An Adaptive Information System for an Empathic Robot using EEG Data

Dominic Heger*, Felix Putze, and Tanja Schultz

Cognitive Systems Lab (CSL)
Karlsruhe Institute of Technology (KIT), Germany
{dominic.heger, felix.putze, tanja.schultz}@kit.edu

Abstract. In this paper we introduce a speech-based information system for a humanoid robot that is able to adapt its information presentation strategy to different 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 selects an information presentation style that best matches the detected patterns. The complete system of recognition and adaptation is tested in an evaluation study with ten participants. We achieve a mean recognition rate of 80% and show that an adaptive information presentation strategy improves user satisfaction in comparison to static strategies.

1 Introduction

Machines play an important role in our everyday lives as matters of communication, work, and entertainment. However, nearly all systems are completely insensitive to the current situation and actions in their environment. Especially in the interaction with humans, machines neglect the internal states of their users, with the consequence of unnatural interaction, inadequate actions and inefficient user performance. This is especially true for humanoid robots which are designed to integrate into the daily life of their users and therefore need to interact with them in a social and empathic way. One of the most important internal states is mental workload. Workload influences how humans process information, their memory span and other factors of cognition which heavily influence the course of interaction between robot and its user.

This paper describes the setup and evaluation of the adaptive humanoid robot Robert. His task is to present gathered information to the user via speech. During the course of one session, the user experiences different levels of mental workload, induced by an external secondary task over which the robot has no control. Therefore, Robert uses electroencephalographic (EEG) signals recorded from the user to recognize different brain activity patterns. Those brain patterns correspond to conditions of low and high mental workload. This information allows

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

Robert to adapt its information presentation strategy to optimally serve the user's needs in the current situation.

2 Related Work

According to Breazeal [3], the design of sociable robots requires the development of human-awareness, which comprises the concept of empathy, i.e. the understanding of internal states of the human, as an important factor. This information can be used to establish behavior strategies that are adequate for the particular situation and environment. In the last decade, the development of adaptive social robots gained rising attention. Researchers identified a number of user states that have an impact on the optimal interaction behavior of robots. There exist a number of implemented systems which can detect and adapt to those internal states. Most of these systems are evaluated in studies with real humans to show the effectivity of the implemented adaptation measures.

For example, in [10], Torrey et al. evaluate a humanoid robot that adapts its dialog behavior in the kitchen domain to the user's expertise. This is done by modifying the vocabulary and language style of the robot. This way, the robot can act more helpful for novice users and more efficient for expert users. The authors show that this adaptation improves performance measures and subjective perception of the robot. Liu et al. [9] 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 classify between three levels of anxiety using regression trees. For most of their participants they could show improvements of perceived anxiety and user performance by real-time adaptation of the task difficulty level according to the detected level of anxiety. In [2], Bonarini et al. describe a stress recognition algorithm for a rehabilitation robot on the basis of biosignals such as blood volume pressure, galvanic skin response and others. Different stress levels are induced by adding noise to the user's control over the robot. For the discrimination of six different states, the system achieves a recognition rate of 88%. The authors claim that this recognition system can be used to adapt the behavior of the robot, e.g. by adjusting the difficulty of the training program.

There exists a whole corpus of adaptive dialog and information presentation systems outside the domain of humanoid robots. For example, Chen and Vertegaal [4] proposed a system for improving context awareness for mobile devices. They developed a recognition system for different attentional states of the user. The authors detected two levels of motor activity using EEG signals and two levels of mental load determined by heart rate variability. With their system they could adapt a mobile cell phone to an notification level appropriate to the classified user state. Kohlmorgen, et al. [8] measured workload using EEG while driving for the online adaptation of in-car systems. They used spatial filters and classification by Liner Discriminant Analysis. Their work shows improvements in the reaction time for most subjects due to mitigation of high workload situations for the driver.

To our best knowledge, this paper presents the first implementation and evaluation of an adaptive information system based on EEG in the domain of humanoid robots.

3 Adaptive Information System

The task of Robert in this scenario is to provide information about students he met to the user. This information forms a database with multiple entries and consists of attributes like name, id, and phone number. The information is reported to the user via synthesized speech. The information system is implemented as a finite state machine in a general purpose dialog management engine developed at the Cognitive Systems Lab. It iterates over all entries of the database and reports them one after another.

The style with which information is presented can be adapted according to the brain patters recognized from the EEG data (see section 4). For that purpose, Robert has two different *behavior styles* between which it can switch at any point between two utterances: The first one (called Low) is selected for brain patterns which correspond to a low mental workload and the second one (called High) is designed for brain patterns corresponding to high workload conditions. While the content which is presented is identical for both, they differ in the style of presentation: Low focuses on a high information throughput. Therefore, it makes only short pauses between utterances and between different database entries. It also uses a blockwise style for number presentation and combines multiple items for one entry into one utterance, if possible. As we implement the information system for a social humanoid robot, maximized efficiency is not the only goal: Robert takes the time to convey the information in complete sentences.

The High behavior style on the other hand is tuned towards situations in which the user is not fully attentive due to a secondary task which he executes in parallel. This results for example in a reduced memory capacity, which is accounted for in this behavior style: Information is presented in a more isolated fashion, only giving one attribute at a time and presenting numbers by reporting each digit separately. To give the user more room to deal with the secondary task, pauses are extended between utterances and even more so between database entries. Time is saved by reducing utterances to the pure minimum, only giving an attribute name and its value.

The actual strategy of the information system defines how it switches between both behavior styles during the course of a session. For the current experiment, we designed four different strategies:

The AlwaysHigh and the AlwaysLow strategies define baseline systems which always use one of the two behavior styles, independent of the current state of the user. The EegAdaptive strategy is connected to the EEG-based recognition system and makes use of the recognized brain patterns to select

	Low	High
pause duration	short	long
number presentation	blockwise	isolated
items per utterance	multiple	one
formulations	complete	concise

Table 1. Two behavior styles for information presentation.

an appropriate behavior (i.e. High when brain patterns corresponding to high mental workload are detected and Low else). As a gold standard, we also define the Oracle strategy which is also adaptive but which has direct access to information on the secondary task. Instead of potentially noisy information from EEG data, it uses this context information to select the optimal behavior for each utterance.

4 Real-Time Brain Pattern Recognition

In order to estimate the level of workload to adjust the Adaptive strategy, we developed a real-time recognition system based on our online EEG workload recognition system [7]. It uses an active EEG-cap (BrainProducts actiCap) to assess the subjects' brain activity by 16 channels with electrodes positioned according to the international 10-20 system. BiosignalsStudio (BSS) [6] forms the input layer for EEG data acquisition, which is a flexible framework for multimodal biosignal recording that has recently been developed at the Cognitive Systems Lab.

We derive spectral features between 4 and 45 Hz from one second long chunks of EEG signals (overlapping by 0.5 seconds). We then employ Support Vector Machines (SVMs) with radial basis function kernels to discriminate different brain patterns corresponding to the two different mental task demands: with and without secondary task. The resulting binary classifications of the SVM are integrated by averaging over the past 10 seconds (linear temporal smoothing), which increases the stability of the recognition results and provides a task specific load value that expresses smooth trends of the estimated workload.

The estimated task load estimation (rational valued) needs to be thresholded to determine the behavior style for the EegAdaptive strategy. Therefore, we determine a subject specific threshold from recognition of the training session, calculated as the mean value of the average task load estimation during the training parts without secondary task, and the training parts with both tasks. If the estimated load level is below the threshold, the robot uses the Low behavior style, if it is above it uses the High behavior style.

To train the recognition system, one training session has been recorded for each subject, consisting of four parts, in the same way as the experimental sessions described in section 5. Each combination of the two robot behavior styles and the two task conditions (with and without secondary task) is performed during training.

Artifacteous data is a serious problem for training of the system. Predominantly eye movement artifacts and muscular artifacts are present in the EEG signals recorded during the experiment. Therefore, we applied a heuristic approach for artifact detection using thresholds on the signal power and its slope to identify artifacts in each data segment of one second length. Contaminated data segments are dropped and not used in the training of the system.

5 Experimental Setup

To test the different information presentation strategies and the effectivity of the brain pattern recognizer, we designed the following evaluation study: Participants were given the task to fill out a form with the information they acquired by listening to the robot. To induce different levels of mental workload, we included a secondary task the participants had to execute in parallel to the primary task for some parts of the experiments. Sections with secondary task induce a higher mental workload than sections without secondary task and we evaluate whether the system is able to recognize varying brain patterns for both situations and to react accordingly.

During the experiment, each participant completed five sessions of information presentation. For each session, a different information presentation strategy was used. The first session was specifically designed to collect EEG data for person dependent training of the brain pattern classifier. In subsequent sessions, the strategies AlwaysHigh, AlwaysLow, EegAdaptive, and Oracle were applied in random order. Robert was present in form of the head of a humanoid robot [1] which talked to the participant. Next to them, participants had sheets of paper to fill in the information presented by the robot and a computer for the execution of secondary tasks.

Each session consisted of a fixed sequence of two sections without secondary task and two sections with secondary task. Each section approximately lasted one minute and the transition between sections was marked using a signal sound. As a secondary task, the particpants processed the cognitive Flanker task in parallel to the robot information task: During the Flanker task, different horizontal arrays of five arrows are displayed (e.g. <<><<). Subjects respond as quickly as possible to the orientation of the middle arrow by pressing the corresponding left or right key.

Using this setup, ten subjects from age 24 to 29 took part voluntarily in the study and signed an consent form. All of them are students or employees of the Institute for Anthropomatics at Karlsruhe Institute of Technology (KIT).

Along objective scores generated from evaluation of the robot information task and the secondary Flanker task, we also collect subjective judgements of the users themselves. This information helps us to evaluate whether any efficiency or quality gains for the tasks where perceived by the user as such, to what degree the users noticed the adaptation efforts of the EegAdaptive and Oracle strategy and to what degree the different strategies and behavior styles influenced the subjective user experience.

Fig. 1. Picture of the recording setup with the robot to the left, the computer for the secondary task in the center and a participant wearing an EEG cap to the right.

Table 2 lists the questions the participants answered immediately after each session in the experiment. Each item was assigned a 6-point scale. Items Q8 to Q11 were adopted from a subset of the Nasa TLX scale [5] to evaluate the experienced workload in several dimensions.

6 Results

6.1 Recognition Results

The brain pattern recognition system was effectively able to determine the mental workload states of the participants in the experiment. Figure 2 shows an example of the time course of the estimated load level during the EegAdaptive session of one subject. The red horizontal line marks the user specific threshold level. Vertical lines indicate borders of the different task conditions: The sections where the participant processed both, the robot information task and the secondary Flanker task, are marked with yellow background color.

The output of the recognizer clearly reflects the different task conditions. To determine the accuracy of the recognition system, we calculate the percentage of recognition outputs where task condition and thresholded workload estimation result match. This is the number of recognition outputs below the threshold, while performing only the robot information task plus the number of recognition outputs above the threshold, while performing both the Flanker task and as the robot information task, divided by the total number of recognition outputs. This results in recognition rates between 69.7% and 90.4% for the EegAdaptive sessions of the 10 subjects with a mean of 80.2% (sd=7.1).

Q1	How much 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 2. Questionnaire for subjective evaluation of different information presentation strategies. Items marked with * are extracted from the Nasa TLX workload scale.

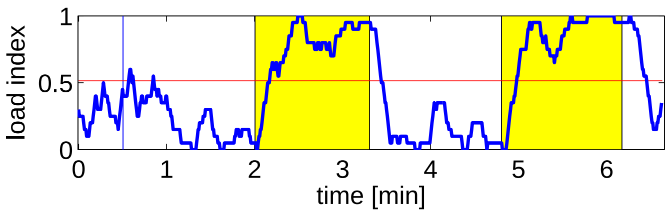

Fig. 2. Task load estimation and threshold for the adapted session of one subject.

Figure 2 also indicates that there is a small delay of the recognition results as the task demand changes. This can be explained by the temporal smoothing of the recognition system, furthermore a switch of the task condition might not have an immediate impact on a person's mental state. Therefore, the robot's behavior in the EegAdaptive sessions is different from the behavior in Oracle sessions due to deferred switching of behavior styles and noise.

6.2 Subjective Evaluation

Table 3 summarizes the results of the questionnaires for the subjective evaluation of the behavior of the robot and the experienced mental workload.

Item	Scale from ... (1) to ... (6)	AlwaysLow	AlwaysHigh	EegAdaptive	Oracle
Q1	not adaptive – very adaptive	1.6 (0.97)	2.3 (1.70)	4.3 (1.42)	5.5 (0.71)
Q2	not appropr. – very appropr.	4.9 (1.10)	3.8 (1.87)	4.7 (1.42)	4.7 (1.25)
Q3	not appropr. – very appropr.	1.9 (0.88)	4.3 (1.16)	3.7 (1.49)	5.0 (0.82)
Q4	don't work with – work with	1.7 (1.06)	2.9 (0.99)	3.4 (1.26)	4.6 (0.69)
Q5	not friendly – very friendly	3.1 (1.20)	3.8 (0.79)	3.5 (1.35)	4.4 (0.84)
Q6	not empathic – very empathic	2.0 (0.82)	2.6 (1.17)	3.4 (1.17)	4.2 (0.78)
Q7	very bad – very good	2.5 (0.85)	3.8 (0.63)	3.5 (0.92)	4.6 (0.70)
Q8	low pressure – high pressure	5.5 (0.71)	2.9 (1.20)	3.9 (0.86)	3.2 (1.32)
Q9	low accomp. – high accomp.	2.9 (1.20)	3.9 (1.20)	3.6 (1.07)	3.7 (0.95)
Q10	low effort – high effort	5.2 (0.79)	3.3 (1.23)	4.2 (0.92)	3.9 (1.29)
Q11	low frustr. – high frustr.	4.3 (1.06)	2.6 (1.07)	3.0 (1.05)	2.6 (0.52)

Table 3. Results of questionnaires for subjective evaluation of the behavior of the robot and experienced mental workload (standard deviations in parentheses).

The result for item Q1 shows that both strategies which are designed as adaptive (EegAdaptive and Oracle) are also perceived as such by the participants. This observation is in accordance with the objective effectiveness of adaptivity measured by the recognition rate of the brain pattern classifier (see section 6.1).

For appropriateness of behavior, we differentiate between behavior in absence of a secondary task (Q2) and behavior in presence of a secondary task (Q3). For Q2, the relative drop from the best to the worst strategy is as small as 22.4%. For Q3, i.e. sections with secondary task, the participants more clearly prefer the High behavior: The gap between the worst and the best ranked strategy increases to 62%. 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 Q2 is as expected: AlwaysLow has the highest score, EegAdaptive and Oracle follow with almost identical scores and the slow AlwaysHigh strategy is ranked last. For Q3, the EegAdaptive strategy scores slightly worse than

Oracle and AlwaysHigh which perform both optimally in sections with secondary task. EegAdaptive usually switches to the correct strategy but with an delay determined by the window size of temporal integration in the classifier and the fact that a switch of behavior styles only takes place between complete utterances. A quicker classification mechanism, a more flexible adaptation scheme or scenarios with longer sections of constant mental workload will mitigate this effect.

Two items (Q4 and Q7) define a metric for overall quality. Both items are strongly correlated ($r = 0.89$). While we see clear gap between the scores of AlwaysLow and the other strategies, the differences between the AlwaysHigh, EegAdaptive and Oracle are much smaller. We explain this observation by the fact that participants mostly rated their ability to cope with the given tasks and that the increased throughput and the more pleasant communication style of the High behavior style did not make a huge difference.

To further investigate how the different strategies were perceived by the participants, Q5 and Q6 asked for how friendly and empathic the behavior during one session was. Q6 reveals that the adaptive strategies were indeed perceived as most empathic and that adaptivity and perceived empathy are correlated ($r = 0.77$ 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, we see no relevant differences in score for the different strategies. We ascribe this to the fact that both behavior styles have aspects that could lead to a perception of friendliness: While High speaks in complete sentences instead of minimal phrases as Low does, the latter is probably perceived as more considerate in the light of the stressful tasks.

Questions Q8 to Q11 investigate the experienced workload for the different experimental sections. For the dimensions time pressure (Q8), achievement (Q9), frustration (Q11), and effort (Q10), we see a similar pattern: AlwaysHigh expectedly performs best and receives 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 has the flexibility to use the participant's cognitive resources when available in sections without secondary task. AlwaysLow is indisputably much worse in all regards than those two strategies. EegAdaptive approaches the lower workload bound and is (with exception of Q10) closer to the score of AlwaysHigh than to the one of AlwaysLow. This indicates that this strategy is a reasonable approximation to the Oracle strategy.

The EegAdaptive strategy depends on the recognition rate of the brain pattern classification to generate satisfying results. This dependency is for example expressed in the higher standard deviation of most items for EegAdaptive than for Oracle (which works in a deterministic way). The correlation coefficient between the difference between both strategies in overall score Q4 and the recognition rates for individual participants (see section 6.1) is $r = -0.73$ ($p < 0.05$). This observation supports the hypothesis that further improvement of the biosignal classification will directly translate to improvements of user satisfaction.

6.3 Task Performance

Item	AlwaysLow	AlwaysHigh	EegAdaptive	Oracle
Info: Correctness Rate	0.85	0.96	0.96	0.94
Info: Completetion Rate	0.98	0.56	0.85	0.82
Flanker: Correctness Rate	0.72	0.87	0.82	0.87

Table 4. Average scores for the robot information task and the Flanker task.

Table 4 shows task performance metrics for the robot information task and the secondary Flanker task. For the former, we have two different metrics: Correctness rate is the number of correctly noted items divided by the number of completed items, while the completion rate measures the number of completed items divided by the total number of available items. We see that AlwaysLow has the highest completion rate due to the high throughput while the slow AlwaysHigh strategy only manages to complete about half of the items. AlwaysLow pays this high completion rate with a lower correctness rate. The two adaptive strategies are able to maintain a reasonable completion rate while keeping the correctness rate as high as the conservative AlwaysHigh strategy. For the task performance concerning the Flanker task, we see a similar pattern: The performance of the AlwaysLow is below the performance of the other strategies which all score comparably. We conclude that the adaptive strategies can improve the information presentation by switching behavior styles without hurting task performance.

7 Conclusion & Future Work

In this paper, we described the design and evaluation of an adaptive speech-based information presentation system for a humanoid robot. This system uses EEG-based brain pattern recognition to switch between two different behavior styles. Those behavior styles are designed to accommodate different conditions of mental workload of the user. An evaluation study with ten participants was conducted. We presented a mean recognition rate of 80% and showed that an adaptive strategy improves user satisfaction in comparison to static ones. We further see that the EEG-based adaptation is a promising approximation to the optimal adaptive strategy.

For further investigations, we aim for an extension of the adaptive information presentation system to a full dialog system for a humanoid robot. This system will be evaluated in scenarios which require both the user and the robot to actively participate in the interaction. There are two main challenges for this endeavor: Firstly, speech is a process which influences brain patterns due to language production and articulation activity. Secondly, a true dialog will also show adaptation of the user to changing workload conditions. This requires more elaborate interaction and adaptation strategies.

References

1. Asfour, T., Welke, K., Azad, P., Ude, A., Dillmann, R.: The karlsruhe humanoid head. In: 8th IEEE-RAS International Conference on Humanoid Robots (Humanoids). pp. 447–453 (2008)
2. Bonarini, A., Mainardi, L., Matteucci, M., Tognetti, S., Colombo, R.: 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)
3. Breazeal, C.: Designing sociable robots. The MIT Press (2004)
4. Chen, D., Vertegaal, R.: Using mental load for managing interruptions in physiologically attentive user interfaces. In: CHI'04 extended abstracts on Human factors in computing systems. pp. 1513–1516 (2004)
5. Hart, S. & Staveland, L.: Human mental workload, chap. Development of NASA-TLX (Task Load Index), pp. 139–183 (1988)
6. Heger, D., Putze, F., Amma, C., Wand, M., Plotkin, I., Wielatt, T., Schultz, T.: BiosignalsStudio: A flexible Framework for Biosignal Capturing and Processing. In: 33rd Annual German Conference on Artificial Intelligence (KI2010) (submitted)
7. Heger, D., Putze, F., Schultz, T.: Online Workload Recognition from EEG data during Cognitive Tests and Human-Machine Interaction. In: 33rd Annual German Conference on Artificial Intelligence (KI2010) (submitted)
8. Kohlmorgen, J., Dornhege, G., Braun, M., Blankertz, B., Muller, K., Curio, G., Hagemann, K., Bruns, A., Schrauf, M., Kincses, W.: Improving human performance in a real operating environment through real-time mental workload detection. In: Toward Brain-Computer Interfacing, pp. 409–422. The MIT Press (2007)
9. Liu, C., Rani, P., Sarkar, N.: Human-Robot interaction using affective cues. In: 15th IEEE International Symposium on Robot and Human Interactive Communication (ROMAN). pp. 285–290 (2006)
10. Torrey, C., Powers, A., Marge, M., Fussell, S., Kiesler, S.: Effects of adaptive robot dialogue on information exchange and social relations. In: 1st ACM SIGCHI/SIGART conference on Human-robot interaction. pp. 126–133 (2006)