International Journal on Social Robotics
Preprint, finally accepted

An EEG Adaptive Information System for an Empathic Robot

Dominic Heger · Felix Putze · Tanja Schultz

Abstract This article introduces a speech-driven information system for a humanoid robot that is able to adapt its information presentation strategy according to brain patterns of its user. Brain patterns are classified from electroencephalographic (EEG) signals and correspond to situations of low and high mental workload. The robot dynamically selects the information presentation style that best matches the detected patterns. The resulting end-to-end system consisting of recognition and adaptation components is tested in an evaluation study with 20 participants. We achieve a mean recognition rate of 83.5% for discrimination between low and high mental workload. Furthermore, we compare the dynamic adaptation strategy with two static presentation strategies. The evaluation results show that the adaptation of the presentation strategy according to workload improves over the static presentation strategy in both, information correctness and completeness. In addition, the adaptive strategy is favored over the static strategy as user satisfaction improves significantly. This paper presents the first systematic analysis of a real-time EEG-adaptive end-to-end information system for a humanoid robot. The achieved evaluation results indicate its great potential for empathic human-robot interaction.

Keywords Human-Robot spoken interaction · Adaptive interaction strategies · Electroencephalography · Mental workload recognition · Human evaluation study

This work has been supported in part by the Deutsche Forschungsgemeinschaft (DFG) within the Collaborative Research Center 588 "Humanoid Robots - Learning and Cooperating Multimodal Robots"

Dominic Heger · Felix Putze · Tanja Schultz
Karlsruhe Institute of Technology
Institute for Anthropomatics
Cognitive Systems Lab (CSL)
Adenauerring 4
D-76131 Karlsruhe, Germany
http://csl.anthropomatik.kit.edu

Dominic Heger
Tel.: +49-721-608-46293
E-mail: dominic.heger@kit.edu

Felix Putze
Tel.: +49-721-608-44328
E-mail: felix.putze@kit.edu

Tanja Schultz
Tel.: +49-721-608-46300
E-mail: tanja.schultz@kit.edu

1 Introduction

Computing machines are about to fundamentally change human's self-understanding and self-constitution as an acting social entity, embodied and embedded in a technological supported environment and process. Also, the accelerating cycles of advancements in information and robotic technologies may soon enable humans to transcend biological limitations and thus transform our lives in ways we cannot imagine. Our current daily life is characterized by interacting with computing machines in different contexts, applications, and environments. We use them for work, for communication, for health purposes, as well as for sports and entertainment in our leisure time. However, in contrast to humans who intuitively sense and react to cognitive or affective states of people in their environment, machines so far interact with us in a way that completely ignores the richness and implications of human internal states. As a result, machines today are still widely insensitive to the context and actions in their environment. This leads to an unnatural form of man-machine communication

and interaction, inadequate system reactions, and inefficient interaction performance. This is in particular a problem for humanoid robots that are designed to blend into the human world and therefore need to interact with us in a social and empathic way. Advancements of this situation require the development of humanoid robots that are - among others - aware of human internal states, such as emotions, vigilance, attention, and mental workload. The latter internal state is very crucial as it influences how humans process information. Also it impacts the human memory span and other factors of cognition which significantly affect the course of interaction between a robot and its user. Therefore, we concentrate in this work on the automatic recognition of mental workload and the appropriate adaptation of the robot's behavior to the recognized human mental workload.

This article describes the setup and evaluation of the adaptive humanoid robot Robert as previously introduced in [1]. His task is to present information to the user via speech. During the course of interaction, the user experiences different levels of mental workload, induced by an external secondary task over which the robot has no control. Robert uses electroencephalographic (EEG) signals recorded from the users to recognize their brain activity patterns, which correspond to conditions of low and high mental workload. The detected brain activity patterns allow Robert to adapt its information presentation strategy to optimally serve the user's needs in the given situation. We presented the first implementation and evaluation of an adaptive information presentation system based on EEG in the domain of humanoid robotics in [1]. The present article improves the reliability of our former work by doubling the number of subjects in the human evaluation study. The results confirm our earlier findings. Furthermore, we elaborate on the details of the dialog system in this study.

2 Related Work

According to Breazeal [2], the design of sociable robots requires human-awareness, which comprises the concept of empathy as an important factor, i.e. the understanding of human internal states and the propper reaction to them. In the last decade, the development of adaptive social robots gained rising attention. Researchers identified a number of user states that the robot needs to consider for optimizing its interaction behavior. Some systems, which detect and adapt to internal states, were implemented and validated in human evaluation studies to show the effectiveness of the adaptation techniques [3][4][5].

Torrey et al. [3] evaluated a humanoid robot that adapts its dialog behavior to the user's expertise. By modifying the vocabulary and language style, the robot better supports novice users and is more efficient for expert users. The authors showed that adaptation increased performance measures and improved subjective perception of the robot. Liu et al. [4] developed a closed-loop human-robot interaction framework in a basketball training scenario. They used various features from cardiac activity, heart sound, bioimpedance, electromyographic activity, and body temperature to discriminate three levels of anxiety using regression trees. Most participants performed better and showed less anxiety when the task difficulty level was adapted to the detected anxiety level. Bonarini et al. [5] described a stress recognition algorithm for a rehabilitation robot on the basis of biosignals such as blood volume pressure, and galvanic skin response among others. Different stress levels were induced by perturbing the user's control over the robot. For discrimination of six different states, the system achieved a recognition rate of 88%. The authors claimed that the recognition system could be applied to adapt the robot's behavior, e.g. by adjusting the difficulty of the training program.

Electroencephalography (EEG) is commonly used to actively control human-machine interfaces by brain signals. EEG directly measures the brain's electrical activity in contrast to other non-invasive measurement methods, such as functional magnetic resonance imaging (fMRI) or magnetoencephalography (MEG). It benefits from a high temporal resolution, low-cost data acquisition equipment, and is useful for mobile applications. Additionally, modern wireless EEG-devices allow for rather unobtrusive data acquisition. Brain Machine Interfaces (BMIs) [6] usually use neural effects that have distinct activation patterns in the brain and are user-controllable. Among the most frequently employed effects are sensorimotor rhythms, slow cortical potentials, evoked potentials, and event related potentials [7]. In robotics, several non-invasive EEG-based BMIs have been proposed to control various kinds of robots [8][9][10][11].

In contrast to the classical BMI paradigm that aims for an explicit command and control of robots, we intend to implicitly recognize spontaneous effects in the EEG that are not actively controlled but correspond to present internal states of the user, such as mental workload.

Several authors have proposed EEG-based workload recognition systems for use in adaptive dialog and information systems but not in the domain of humanoid robots. Berka et al [12] developed the B-Alert system for monitoring alertness and cognitive workload in opera-

tional environments using a wireless EEG sensor headset. They applied the system to several tasks, such as sleep deprivation studies, military motivated monitoring (Warship Commander Task), reaction and digit sequence identification, as well as to image memorization tasks. They predicted four states of alertness using a linear discriminant function on spectral features in the range of 3-40 Hz derived from one second long epochs of data. Wilson and Russel [13] evaluated an autonomous air vehicle task with adaptive aiding controlled by real-time recognition of low and high workload states. They used five frequency bands between 4 and 43 Hz from EEG and electrooculographic (EOG) signals, as well as interbeat intervals from electrocardiogaphic (ECG) signals. Based on neural network classification they could show that the closed-loop adaptive automation significantly improved performance in a complex military task. Chen and Vertegaal [14] improved context awareness for mobile devices. They developed a recognition system for different attentional user states. The authors detected two levels of motor activity using EEG signals and two levels of mental load determined by heart rate variability. Their system used the classified user state to adapt a mobile phone to the appropriate notification level. Kohlmorgen et al. [15] used EEG to measure workload during driving for online adaptation of in-car systems. They used spatial filters and classification by Linear Discriminant Analysis. Their system improved the reaction time for most subjects due to mitigation of high workload situations for the driver. Gevins and Smith [16] evaluated subjects performing different tasks of computer interaction and sequential memorization of stimuli (n-back tasks). They used spectral features of the theta and alpha frequency bands from segments of four seconds length and applied subject-specific multivariate functions and neural networks to discriminate three workload levels. Honal and Schultz [17] analyzed task demand from EEG data recorded in lecture and meeting scenarios. They used Support Vector Machines (SVMs) and Artificial Neural Networks for classification and regression of short time Fourier transform features (2 second epochs). To improve the convenience of brain activity data acquisition they developed a comfortable headband and compared its performance to a standard EEG cap.

3 Adaptive Information System

Robert's task is to give its users information about students they met. The information is reported to the user via text-to-speech synthesis. Information is compiled in a multiple-entry database, listing the attributes of students, such as name, id, and telephone number. The implemented presentation strategies iterate over all database entries and report them one after another.

The strategy of presenting this information to the users is adapted to their brain patterns recognized from the EEG data (see section 4). Robert has two different *behavior styles* which can be switched seamlessly between two utterances: The Low behavior style is designed for brain patterns which correspond to low mental workload, and the High behavior style is designed for brain patterns corresponding to high workload conditions. Although the style of presentation differs between Low and High, the content of information stays the same:

The Low behavior style focuses on high information throughput, i.e. only short pauses between utterances and between different database entries are made. Whenever possible, multiple information chunks are merged into one utterance and phone numbers are presented in a blockwise fashion. However, as Robert is designed to be a social robot, maximizing efficiency is not the only criterion but will be complemented by politeness. Thus, Robert takes the time to convey information in complete sentences to mimic a polite communication partner.

The High behavior style on the other hand is tuned towards situations in which the user has to divide his cognitive resources into two tasks which he executes in parallel. As this multi-tasking may cause memory capacity reduction, split attention, and limited processing capabilities, the High behavior style accommodates the situation by presenting information in a separated fashion, giving only one attribute at a time and reporting phone numbers as single digits. Furthermore, pauses are extended between utterances and database entries such that the user has more time to deal with the secondary task. Reporting time is conserved by limiting the information to the attribute name and value, thus minimizing utterance duration.

Presentation strategy	**Behavior style**
AlwaysLow	Fixed to Low
AlwaysHigh	Fixed to High
EegAdaptive	Derived from EEG
Oracle	Derived from given information on secondary task

Table 2: Presentation strategy and corresponding behavior style of the information system.

The interaction strategy of the information system defines in which fashion switches take place between the two behavior styles over the course of a session. For the experiments described below, we im-

Robert's Behavior style	Low	High
Pause duration	short (500ms)	long (2000ms)
Number presentation	blockwise	isolated
Items per utterance	multiple	single
Formulations	polite	concise
Example utterances	The name of the next person is Heidi Kundel. Her telephone number is 52-11-66-3.	Heidi Kundel Telephone: 5-2-1-1-6-6-3

Table 1: Low and High behavior styles for information presentation.

plemented four strategies: AlwaysHigh, AlwaysLow, EegAdaptive, and Oracle.

The AlwaysHigh and the AlwaysLow strategies define baseline systems which ignore the current state of the user but rather stick to one behavior style. The EegAdaptive strategy uses the recognized brain patterns to select an appropriate behavior (i.e. High when brain patterns corresponding to high mental workload are detected, and Low otherwise). As a gold standard, we also define the Oracle strategy which switches between behavior styles according to the reference information on the secondary task, i.e. instead of relying on potentially noisy information from EEG data, it selects the optimal behavior for each utterance according to the contextual information of whether the secondary task is currently running or not. Tables 1 and 2 summarize Robert's presentation strategies and corresponding behavior styles.

The information system is implemented in a general purpose dialog management engine developed at the Cognitive Systems Lab. It is based on an Information State Update [18] engine. The dialog strategy is implemented as a set of rules which consist of preconditions and bindings. Preconditions determine if a rule is executed and consist of constraints concerning the dialog state, e.g. the index of the current database entry and the time passed since the last system utterance. From each valid rule, its bindings are executed. Bindings modify the dialog state and generate synthesized speech output using the OpenMary text-to-speech system [19]. Our engine supports rules which are adaptive to a dynamically updated workload estimate: Each rule which produces a system speech output is assigned two alternative speech acts corresponding to the two different behavior styles. Those speech acts can differ in linguistic and phonetic style as well as the amount of content. The selection depends on the employed strategy (see Table 2) and the interaction state at execution time of the speech act, i.e. the decision is delayed as much as possible. Furthermore, the system supports switching points within utterances to guarantee swift reactions to state changes within an ongoing speech act. However, this feature was not applied in the experiments to insure fair comparison of consistent utterances.

4 Real-Time Brain Pattern Recognition

The realization of the EegAdaptive strategy requires a real-time EEG-based recognition system which estimates the level of mental workload. For this purpose we adapted our existing online EEG workload recognition system as described in [20]. The systems applies an active EEG-cap (BrainProducts actiCap) to assess the subjects' brain activity using 16 electrodes placed at positions FP1, FP2, F3, Fz, F4, F7, F8, T3, T4, C3, Cz, C4, P3, P4, Pz, and Oz according to the international 10-20 system [21] with reference to the left mastoid. The impedance of each electrode was kept below 20 kΩ during all sessions. Amplification and A/D-conversion was performed on a 16 channel VarioPort biosignals recording system by Becker Meditec using a sampling rate of 256 Hz. The flexible framework BiosignalsStudio [22], which is designed for multimodal biosignal recording, forms the input layer for EEG data acquisition. Figure 1 depicts the training and application processes of the workload recognition system for the EegAdaptive strategy.

Among the most challenging problems for training and operation of the workload recognition systems are artifacts, such as eye movement and muscular artifacts. Both are predominantly present when EEG signals are recorded under less restricted conditions, as it is the case in the experiments described below. Therefore, we apply fully automatic artifact reduction methods based on the combination of two blind source separation techniques: Independent Component Analysis (ICA) and Canonical Correlation Analysis (CCA). ICA is well-known to be very effective for artifact removal of eye blinks and saccades. The infomax algorithm [23] is applied to the training data to calculate a transformation matrix that decomposes the 16-channel EEG signal into 16 independent components. The components related to eye movement activity are identified by frequency and power characteristics. During the operation

of the brain pattern recognition system the EEG signals are transformed using the precalculated transformation matrix. The components which are identified as eye blinks or saccades are high pass filtered by 16 Hz using a Fast Fourier Transform (FFT)-based filter. For muscular artifacts, we apply blind source separation based on canonical correlation analysis that has shown to be more effective than low pass filtering or ICA-based methods [24]. It leverages the fact that in general muscle activity has a lower autocorrelation than brain activity. After decomposition, those components with an autocorrelation below a certain threshold are set to zero. Hereafter, the signals are recomposed into cleaned EEG signals by back transformation.

The EEG data of all channels are split into windows of two seconds length with an overlap of 1.5 seconds. We applied a Hamming window function and computed the discrete Fourier transform on all windows using the short-time FFT. For each frequency coefficient between 4 and 45 Hz we calculate the logarithmic power, resulting in a feature vector of 83 frequency bins. The dimensionality of the feature space is then reduced by averaging over three adjacent frequency bins. Finally, the features of all channels are concatenated. We employ Support Vector Machines (SVMs) [25] with linear kernels to discriminate different brain patterns corresponding to two different levels of mental workloads, i.e. with and without secondary task. The resulting binary classifications of the SVM are integrated by averaging over the past 10 seconds (linear temporal smoothing). This procedure increases the robustness of the recognition results and provides a task specific load value that displays smooth trends over the estimated workload. The output of this linear temporal smoothing is a rational valued workload estimation ranging between 0 and 1, which is thresholded to control switching between the two behavior styles in the EegAdaptive strategy. To determine a subject specific threshold from recognition results of the training session we calculate the average workload estimation for the training parts without the secondary task (w_1) and the average workload with both tasks (w_2). The subject specific threshold t is calculated as $t = \frac{w_1+w_2}{2}$. Robert applies the Low behavior style if the estimated workload level is below t, and the High behavior style otherwise.

To train the recognition system, we recorded per subject one training session, consisting of four parts, as described in the experimental sessions section 5. During training all combinations of the two behavior styles and the two task conditions (with and without secondary task) are performed. The data of the training session is used to calculate the transformation matrix for the ICA-based artifact removal and to identify independent components that contain ocular artifacts.

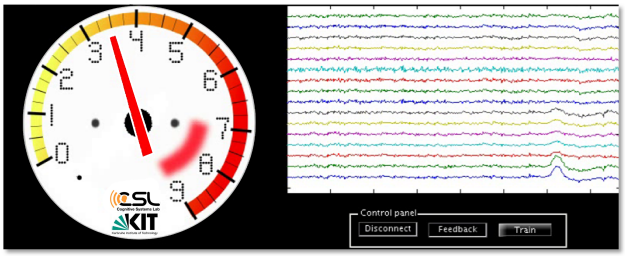

Fig. 2: Speedometer visualizes the recognized workload in real-time.

The described workload recognition system features a speedometer to visualize the recognized workload (see Figure 2). However, we did not reveal this information to the subjects during the experiments to avoid distractions and contaminations on the workload. The complete recognition process is implemented in MATLAB and runs in real-time on a standard desktop computer. More details on the workload recognition system can be found in [20].

5 Experimental Setup

The aim of the experimental study was to evaluate (1) the performance of the real-time EEG-based brain pattern recognizer for discrimination between high and low mental workload, (2) the impact of the presentation strategies on the users' task performance, and (3) the users' overall subjective appeal to the end-to-end system.

For this purpose we designed a multi-level evaluation study in which participants had to perform two tasks, partly in dual-tasking fashion to induce different levels of mental workload. In the primary task participants were asked to manually fill in a paper form according to spoken instructions given by Robert. Performance criteria are correctness and completeness of the information filed on paper. In the secondary task participants processed a variant of the cognitive Eriksen flanker task [26], in which horizontal arrays of five arrows are displayed (e.g. <<><<). Participants were expected to report the orientation of the middle arrow by pressing the corresponding left or right key on the keyboard. Performance criteria are correctness and reporting speed. Apart from the objective performance measures correctness, completeness, and reporting speed, we collected subjective user judgments by a questionnaire. Based on the questions we evaluated how users perceived the interaction quality and efficiency, to what degree users noticed the adaptation

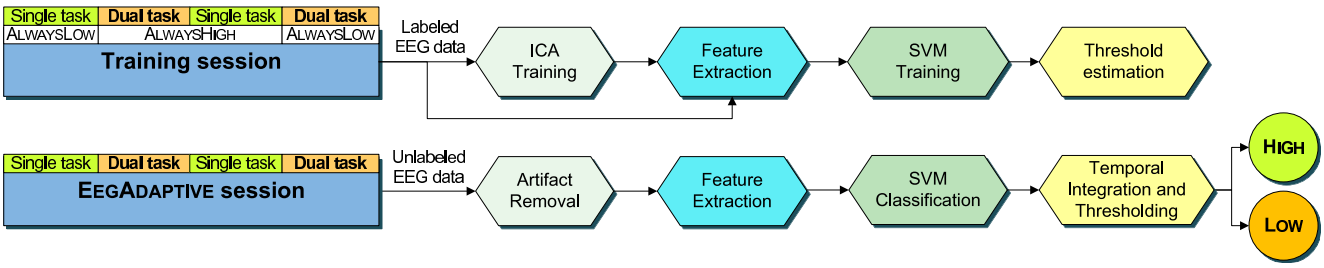

Fig. 1: Workload recognition system architecture. Parameters and models of the system are estimated from EEG data of the training sessions with and without secondary task (top). In the EegAdaptive session the robot's behavior style (High or Low) is determined from unlabeled EEG data (bottom).

of the EegAdaptive and Oracle strategy, and how changes in strategy and behavior style impact the subjective user experience.

Figure 3 shows the experimental setup. Robert was present in form a humanoid robot head [27] which talked to the participants using text-to-speech synthesis. The participants faced paper forms to be filled in as well as a desktop computer to execute the flanker task.

Fig. 3: Recording setup with Robert [27] (left side), the computer for the secondary task (center) and participant wearing an EEG cap (right side).

In total 20 subjects participated in the experiment and completed five sessions which were recorded consecutively in one sitting. In the first session A EEG data were recorded to train a person-dependent brain pattern classifier for each subject (see Figure 1 top). In four subsequent sessions B_1 to B_4 we varied the presentation style in which Robert gives instructions to the subject. In each session B_i one of the strategies AlwaysHigh, AlwaysLow, EegAdaptive, and Oracle was applied consistently throughout the session. To eliminate the impact of bias effects such as fatigue, the order of strategies was randomly chosen. All sessions consist of a fixed sequence of two alternate segments with and without secondary task. Transitions between segments were marked by an acoustic signal. Each segment lasted approximately one minute. Table 3 summarizes the experimental design. Each subject performed five sessions of four minutes duration each, resulting in about 20 minutes data per subject, summing up to about 400 minutes data for all 20 participants.

Prior to the main experiment we performed a pilot study on five subjects to calibrate task difficulty and duration. The main purpose was to ensure that all test conditions significantly differ from each other and do not result in overloaded or underchallenged users. The final study was performed on 20 new subjects between 21 and 29 years old, who participated voluntarily in the study. All participants are students or employees of the Institute for Anthropomatics at Karlsruhe Institute of Technology (KIT). Each participant signed a consent form prior to the experiments. None of the participants had any prior experience with the EEG-based workload recognition system.

Table 4 lists all questions of the questionnaire the participants answered immediately after each session B_i, so for each subject we collected in total four questionnaires. Each question was assigned to a 6-point scale. The items deal with the adaptation capabilities of the robot (Q1), the appropriateness of its behavior (Q2, Q3), its social competence (Q5, Q6) and an overall judgment (Q4, Q7). Items Q8 to Q11 were adopted from a subset of the Nasa TLX scale [28] to evaluate the experienced workload along several dimensions.

6 Experimental Results

This section presents the experimental results of the above described study aiming at the evaluation of the low vs. high workload discrimination performance of the real-time EEG-based brain pattern recognizer (section 6.1), the impact of the presentation strategies on the users' task performance (section 6.2), and the users' overall subjective appeal to the end-to-end system (section 6.3).

For each Subject	Single 1 Minute	Dual 1 Minute	Single 1 Minute	Dual 1 Minute	Total
Session A	EEG brain pattern training				4 min
Session B_1	AlwaysHigh	AlwaysHigh	AlwaysHigh	AlwaysHigh	4 min
Session B_2	AlwaysLow	AlwaysLow	AlwaysLow	AlwaysLow	4 min
Session B_3	EegAdaptive	EegAdaptive	EegAdaptive	EegAdaptive	4 min
Session B_4	Oracle	Oracle	Oracle	Oracle	4 min
Total					20 min

Table 3: Experimental Setup and amount of Evaluation Data

Q1	How strongly did the robot adapt to the switch between the conditions with and without secondary task?
Q2	How appropriate was the behavior of the robot in conditions without secondary task?
Q3	How appropriate was the behavior of the robot in conditions with secondary task?
Q4	Would you like to work together with a robot with this behavior?
Q5	How do you judge the behavior of the robot concerning "friendliness"?
Q6	How do you judge the behavior of the robot concerning "empathy"?
Q7	How do you judge the behavior of the robot in general?
Q8	Experienced time pressure*
Q9	Experienced accomplishment*
Q10	Experienced effort*
Q11	Experienced frustration*

Table 4: Questionnaire for subjective evaluation of presentation strategies. Items marked with * are extracted from the Nasa TLX workload scale.

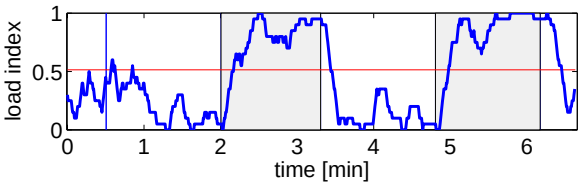

Fig. 4: Individual Workload estimation (blue) and Threshold level (red) in single-task (white background) and dual-task (gray background) mode over a EegAdaptive session.

6.1 Workload Recognition Results

The EEG-based workload recognition system was trained on the 4 minutes data of session A for each of the 20 participants. We applied fully automatic artifact removal based on ICA and CCA as described in section 4. The resulting cleaned EEG signals of all channels were used to train person-specific SVM classifiers and thresholds to discriminate the brain patterns corresponding to low and high mental workload. Figure 4 displays an example of the estimated workload level over time during the EegAdaptive session of one subject. The blue curve shows the recognizer's output, the red horizontal line marks the user specific threshold level. The segments in which the participants performs single-tasking (paper form filling according to Robert's instructions) are indicated by a white background, the dual-tasking segments (subject performs the secondary flanker task in addition) are indicated by a gray background. The vertical lines mark the ground truth.

The blue curve of the recognizer's output in Figure 4 clearly reflects the different task conditions, i.e. the recognizer discriminates well between single and dual-task conditions. However, the graph also shows a delay between changes in task demand and recognized workload. This can be explained by the temporal smoothing implemented in the recognition system. Furthermore, a switch of the task condition might not have an immediate impact on a person's mental state. As a result, Robert's behavior in the EegAdaptive sessions is slightly different from the Oracle sessions due to this delayed switching and due to some recognition errors.

To determine the recognizer's accuracy, we calculated the percentage of recognition outputs (provided for each window, i.e. one output every 0.5 seconds) where task condition and thresholded workload estimation results match. The accuracy is given by the sum of the number of recognition outputs below the threshold during single-tasking plus the number of recognition outputs above the threshold during dual-tasking, divided by the total number of recognition outputs. Figure 5 gives the breakdown of the recognition rates for the EegAdaptive sessions over all 20 subjects. The recognition accuracy ranges from 70.8% to 94.0% with a mean of 83.5% (sd=6.5).

Note that the recognition accuracy averaged over all time slices is not a priori a measure of the end-to-end system's quality in terms of strategy adaptation. This is due to the fact that only a fraction of the recognized workload decisions have an impact on the system's

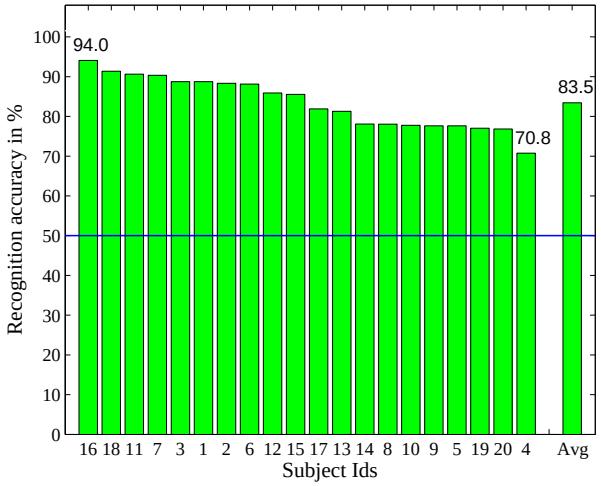

Fig. 5: Recognition accuracies for the EegAdaptive sessions of 20 subjects in descending order. Chance level is at 50%.

strategy changes. Only the recognition results at the utterance boundaries influence the strategy since within utterances the behavior remains unchanged. Therefore, we performed an experiment in which we limited the recognition accuracy calculation to the relevant decision points. This resulted in an average accuracy of 81%, which is reasonably close to the overall recognition accuracy of 83.5%. We therefore conclude that the overall performance of the recognizer is indeed a robust estimator of the system performance in the adaptation task.

6.2 Task Performance

To study the impact of the presentation strategies on the users' task performance, we applied two task performance metrics, correctness and completion rate. The former was calculated for both, the primary robot instruction task and the secondary Eriksen flanker task, while the latter was only applicable to the primary task. In case of the primary robot instruction task, the correctness rate was calculated as the ratio between the correctly quoted items and the number of completed items which were noted down on paper by the subjects who listened to the robot's instruction. The completion rate was calculated as the ratio between completed and total number of items in the robot's database. In case of the Eriksen flanker task, the correctness rate was calculated as the number of keys correctly pressed, divided by the number of presented stimuli.

Table 5 gives the correctness and completion rates of robot and flanker task performance averaged over all 20 subjects for the four presentation strategies. The numbers show that the presentation strategy AlwaysLow outperforms all other strategies in terms of completion rate due to the high throughput. In contrast, for the AlwaysHigh strategy subjects only manage to complete about half of the items. However, AlwaysLow trades this high completion rate with a dramatically lower correctness rate. Since AlwaysLow leaves only few resources for the subjects to properly carry out the secondary flanker task, AlwaysLow is outperformed by the other strategies in terms of flanker correctness rate. In comparison, both adaptive strategies EegAdaptive and Oracle are able to maintain a reasonable completion rate while keeping the correctness rate at the same level as the conservative AlwaysHigh strategy. Furthermore, it can be observed that the fully automatic strategy adaptation applying EegAdaptive compares favorably with the Oracle strategy, indicating that the EEG-based recognition of brain patterns results in a fairly reliable switching behavior. Overall, we conclude that adaptive strategies improve the information presentation by switching behavior styles without hurting task performance.

Strategy	Correctness (robot)	Completion (robot)	Correctness (flanker)
AlwaysLow	86%	98%	69%
AlwaysHigh	96%	58%	87%
EegAdaptive	96%	85%	82%
Oracle	94%	85%	86%

Table 5: Average completion and correctness rates for the robot instruction and the Eriksen flanker task.

6.3 Subjective Evaluation

The aim of the subjective evaluation was to assess the users' subjective appeal to the end-to-end system, i.e. the experienced robot behavior and the mental workload. For this purpose the same questionnaire consisting of 11 questions was presented after each of the B_i sessions. Table 6 summarizes the questionnaire results averaged over all 20 participants.

The result for question Q1 shows that both adaptive strategies (EegAdaptive and Oracle) are indeed perceived as being adaptive. This observation is in accordance with the objective effectiveness of adaptivity measured by the EEG-based brain pattern recognition rate (see section 6.1).

For appropriateness of behavior, we differentiate between behavior in absence of a secondary task, i.e. single-tasking (Q2) and in presence of a secondary task, i.e. dual-tasking (Q3). For single-tasking, the relative drop from the best to the worst strategy is as small as

Item	Scale from ... (1) to ... (6)	AlwaysLow	AlwaysHigh	EegAdaptive	Oracle
Q1	not adaptive – very adaptive	2.0 (0.97)	2.5 (1.66)	4.5 (1.10)	5.4 (1.09)
Q2	not appropriate – very appropriate (single task)	4.9 (1.00)	4.1 (1.75)	4.9 (1.14)	5.1 (1.07)
Q3	not appropriate – very appropriate (dual task)	2.3 (1.10)	4.3 (1.03)	3.9 (1.25)	5.1 (0.89)
Q4	don't work with – work with	2.2 (1.15)	3.3 (1.18)	3.6 (1.14)	4.8 (0.69)
Q7	very bad – very good	2.8 (0.95)	4.0 (0.71)	3.9 (0.87)	4.8 (0.61)
Q5	not friendly – very friendly	3.1 (1.15)	3.8 (0.80)	3.7 (1.22)	4.3 (0.86)
Q6	not empathic – very empathic	2.2 (0.93)	2.6 (1.19)	3.4 (0.99)	4.4 (0.87)
Q8	low pressure – high pressure	5.3 (0.66)	3.2 (1.14)	4.0 (0.99)	3.5 (1.23)
Q9	low accomplishment – high accomplishment	3.0 (1.19)	3.8 (1.16)	3.7 (1.04)	4.0 (1.27)
Q10	low effort – high effort	5.1 (1.05)	3.5 (1.12)	4.4 (0.75)	4.0 (1.09)
Q11	low frustration – high frustration	4.0 (1.25)	2.5 (1.05)	3.0 (1.00)	2.5 (0.61)

Table 6: Subjective evaluation of the robot's behavior and experienced mental workload; average score (standard deviations).

24.1%. For dual-tasking, the participants clearly prefer the High behavior: The gap between the worst and the best ranked strategy increases to 54.9%. We explain this observation by the fact that the benefit of both behavior styles is perceived asymmetrically: While High improves throughput and convenience of the information presentation, Low can make the difference between successful task completion and mental overload. Still, the order of strategies for single-tasking is as expected: AlwaysLow, EegAdaptive and Oracle have very similar scores with non-significant differences while the slow AlwaysHigh strategy is perceived worst. For dual-tasking, the EegAdaptive strategy scores slightly worse than Oracle and AlwaysHigh which perform both optimally in dual-tasking segments (AlwaysLow is indisputably the worst strategy). EegAdaptive usually switches to the correct strategy but with a small delay. As described above, this delay is determined by the window size of temporal integration in the classifier and the fact that a switch of behavior style takes place only between utterances. We assume that a more immediate classification mechanism, a more flexible adaptation scheme and scenarios with longer segments of constant mental workload will mitigate this effect.

The two questions Q4 and Q7 define a metric for overall perceived quality of the system. Both items are strongly correlated ($r = 0.86$). The results reveal a clear quality gap between AlwaysLow and the other strategies. While Oracle outperforms the others by far, the average difference between AlwaysHigh and EegAdaptive is much smaller. This observation is somewhat surprising given the significant differences in objective performance criteria. However, it can be explained by the fact that the EegAdaptive strategy depends solely on the recognition performance of the brain pattern classification. This dependency is expressed in higher standard deviations of most items for EegAdaptive compared to Oracle (which works in a deterministic way). Table 7 further investigates this issue. Most of the items are significantly correlated with recognition accuracy. When splitting the data into two groups according to the session's recognition rate (below average vs. above average, denoted $acc. \leq \varnothing$ and $acc. > \varnothing$), the distance to the scores of Oracle is reduced for the better sessions and thus the gap between EegAdaptive and AlwaysHigh increases. In summary, we observe a distinct user preference for EegAdaptive over the non-adaptive strategies given a sufficiently high recognition accuracy. This observation supports our assumption that recognition performance is a key factor in subjective perception and that further improvement of brain pattern classification will directly translate to improvements of user satisfaction.

To further analyze the perception of the four presentation strategies, Q5 and Q6 asked for how friendly and empathic the behavior was perceived over the session. Q6 reveals that the adaptive strategies (EegAdaptive and Oracle) were indeed perceived as most empathic. Adaptivity and perceived empathy are highly correlated ($r = 0.73$ between Q1 and Q6). This indicates that developing adaptive strategies for human-robot communication is an important step towards the implementation of truly social robots. For friendliness, no significant differences between strategies were observed. We ascribe this to the fact that both behavior styles could lead to a perception of friendliness: While High speaks in complete and thus more polite sentences, Low produces minimal phrases which might be perceived as more considerate given the stressful tasks.

Questions Q8 to Q11 investigate the experienced workload in single- and dual-tasking segments. The dimensions time pressure (Q8), accomplishment (Q9), effort (Q10), and frustration (Q11) show similar patterns: AlwaysHigh expectedly performs best, receiving scores which indicate relatively low workload. Oracle gets very close to those bounds. This shows that an adaptive strategy is able to reach near-optimal workload levels while it flexibly makes the most of cognitive resources whenever available in single-task situa-

Item	Scale from ...(1) to ...(6)	ρ between acc. and E-O	d_{low} := O-E ($acc. \leq \varnothing$)	d_{high} := O-E ($acc. > \varnothing$)	d_{low}-d_{high}
Q1	not adaptive – very adaptive	-0.40	1.1	0.6	0.5
Q2	not appropriate – very appropriate (single task)	-0.18	0.4	0.1	0.3
Q3	not appropriate – very appropriate (dual task)	-0.51*	1.5	0.8	0.7
Q4	don't work with – work with	-0.35	1.3	1.1	0.2
Q7	very bad – very good	-0.51*	1.3	0.6	0.7*
Q5	not friendly – very friendly	-0.74*	0.8	0.5	0.3
Q6	not empathic – very empathic	-0.54*	1.5	0.4	1.1*
Q8	low pressure – high pressure	0.29	-0.8	-0.1	-0.7
Q9	low accomplishment – high accomplishment	-0.24	0.3	0.3	0.0
Q10	low effort – high effort	0.46*	-0.9	0.1	-1.0*
Q11	low frustration – high frustration	0.24	-0.6	-0.3	-0.3

Table 7: Relation of recognition accuracy (acc.) and the difference between user ratings of Oracle and EegAdaptive ($O-E$). The two columns d_{low} and d_{high} show the mean difference between Oracle and EegAdaptive ($O-E$) for sessions below resp. above average workload recognition accuracy. * marks statistical significance (Pearson correlation ρ) or significantly different from each other (d_{low}-d_{high} one-tailed t-tests) with significance level $\alpha = 0.05$.

tions. AlwaysLow is indisputably much worse in all regards compared to adaptive strategies. EegAdaptive approaches the lower workload bound and performs (with exception of Q10) more similar to AlwaysHigh than to AlwaysLow. This indicates that the fully automatic adaptive strategy EegAdaptive is a very reasonable approximation to the Oracle strategy.

7 Adaptive Human-Robot Communication

To further exploit the above described brain pattern recognition and adaptive presentation strategy in a real-life setup, we integrated both components into an interactive spoken dialog system for a humanoid robot. The robot acts as agent which can be queried by the user for information and services via naturally spoken commands. Our demonstration platform is equipped with a monitor displaying the ThinkingHead[1], a virtual, morphable head. The display indicates the detected user state (high or low workload) by changing the background colors and thus makes the robot's behavior changes transparent to the user. The EEG-cap for data acquisition is replaced by an Emotiv EPOC[2] headset to improve user comfort. The wireless EPOC device (originally developed as an advanced gaming device) uses 16 saline electrodes attached to flexible arms. Even for non-expert users the device is ready for use after a very short setup time, as a time consuming application of electrode gel is not necessary. Supporting this new device required only minor parameter modifications to the described recognition system (number of electrodes, sampling rate).

Human-robot communication is controlled by a light-weight rule-based dialog strategy. Similar to the EegAdaptive strategy, the dialog strategy allows to switch between two distinct interaction behavior styles based on the measured brain patterns corresponding to low or high workload. The goal of the low workload behavior style is to provide as much information as possible while keeping the user engaged in conversation. In this mode, the robot employs a talkative, verbose speaking style, jokes occasionally and uses expressive mimics. The goal of the high workload behavior style is different: In this mode, the robot tries to use the limited resources of the user as efficiently as possible by minimizing the amount of non-critical information and emphasizing important aspects by speaking loud and slow.

This real-life human-robot communication setup differs in one main aspect from the setup described earlier, i.e. the user speaks to control the robot. This has two major implications for the system. On the one hand, the interaction strategy has to deal with the concept of initiative. In the current system, this is performed in a straight-forward rule-based fashion, where the high workload behavior takes more control of the conversation than the low workload behavior, relieving the user from non-critical decisions. On the other hand, the recognition system must handle brain activity patterns and muscle artifacts generated during speech production. This effect needs to be compensated by training the brain pattern recognition system with data which include speech segments.

8 Conclusion

This article describes the design, implementation, and evaluation of a speech-driven information system for human-robot interaction. The humanoid robot is able to adapt its behavior styles according to brain patterns

[1] http://thinkinghead.edu.au/
[2] http://www.emotiv.com/

of its user. Those behavior styles are designed to accommodate different conditions of mental workload of the user. In an evaluation study with twenty participants we achieved a mean recognition rate of 83.5% for the discrimination between low and high mental workload. Furthermore, we showed that the adaptive strategy based on the brain pattern recognition improves user satisfaction in comparison to static interaction strategies. Comparison to the oracle strategy with respect to the users' task performance and subjective evaluation show that the EEG-based adaptation is a promising approximation to the optimal adaptation strategy. The achieved evaluation results indicate its great potential for empathic human-robot interaction.

In future investigations, we aim for a systematic and quantitative evaluation of an adaptive spoken dialog system for a humanoid robot. The system will be evaluated in more natural scenarios which require both the user and the robot to actively participate in the interaction. We foresee two main challenges: Firstly, we need to identify a scenario which allows for robust quantitative evaluation of natural interaction without repetitive tasks. Secondly, a natural dialog will show the users' adaptation to both, changing workload conditions and changing robot behavior. Therefore, the envisioned scenario requires advances in interaction and adaptation strategies, and needs to consider less evident side effects of adaptive behavior.

References

1. D. Heger, F. Putze, T. Schultz, An adaptive information system for an empathic robot using EEG data, in *International Conference on Social Robotics* (Springer, 2010), pp. 151–160
2. C. Breazeal, *Designing sociable robots* (The MIT Press, 2004)
3. C. Torrey, A. Powers, M. Marge, S. Fussell, S. Kiesler, Effects of adaptive robot dialogue on information exchange and social relations, in *1st ACM SIGCHI/SIGART conference on Human-robot interaction* (2006), pp. 126–133
4. C. Liu, P. Rani, N. Sarkar, Human-Robot interaction using affective cues, in *15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (2006), pp. 285–290
5. A. Bonarini, L. Mainardi, M. Matteucci, S. Tognetti, R. Colombo, Stress recognition in a robotic rehabilitation task, in *Robotic Helpers: User Interaction, Interfaces and Companions in Assistive and Therapy Robotics, a Workshop at ACM/IEEE HRI* (2008), pp. 41–48
6. J. Wolpaw, N. Birbaumer, D. McFarland, G. Pfurtscheller, T. Vaughan, Brain-computer interfaces for communication and control, Clinical neurophysiology **113**(6), 767 (2002)
7. S. Mason, A. Bashashati, M. Fatourechi, K. Navarro, G. Birch, A comprehensive survey of brain interface technology designs, Annals of Biomedical Engineering **35**(2), 137 (2007)
8. C. Bell, P. Shenoy, R. Chalodhorn, R. Rao, Control of a humanoid robot by a noninvasive brain–computer interface in humans, Journal of Neural Engineering **5**, 214 (2008)
9. J. Millan, F. Renkens, J. Mouriño, W. Gerstner, Noninvasive brain-actuated control of a mobile robot by human EEG, Biomedical Engineering, IEEE Transactions on **51**(6), 1026 (2004)
10. D. McFarland, J. Wolpaw, Brain-computer interface operation of robotic and prosthetic devices, Computer **41**(10), 52 (2008)
11. D. McFarland, W. Sarnacki, J. Wolpaw, Electroencephalographic (EEG) control of three-dimensional movement, Journal of Neural Engineering **7**, 036007 (2010)
12. C. Berka, D. Levendowski, M. Cvetinovic, M. Petrovic, G. Davis, M. Lumicao, V. Zivkovic, M. Popovic, R. Olmstead, Real-time analysis of EEG indexes of alertness, cognition, and memory acquired with a wireless EEG headset, International Journal of Human-Computer Interaction **17**(2), 151 (2004)
13. G. Wilson, C. Russell, Performance enhancement in an uninhabited air vehicle task using psychophysiologically determined adaptive aiding, Human Factors: The Journal of the Human Factors and Ergonomics Society **49**(6), 1005 (2007)
14. D. Chen, R. Vertegaal, Using mental load for managing interruptions in physiologically attentive user interfaces, in *CHI'04 extended abstracts on Human factors in computing systems* (2004), pp. 1513–1516
15. J. Kohlmorgen, G. Dornhege, M. Braun, B. Blankertz, K. Müller, G. Curio, K. Hagemann, A. Bruns, M. Schrauf, W. Kincses, in *Toward Brain-Computer Interfacing* (The MIT Press, 2007), pp. 409–422
16. A. Gevins, M. Smith, Neurophysiological measures of cognitive workload during human-computer interaction, Theoretical Issues in Ergonomics Science **4**(1), 113 (2003)
17. M. Honal, T. Schultz, Determine Task Demand from Brain Activity, Proceedings of the First International Conference on Biomedical Electronics and Devices, BIOSIGNALS 2008, Funchal, Madeira, Portugal, January 28-31, 2008, Volume 1 pp. 100–107 (2008)
18. S. Larsson, D. Traum, Information state and dialogue management in the TRINDI dialogue move engine toolkit, Natural language engineering **6**(3&4), 323 (2000)
19. M. Schröder, J. Trouvain, The German text-to-speech synthesis system MARY: A tool for research, development and teaching, International Journal of Speech Technology **6**(4), 365 (2003)
20. D. Heger, F. Putze, T. Schultz, Online Workload Recognition from EEG data during Cognitive Tests and Human-Machine Interaction, in *33rd Annual German Conference on Artificial Intelligence (KI2010)* (2010), pp. 410–417
21. H. Jasper, The 10-20 electrode system of the International Federation, Electroencephalography and Clinical Neurophysiology **10**, 371 (1958)
22. D. Heger, F. Putze, C. Amma, M. Wand, I. Plotkin, T. Wielatt, T. Schultz, BiosignalsStudio: A flexible Framework for Biosignal Capturing and Processing, in *33rd Annual German Conference on Artificial Intelligence (KI2010)* (2010), pp. 33–39
23. S. Makeig, A. Bell, T. Jung, T. Sejnowski, et al., Independent component analysis of electroencephalographic data, Advances in neural information processing systems pp. 145–151 (1996)

24. W. De Clercq, A. Vergult, B. Vanrumste, W. Van Paesschen, S. Van Huffel, Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram, Biomedical Engineering, IEEE Transactions on **53**(12), 2583 (2006)
25. C. Chang, C. Lin. LIBSVM: a library for support vector machines (2001)
26. C. Eriksen, D. Schultz, Information processing in visual search: A continuous flow conception and experimental results, Attention, Perception, & Psychophysics **25**(4), 249 (1979)
27. T. Asfour, K. Welke, P. Azad, A. Ude, R. Dillmann, The Karlsruhe Humanoid Head, in *8th IEEE-RAS International Conference on Humanoid Robots (Humanoids)* (2008), pp. 447–453
28. S. Hart, L. Staveland, *Human mental workload* (1988), chap. Development of NASA-TLX (Task Load Index), pp. 139–183

Dominic Heger received his Diploma in Informatics from Universität Karlsruhe (TH) in 2009. He is a research assistant and Ph.D. student at the Cognitive Systems Lab (CSL) at the Karlsruhe Institute of Technology. He is currently working in the DFG funded Collaborative Research Center CRC 588 "Humanoid Robots - Learning and Cooperating Multimodal Robots". His research interests include human-robot interaction, biosignals processing, and cognitive modeling.

Felix Putze received his Diploma in Informatics from Universität Karlsruhe (TH) in 2008. He is a research assistant and Ph.D. student at the Cognitive Systems Lab (CSL) at the Karlsruhe Institute of Technology. He is currently working on a project developing a cognitive interaction system. His research interests include cognitive modeling, dialog modeling and recognition of human inner states.

Tanja Schultz received her Ph.D. and Diploma degrees in Informatics from University Karlsruhe, Germany in 2000 and 1995 respectively. Since 2000 she is a Research Scientist at Carnegie Mellon in the Language Technologies Institute and since 2007 also holds a Full Professorship position at the Karlsruhe Institute of Technology (KIT) in Germany. She is the director of the Cognitive Systems Lab, which focuses on the development of human-centered technologies and intuitive human-machine interfaces based on biosignals, by capturing, processing, and interpreting signals such as speech, muscle, and brain activity.