An approach towards multitasking in ACT-R/PM

Jeronimo Dzaack, Juergen Kiefer, Leon Urbas
Research Group User Modeling in Dynamic Human-Machine-Systems
Technische Universität Berlin
Jebensstr. 1, J2-2
10623 Berlin, Germany
{jdz, jki,lur}@zmms.tu-berlin.de
+49 30 314 7200

ABSTRACT

In human-machine-interaction, interruptability and resumption of different tasks are common aspects to influence human performance. To understand human behavior in such situations, we conducted a multitasking experiment where subjects had to perform a test of attention when driving. In this paper, we present an ACT-R/PM model on how people perform the test of attention (secondary task of multitasking scenario). This first model serves as basis for our long-term-objective: to model multitasking, i.e. to simulate how people interrupt a task and recover, considering and integrating individual differences in human performance models. The reported study is a first attempt and smoothes the way for ongoing studies.

INTRODUCTION

In early stages of the design of technological systems, aspects of users more and more call for attention. Computer simulations, in our eyes, seem to be an appropriate method to consider these aspects. The resulting simulations are based on empirical, psychological results. One main feature we are interested in is multitasking.

Multitasking is often seen as natural ability. Interruptability and resumption of tasks are implicit components of multitasking, constituting a need which, so far, has not been fully considered in high-level cognitive architectures. Some, for instance, call for a single model to capture performance on multiple tasks. This reflects a part of the aim of our research group: we strive for a cognitive architecture to simulate multitasking, i.e. performance during interruption and resumption. Most models focus on single tasks like the Sternberg task or the Sperling task. In this case, interruption is forced by a self-defined break. We therefore investigated how people handle interruptability and resumption by using a multitasking scenario. The overall aim is to integrate these aspects in cognitive architectures. This paper is intended to illustrate our approach: we present an ACT-R/PM-model of the secondary task performed in the presented multitasking experiment conducted in a driving simulator at the Technische Universität Berlin.

EXPERIMENTAL SETTING

For the experimental setting, the situation while driving was used as multitasking scenario: driving was taken as main task. For the secondary task we referred to the D2 test of attention by Brickenkamp (2001), investigating individual differences on attention and concentration. The aim is to identify a pattern as correct according to Brickenkamp´s specification. As in real-life-scenarios of human-machine-interaction, e.g. when managing a navigation system or switching the radio, the used test acquires visual attention and, on an abstract level, can be seen as model for a driver infotainment system. Hence, it is referred to as D2-Drive test (Urbas et al., 2005).

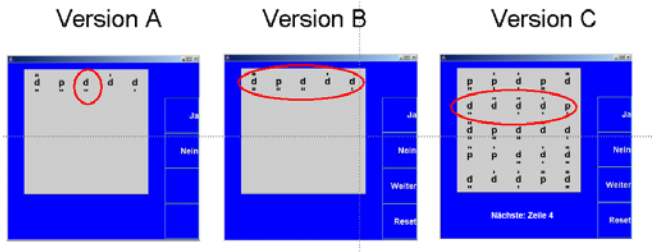

Figure 1: Test of attention

Three versions of the D2-Drive test were developed (see figure 1): the first relies on (static) visual search of the pattern in the middle (version ***A***). The second maintains sequential, horizontal visual search of multiple patterns in a row (equivalent to reading from left to right, version ***B***). The third is a combination of horizontal and vertical visual search (i.e., version ***A*** and ***B***) with a recall task (called version ***C***): within a field of information, one specific pattern (determined by row and column) must be found. We assume our electronic implementation to be cultural independent requiring no previous knowledge or special expertise. The configuration of our realization of the original paper and pencil test is appropriate to investigate interruptability and resumption, in a multitasking in-car-scenario. Consequently, each subject was asked to perform

the test of attention as pre-test (single task, baseline), real-test (while driving) and post-test (single task) after the experiment.

After a training phase in driving and a pre-test of the D2-Drive test, at three task-switching points a D2-Drive test had to be performed. The duration for each test was one minute. Subjects were instructed to attend driving as main task (high priority): they were asked to pay attention and to drive safely.

While the task scenario (single task vs. multitasking) was treated as within-subjects-design (i.e., each subject performed single and multitasking condition), we used a between-subjects-design to investigate the performance of subjects in the three different test versions (version ***A*** vs. version ***B*** vs. version ***C***).

Twenty-four subjects joined the study. Regarding to the D2-Drive test, they were told to work carefully but concurrently to complete as many patterns as possible.

HYPOTHESIS: D2-DRIVE (SINGLE TASK)

In this paper, we focus on performance in the test of attention. Please note that performance was measured by number of items per minute and **not** by number of correct items because the error rate approximates 0.

Assuming different levels of complexity, we expected the number of resulting patterns to be different in the three versions for the single task condition. Using performance rate **r**, we hypothesize

H1: **r** (A) > **r** (B) > **r** (C)

More concrete, we expected an improved performance of version ***A*** to ***B*** to ***C*** (a higher number of processed items).

RESULTS: D2-DRIVE (SINGLE TASK)

We observe a significant better performance in version ***A*** compared to ***B*** as well as compared to ***C*** ($p < .05$ for both), but there is no difference between ***B*** and ***C***. But a huge range in ***C*** can be observed: ***C*** evokes individual performance contrary to ***A*** where most of the subjects seem to perform approximately the same number (see figure 2).

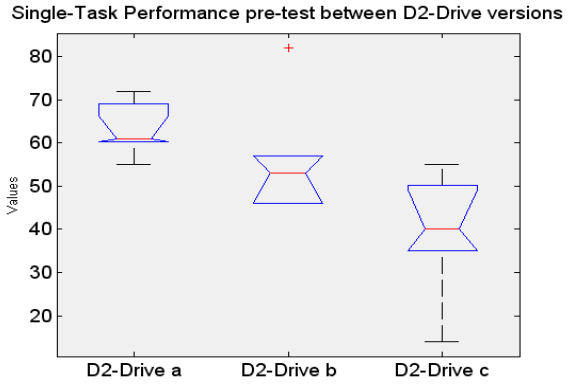

Figure 2: Task Difficulty

In version ***A***, most people reach between 60 and 70 items per minute. We therefore assume that one second for each pattern is a valid time prediction for subjects´ performance in version ***A***. Observing the data more precisely we found different strategies of performing the given task of identifying a pattern as correct or not. Over all subjects we see a significant difference ($x^2 = 26.7$, $p < .001$) in the response time for *d*-patterns or *p*-patterns. Observing the data individually per subject we find that the response times of more than 70% of the subjects show at least a small advantage for *p*-patterns in respect to response time, but only about 30% of the subjects show a significant difference ($p < .05$) in the time for judging a *p*- or *d*-pattern. 50 % of their response times are between 562 and 688 ms for a p-pattern (median = 610), 625-782 for a d-pattern (median = 687). We further identified other strategies: rhythmic key-presses with a quite impressive low deviation between response times, response-bursts for two or three elements due to separation of encoding and answering (variant B and C only). Finally 10% of the subjects show a clear advantage of the d-pattern in respect to response time.

A FIRST ACT–R/PM – MODEL OF D2-DRIVE

The goal of our research project is to predict users´ performance in human machine interaction. To do this, we use the ACT-R architecture (Anderson, 2004). Based on the described experimental study an ACT-R/PM model of human performance in the test of attention (D2-Drive) was derived.

In what follows we introduce the ACT-R/PM model of D2-Drive used in the experimental setting to collect empirical data. We start with the given assumptions which form the basis of our model. Subsequently the structure of the model is explained. We prove the correctness of this approach in comparing the model with data of the experiments. The results are discussed at the end.

Assumptions

To implement a model that predicts user behavior means reducing human beings to specific elements (e.g., goal-oriented, emotionless, perfect) because of the interference and the complexity. Not every aspect of human behavior can be integrated, and at the current state there is no need to do so. To predict user behavior a wide understanding of psychological theories to build and empirical data to verify the model is needed. The underlying concepts have to be outlined and, consequently, integrated within the implementation of the model.

The implementation of the ACT-R model of the secondary task – the interaction of humans with D2-Drive – implies assumptions that need to be integrated.

Declarative Memory

The main task of the D2-Drive test is to identify a given pattern as a correct (D2-) pattern. If the letter in the center position is a *d*, only nine arrangements of the determined alphabet (i.e., nothing: 0, one stroke: 1, two strokes: 2) are

possible to derive the conclusion. To enable the model to identify a pattern with a *d* as correct or not correct, we put nine chunks in the declarative memory concerning all possible statements (e.g., 00 → No, 01→No, 11→Yes, 02→Yes). Retrieving the chunk connected with the given signs allows the model to derive a conclusion (yes or no).

Strategies

The results of the structured interviews at the end of the experiments and the supervision of subjects show that there are several strategies to solve the problem of identifying the pattern as correct or not as described before. For the implementation of this model, we used one strategy for all three cases of the D2-Drive test based on own experience and empirical evidence (as can be read in the results section of this paper).

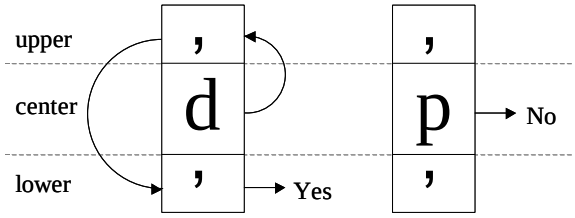

Figure 3: D2-pattern and encoding strategy

Each pattern is cut into three segments (upper: sign, center: letter, lower: sign) that are treated separately (see figure 3). A pattern is identified as correct D2-pattern if there is a *d* **and** two strokes.

1. SET VISUAL ATTENTION to center segment
2. ENCODE letter
    a. p → D2-pattern: no → END
    b. d → SET VISUAL ATTENTION to upper segment
3. ENCODE sign
4. SET VISUAL ATTENTION to lower segment
5. ENCODE sign
6. RETRIEVAL IN DM (both signs)
    a. is a D2-pattern → D2-pattern: yes → END
    b. no D2-pattern → D2-pattern: no → END

Figure 4: Algorithm to identify a specific pattern

The algorithm to identify a pattern as correct or not correct works as described (see figure 4):

The visual attention is set to the center segment of the pattern and the letter is encoded. If the letter is a *p*, the pattern cannot be a D2-pattern per definition. The procedure stops and the pattern is encoded as **no** D2-pattern. Otherwise the letter is a *d* and the visual attention is set to the upper segment to encode the upper sign and afterwards to the lower segment to encode the lower sign. In the next step a retrieval with both encoded signs is set to the declarative memory to identify the pattern. Both signs have to be encoded, because in all nine possible cases it is necessary to control both signs to identify a pattern as correct or not correct.

Visual Search

The first visual search in the model has to be treated separately because the empirical data shows a gap of orientation between the first and the subsequent tasks.

The visual focus is always on the considered pattern and changes to the next one when the (keyboard-) key is pressed to enter the conclusion of the model.

In versions ***B*** and ***C***, the model returns after 5 (9, 13) patterns to the beginning of a row. If the "Eye" of ACT-R cannot see anything (i.e., visual-location throws an error) the "Eye" has to be redirected to the (next) starting point because the end of the row is reached and the visual-location points an empty space.

In version ***C*** the number of the next row has to be stored. We assume this to happen before pressing the button the last time in a row.

General

The trial of performing the D2-Drive test is determined by the version of D2-Drive and a given time. Both parameters are set with the call of the start-function *s* of the model (i.e., s "a" 10).

Because of the re-use of specific structural elements all three versions are integrated in **one** single model.

Structure

The structure of the implementation of the ACT-R/PM model of D2-Drive is quite simple (see figure 5). It is a loop of separated tasks starting with a first visual search for orientation reasons. The loop consists of reading the pattern, interpretation of the pattern (both steps are described as identifying a pattern above), resetting the visual component (visual-location) and pressing the key.

1. SET STARTing point
2. READING pattern
3. INTERPRET pattern
4. MOVE VISUAL-LOCATION
5. PRESS-KEY

Figure 5: Structure of the ACT-R model

To cope with different settings of the three versions, the part *resetting the visual component* has to be altered for each version of D2-Drive because every version requires different coordinates for the observed pattern (see figure 6).

In version ***A***, only the middle pattern is observed. Thus, the visual component has to be reset to the center segment of the middle pattern after identifying one pattern. In version B, one row is observed successively and is then started again with a changed pattern. Hence the X-coordinate of the visual component has to be changed to step from one pattern to the next pattern. If the visual-location is empty (i.e., throws an error) the end of a row is reached and the visual component has to be reset to the starting coordinates of the row. In version ***C,*** additionally the number of the next row to be observed has to be stored. Thus the number has to be stored if the end of a row is reached. Changing the coordinates means to change the X- as well as the Y-coordinates of the visual component after reaching the end of a row.

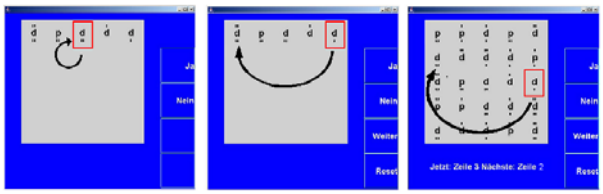

Figure 6: What to do at the end of a row?

Compare: Model vs. Reality

The output of the implemented ACT-R/PM model shows that a *d*-pattern requires 950 milliseconds and a *p*-pattern requires 750 milliseconds. In comparison with the data of the experiment, we conclude that the chosen strategy can be managed for this purpose. The amount of processed pattern in 60 seconds in all three versions (A: 75, B: 73, C: 56) is comparable with the results from the experiment regarding the upper third of the derived data. This can be explained by the assumptions of cognitive models to be perfect (i.e., no interfering variables). The hypothetical assumptions on performance (A > B > C) are not approved by the model. In version A and version B, nearly the same amount of patterns are processed. The model predicts the same performance for version A and version B, further it predicts for version C to be the less effective one (i.e., A = B, A > C).

All together the model shows a slightly over-estimation of performance. But the relative tendency of the tree different versions can be predicted by the model (A > B > C).

Using the Stand-Alone-Version of ACT-R 5.0, the performance times predicted by the models were close to subjects real performance. Integrating visual attention and motor action, ACT-R/PM turned out to be appropriate for our attempt to model user behavior in each of the three versions of the intended test-version.

DISCUSSION

The model presented in this paper is an attempt in modeling the secondary task and a first step of our vision to simulate interruptability and resumption of tasks in the driving simulator scenario we used.

In our model we implemented one strategy observed by most subjects. A next step in understanding individual behavior therefore must be the extension to a model including individual differences in strategic behavior.

The final aim is modeling multitasking in cognitive architectures. Thus we have to combine the developed cognitive model with another model to observe the model-behavior in multitasking and compare the results with the data from the experiment.

Individual differences

Based on subjects statements in the feedback questionnaire as well as on observations measured by eye movements, this section attends the importance of individual differences of subjects. People differ in their performance, behavior and (working memory) capacity (see Jongman et. al., 1999). Work by Daily et al. (2001), for instance, suggests an individual component of working memory capacity. Rehling et al. (2004) refer to individual difference factors in a complex task environment. All this recommends various derivatives of the starting ACT-R/PM model we derived. Please keep in mind that the focus of this paper is only on performing the test of attention in a single task condition. Ongoing research will investigate how to approach a multitasking ACT-R/PM model by questioning how, or if at all, to handle this complexity.

Another observation on how subjects process is the dimension of steps each one uses: in version ***B***, it seems to be of advantage not to compare and to press a key (y/n) after each pattern but to keep in mind the answer and then insert as many answers as can be kept in memory. This strategy has not been considered so far.

A general executive for multitasking

The next step in our research group is to combine our model with the car-driving scenario to analyze the effects of multitasking. To do so, we will use the *General Executive* described in Salvucci (2004). He proposes a general executive for multitasking suggesting to allow concurrent goals stored in a goal set. Because of the serial processing of ACT-R, only one single goal can be executed at the same time. Two heuristics define when to switch between goals. To determine which goal is chosen next the urgency of the concurrent goals is calculated and the most urgent is chosen.

In this case, we want to use the scenario of Salvucci to represent the primary task of our experiment and the ACT-R model of D2-Drive as secondary task.

Conclusion

We are aware of the limits of our model, although it is a starting point in our research. A next step concentrates on the question whether existing multitasking models in ACT-R are appropriate for our purpose (for instance, see Salvucci, 2001). Models of driving as well as of D2-Drive, taken together, will enlighten our way of modeling

interruptability and resumption in human machine interaction.

ACKNOWLEDGEMENT

This work was funded by grants of VolkswagenStiftung (research group user modeling), DFG (Research training group GRK 1013 prometei) and IBB PROFit (HMI Engineering for networked driving).

REFERENCES

Anderson, J., Bothel, D., Byrne, M.D., Douglass, S., Lebiere, C., and Qin, Y (2004) An Integrated Theory of the Mind. Retrieved from http://act-r.psy.cmu.edu/-papers/403/IntegratedTheory.pdf.

Brickenkamp, R. (2001). Test d2 Aufmerksamkeits-Belastungs-Test. 9., überarbeitetet und neu normierte Auflage. Hogrefe Verlage. Bern, Schweiz.

Daily, L. Z., Lovett, M. C., & Reder, L. M. (2001). Modeling individual differences in working memory performance: A source activation account in ACT-R. Cognitive Science 25, 315-353.

Heinath, M., Dzaack, J., Kiefer,J. Urbas, L. (unpublished) Handbook of D2-Drive. Technical University of Berlin. Berlin 2005.

Jongman, L. & Taatgen, N. A. (1999). An ACT-R model of individual differences in changes in adaptivity due to mental fatigue (pp. 246-251). In Proceedings of the twenty-first annual conference of the cognitive science society. Mahwah, NJ: Erlbaum.

Rehling, J., Lovett, M., Lebiere, C., Reder, L. M., & Demiral, B. (2004) Modeling complex tasks: An individual difference approach. In proceedings of the 26th Annual Conference of the Cognitive Science Society (pp. 1137-1142) . August 4-7, Chicago, USA

Salvucci, D. D., Chavez, A. K., & Lee, F. J. (2004) Modeling effects of age in complex tasks: A case study in driving. In proceedings of the 26th Annual Conference of the Cognitive Science Society (pp. 1197-1202) . August 4-7, Chicago, USA

Salvucci, D. D., Boer, E. R., & Liu, A. (2001). Toward an integrated model of driver behavior in a cognitive architecture. Transportation Research Record, 1779, .

Urbas, L., Schulze-Kissing, D., Leuchter, S., Dzaack, J., Kiefer, J., Heinath, M. (2005) *Programmbeschreibung D2-Drive-Aufmerksamkeitstest [Manual for D2-Drive Test of Attention]*. Berlin: ZMMS