On the Role of the NIMITEK Corpus in Developing an Emotion Adaptive Spoken Dialogue System

Milan Gnjatović, Dietmar Rösner

Otto-von-Guericke-University Magdeburg, Department of Knowledge Processing and Language Engineering
P.O. Box 4120, D-39016 Magdeburg, Germany
{gnjatovic, roesner}@iws.cs.uni-magdeburg.de

Abstract

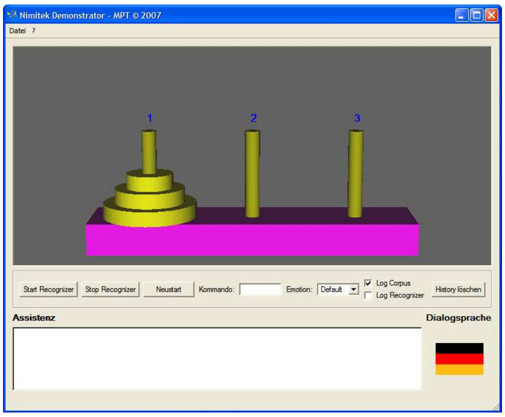
This paper reports on the creation of the multimodal NIMITEK corpus of affected behavior in human-machine interaction and its role in the development of the NIMITEK prototype system. The NIMITEK prototype system is a spoken dialogue system for supporting users while they solve problems in a graphics system. The central feature of the system is adaptive dialogue management. The system dynamically defines a dialogue strategy according to the current state of the interaction (including also the emotional state of the user). Particular emphasis is devoted to the level of naturalness of interaction. We discuss that a higher level of naturalness can be achieved by combining a habitable natural language interface and an appropriate dialogue strategy. The role of the NIMITEK multimodal corpus in achieving these requirements is twofold: (1) in developing the model of attentional state on the level of user's commands that facilitates processing of flexibly formulated commands, and (2) in defining the dialogue strategy that takes the emotional state of the user into account. Finally, we sketch the implemented prototype system and describe the incorporated dialogue management module. Whereas the prototype system itself is task-specific, the described underlying concepts are intended to be task-independent.

1. Introduction

One of the widely accepted postulates of human-machine interaction (HMI) is that it should be as natural as possible. Watt (2004) introduces the term *habitable language* to denote a language in which users can express themselves naturally. Considering advisory systems, Guindon (1988, 191–2) extends this definition to apply to the *habitable natural language interface*. Carbonell (1986, 162) introduces similar criteria for *robust natural language interfaces*. In summary, user habitability is defined in terms of naturalness of interface language, little conscious effort invested by the user to avoid uttering sentences that would not be recognized by the system, linguistic coverage, robustness of the system's behavior, speed of response, informative error messages, etc. However, the issue of naturalness considers more than just the interface. In order to achieve a higher level of naturalness, the user has to be convinced that she participates in the communication. To achieve this, an appropriate dialogue strategy applied by the system should be combined with a habitable natural language interface. Such a dialogue strategy should take various interaction features into account, including the emotional state of the user.

This kind of research is essentially supported by corpora of HMI (Douglas-Cowie et al., 2000). Representing an integration of several lines of our previous research, this paper reports on the creation of the multimodal NIMITEK corpus of affected behavior in HMI and its role in the development of an already implemented spoken dialogue system. The NIMITEK prototype system is a spoken dialogue system for supporting users while they solve problems in a graphics system (e.g., the Tower-of-Hanoi puzzle). The central feature of the system is adaptive dialogue management. As described in the paper, the system dynamically defines a dialogue strategy according to the current state of the interaction (including also the emotional state of the user).

The dedicated prototypical task implemented in our prototype system is a 3-disks version of the *Tower of Hanoi* puzzle introduced by Édouard Lucas in 1883. The puzzle consists of three pegs and three disks of different sizes. At the start of the game, the disks are stacked in order of size on the leftmost peg, as shown in Figure 1. The goal of the puzzle is to move the entire stack to the rightmost peg moving according to the following rules: only one disk can be moved at a time, all three pegs can be used, and no disk may be placed on top of a smaller disk. In the NIMITEK system, users are allowed only to verbally address the system (i.e., there is no mouse or keyboard, etc.).



Emotion: Default
☑ Log Corpus
□ Log Recognizer

Figure 1: Tower-of-Hanoi Puzzle: Screen display of the NIMITEK prototype system

In the first part of the paper, we report about the collecting of the NIMITEK multimodal corpus of affected behavior (Gnjatović and Rösner, 2006), its evaluation in terms of definition of a data-driven model of user states, and the annotation of spoken dialogue acts. In the second part, we explain how we used the NIMITEK corpus to achieve a more habitable language interface and to develop a dialogue strategy for supporting users. Finally, we report about the implementation of the dialogue management module incorporated in the system.

2. The NIMITEK Corpus

The NIMITEK corpus was collected using the refined Wizard-of-Oz (WOZ) technique introduced by Gnjatović and Rösner (2006). This refinement addresses the problem of role-playing subjects (Batliner et al., 2000; Pirker and Loderer, 1999) in such a way that a WOZ scenario could result in *ecologically valid* data (Douglas-Cowie et al., 2000). Subjects in the WOZ experiment were asked to undertake a test of both intelligence and communication abilities supported by the spoken natural language dialogue system. In fact they were confronting a set of graphically based tasks specified with the intention to stimulate the verbal interaction between subjects and the system. The subjects were only allowed to give spoken instructions to the system. They were given an illusion that they communicate with the system, while the human operator in the other room played the role of the system. The set of instructions accepted by the system was not predefined—their determination and formulation was imputed to be a part of the test. Stimuli used for an emotional response were intentional misunderstanding of subject's request and performing an incorrect operation, pretending not to understand subject's request and asking for a repetition, confronting subjects to unsolvable tasks, etc. Ten healthy native German speakers (7 female, 3 male, in age from 18 to 27; mean 21.7) participated in the experiment. Almost 15 hours of session time is recorded. The language used in the experiment was German. In addition to the video and audio streams, the desktop of the subject's PC (where the tasks, e.g., Tower-of-Hanoi, were solved) was also recorded. The corpus was gathered in 2006 and annotated in 2007.

2.1. Evaluation of the Emotional Content

The evaluation of the emotional content of the NIMITEK corpus was performed in two phases. The first phase of the evaluation process (Gnjatović and Rösner, 2006) had the primary aim to (1) assess the level of ecological validity of the NIMITEK corpus and to (2) define a data-driven model of user states. Three types of evaluators participated in this process. The first group (three German native speakers) was allowed only to hear audio recordings. These evaluators were influenced by lexical meaning as well. The second group (three non-German speakers) was also allowed only to hear audio recordings. These evaluators—two Serbian native speakers and a Hungarian native speaker—did not have knowledge of German language, have never lived in a German speaking environment, and did not have any contact with German language in everyday life. For this group the lexical meaning was missing and thus the prosody became central for evaluating emotions. Finally, one additional German native speaker was allowed to simultaneously hear and see video recordings.

Four randomly selected sessions were evaluated in complete duration (approximately five hours). The evaluation unit was a dialogue turn or a group of several successive dialogue turns. Only subjects' expressions were evaluated. Evaluators performed this perception test independently from each other. Evaluators assigned one or more labels to each evaluation unit. The choice of labels was data driven—the evaluators were allowed to introduce labels according to their own perception. Introduced labels relate to emotion, emotion-related state of or talk style of the subject. Recordings evaluated as emotional were further graded with respect to their intensity (low,medium, or high).

This evaluation phase demonstrated a satisfying level of ecological validity of the corpus: genuine emotions of different intensities were elicited, subjects signaled them overtly and their emotional expressions were extended both in time and modality (voice and mimic). Detailed evaluation results of this phase and a more elaborative discussion related to ecological validity of the corpus will be given in another paper. Here, we concentrate on evaluation aspects that are particularly important for the implementation of the NIMITEK prototype system—the definition of a data-driven model of user states. It should be noted that some of the introduced labels represent different but closely related emotions or emotion-related states(e.g., *confused*, *insecure*, and *fear*; *disappointed* and *sadness*; or *pleased*, *joy* and *surprised*). Keeping in mind the purpose for which our prototype system was planned, it was necessary to reduce the number of different user states. In order to define a usable data-driven model of user states, we grouped labels that relate to similar or mixed emotions or emotion-related states. Following clarifications collected from the evaluators, we mapped these labels onto six classes that form the ARISEN model of user states, given in Table 1.

Class	***Mapped labels***
Annoyed	anger, nervousness, stressed, impatient
Retiring	fear, insecure, confused
Indisposed	sadness, disappointed, accepting, boredom
Satisfied	joy, contentment, pleased
Engaged	thinking, surprised, interested
Neutral	neutral

Table 1: The ARISEN model of user states

The aim of the second evaluation phase was to prove the the appropriateness of this mapping. The experimental sessions evaluated in the first phase were re-evaluated by a new group of evaluators. The evaluators could use only labels from the ARISEN model. In comparison with the first phase, the re-evaluation was performed over smaller evaluation units. The same evaluation material was divided in 2720 evaluation units. Each evaluation unit was evaluated by four or five German-speaking evaluators. Independently from each other, they assigned one or more labels to each evaluation unit. We used majority voting in order to attribute labels to evaluation units. If at least three evaluators agreed upon a label, it was attributed to the evaluation unit. The evaluation results of the second phase are given in Table 2.

We comment the results of the second phase in light of the purpose for which the system was planned. Among our

Evaluation units	*Number*
with no majority voting	315 (11.58%)
with one assigned label	1907 (70.11%)
with two assigned label	476 (17.5%)
with three assigned label	22 (0.81%)
total	2720 (100%)
Label	***Nr. of evaluation units attributed with the label***
Annoyed	487 (17.9%)
Retiring	111 (4.08%)
Indisposed	156 (5.74%)
Satisfied	106 (3.9%)
Engaged	1548 (56.91%)
Neutral	517 (19.01%)

Table 2: Results of the second evaluation phase

Speech function	*Example*
Command	”Rotate to the left.”
Offer	-
Question	”What are names of these rings?”
Statement	”You are not doing what I say.”

Table 3: Subjects’ utterances illustrating speech functions

Speech function	*Number of occurrences*
Command	6798 (76.25%)
Question	390 (4.37%)
Statement	1727 (19.37%)
Total	8915 (100%)

Table 4: Annotation of dialogue acts

research plans, we intended to define and implement a dialogue strategy applied by the system that addresses a negative state of the user. The results of the second evaluation phase gives us a better insight in what a negative user state in our scenario means. We give briefly three different explanation. First, the user can be frustrated due to problems that occurred in the communication, e.g., the system misunderstands the subject’s request and performs an incorrect operation. Second, the user can be discouraged because she does not know how to solve a given task. Finally, there can be a lack of interest in the user’s attitude to solve the task. The negative state comprises the user states *Annoyed*, *Retiring* and *Indisposed*. An appropriate dialogue strategy that addresses these points is discussed in Subsection [X]. In contrast, the positive state relates to a user that is motivated to solve the task and/or satisfied with the communication. This includes the user states *Engaged* and *Satisfied*.

2.2. Annotation of Dialogue Acts

Considering the nature of dialogue, Halliday (1994, 68–71) suggests an interpretation of the clause in its function as an *exchange*. He distinguishes between two fundamental types of speech role—giving and demanding— as well as between two basic types of the exchange commodity—verbal (*information*) and nonverbal (*goods-&-services*). The role in the exchange and the exchange commodity define the four primary speech functions of: command (demanding goods-&-services), offer (giving goods-&-services), question (demanding information) and statement (giving information). We adopt this classification. Table 3 comprises subjects’ utterances from the NIMITEK corpus that illustrate the speech function. One entry in the table is empty. According to the experimental settings, the subjects were allowed only to verbally address the system. Thus, no offers produced by the subjects were annotated. It does not mean that there are not clearly marked body and mimic gestures produced by the subjects in the NIMITEK corpus—they are just not in the focus of our attention in this paper. The summarized results of the annotation of dialogue acts are given in Table 4. The given numbers relate only to utterances that were spontaneously produced by the subjects (i.e. utterances that were not predefined).

We mention two important implications of these results. First, 76.25% of subjects’ dialogue acts are commands. Therefore, we devote a particular attention to this class. Processing of users’ commands is discussed in more details in subsection 3.1.

The second implication relates to the dialogue strategy applied by the system. According to the experimental settings, problems in the communication were caused on purpose and the evaluation of emotional content demonstrated that subjects expressed their emotions overtly. Still, questions make only 4.37% of all utterances produced by the subjects. In addition, the subjects demanded support from the system in only 12 of 6798 commands, although the human operator playing the role of the system offered support 59 times explicitly using the word *help*, e.g., *Do you need help?* (in German: *Brauchen Sie Hilfe?*). Thus, a dialogue strategy aimed to support the user to overcome problems that occur in the communication must not rely on the assumption that the user will clearly state a need for a support. The system should rather detect such a need and be initiator and carrier of provided support. This issue is further discussed in Subsection 3.2.

3. Towards a More Natural HMI

In the following subsections, we describe how the NIMITEK corpus was used to achieve a more habitable language interface and to define a dialogue strategy for supporting users.

3.1. Model of the Attentional State

Analysis of commands from the NIMITEK corpus shows that subjects often produced elliptical or minor sentences, as well as context dependent and relative sentences. This effect is due to the fact that users omit to utter information that they believe is known by the system and, in the same time, they bring new information in the focus of attention. Their belief about an additional non-linguistic context shared between subjects and the simulated system was supported by the desktop of their PC. Subjects considered it to be a reliable source of information. As a small illustration of this claim, in such cases when wizard’s actions were in a collision with the actual state on the desktop, subjects often tried to refer first to the desktop (Gnjatović and Rösner, 2006, 63). Here are some examples of users' commands from the NIMITEK corpus:

- *the second smallest ring on the two,*
- *number one on the number two,*
- *the next ring on the two,*
- *on the two,*
- *back.*

Although these commands may have the same propositional content, only the first command provides sufficient information necessary for its interpretation. To interpret the rest of these commands, additional information (e.g., contextual information, dialogue history, etc.) has to be taken into account. In order to achieve a higher level of naturalness of the communication, the users have to be allowed to flexibly construct their commands. There are at least three interaction features that are important for processing of such users' commands—the state of the task, the history of interaction, and the attentional information. The first two features are easy to define and implement for the given scenario. (e.g., a 3-disk version of the Tower-of-Hanoi puzzle gives a set of $27(3^3)$ possible states of the task). However, to model the third feature, we applied a more general approach.

Attentional information is already recognized as crucial for processing of utterances in discourse by Grosz and Sidner (1986, 175). We model attentional information on the level of a users command. Inspection of the commands from the NIMITEK corpus resulted in a set of focus classes whose instances form attentional information. These classes can be ordered from the most general to the most specific:

- task focus—relating to the currently ongoing task,
- object focus— relating to the currently selected object,
- action focus—relating to the action that is to be performed over the selected object,
- direction focus—relating to further specification of the action to be performed.

For example, the commands *the second smallest ring on the two* contains two phrases that determine two focus instances:

- the phrase *the second smallest ring* relates to the middle disk that belongs to the object focus class,
- the phrase *on the two* relates to the second peg that belongs to the direction focus class.

Instances of these classes are interrelated—an instance of a more specific focus class is a sub-focus of an instance of the immediately preceding more general focus class. We map focus instances onto *the focus tree*: each instance is represented by a node in the focus tree, instances from the same focus class are placed on the same tree level, and each node, except the root node, represents a sub-focus of its parent node. The focus tree for the 3-disk version of the Tower-of-Hanoi puzzle is given in Figure 2. To each focus instance in the focus tree a set of phrases that represent it is assigned. At any given point, the current focus of attention is placed on exactly one node from the focus tree. Please note: In the Tower-of-Hanoi puzzle there is only one instance in the action focus class—*move*. The action focus level in Figure 2 is represented only for the purpose of generality, although it may appear to be unnecessary. However, in other graphical tasks that are also included in the NIMITEK corpus (such as Tangram, etc.) we can differentiate between several instances of the action focus class (e.g., translation and rotation).

During the processing of a user's command, focus instances comprised in it are automatically extracted and mapped onto the focus tree with respect to the position of the current focus of attention. The modeling method for the focus tree and the rules for transition of the focus of attention are intended to be task independent and are introduced in more detail in (Gnjatović and Rösner, 2007a). We state some advantages gained in our prototype system from such a modeling of attentional information with respect to processing of users' commands:

- Instead of predefining a grammar for accepted users commands, we allow flexible formulation of commands. Implementation of the natural language understanding module in the NIMITEK prototype was demonstrated to work well for different syntactic forms of user's commands: (i) elliptical commands (e.g., *on the two*, etc.), (ii) complex commands (i.e., commands containing words that are not a part of the vocabulary recognized by the speech recognition module); and (iii) context dependent commands (e.g., *the next ring*, *back*, etc.)
- Processing of commands is independent from the predefined grammar used in the speech recognition module.
- Processing of commands is independent from the size of vocabulary (i.e. the sets of predefined phrases), and it was demonstrated to function for German and English.

3.2. Dialogue Strategy

Although the importance of a timely recognition of problems occurring in HMI has already been recognized (Batliner et al., 2000), less attention is devoted to another aspect of the dialogue management— the resolution of these problems. The dialogue strategy introduced in (Gnjatović and Rösner, 2007b; Gnjatović and Rösner, 2008a) is aimed to address the latter aspect. The main idea is that according to the current state of the interaction, the system dynamically defines an appropriate dialogue strategy.

The state of interaction is defined as a composite of five interaction features:

- *State of the task*—determined by current positions of the disks.

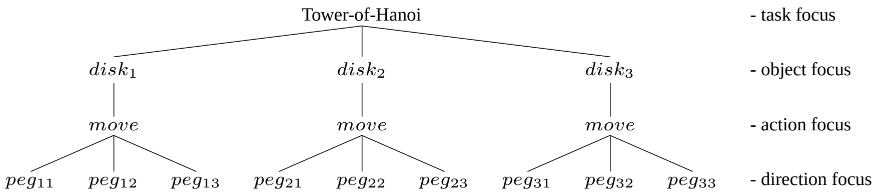


Figure 2: The focus tree for Tower-of-Hanoi puzzle

- *User's command*—each user command is assigned to one of the following classes {*valid command, illegal command, semantically incorrect command, help command, switching between interface languages, unrecognized command*}.
- *Focus of attention*—discussed in Subsection 3.1.
- *State of the user—negative, neutral*, or *positive*, as discussed in Section 2.1.
- *History of interaction*—a collection of previous values of other interaction features, previously applied dialogue strategies, and time stamps.

Although this paper primarily addresses the implementation of the dialogue management module incorporated in the NIMITEK prototype system, it should be noted that these features are of a general nature. The same holds for the dialogue strategy. It includes three decision making processes related to the way of providing support to users. These processes are also of a general nature, but, for the purpose of better clarity, we describe them with respect to the existing implementation of the dialogue management module in the NIMITEK prototype system.

Decision 1: When to provide support to the user? The system provides support when one of the following four cases is detected:

- user is misguided (e.g., the history of interaction shows that the state of the task has been pushed away from the solution, etc.),
- user instructs an illegal move (e.g., placing a bigger disk on top of a smaller disk, etc.),
- user instructs a semantically incorrect command (e.g., trying to move a peg instead of a disk, etc),
- users command is not recognized,
- user is inactive,
- user asks for support.

Decision 2: What kind of support to provide? Two kinds of support can be provided to the user:

- *Task-Support*—explaining the rules of the puzzle and helping to find its solution.
- *Interface-Support*—helping to formulate a valid command.

Task-Support is provided in cases when the user instructs an invalid command or when the history of interaction shows that the state of the task either draws back from the expected final state or does not make any significant progress towards the final state. Interface-Support is provided in cases when the user instructs a command that cannot be recognized or a semantically incorrect command, etc.

Decision 3: How to provide support? The manner of providing support is determined by the state of the user. A user in negative emotional state is provided with a more informative support than the user in positive or neutral emotional state. Thus, support is provided in two manners:

- ***Low intensity*** of support for users in positive or neutral emotional state.
- ***High intensity*** of support for users in negative emotional state.

Low intensity of Task-Support means to inform the user that her last move pushed her away from the final solution of the puzzle or that her last move violates the rules of the game. High intensity of Task-Support is to inform the user as well, but also to propose the next move. Providing low intensity of Interface-Support, the system guides the user to complete the started command by stating iterative questions (e.g., which disk should be selected?; where to move the selected disk?; etc.). High intensity of Interface-Support is to check whether the started command can be completed in such a way that it pushes the state of the task towards the final solution. If so, the system proposes such a completion of the command to the user. Otherwise, the system warns the user that the started command is not appropriate.

3.3. Dialogue Management Module

In this section, we describe the dialogue management module incorporated in the spoken dialogue system for supporting the users while they solve Tower-of-Hanoi puzzle (Figure 3). The speech recognition module, the emotional classifier and the graphical platform are also represented as interacting components. However, these components are not discussed in the paper. More details about emotion recognition from speech are given by Vlasenko et al. (2007), while emotions recognition from mimics is discussed by Niese et al. (2007). During the uninterrupted processing of users commands the dialogue management module takes as input the textual version (German, English) of a users command outputted from the speech recognizer. The focus instances are automatically derived from the given command. Then the following actions are undertaken:

- change of the state of the task: the command is recognized. This information is sent to the graphical platform for display (i.e., the command is performed),

- change of the focus of attention: the command is mapped onto the focus tree and the focus of attention is updated.

Then, if needed, the system applies a dialogue strategy according to the current state of the interaction, as described in Subsection 3.2.. In general, support information may contain a proposed move and a message for the user. To illustrate: For a misguided user in a positive emotional state, only an audio message informing her that her last move was wrong may suffice, while for a user in a negative emotional state it might be more appropriate if the system proposes the next correct move also. The support information is used in two ways:

- it is sent to the graphical platform for display. Audio messages are played back, and, in addition, their textual content is visually displayed on the screen. If there is also a move proposed by the system, it is graphically displayed. The disk to be moved is marked in red to make also visually clear—beside the audio and textual messages—that this currently performed move is a proposal of the system.

- it is internally used to actualize the state of the task, the focus of attention, and the history of interaction.

The NIMITEK prototype system is sketched in Figure 3.

4. Ongoing Work: Adequate Intonation of the Synthesized Speech Output

Considering information-seeking HMI, Bateman et al. (1998) place particular emphasis on achieving adequate intonation of the synthesized speech output of the system. They state that in such kind of interaction *intonation often is even the only means to distinguish between different dialogue acts, thus making the selection of the appropriate intonation crucial to the success of the information-seeking process* (Bateman et al., 1998).

In an ongoing experiment, the audio recordings from the NIMITEK corpus are used as stimuli in a functional magnetic resonance imaging (fMRI) study of prosody processing. Wendt and Scheich (2002, 699) state that one reason for the inconsistency in previous studies on prosody perception concerning the functional role of the hemispheres is due to the language material used (e.g., the stimulus material was not sufficiently evaluated, etc.). Therefore, in this experiment, particular attention is devoted to the choice of appropriate stimuli. Audio recording of utterances from the NIMITEK corpus that have the same or similar lexical meaning, but different prosodic intonation of functional elements in utterances (i.e., one of two focus instances comprised in a command is prosodically marked) are played back to subjects in the fMRI scanner. The task of the subjects was to detect which of two focus instances carries prosodic intonation. This study is expected to provide a better insight in how users percept prosodically intoned spoken output of the system. We plan to to apply that knowledge to enable the NIMITEK prototype system to automatically synthesize audio output with an appropriate intonation that will increase the level of naturalness of the communication.

5. Conclusion

In this paper, we reported on the role of the multimodal NIMITEK corpus of affected behavior in HMI in developing an emotion adaptive spoken dialogue system. The corpus was collected using the refined Wizard-of-Oz (WOZ) technique. This refinement addresses the problem of role-playing subjects in such a way that a WOZ scenario could result in ecologically valid data. The evaluation of emotional content demonstrated a satisfying level of ecological validity of the corpus: genuine emotions of different intensities were elicited, subject signaled them overtly and their emotional expressions were extended both in time and modality (voice and mimic). With respect to the signaled emotions, this corpus is unique for the German language.

Our intention was to develop a spoken dialogue system for supporting users while they solve problems in a graphics system (e.g., the Tower-of-Hanoi puzzle)—the NIMITEK prototype system. The central feature of the system is adaptive dialogue management. The underlying idea is that the system dynamically defines a dialogue strategy according to the current state of the interaction (including also the emotional state of the user). The application domain planned for the prototype system was also used in the WOZ simulation conducted to collect the NIMITEK corpus (i.e., a graphics system). Thus, we resorted to the NIMITEK corpus in order to get a better insight in emotion and emotion-related states of the users.

We concentrated on the development of the dialogue management module incorporated in the NIMITEK prototype system. The particular attention was devoted to the level of naturalness of interaction. We discussed that a higher level of naturalness can be achieved by combining a habitable natural language interface and an appropriate dialogue strategy. This paper reported the twofold role of the NIMITEK multimodal corpus of affected behavior in HMI in achieving these requirements: (1) in developing the model of attentional state on the level of users commands that facilitates processing of flexibly formulated commands, and (2) in defining the dialogue strategy that takes the emotional state of the user into account. Although the importance of a timely recognition of problems occurring in HMI has been already recognized, less attention is devoted to another aspect of the dialogue management—the resolution of these problems. The introduced dialogue strategy addresses the latter point. Finally, we sketched the implemented prototype system and described the incorporated dialogue management module.

The prototype system was primarily implemented for the purpose of demonstrating the described theoretical concepts (particularly the model of the attentional state and the adaptive dialogue strategy). We emphasize the fact that, whereas the system itself is task-specific, the described underlying concepts are intended to be task-independent.

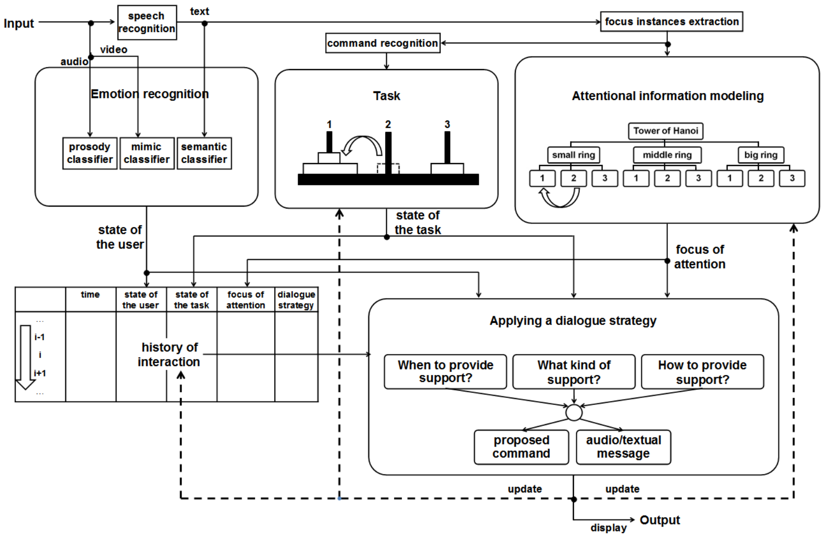


Figure 3: The simplified schema of the NIMITEK prototype system

Thus, changes of the graphical task, the focus tree, the vocabulary or the set of user states require only a minimal change in the implementation of the system.

6. Acknowledgements

The presented study is performed as part of the NIMITEK project (http://wdok.cs.uni-magdeburg.de/nimitek), within the framework of the Excellence Program ”Neurowissenschaften” of the federal state of Sachsen-Anhalt, Germany (FKZ: XN3621A/1005M). The responsibility for the content of this paper lies with the authors.

7. References

J.A. Bateman, E. Teich, and A. Stein. 1998. Speech Generation in a Multimodal Interface for Information Retrieval: The SPEAK! System. In P. Fankhauser and M. Ockenfeld, editors, *Integrated Publication and Information Systems. 10 Years of Research and Development at IPSI*, pages 149–168. Samkt Augustin: GMD Forschungszentrum Informationstechnik.

A. Batliner, K. Fischer, R. Huber, J. Spilker, and E. Nöth. 2000. Desperately Seeking Emotions: Actors, Wizards, and Human beings. In *Proceedings of the ISCA Workshop on Speech and Emotion: A Conceptual Framework for Research*, pages 195–200.

J. Carbonell. 1986. Requirements for robust natural language interfaces: the LanguageCraft (TM) and XCALIBUR experiences. In *Proceedings of the COLING-86*, pages 162–163, Washington, D.C., USA.

E. Douglas-Cowie, R. Cowie, and M. Schröder. 2000. A new emotion database: Considerations, sources and scope. In *Proceedings of the ISCA Workshop on Speech and Emotion*, pages 39–44, Northern Ireland.

M. Gnjatović and D. Rösner. 2006. Gathering Corpora of Affected Speech in Human-Machine Interaction: Refinement of the Wizard-of-Oz Technique. In *Proceedings of the International Symposium on Linguistic Patterns in Spontaneous Speech (LPSS 2006)*, pages 55–66, Academia Sinica, Taipei, Taiwan.

M. Gnjatović and D. Rösner. 2007a. An approach to processing of users commands in human-machine interaction. In *Proceedings of the 3rd Language and Technology Conference (LTC'07)*, pages 152–156, Adam Mickiewicz University, Poznan, Poland.

M. Gnjatović and D. Rösner. 2007b. A Dialogue Strategy for Supporting the User in Spoken Human-Machine Interaction. In *Proceedings of the XII International ”Conference Speech and Computer” (SPECOM'2007)*, pages 708–713, Moscow State Linguistic University, Moscow, Russia.

M. Gnjatović and D. Rösner. 2008a. Emotion Adaptive Dialogue Management in Human-Machine Interaction. In *Proceedings of the 19th European Meetings on Cybernetics and Systems Research (EMCSR 2008)*, University of Vienna, Vienna, Austria.

B.J. Grosz and C.L. Sidner. 1986. Attention, Intentions, and the Structure of Discourse. *Computational Linguistics*, 12(3):175–204.

R Guindon. 1988. How to Interface to Advisory Systems? Users Request Help With a Very Simple Language. In *Proceedings of ACM Conf. on Computer Human Interaction (CHI88)*, pages 191–196, Washington, D.C., USA.

M.A.K. Halliday. 1994. *An Introduction to Functional Grammar*. Edward Arnold, London New York, Second edition.

E. Lucas. 1959. La tour de Hanoï et la question du Tonkin. In *Récréations Mathématiques*, pages 285–286. Reprinted by Blanchard, Paris, France. Original published by Gauthier-Villars, Paris, 1884.

R. Niese, A. Al-Hamadi, and B. Michaelis. 2007. A Novel Method for 3D Face Detection and Normalization. *Journal of Multimedia*, 2(5):1–12.

H. Pirker and G. Loderer. 1999. I said ”two ti-ckets”: How toTalk to a Deaf Wizard. In *Proceedings of the ESCA Workshop on Dialogue and Prosody*, pages 181–185.

B. Vlasenko, B. Schuller, A. Wendemuth, and G. Rigoll. 2007. Frame vs. Turn-Level: Emotion Recognition from Speech Considering Static and Dynamic Processing. In

Proceedings of 2nd International Conference on Affective Computing and Intelligent Interaction (ACII 2007), pages 139–147, Lisbon, Portugal.

W.C. Watt. 2004. Habitability. *American Documentation*, 19:338–351.

B. Wendt and H. Scheich. 2002. The ”Magdeburger Prosodie-Korpus”. In *Proceedings of the Speech Prosody 2002 Conference*, pages 699–701, Laboratoire Parole et Langage, Aix-en-Provence.