Testing a Quantitative Model of Time Estimation in a Load-Switch Scenario

Nele Pape (nele.pape@zmms.tu-berlin.de)
GRK Prometei, ZMMS, Technische Universität Berlin
Franklinstr. 28-29, FR 2-7/2, 10587 Berlin Germany

Leon Urbas (leon.urbas@tu-dresden.de)
Dept. of. Electrical Engineering and Information Technology,
Technische Universität Dresden
01069 Dresden Germany

Abstract

A model of prospective time estimation was tested in two experimental variations which examine the influence of load switch in task demands on time estimation. The model predicts these influences on time estimates by means of memory processes such as spreading activation. The approach was integrated into a cognitive architecture and has previously been tested successfully. In two experiments participants had to work on a counting task with different levels of working memory demands (High/Low). The participants had to stop each trial after a perceived duration of a previously presented sample of 100 seconds (altered reproduction method) and received feedback. In the Low group most trials were performed in low load and one or two trials in high load (load switch), and vice versa for the High group. For the Low group the model predicts overestimations at load switches, but underestimations for the High group. We found that the model predictions in the first experiment only match the experimental results for the Low group, most probably due to the experimental design. In the second experiment, the design was therefore slightly changed and the timing task was embedded into a manual control task within a microworld environment. In this setting the model predictions match the time estimates for both groups. The series of experiments reported give strong evidence that the model is able to capture and to predict influences of task demands on time-estimates. The timing model may be used as a base for modeling subjective temporal reasoning and the timing of interaction with a dynamic system.

Keywords: Time estimation, cognitive modeling, coordinative working memory, memory processes, spreading activation, feedback.

Introduction

People can be good at estimating time and they sometimes rely on their estimates even when they are part of a safety-critical system. However, in stressful situations or in the course of demanding tasks, time estimates might be distorted to a large degree.

Time perception is crucial for everyday purposes and especially in the area of human-machine-interaction. In the context of operator performance, supervision of processes is a time critical task that might be prone to human errors, if other task demands rise suddenly.

The influence of task demand on time estimation has been examined thoroughly. A number of factors that are said to have an influence on time estimation are discussed in the literature. The most frequently mentioned factors are: attention (Block & Zakay, 1996; Zakay, 1993; Byrne, 2006), memory load (Brown, 1997; Brown & West, 1990; Dutke, 2005), or simply forgetting to estimate time if the task gets more demanding (Taatgen et al., 2007).

The most prominent model is the Attentional Gate Model (Block & Zakay, 1996). This assumes that a mental pacemaker regularly generates pulses to measure time. If a person directs attention to the course of time, a gate opens and the pulses are accumulated in a cognitive counter. When attention is distracted by a secondary task, the gate remains closed, pulses are not accumulated and the time-estimation is distorted. This way estimations turn out to be shorter whenever attentional ressources are captured by demanding secondary tasks.

A serious shortcoming of the Attentional Gate (ATG) Model is that it does not differentiate between specific and overall task demands. The model proposes influence of general attention but does not capture differences of specific task properties. Dutke (2005) therefore designed a counting task experiment to investigate the influence of two different working memory demands (sequential and coordinative) on time estimation. According to the ATG Model both demands would equally influence time estimation because attention is needed in both cases. However, Dutke's results show that both factors influence task-performance, but only high coordinative working memory demands distort time-estimates.

For the domain of human-machine-interaction, the susceptibility to workload induced distortions of time estimation is of high importance because operators do experience strong changes in workload (see e.g. Decortis and Cacciabue 1999). This might eventually lead to mishandling of the system due to a wrong timing of action. Furthermore, one can observe that most often time estimates need to be given under the very same general conditions as the reference time representations have been acquired before. Therefore we chose to set up a model that is designed for reproduction of time estimates (e.g. instead of giving time estimates verbally).

In the following we first sketch our computational implementation of a variant of the ATG Model which is prone to different task demands. We then introduce shortly the counting task and its specific task demands that may distort time estimation. Finally we show a series of two experiments that have been designed to challenge the models predictions.

A Computational Model of Time Estimation Involving Memory Processes

The idea behind the proposed model (Pape & Urbas, 2008) resembles some broadly accepted components of the ATG Model (Block & Zakay, 1996) with a pacemaker that generates pulses, an accumulator and an estimator, but without an associated gate. The main difference to the ATG model is a specific working memory account, which is realized by a mechanism to provide short-estimates between meaningful events (or "contextual changes" in the words of Block & Zakay, 1996) and an updating or construction process that integrates these short-estimates into a time estimate of the whole episode.

Figure 1 sketches the basic idea of the model: The vertical dashed line represents the pulses generated by the pacemaker as time goes by. The accumulator collects these pulses until a meaningful event occurs (depicted by an 'X' on the dashed line). The count of collected pulses together with some contextual information is stored in a temporal chunk (the short-estimate) that may be understood as an element of episodic memory (Tulving, 2002). The updating process then constructs a new episode-estimate by retrieval of the latest episode-estimate from memory and adding of the short-estimate. For instance, at the second event in the example shown in Figure 1, the episode-estimate, which carries 5 pulses, is retrieved and the newly accumulated 6 pulses in the short-estimate are summed up. A new episode-estimate with 11 pulses is stored in memory while the former remains. With a perfect memory, this new episode-estimate will be retrieved when the next event occurs, because it is the most recently generated chunk (with the highest activation). Additional memory activities might influence the activation level of two consecutive episode-estimates in a way that the wrong episode is retrieved instead of the latest episode-estimate (see dash-dotted line in Figure 1). So in our example instead of a final time representation with 24 pulses, a representation with 20 pulses is stored in memory. Therefore, demanding tasks cause time representations with fewer pulses than less demanding tasks. This mechanism generates shorter time representations only. Overestimations occur when the generated time representation is longer than a former time-representation. Contrary to other timing-models, this model needs no additional elements for pacemaker and accumulator variance and no attentional gate. Distortions and distribution of time-representations emerge naturally by means of variance in memory processes.

This approach was integrated into the cognitive architecture ACT-R (atomic components of thought – rational analysis; Anderson et al., 2004) and is called TaSTE (Task Sensitive Time Estimation) Module.

We utilized the sub-symbolic declarative memory mechanisms proposed and implemented in ACT-R without changes. The activation level A_i of a chunk i is calculated by the base-level, a noise component ε (set to 0.1) and a context component which is not shown in equation 1. For base-level activation the number of presentations n for chunk i and the time since the *jth* presentation are taken into account. The decay of activation is calculated with d (set to 0.4)

$$A_i = \ln\left(\sum_{j=1}^{n} t_j^{-d}\right) + \varepsilon$$

Equation 1: Calculating activation of chunk i.

Activation spreading from the current goal towards the episode-estimates is enabled via the above-mentioned contextual information and helps to keep the episode-estimates retrievable. The parameter association strength was modified to s = 6.

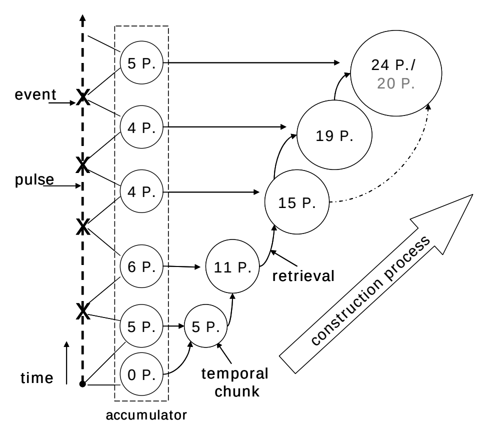

Figure 1: Construction process during an interval with several events. (Dotted arrow indicates the retrieval of an old instead of last time representation)

The Counting Task

The integrated timing-model was tested within a counting task (Dutke, 2005) with varying demands (sequential and coordinative) to compare human data to the predictions of the model. Sequential complexity refers to task variations that affect the number of simple and independent processing components and is demanding general intentional resources. Coordinative complexity refers to tasks in which the information flow between interrelated processing components needs to be coordinated (Mayr et al., 1996) and demands working memory resources.

In the counting task, the participants were asked to search lists of ten two digit numbers for either one or three targets (for low coordinative demand "16"; for high coordinative demand: "16", "38", and "67"). The sequential demand was varied with the overall number of targets contained in the lists (either 14 or 27 targets can be found within 40 lists).

The subjects have to count how often the different targets appear. On every third encounter of a target, the appropriate answer is given by pressing a specific key (e.g. labeled "18"), in all other cases the key marked with "No" should be pressed. After 400 sec., subjects were asked to reproduce the perceived duration by pressing a key to indicate the start and the end of the interval. Participants were randomly assigned to the four experimental conditions that result from the 2x2 between-subjects design (two levels of coordinative demands, two levels of sequential demands). Almost all participants underestimated the duration of the counting task. High coordinative demands produced larger reproduction errors and shorter estimates than low coordinative demands. For increased sequential demands the reproduction error was unaffected by the manipulation.

The model estimates showed the same effects of these demands as the human data (Pape & Urbas, 2008), because in the high coordinative condition more additional information has to be maintained. Both, simulation results in task performance and time estimations reveal comparable variability to each condition to empirical data, because the task model and the time module rely on retrieval processes where slight changes in activation lead to differences in results.

The load switch scenario

To adequately test the validity of the timing module we could either change the task or the scenario around the task as well as the estimation method. But, because with a new task it could be argued that the model data is dependent on the way the task was modeled and does not necessarily mirror the estimation processes that are assumed, we changed the task scenario and estimation method. This way we were able to reuse the model of the counting task that showed comparable performance to empirical data before (Pape & Urbas, 2008).

For the experiments reported here we also changed the interval duration to 100 seconds to check whether the model also holds for shorter intervals. Furthermore we modified the reproduction method. Instead of simply waiting, the participant had to work on the same task as in the encoding phase. We used repetitive timing to ensure that people were able to build up a good time representation before the load of the task switched (see Altman & Gray, 2008 for task switching scenarios) after several trials to either higher or lower coordinative demands.

Model runs

The model ran 22 times for each of two conditions representing the two groups used in the experiment for four different trials. To provide a reference the first trial was stopped after 100 seconds, the model thereafter used the built up representation as a reference to stop the next trial. In the case of the high condition group the first trial started with high coordinative demands which means the model had to cope in counting the occurrences of three targets and meanwhile building up a time representation. In case of the low condition group there was just one target to count. The time representation was used in the subsequent model run (a trial of equal load) for comparison to the new constantly updated representation. The task was stopped after an equivalent number of pulses had been collected. Because we assume that people build up a robust representation after a number of trials of equal load, we took the mean of the accumulated pulses for the interval and used it as time representation for the load-switch trial. In this trial the coordinative load changed compared to the previous which means low load in case of the high load group and vice versa.

This way we ended up with reproductions either derived in **inload** trials (trials according to the group condition) or **switch** trials for both groups (High/Low) (see model data figure 3 and 5).

No main effects in reproductions were found, but a significant interaction inload/switch*Group ($F(1,42)=7.5; p<.01; \eta^2=0.15$) show the different switch effects for the two groups. The model reproductions in the High group were much shorter in the switch trial and in the High group much longer than in the normal inload trials.

Experiment one

Our hypotheses generated by the model predictions were (1) reproductions performed in the same condition as experienced in the sample will be distributed around 100 seconds for both groups. (2) The load switch trial causes underestimations for the High group and overestimations for the Low group.

Participants

Forty-two participants (aged 21-48 years; Mean=26.05, SD=5.63) took part in the main experiment. The volunteers (25 male, 17 female) were paid 10 euros for participation.

Apparatus and setting

A standard keyboard was adapted as the entry device for the participants. Four keys on the number pad were covered with green tape that read 18, 34, 59, and also N and further apart another key marked Y. No sources of temporal information were available in the room.

Procedure

The participants were randomly assigned either to the High or Low group. Every experimental session began with the presentation of the sample duration. In every trial including the sample in the beginning, participants had to count the number of targets that appeared within the lists. Lists of 5 to 12 items (two digit numbers) appeared one after another in the middle of the screen for a time according to the number of items (3 to 10 seconds). Between lists the monitor was blank for 2 seconds. After the duration of 100 seconds, which was unknown to the participants, the task stopped and an instruction appeared on screen that the participant had to reproduce the experienced duration by starting and stopping the next trial by using the 'Y' button.

Session structure

The same instructions and training trials were given to all participants. After completing a demographic questionnaire the participants were informed about the counting task. After a training trial, the participants read another instruction about the experimental procedure and the reproduction procedure. Furthermore, they were informed that the length of lists, the number of lists, and the number of targets vary. Before a new trial started, the participants were informed look either for all three targets 18, 34, 39 or for just one target. No further targets were to appear than those mentioned.

After the 1st, 6th and 8th (last) reproduction the participants had to fill out a NASA-TLX questionnaire (Hart & Staveland, 1988) that measures workload.

Immediately following the experimental trials we conducted a structured interview to learn about the time estimation strategy, the difficulties of the tasks, and their strategy for the counting task the participants had applied.

Testing

After the sample-duration-trial (of 100 seconds), participants had to reproduce the duration 8 times with subsequent feedback about the quality of their reproduction (figure 2). A horizontal bar indicates the correspondence between sample duration and reproduced duration. If the horizontal bar is located below the middle area, the duration has been underestimated.

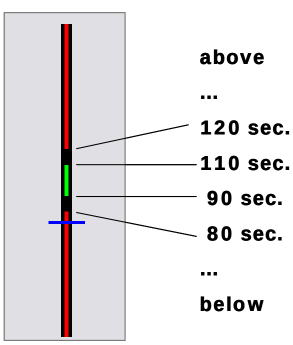

Figure 2: The feedback given after each reproduction (here the feedback indicates a strong underestimation).

No information about the assigned condition was given to the participants. Before every trial, participants were informed about the targets they had to count. When the trial was not stopped by the participant after 140 seconds, a message appeared on screen saying that no more lists are going to show up and the 'Y' button is to be pressed.

Results and comparison of experiment one

For the scores on the NASA TLX (1st and 2nd measures in load, the 3rd after the switch) a one-way repeated ANOVA revealed a significant interaction effect of NASA-TLX score and group (Low, High) ($F(2,76)=13.8$, $p<.01$; $\eta^2=0.267$). Planned contrasts showed that the first two measures in the NASA-TLX changed significantly to the third (group Low: $F(1,38)=15.6$, $p<.01$; $\eta^2=0.291$; group High: $F(1,38)=24.1$; $p<.01$; $\eta^2=0.388$). Therefore, the load-switch in the last trial seemed to have had the expected effect.

For the eight time-reproductions of the empirical data, the repeated ANOVA revealed a significant effect for reproductions ($F(7,238)=8.86$, $p<.01$, $\eta=.46$). There was a significant difference between inload reproductions to switch reproductions ($F(2,43)=7.78$;$p<.05$;$\eta^2=0.35$). But the predicted interaction between group and trial condition did not reach significance. Planed contrasts reveal that for the low group most reproductions in the inload condition were significantly shorter than the final one. Therefore just the low group showed the predicted switch effect, as shown in figure 3.

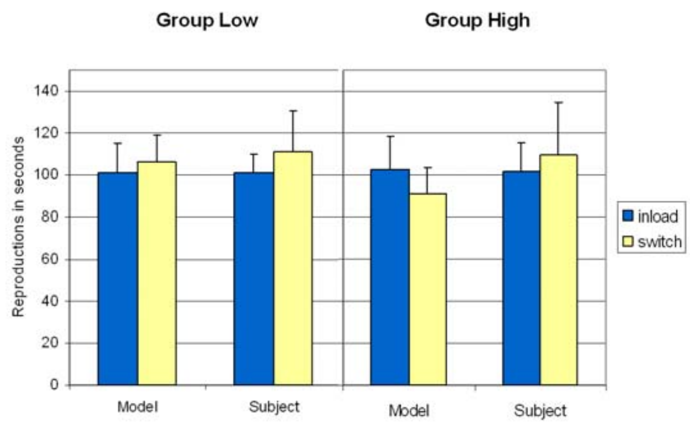

Figure 3: Model reproductions compared to empirical data in experiment one.

We had four possible explanations for the results. First, subjects in the High group reported in the interview that they were aware that their time perception would change after the switch to low load and therefore waited longer until they stopped the trial. Subjects in the Low group were too busy in the last high condition trial to reason about these things.

Second, some authors (Sturmer, 1966; Wearden et al., 1999) report that repetitive time estimations in a monotonic task with no background activity and no feedback reveal a lengthening effect, which means that estimates get longer the more estimates were made. We tried to avoid this by giving feedback but this might not have helped to totally prevent the effect. Third, the NASA TLX might have interfered with the estimates because after presenting the questionnaires participants showed a slightly longer reproduction.

Forth, the single switch in load after 7 inload trials might have been unexpected, causing participants to overestimate although participants were trained in both conditions. Therefore we conducted a second experiment that avoids the assumed factors.

Experiment two

For the second experiment subjects experienced four inload trials including the sample trial without reproduction. Then a first switch trial occurred. After that another four inload trials including the sample trial had to be completed before the second switch trial occurred.

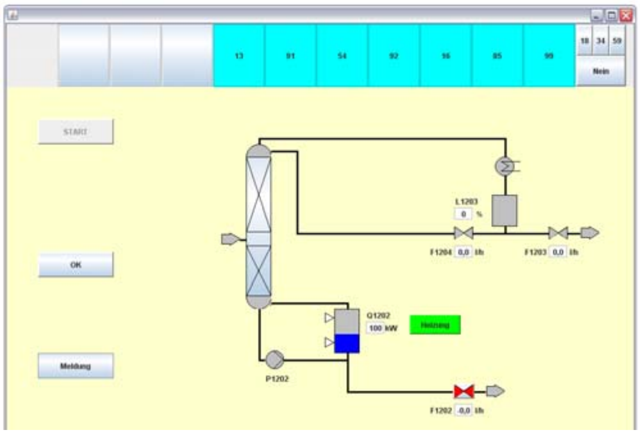

Figure 4: In the microworld scenario the dark blue liquid had to stay between the white triangles by opening the red valve below. At the same time the blue alarms had to be handled.

To add more background activity we used a microworld environment of an operator task in which the level of some liquid had to be maintained within a certain range and alarms had to be responded to (see figure 4). The operator has to handle certain important alarms which need to be counted, and ignore the remaining alarms. The alarm task resembled the counting task and for every new trial the participant in the role of the operator was informed about the important upcoming alarms just as in the previous experiment.

We assumed that the high workload of this multitasking set effectively hinders the participants to post-hoc reason about their way of time perception and compensate. Furthermore, we hoped to reduce the lengthening effect by inducing the first switch earlier and 'start anew' with a second sample trial afterward. Finally we eliminated the NASA TLX to avoid additional interference effects.

Participants second experiment

Fifty-three participants (aged 21-40 years; M=26.43, SD=4.84) took part in the second experiment. The volunteers (28 male, 25 female) were paid 10 euros for participation.

Procedure, structure and testing

The second experiment resembled the first experiment with the above mentioned differences. The participants received extra training for the operator task and had to interact with the mouse in the microworld environment instead of with the keyboard.

Results and comparison

A main effect for reproductions was found ($F(3,153)=4.382$; $p<.01$; $\eta^2=.079$). Furthermore for the second half of the experiment we found a significant interaction between reproductions and group ($F(3,153)=2.6$; $p=.053$; $\eta^2=0.049$) and a linear trend in increasing estimates by means of a planned contrast ($F(1,51)=10.7$; $p<.01$; $\eta^2=.173$).

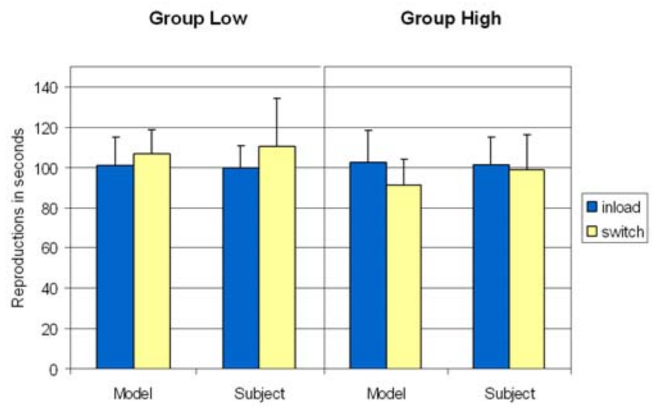

Figure 5: The comparison of the model predictions and the empirical reproductions of experiment two.

This time we find the predicted significant interaction for group*inload/switch as predicted by the model. Figure 5 shows the differences between inload and switch reproductions for the second half of the experiment with the significant interaction inload/switch*group ($F(1,51)=5.3$, $p<.05$; $\eta^2=.094$). Furthermore, not only the means of experiment and simulation resemble each other pretty much, but also the measures for the distributions are comparable (see table 1).

Table 1: The means and standard deviation in brackets of time reproductions for model and experimental data.

	Low		High	
	Model	Participants	Model	Participants
inload	100.9 (14.2)	99.8 (11.3)	102.4 (15.9)	101.6 (13.3)
switch	106.5 (12.6)	110.2 (23.9)	91.2 (12.6)	98.8 (17.6)

Discussion

The two experiments show that the TaSTE Module is able to predict human time estimates even under changing task demands not just in respect to the mean of the estimates but also in terms of distribution. Other current timing modules for ACT-R are not able to predict these task demand induced differences. The module presented by Taatgen et al. (2007) which has been designed for short term estimates assumes that distortions emerge from people “forgetting” to estimate time and restarting their timer. This would indeed result in shorter estimates. But the probability for restarting the timer has to be estimated for each task. Therefore it is only possible to replicate but not to predict distortions in time estimates. Byrne’s (2006) timing module assumes that attention factors cause distortions. In the case of the counting task the same amount of time is available for attention to time under difficult and easy conditions. In the case of experiment two, hardly any time is given for attention to time, because of the supervision task for the level of the liquid and the alarms. Byrne's timing module therefore predicts no difference for the load switch but a high difference for experiment one and two.

Nevertheless our model still needs further work, because additional factors seem to influence time estimation. These are (1) the lengthening effect of repetitive estimates, (2) additional questionnaires that might also lengthen estimates such as the NASA TLX, and (3) people are aware of their time distortions and counteract if they have the resources to do so.

At least for the lengthening effect there might be some explanation in the implemented model: More temporal chunks will reduce the activation spread to the distinct chunks and more confusion will occur during the updating process of the time representation.

Conclusion

The results of the experiments show that variance and distortion of human time estimation may be modeled by basic memory mechanisms as implemented in ACT-R. In this sense the TaSTE module is an integrated model that builds upon principles that are found in other cognitive domains. This does not imply that time estimating processes have to work they way sketched here. But formalizing a quantitative model allows evaluating different mechanisms in different task setting.

Next steps are to analyze the sensitivity of the model against different kind of tasks. The limits of the model predictions concerning the durations between events and the influence of the structure of short-estimates should be investigated further.

Acknowledgments

We thank Liubov Cramer, Rebecca Wiczorek, Julia Niemann and Thomas Nicolai for their help in preparing and conducting the experiments, our cooperation partners at UCIT (Finland) for discussing this paper, and Richard Holmes, Berlin for comments on language. This work was sponsored by Volkswagen Stiftung (Research Group Modelling of User Behaviour in Dynamic Systems) and Deutsche Forschungsgemeinschaft (DFG Research Training Group Prospective Design of Human Technology Interaction, GRK 1013).

References

Altmann E.M., & Gray, W.D. (2008). An Integrated Model of Cognitive Control in Task Switching. *Psychological Review,* 115, 602-639.

Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An Integrated Theory of the Mind. *Psychological Review,* 111, 1036–1060.

Block, R.A., & Zakay, D. (1996). Models of Psychological Time Revisited. In H. Helfrich (Eds.), *Time and mind. Proceedings of the International Symposium on Time and Mind.* Kirkland, WA: Hogrefe & Huber, 171-195.

Brown, S.W. Attentional Resources in Timing: Interference Effects in Concurrent Temporal and Nontemporal Working Memory Tasks. *Perception & Psychophysics,* 59 (1997), 1118-1140.

Byrne, M.D. (2006). An ACT-R Timing Module Based on the Attentional Gate Model. ACT-R Workshop. Retrieved from http://chil.rice.edu/projects/agm/index.html [2009-03-08].

Brown, S.W., & West, A.N. (1990). Multiple timing and the allocation of attention. *Acta Psychologica,* 75, 103-121.

Cowan, N. (2005). *Working memory capacity*. New York, NY: Psychology Press.

Decortis, F. & Cacciabue, P. C. (1988) Temporal dimension in cognitive models. In E. W. Hagen (Ed). *Proceedings of the 4th IEEE Conference on Human Factors and Power Plants.*

Dutke, S. (2005). Remembered duration: Working memory and the reproduction of intervals. *Perception & Psychophysics,* 67, 1404-1422.

Hart, S.G., & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* 139-183. Amsterdam: Elsevier Science Publishers.

Lovett, M. C., Reder, L. M., & Lebiere, C. (1999). Modeling working memory in a unified architecture: An ACT-R perspective. In Miyake, A. & Shah, P. (Eds.) *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control.* New York: Cambridge University Press.

Mayr, U., Kliegl, R., & Krampe, R.Th. (1996). Sequential and coordinative processing dynamics in figural transformations across the life span. *Cognition,* 59, 61-90.

Pape, N., & Urbas, L. (2008). A Model of Time-Estimation Considering Working Memory Demands. *Proceedings of the 30th Annual Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Sturmer, G. (1966). Stimulus variation and sequential judgements of duration. *Quarterly Journal of Exp. Psychol.* 18, 4, 354-7.

Taatgen, N.A., van Rijn, H., & Anderson, J.R (2007). An Integrated Theory of Prospective Time Interval Estimation: The Role of Cognition, Attention and Learning. *Psychological Review,* 114, 3, 577-598.

Tulving E. (2002). Episodic memory: from mind to brain. *Annual review of psychology* 53, 1–25.

Wearden, J.H., Pilkington, R., & Carter E. (1999). "Subjective lengthening" during repeated testing of a simple temporal discrimination. *Behavioural Processes,* 46, 25-38.

Zakay, D. (1993). Relative and absolute duration judgments under prospective and retrospective paradigms. *Perception & Psychophysics* 54, 656–664.