Dummy Model based Workload Modeling

Felix Putze, Tanja Schultz
Karlsruhe Institute of Technology
Cognitive Systems Lab
Karlsruhe, Germany
{felix.putze, tanja.schultz}@kit.edu

Robert Pröpper
Technical University of Berlin
Neural Information Processing Group
Berlin, Germany
robert.proepper@mailbox.tu-berlin.de

***Abstract*—In this paper, we show how a model of human cognition based on ACT-R can be improved to accurately predict cognitive performance under different workload levels. For this purpose, we propose a novel approach which uses an EEG-based workload model to (de-)activate a dummy model which runs in parallel to the actual task model. The dummy model consumes cognitive resources to reflect the effect of workload on behavior and performance. We evaluate the approach in two user studies with different tasks and show a significant reduction of prediction error.**

***Index Terms*—Cognitive model, EEG, Workload, Adaptation**

I. Introduction & Related Work

Cognitive Modeling is a technique to predict human behavior and performance using models which are based on psychological theories and optimized from empirical data. While cognitive modeling is a tremendously useful tool to understand processes in the human mind, it was also applied very successfully in research on Human-Computer Interaction (HCI) for simulating users during early-stage usability evaluation, for example to automatically identify efficiency bottlenecks or optimal interface designs.

One example of a mature cognitive architecture is ACT-R (Adaptive Control of Thought – Rational) [1]. ACT-R has been used for cognitive user simulation. Amant et al. [2] used an ACT-R user model to simulate operation of a cell phone menu in ACT-R for predicting response time. The authors showed that the computational cognitive model outperformed a traditional non-cognitive model based on Fitt's Law in prediction accuracy. [3] used a similar methodology to evaluate interfaces for smart homes for users with cognitive disabilities. The authors showed that for their use case of a contextual assistant, the ACT-R model yielded an accurate prediction of user behavior. The DISTRACT-R model [4] used the ACT-R mechanism of threaded cognition to combine a complex model of car driving [5] with adjustable models of interfaces for in-car systems like radio, and phone. The software is able to model the impact of system operation on driving performance for different interface configurations for rapid prototype evaluation.

While the use of cognitive models in simulation is already helpful during development, our goal is to enable real-time "model tracing" [6], i.e. the prediction of a a single person's cognitive state while performing a task. Model tracing would allow an interactive system to continuously predict the cognitive state of its user to adapt the system behavior to this prediction, for example by adjusting interface complexity. One central requirement for model tracing in realistic application scenarios is the ability to accommodate the effect of different user states which influence the user's cognition. For example, a central executive function which is involved in nearly every complex cognitive task is memory. The ACT-R memory model accounts for the retrieval probability and latency for a memorized item given how often and how recent it was presented to the agent. This memory model was designed to model average human performance with no distracting tasks and a low workload level. However, cognitive performance depends to a large extend on the current workload level: A higher workload level – for example induced when humans perform several tasks at the same time – influences information retrieval, activation decay, etc. When applying a standard computational cognitive model with default parameters to predict performance in a cognitive task under high workload, the prediction of memory performance will be overly optimistic. This will lead to sub-optimal adaptation choices by the system, for example causing information overload or over-complex interface designs.

There are few publications available which deal with cognitive modeling of non-standard conditions. In [7], the authors model the impact of arousal on memory performance by making the decay parameter scale depending on the arousal level to represent an inverted u-shaped performance curve. The authors of [8] propose the technique of overlays. Overlays modify the basic mechanisms of the architecture to model the effect of certain cognitive conditions. The proposed overlays range from simple parameter adjustments to the addition of a "worry" task to consume cognitive resources. The authors do not provide a detailed evaluation of their proposed approaches. [9] proposes a mechanism to develop models which predict performance under sleep deprivation. A modification of utility calculation results in higher probability of executing no rule, reflecting the effect of microlapses. The CO-JACK architecture [10] incorporates modulation of cognitive processes to model variability due to physiological factors and affect. The hybrid architecture ACT-RΦ combines ACT-R with the HumMod model of physiology and affect [11]. Modeled affective states and physiological conditions influence the simulated cognition by modifying utility values of production rules, for example to model the impact of thirst on decision making.

To our best knowledge, there are no papers which represent the user state workload within a cognitive architecture.

Another, more principled limitation of the current state of the art is the unsolved challenge of measuring and injecting the user states at runtime into the model. Typically, the user state is defined by the modeler and then remains unchanged. This method is however not suited for dynamically changing user states. Furthermore, most of the cited works define additional parameters beyond the parameters of the architecture itself. This increases the complexity of each model and thus the danger of overfitting. In this paper, we propose the *dummy model approach* which addresses those limitations by measuring a person's workload from EEG and using the dual-tasking mechanism of ACT-R to implicitly represent the effect of increased workload on performance and behavior. Works like [12] or [13] show that the combination of brain activity measurement and models of cognitive processes can result in predictions which are more precise than predictions using only one of those two approaches.

II. The Dummy Model Approach

Our goal in this work is to represent different workload levels within ACT-R. Instead of an explicit representation, we use the ACT-R mechanism for dual-tasking to switch a "dummy model" on and off. When switched on, this dummy model runs in parallel to the actual task model to occupy cognitive resources. This implicitly models the effect of high workload on cognitive task performance.

Our proposed architecture consists of four components: an EEG-based workload model, a main task model, a dummy model and a switching strategy. The interaction between those components is displayed in Figure 1: During the execution of a task, the EEG-based workload model continuously evaluates features extracted from an EEG signal stream which is recorded live. It yields a workload estimate on a scale from 0 to 1 which is propagated to the ACT-R model. The ACT-R model then decides based on the workload estimate to either activate or deactivate a dummy module. This dummy module represents a separate task thread in ACT-R (using the Threaded Cognition mechanism [14], which allows the quasi-parallel execution of multiple models) which is an abstract model of cognitive activity caused by a secondary task. Executing the dummy model in parallel to the main task model will cause cognitive resources to be occupied for the actual task model, potentially resulting in reduced task accuracy or increased response time.

In this paper, we will explain the details of the dummy-model approach and validate the approach using data from two user studies. To our best knowledge, this is the first implementation and systematic evaluation of an ACT-R model which is able to predict behavior and performance under variable workload levels. The following subsections describe the components required for the dummy model approach to represent the effect of workload on behavior and performance.

A. Main Task Model

The employed main task model is a regular ACT-R model of the main task. It can be used to predict behavior and performance of a person executing that task. In this paper, our example main task is the Paired Associate (PA) task. The PA is a task of learning and recalling the pairing of two items – a stimulus and a response. First, a list of such pairings is learned by a participant. Then, a sequence of stimuli is presented and the participant is asked to give the associated response. Associative learning is an important aspect of intelligence and required by many cognitive tasks in an HCI context (e.g. learning the correct input command in response to certain system outputs). Additionally, our implementation of the PA represents the operation of a GUI (visual stimuli on a screen and manual responses on a keyboard). For those two reasons, the PA task is a representative abstraction of many relevant tasks in real life. For the ACT-R model, we used a slightly modified version of the PA model provided as part of the ACT-R distribution. Input and output to the model were directly provided from and to the GUI which participants used during the experiments[1].

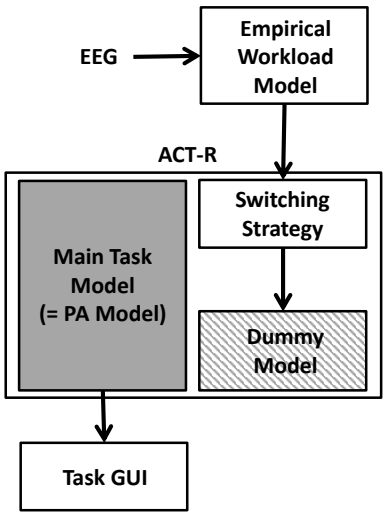


Fig. 1. System setup for workload adaptation of ACT-R based on a dummy model.

The goal of model tracing in individual sessions implies that we need to individualize modeling parameters for each participant, as we are predicting performance in one concrete situation, not on average. Therefore, we determined optimal model parameters for each participant. To avoid overfitting of the models to limited training data, we restricted our optimization to sequential adjustment of only two parameters of the declarative memory model (the retrieval threshold τ and the retrieval latency parameter F). There were no additional tuning parameters for our workload adaption approach, which is therefore parameter-free, i.e. it can be applied to new participants without adjustment.

[1]To have the ACT-R model interact with the task GUI, which was implemented in Python, we implemented the `Hello Python` module. Similarly to the existing `Hello Java` module [15], this component allows interfacing Python applications with ACT-R models.

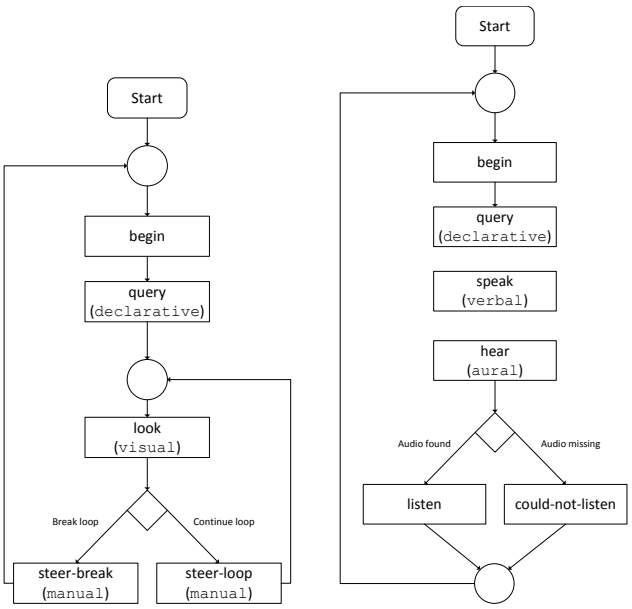
B. EEG-based Workload Model

To provide an EEG-based workload model, we employed the model as described in [16], using frequency features calculated on windows of 2 s length with an overlap of 1.5 s. For data acquisition, we applied an active EEG-cap (BrainProducts actiCap) to measure the participants' 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 [17] 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. A Support Vector Machine classifies each feature vector as either low or high workload. After temporal smoothing, the workload model yields a workload estimate on a scale from 0 to 1. For more details on the workload model, please refer to [16].

C. Dummy Model & Switching Strategy

The dummy model is an ACT-R model which runs in parallel to the main task model (using the Threaded Cognition mechanism). It abstractly models the cognitive processes involved in the secondary task. In contrast to the main task model, the dummy model is not a detailed model of a valid human solution strategy of the secondary task. Instead, it contains a sequence of requests to ACT-R modules associated to the task, for example the declarative module or the visual module. This sequence is repeated infinitely while the model is activated. It is possible to activate and deactivate the model at runtime. In deactivated mode, the model does not perform any module requests. In activated mode, the repeated module requests cause exclusive or limited cognitive resources to be temporarily blocked for the main task model, potentially resulting in longer response times or even task failures. The activation of the dummy model is performed on the basis of the workload model, i.e. when high workload is detected. This reflects the degradation of human performance caused by multi-tasking.

As different secondary tasks may have different characteristics of resource usage, we implemented different dummy models corresponding to different types of secondary tasks. For this work, we chose two different paradigms of secondary tasks to explore the possibilities of the presented approach: The Sternberg memory task (SB) [18] and the Lane Change Task (LCT) [19]. The Sternberg task generates heavy memory load and is therefore expected to interfere with the memory demands of the Paired Associate main task. We employed a purely acoustical version of the task where the stimuli were read to the participant and responses were given verbally. Therefore, we did not expect interference with the visual input and motor output of the PA task. The corresponding dummy module `Sternberg-Dummy` performs periodic requests to the aural, the verbal and the declarative module. The LCT on the other hand is a driving task executed in a driving simulator. It requires the participant to change lanes on a three-lane highway as indicated by road signs. The memory load of this task is low, however input and output modalities interfere with the PA task. The corresponding dummy module `LCT-Dummy` performs periodic requests to the visual and the motor module. Figure 2 shows flow charts which describe the sequence of operations performed by the two dummy models. Each block corresponds to one production rule occupying a corresponding ACT-R module (e.g. "steer" occupies the manual module). Note that corresponding real ACT-R models would be much more complex, especially for the LCT: Considering how complex even a model of lane and distance keeping is [5], we save much modeling effort by introducing the dummy model. We achieve this reduction in effort because the dummy model concentrates on its main goal which is to give an abstract representation of the cognitive resources required for the execution of the secondary task.



(a) Lane Change Task (b) Sternberg task

Fig. 2. Flowcharts of two ACT-R dummy models `LCT-Dummy` (left) and `Sternberg-Dummy` (right). Each box corresponds to one ACT-R rule. In parentheses we give the ACT-R modules which are involved in the execution of the corresponding rule.

Note that cognitive resources are not permanently occupied by the secondary tasks, as those also contain pause segments. For this reason, the dummy model is not processed all the time during high workload phases. Instead, the model is randomly activated during those phases with a certain probability. This probability is 50% for the `Sternberg-Dummy` and 25% for the `LCT-Dummy`. Those numbers correspond to the ratio of task processing to pause segments for those two secondary tasks.

III. Experimental Setup

For evaluation of the dummy-model approach, we recorded two data sets with the same main task (PA) and two different secondary tasks (SB, LCT). In both data sets, participants performed the PA task multiple times with and without secondary task. Each condition (e.g. with secondary task (`HIGH`)

and without secondary task (LOW)) was repeated twice for the LCT data set and four times for the SB data set. Additionally, there was a training session for each condition which was not recorded.

The LCT data set was recorded in the driving simulator. For the PA task, we showed sequences of words for learning and query on a 7“ computer screen in the driver's cockpit. All presented words were in German language with four letters and associated with numbers from 1 to 9. A learning phase consisted of 16 words. During the query phase, each word was queried three times in randomized order. The response window was 4 s, correct answers were always shown afterwards for 2 s. Participants gave their response using a numeric keypad strapped to their right leg.

The LCT task was performed using a force-feedback steering wheel and a gas and brake pedal for controlling a virtual car on a large projection screen. Participants were instructed to drive at a constant speed of 180 km/s and had the task to follow lane changing instructions given visually on road signs which appeared at fixed intervals. One run of the LCT task this setup lasted for five minutes.

For the SB data set, recording was performed on a standard desktop computer and screen, on which the PA task was performed. For the Sternberg task, sequences of five short phonetic strings (e.g. ”omo”) without semantic meaning were read to the participant during a learning phase. In a subsequent query phase, target strings from the learning phase had to be discriminated from distractors. Responses were given verbally by the participant in a subsequent response window of 3 s. In total, one query phase contained 20 phonetic strings. One run of the Sternberg task consisted of four pairs of training and query phases and lasted five minutes.

Using this setup, we recorded a total of nine sessions in the SB data set and nine sessions in the LCT data set (one session per part). All 18 participants were university students with a mean age of 23 ($\sigma = 2.6$). Overall, 6 of the participants were female, 12 were male. All participants gave their written consent to their participation in the study.

IV. Evaluation of the Dummy Model Approach

The evaluation of the dummy model approach consists of three parts: First, we evaluate the non-adapted model (i.e. without dummy model) for its ability to predict human performance for the PA runs with and without secondary task. Second, we repeat this analysis with an oracle-adapted model (i.e. with dummy model activated by perfect workload recognition). Third, we evaluate the EEG-adapted model (i.e. with dummy model activated by the EEG-based workload model).

A. Evaluation of Non-Adapted Model

In this section, we first look at the experimental performance metrics of the participants in LOW and HIGH conditions. Second, we compare the performance of the non-adapted model (i.e. without dummy model) to the human performance in the two data sets SB and LCT.

Table I summarizes two performance metrics of the participants on both data sets for LOW and HIGH workload conditions in the PA task. We see that response time and response accuracy rise significantly ($p < 0.01$ for two-sided t-tests on all data sets and performance metrics) for both data sets when an additional secondary task is processed. A non-adaptive model which is designed to predict such performance metrics assuming full concentration on the main task will therefore greatly overestimate performance of human participants when a secondary task is actually present. By inspecting the standard deviations, we also see that the individual differences are large, stressing the need for model individualization to perform real-time prediction.

Condition	Accuracy [%]	Response Time [s]
LOW (SB data set)	32.3 (3.34)	1.60 (0.29)
HIGH (SB data set)	15.75 (6.29)	2.02 (0.49)
difference HIGH vs. LOW	-16.55*	0.42*
LOW (LCT data set)	40.89 (5.78)	1.51 (0.32)
HIGH (LCT data set)	37.95 (7.46)	1.75 (0.38)
difference HIGH vs. LOW	-2.94*	0.24*

TABLE I
Average PA performance of all test participants for different workload levels in two data sets. Standard deviation given in parentheses. An asterisk denotes a significant difference between HIGH and LOW at $\alpha = 0.01$.

Moreover, there is a difference between the two data sets in the quality of performance impact: While the performance differences between LOW and HIGH workload conditions are statistically significant on average, the measured effect size is not the same for both data sets. We have a strong impact on the accuracy metric only for the SB data set. For the LCT data set, only reaction time shows a strong degradation for the HIGH condition compared the corresponding LOW condition. This observation is consistent with the types of distraction which are caused by the two secondary tasks: The SB task incurs strong working memory load and therefore turns the declarative module into a bottleneck. This behavior harms both PA accuracy and PA response time. On the other hand, the LCT task only marginally influences PA accuracy, as it does not occupy the declarative module. Instead, occupying visual and manual module leads to delays in stimulus processing and response execution, therefore to an increase in response time.

For evaluation of model prediction, we use the prediction error (PE) metric, which is the average absolute difference between empirically measured and predicted value. Often, we look at the relative PE, i.e. the ratio between PE and the human performance. Table II shows the PE of the non-adapted PA model when predicting the performance metrics of the individual participants for LOW and HIGH workload level. For this evaluation, we look at the metrics response time and response accuracy of the Paired Associate task. We report the absolute PE as well as PE relative to the respective metric measured for the human participants. We compare model predictions to the human performance in the final run in a session for each workload level, as this run contains the weakest learning effects. Unsurprisingly, the individualized PA model can very reliably predict performance

for the LOW condition (this is not trivially true as individual parameters were estimated on a different run of the PA task): PE averaged across all participants is below 2% for both data sets. However, PE increases substantially when transferring this model to one of the HIGH workload conditions. For the SB data set, predicted response accuracy is nearly twice as large as measured empirically (92% PE for response accuracy in the HIGH condition of the SB data set). For both data sets, the average response time of actual test participants was more than 20% higher than predicted by the non-adapted model (21.4% for the SB data set, 25.8% for the LCT data set). This large prediction error for the HIGH workload level mandates the use of a workload-adaptive model.

Condition	PE Accuracy	PE Response Time
LOW (SB)	0.39 (=1.12%)	0.02 (=1.25%)
HIGH (SB)	14.49 (=92.0%)	0.43 (=21.4%)
LOW (LCT)	0.21 (=0.51%)	0.02 (=1.32%)
HIGH (LCT)	6.3 (=15.4%)	0.39 (=25.8%)

TABLE II
AVERAGE ABSOLUTE AND RELATIVE PE FOR APPLYING THE **NON-ADAPTED** PA MODEL TO DATA FROM DIFFERENT CONDITIONS.

B. Evaluation of the Oracle-Adapted Model

To evaluate the benefit of workload-adaptive models, we start by analyzing the dummy-model approach using a workload oracle. The workload oracle directly derives the correct workload level (high or low) from the task condition and propagated this value to the ACT-R model to switch the dummy model on or off. Table III presents average absolute and relative prediction error when applying the workload-adaptive model using workload oracle. As we assume perfect workload recognition, performance prediction in both LOW conditions was identical to the performance of the non-adapted model in Table II (which always operated without dummy model). When looking at the results for HIGH conditions, we see that the prediction error compared to the non-adapted model was reduced for all situations and metrics. For easier comparison, we report the difference between non-adapted and oracle-adapted prediction relative to the PE of the non-adapted model. For the SB data set, the result is most convincing, with a reduction of 57.0% absolute (61.96% relative) in prediction error for response accuracy. PE for response time was reduced by 8.0% (37.38% relative). Both reductions in prediction error for the SB data set were statistically significant ($p \leq 0.001$ for both response accuracy and reaction time).

For the LCT data set, we also observe a reduction in PE, at least for response time (absolute reduction of PE by 9.7%, which is 37.6% relative). This reduction barely not significant at $p = 0.07$. Regarding response accuracy, we did not expect to observe an effect of using the oracle-adapted model, as the LCT dummy model mostly affected stimulus perception and response generation, but not memory retrieval. The fact that the reduction in response time was not significant is caused by the high inter-person variability for the LCT data set; for

Condition	PE Accuracy	Δ Acc	PE Resp. Time	Δ RT
LOW (SB)	0.39 (=1.12%)	-	0.02 (=1.25%)	-
HIGH (SB)	5.52 (=35.0%)	-57%	0.27 (=13.4%)	-8.0%
LOW (LCT)	0.21 (=0.51%)	-	0.02 (=1.32%)	-
HIGH (LCT)	6.3 (=15.4%)	-0.0%	0.30 (=16.1%)	-9.7%

TABLE III
AVERAGE PE FOR APPLYING THE **ORACLE-ADAPTED** PA MODEL TO DATA FROM DIFFERENT CONDITIONS. "Δ ACC" IS THE REDUCTION IN RELATIVE PE COMPARED TO THE PREDICTION OF THE NON-ADAPTED MODEL. "Δ RT" IS THE SAME FOR RESPONSE TIME.

one participant in this data set, response time even decreases in the HIGH condition. The current parameter-free model (i.e. a model which has no individual parameters which allow an individual adjustment the dummy model) is not able to predict the occurrence of such "paradox" performance beforehand. If we remove the participant with "paradox" performance from the analysis, the reduction in prediction accuracy (which then increases to 12.0%, which is 46.51% relative) becomes significant ($p = 0.01$).

The quantitative difference between both data sets also shows that the implementation of a generic, task independent dummy model is not feasible: For example, a model which would accurately predict the effect of the LCT secondary task would inevitably underestimate the impact of the Sternberg secondary task on response accuracy. A consequence of this observation is that at least a rough estimate of the cognitive resource requirements of the secondary task is necessary to select an appropriate dummy model, as well as an estimate on the task intensity (i.e. the probability with which the dummy model is active during HIGH workload).

C. Evaluation of the EEG-Adapted Model

Up to this point, we have assumed a perfect workload oracle to differentiate between LOW and HIGH workload condition for triggering the dummy model. This is of course a best-case assumption as workload prediction based on EEG-based models will always be prone to classification errors. Such errors may reduce the benefit of modeling high workload situations. Additionally, they may lead to the activation of the dummy model in low workload conditions. In the following, we quantify the effect of replacing the workload oracle with a realistic EEG-based model.

As a prerequisite, we need to evaluate the ability of the EEG-based workload model to discriminate LOW and HIGH workload level. For this purpose, we evaluate the classification accuracy after temporal smoothing. Averaged over all participants, we achieved a person-dependent classification accuracy of 84.1% ($\sigma = 14.1$) for discriminating between LOW and HIGH for the LCT data set and a mean person-dependent accuracy of 64.9% ($\sigma = 15.3$) for for discriminating between LOW and HIGH for the SB data set. The substantially lower accuracy for the latter evaluation can be explained by the non-continuous workload induced by the Sternberg task (learning and query phase) in comparison to the LCT: On average across all participants, the temporally smoothed workload level for

Condition	PE Accuracy	Δ Acc	PE Resp. Time	Δ RT
LOW (SB)	3.44 (=9.92%)	+8.8%	0.11 (=6.9%)	+5.65%
HIGH (SB)	7.34 (=46.6%)	-45.4%	0.26 (=12.9%)	-8.5%
LOW (LCT)	0.34 (=0.83%)	+0.32%	0.08 (=5.31%)	+3.99%
HIGH (LCT)	6.1 (=14.8%)	-0.6%	0.32 (=17.2%)	-8.6%

TABLE IV
AVERAGE PE FOR APPLYING THE **EEG-ADAPTED** PA MODEL. "Δ ACC" IS THE REDUCTION IN RELATIVE PE COMPARED TO THE PREDICTION OF THE NON-ADAPTED MODEL. "Δ RT" IS THE SAME FOR RESPONSE TIME.

the query phases of the Sternberg task is 0.15 higher than the recognized workload level for learning phases.

When applying the EEG-adapted model, we modify the fraction of time during which the dummy model is activated for each condition, based on the workload estimates of the individual participants. When using a workload oracle, the dummy model is activated exactly at the desired ratio r (50% for SB−Dummy and 25% for LCT−Dummy, see Subsection II-C) during HIGH conditions and turned of completely during LOW conditions. For an EEG-adapted model, those ratios changes depending in the classification accuracy a of the EEG-based workload model: The dummy model is then activated $a \cdot r$ of the time during HIGH conditions and $(1 - a) \cdot r$ of the time during LOW conditions. Table IV summarizes the prediction error of the adapted models using those individual workload performance values. We see that while prediction for the HIGH conditions is less accurate than when using a workload oracle, we still outperform the non-adapted model. For the LOW condition, prediction error never increases by more than 8.8% compared to the non-adapted model. When comparing the magnitude of improvement in prediction error for the HIGH results with the magnitude of deterioration for the LOW condition, we see that on average, we still achieve a substantial net benefit using the EEG-adapted models (assuming equal distribution of both conditions) instead of the non-adapted models. For example, on the SB data set, we have an improvement of 45.4% for the HIGH condition (EEG-adapted model vs. non-adapted model) compared to a degradation of 8.8% for the LOW condition.

Those results indicate that it is possible to combine EEG-based modeling of human workload with computational modeling of task execution in a cognitive architecture to provide adaptivity to varying conditions. All presented components (ACT-R model, EEG data acquisition and workload modeling) are available as real-time capable components. This allows the system to predict performance of a participant during the actual operation of the task, even in the case of dynamically changing workload conditions.

V. DISCUSSION

In the evaluation, we showed the importance of workload adaptation for the purpose of model tracing in situations of variable workload. For both the SB and the LCT data set, we could show a significant reduction in prediction error for the workload-adaptive model compared to the non-adaptive one. This is an important finding which is new to the research community. This improvement still holds if workload prediction is not perfectly accurate but derived from EEG. Additionally, we saw that the impact of the two secondary tasks on the performance in the main task was different, which means that the approach requires the availability of multiple dummy models for different cognitive resource profiles.

REFERENCES

[1] J. R. Anderson, D. Bothell, M. D. Byrne, S. Douglass, C. Lebiere, and Y. Qin, "An Integrated Theory of the Mind," *Psychological Review*, vol. 111, no. 4, pp. 1036–1060, 2004.

[2] R. S. Amant, T. E. Horton, and F. E. Ritter, "Model-based evaluation of expert cell phone menu interaction," *Transactions on Computer-Human Interaction*, vol. 14, no. 1, 2007.

[3] B. Chikhaoui and H. Pigot, "Towards analytical evaluation of human machine interfaces developed in the context of smart homes," *Interacting with Computers*, vol. 22, no. 6, pp. 449–464, 2010.

[4] D. D. Salvucci, M. Zuber, E. Beregovaia, and D. Markley, "Distract-R: rapid prototyping and evaluation of in-vehicle interfaces," in *Proceedings of the Conference on Human Factors in Computing Systems*, New York, USA, 2005, pp. 581–589.

[5] D. D. Salvucci, "Modeling Driver Behavior in a Cognitive Architecture," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 48, no. 2, pp. 362–380, 2006.

[6] W.-T. Fu, D. Bothell, S. Douglass, C. Haimson, M.-H. Sohn, and J. Anderson, "Toward a real-time model-based training system," *Interacting with Computers*, vol. 18, no. 6, pp. 1215–1241, Dec. 2006. [Online]. Available: http://iwc.oxfordjournals.org/content/18/6/1215

[7] R. E. Cochran, F. J. Lee, and E. Chown, "Modeling emotion: Arousals impact on memory," in *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, 2006.

[8] F. E. Ritter, A. L. Reifers, and M. J. Schoelles, "Lessons from Defining Theories of Stress," in *Integrated Models of Cognitive Systems*, 2006.

[9] G. Gunzelmann, L. Richard, K. A. Gluck, P. A, and D. F. Dinges, "Fatigue in sustained attention: Generalizing mechanisms for time awake to time on task," in *Cognitive fatigue: Multidisciplinary perspectives on current research and future applications*, ser. Decade of Behavior/Science Conference. Washington, DC, USA: American Psychological Association, 2011, pp. 83–101.

[10] R. Evertsz, F. E. Ritter, P. Busetta, M. Pedrotti, and J. L. Bittner, "CoJACKAchieving principled behaviour variation in a moderated cognitive architecture," in *Proceedings of the 17th conference on behavior representation in modeling and simulation*, 2008, pp. 80–89.

[11] C. L. Dancy, "ACT-R: A cognitive architecture with physiology and affect," *Biologically Inspired Cognitive Architectures*, vol. 6, pp. 40–45, 2013.

[12] J. P. Borst and J. R. Anderson, "Using the ACT-R cognitive architecture in combination with fMRI data," in *An Introduction to Model-Based Cognitive Neuroscience*. Springer, 2015, pp. 339–352.

[13] F. Putze, D. V. Holt, T. Schultz, and J. Funke, "Model-based Identification of EEG Markers for Learning Opportunities in an Associative Learning Task with Delayed Feedback," in *Proc. of 24th International Conference on Artificial Neural Networks*, Hamburg, Germany, 2014.

[14] D. D. Salvucci and N. A. Taatgen, "Threaded cognition: An integrated theory of concurrent multitasking," *Psychological Review*, vol. 115, no. 1, pp. 101–130, 2008.

[15] P. Bttner, "Hello Java: Linking ACT-R 6 with a Java Simulation," in *Proceedings of the 10th International Conference on Cognitive Modeling*, Philadelphia, USA, 2010, pp. 289–290.

[16] J. Jarvis, F. Putze, D. Heger, and T. Schultz, "Multimodal Person Independent Recognition of Workload Related Biosignal Patterns," in *Proceedings of the 13th International Conference on Multimodal Interfaces*, ser. ICMI '11, Alicante, Spain, 2011, pp. 205–208.

[17] H. H. Jasper, "The ten twenty electrode system of the international federation," *Electroencephalography and clinical neurophysiology*, vol. 10, pp. 371–375, 1958.

[18] S. Sternberg *et al.*, "High-speed scanning in human memory," *Science*, vol. 153, no. 3736, pp. 652–654, 1966.

[19] S. Mattes, "The lane-change-task as a tool for driver distraction evaluation," *Quality of Work and Products in Enterprises of the Future*, pp. 57–60, 2003.