Paris, and C. Tasso, editors. *User Modeling: Proceedings of the Sixth International Conference, UM97*. Springer Wien New York, Vienna, New York, 1997. Available on-line from http://www.cs.uni-sb.de/UM97/ and http://um.org/um_97/proceedings-overview.html.
14. F. Jensen, F. V. Jensen, and S. L. Dittmer. From influence diagrams to junction trees. In R. Lopez de Mantaras and D. Poole, editors, *Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence*, pages 367–373. Morgan Kaufmann, San Francisco, 1994.
15. M. A. Just and P. A. Carpenter. A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99:122–149, 1992.
16. A. Kashihara, T. Hirashima, and J. Toyoda. A cognitive load application in tutoring. *User Modeling and User-Adapted Interaction*, 4:279–303, 1995.
17. E. Marx. *Über die Wirkung von Zeitdruck auf Sprachproduktionsprozesse [The Effect of Time Pressure on Speech Production Processes]*. PhD thesis, University of Münster, Germany, 1984.
18. Norsys Software Corp., Vancouver, BC, Canada. *Netica API Programmer's Library Reference Manual*, 1997.
19. S. Oviatt. Multimodal interactive maps: Designing for human performance. *Human-Computer Interaction*, 12:93–129, 1997.
20. J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, CA, 1988.
21. C. Roßnagel. Kognitive Belastung und Hörerorientierung beim monologischen Instruieren. *Zeitschrift für Experimentelle Psychologie*, 42:94–110, 1995.
22. R. Rummer. Verarbeitungsbeschränkungen bei der Sprachproduktion: Ein experimenteller Ansatz zur Erforschung sprachlicher Makroplanungsprozesse. In R. H. Kluwe, editor, *Strukturen und Prozesse intelligenter Systeme*, pages 41–63. Deutscher Universitäts-Verlag, Wiesbaden, 1997.
23. R. Schäfer and T. Weyrath. Assessing temporally variable user properties with dynamic Bayesian networks. In A. Jameson, C. Paris, and C. Tasso, editors, *User Modeling: Proceedings of the Sixth International Conference, UM97*, pages 377–388. Springer Wien New York, Vienna, New York, 1997.
24. R. D. Shachter. Evaluating influence diagrams. *Operations Research*, 34:871–882, 1986.
25. T. Weyrath. Erkennung von Arbeitsgedächtnisbelastung und Zeitdruck in Dialogen – Empirie und Modellierung mit Bayesschen Netzen [Recognition of working memory load and time pressure in dialogs: Empirical studies and modeling with dynamic Bayesian Networks]. Master's thesis, Department of Computer Science, University of Saarbrücken, 1998.