LAURA BORGHANS & MYRTHE PRINCEN

The stability of memory performance using an adapted version of the Delayed Matching To Sample task: an ERP study

Original Paper

This study examined the stability of performance using an adapted version of the Delayed Matching to Sample memory test (DMTS). With this test the stability of visual working memory can be measured. In the DMTS participants have to memorize a cue for a certain delay, after which a probe is presented. This is an often used test, but it is not yet clear if the performance of this test is also stable over time. In present paradigm stability was assessed mixing short (10-20s) and long delays (140-200s) and presenting cues during the cue-probe interval. Additionally, EEG was measured during testing. For the memory-related components, the P300 and P600, amplitude and latency were compared between sessions and the same was done for accuracy and mean reaction time of behavioral data. Faster reaction times for long delay stimuli were found in Session 2. No effects in accuracy were found. Longer P600 latencies were found for long delay stimuli in the first session as compared to the second. High correlations between sessions were found for almost all other measures. These high correlations suggest that test-retest scores were stable. Therefore, we conclude that the new DMTS paradigm has high stability.
Keywords: Delayed Matching to Sample; ERP; memory stability, EEG

Laura Borghans; Bachelor student year 3
Myrthe Princen; Bachelor student year 3
Maastricht University, Maastricht, The Netherlands

l.borghans@student.maastrichtuniversity.nl
mm.princen@student.maastrichtuniversity.nl

INTRODUCTION

The DMTS is a computerized test that has frequently been used to assess visual working memory in a variety of settings, such as in animal (see Dudchenko, 2004, for review), clinical (e.g., Chamberlain et al., 2011), and pharmacological studies (e.g., Turner et al., 2003). In the DMTS, participants hold an item (the cue) in memory for a certain delay, after which they receive a probe and have to choose the remembered item amongst a number of distractor items. In the typical DMTS, for instance the one developed by Cambridge Cognition (i.e. the CANTAB, see also Turner et al., 2003), the cue-probe interval usually varies from 0-12 s (Chamberlain et al., 2011; Turner et al., 2003). By possibly varying the delay periods, it makes the test perfectly suitable to investigate the various memory storage phases, i.e. encoding and consolidation. In this study, the stability of memory in an extensively modified DMTS was examined in order to establish whether repeated testing makes sense using this paradigm.

For intact learning and comprehension, working memory is needed. Baddeley (2003) proposed a model of working memory, which consists of the phonological loop, the visuo-spatial sketchpad and the episodic buffer. These components are supervised by the central executive that directs attention to relevant information. The visuo-spatial sketchpad is assumed to hold visual and spatial information. Because the visuo-spatial sketchpad is engaged when performing visual tasks we were particularly interested in this aspect. Here we used a DMTS with two intervals, namely 10-20 and 140-200 s, mixing the trials with short and long delays. The 10-20 interval represents memory encoding and this interval is used often in studying the DMTS. In our paradigm a new interval was additionally presented. This 140-200 s was used to examine the stage of memory consolidation. Evidence from animal studies suggests that the longer the interval, the less the accuracy (Grant, 1991). Also in humans an impaired accuracy was found in a visual array task, in which longer delays resulted in lower accuracy regardless of cognitive load (Ricker, 2010). This would suggest that in the present paradigm the performance of the participants should be decreased when the delay is longer. However, due to another modification, namely presenting 2 probe items instead of 4, the general load of the test was decreased to some extent, making the load more equal to standard DTMS tasks. Typically, no other stimuli are presented during the intervals between a cue and a probe. Here, however, cues were presented during the cue-probe intervals. This could reduce the test by around 15 minutes. To conclude, a strongly modified version of the DMTS was used in this study.

In a new paradigm, it is important to assess the stability of responding, i.e., whether scores are similar on various test days. Examining this issue was the main aim of this study, especially since stability has, to our knowledge, not critically been evaluated so far regarding the DMTS, especially not in a modified version as used here.

Electroencephalography (EEG) is a method to measure electrical brain activity. Electrodes placed on the scalp can record small changes in overall electrical activity in response to a stimulus. Averaging the responses evoked by a stimulus, an event related potential (ERP) containing various components can be extracted, which is caused by and time locked to sensory, motor or cognitive processes (Luck, 2005).

ERP components can be described by polarity and order of occurrence by extracting amplitude and latency. In this study, ERPs were recorded in response to the probes, because performance for encoding (i.e., short delay stimuli) and consolidation (i.e., long delay stimuli) can be assessed by looking at the accuracy of responding to probes.

ERP components of interest in this research were the P300 and P600. The human P300 component is a positive wave, elicited approximately 300 ms after stimulus onset. It is affected for example by stimulus probability, attention and expectancy (Kaestner & Polich, 2011), but is also enlarged for memorized compared to unfamiliar items (Mecklinger, 2010). Previous research has shown high test-retest reliability, thus stability, for the amplitude and latency of the P300 (Hall, Schulze, Rijsdijk, Picchioni, Ettinger et al., 2006). A P600 component also has a positive amplitude with a peak approximately 600 ms after stimulus presentation. The P600 is specifically related to memory processes (Mecklinger, 2010) and generally increased during memorization of the cue in a DMTS (Klaver et al., 1999.

The aim of this research was to investigate whether visual working memory, measured using a novel version of the DMTS, is stable over time. The stability of performance was established over two sessions by comparing accuracy and mean reaction time for behavioral data, and peak latency and amplitude for the P300 and P600 components. Klein and Fiss (1999) also found high stability for working memory. Therefore, behavioral data (accuracy and reaction time) as well as the ERP components (latency and amplitude) were expected to be equal in both sessions. Additionally, stability for short vs. long delay stimuli would be equally strong. In other words, the stability was expected to be very high for all dependent variables. Finally, it was expected that performance for longer delays would be slightly decreased compared to shorter delays, meaning less items would be remembered for longer delays.

METHODS

Participants

The participants were 20 right-handed students of Maastricht University (17 female), recruited via advertisements at university. The participants' mean age was 20.55 (range 18-27). The study was approved by the ethical committee (Ethische Commissie Psychologie) at Maastricht University and participants signed an informed consent before participation. They were rewarded with 5.5 course credits.

Design

In this study a within subject design was used, with session as within subject factor. The dependent variables were: reaction time (ms), accuracy (%), peak latency (ms) and peak amplitude (μV). To make sure that different versions of the task would not have influence on the participants' performance, two different versions of the test were balanced over participants. In these versions, only the pictures were different.

Procedure

The participants had to come to the lab three times, once for a training session and twice for a test session. The purpose of the first session was to train the participants to become familiarized with the tests. During this session no EEG was recorded. The training session had a duration of 1 hour. The first and the second test session were the experimental sessions, in which EEG was recorded. These sessions had a duration of 1.5 hour each. These sessions were at the same time of the day, but with a delay of one week. In this way, for each participant the stability over the same amount of time was assessed, possibly reducing variance between participants.

Measuring EEG

An EEG cap was used to place a set of 32 EEG electrodes according to the international 10-20 system (Jasper, 1958). A reference and a ground was placed at the linked mastoids and at the forehead, respectively. Eye movements were detected by horizontal and vertical electro-oculogram (EOG) recordings. Before electrode attachment, the positions were slightly scrubbed with a gel in order to provide good conductance. The impedance value was <10 kOhm. Both EEG and EOG were filtered between 0.01 and 100 Hz and sampled at 500 Hz.

DMTS

The test consisted of 50 immediate (interval between encoding and recognition 10-20 s) and 50 delayed (interval between encoding and recognition 140-200 s) recognition trials. Pictures were presented one by one in the middle of the screen. These pictures were mixed with 100 recognition trials, which consisted of the presentation of a new picture and an old picture, presented together in the middle of the screen. There were 50 recognition trials, the probes, that required long-term recognition and 50 trials that required short-term recognition. When a recognition trial was presented, the participants' task was to decide which picture was new and which was old. After recognition, which was done by a button press (left button if the left was the old stimulus and right if it was the right one), the next trial was presented. This could either be a probe or an encoding trial.

All pictures were everyday, easy-to-name objects, presented in grayscale (7 x7 cm). Each encoding picture was presented for 1000 ms. The recognition trial remained on the screen for 2000 ms, unless the participant responded faster. In that case, the stimulus disappeared immediately after response. The interstimulus interval was 1000 ms for each trial type.

Outcome variables were the number of correct recognitions for each delay (accuracy), as well as the mean reaction time of the responses for the behavior. The amplitude and latency of the P300 and P600 were the ERP outcomes.

Analysis

DMTS

Mean reaction time and accuracy were calculated. All behavioral and ERP data were analyzed with GLM repeated measures. Session (1 vs. 2) was taken as within-subject variable, as were for some analyses also stimulus type (short vs. long delay). Also Pearson correlations were calculated for Session.

EEG

Offline, eyeblink activity was removed from the EEG using the ocular correction method in Vision Analyzer 2 (Brain Products GMBH). Furthermore, trials in which movement artifacts occurred were excluded from analysis. Data were filtered with a band pass of 1 – 30 Hz. The EEG fragments within an epoch of 100 ms before stimulus onset and 1000 ms after onset were averaged, using the pre-stimulus interval as baseline. Separate averages for the two intervals for recognition were made for each of the two sessions. Next, grand averages over participants were calculated for each stimulus type (short delay/ long delay, new/old), from which the following ERP components were determined: P300 (200 – 350 ms) and P600 (450 – 750 ms). For these components latency and amplitude were analyzed. In this paper only data for the PZ electrode are presented.

RESULTS

Behavioral data

The mean reaction time and accuracy for all participants are shown in figure 1. A difference is made between first and second session, and between short and long delay stimuli.

Short delay across sessions. A repeated-measures ANOVA with session as repeated measure variable on reaction time and accuracy for short delay stimuli revealed no effect of session (RT: $F(1,19)=1.782$, $p<.199$; Acc: $F(1,19)=.104$, $p<.751$). These results additionally showed a significant correlation between sessions for reaction time ($r=.858$, $p<.001$) and a marginally significant correlation between sessions for accuracy ($r=.429$, $p<.06$).

Long delay across sessions. A repeated-measures ANOVA with session as repeated measure variable on reaction time and accuracy for long delay stimuli revealed no effect of session for accuracy ($F(1,19)=.435$, $p<.519$) and a significant effect for reaction time ($F(1,19)=8.086$, $p<.011$). The results for accuracy revealed a marginally significant correlation between sessions ($r=.428$, $p<.061$). The results for reaction time showed a significant correlation between sessions ($r=.833$, $p<.001$). Participants were significantly faster in the second session ($\mu 1 = 970$ms, $\mu 2 = 886$ms). Short vs. long delay Session 1. A repeated-measures ANOVA with delay (short vs. long) as repeated measure variable on reaction time and accuracy for Session 1 stimuli revealed a significant effect of delay (RT: $F(1,19)=12.030$, $p<.004$; Acc:

F(1,19)=12.388, p<.003). These results were accompanied by significant correlation between delays (RT: r=.913, p<.001; Acc: r=.522, p<.019). Participants were significantly faster and more accurate for stimuli with a short delay, compared to stimuli with a long delay (RT: μ_1 = 890, μ_2 = 970; Acc: μ_1 = .874, μ_2 = .815).

Short vs. long delay Session 2. A repeated-measures ANOVA with delay (short vs. long) as repeated measure variable on reaction time and accuracy for session 2 stimuli revealed a significant effect of delay for accuracy (F(1,19)=9.800, p<.007) and no significant effect of delay for reaction time (F(1,19)=1.518, p<.234). The results additionally showed a significant correlation between delays (RT: r=.894, p<.001; Acc: r=.713, p<.001). Participants were significantly better in responding to short delay stimuli compared to long delay stimuli (Acc: μ_1 = .868, μ_2 = .827).

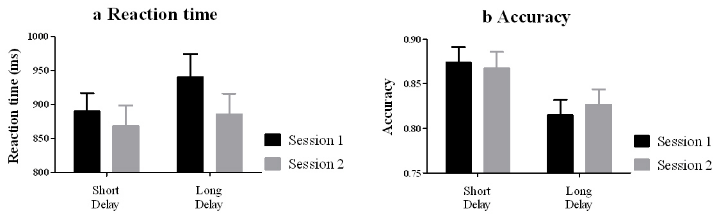

Figure 1. Mean reaction time (ms) and accuracy for long and short delays per session.
Figure 1A: significant longer reaction times for long delay stimuli in Session 1 are present.
Figure 1B: no significant effects of session for accuracy were present.

ERP data

The grand average of the ERP data of this experiment can be found in figure 2. Different lines show different sessions and different delays. The results of the EEG data can be found in figure 3

P300, both delays across sessions. A repeated-measures ANOVA with Session (1 vs. 2) as repeated measure variable on P300 latency and amplitude for short delay stimuli revealed no effect of session (latency: F(1,19)=.236, p<.634; amplitude: F(1,19)=.007, p<.934). Significant correlations between sessions were found (latency: r=.474, p<.036; amplitude: r=.874, p<.001).

A repeated-measures ANOVA with Session (1 vs. 2) as repeated measure variable on P300 latency and amplitude for long delay stimuli revealed no effect of session (latency: F(1,19)=.502, p<.488; amplitude: F(1,19)=1.372, p<.257). The results for amplitude revealed a significant correlation between sessions (r=.855, p<.001). Latency did not significantly correlate between sessions (r=.131, p<.583).

P600, both delays across sessions. A repeated-measures ANOVA with session (1 vs. 2) as repeated measure variable on P600 latency and amplitude for short delay stimuli revealed no effect of session (latency: $F(1,19)=.240$, $p<.631$; amplitude: $F(1,19)=2.737$, $p<.115$). The amplitude showed a significant correlation between sessions ($r=.642$, $p<.003$). The results for latency did not show this correlation ($r=.146$, $p<.541$).

A repeated-measures ANOVA with Session (1 vs. 2) as repeated measure variable on P600 latency and amplitude for long delay stimuli revealed no effect of session for amplitude ($F(1,19)=.442$, $p<.515$) but a significant effect for latency ($F(1,19)=7.181$, $p<.016$). This significant effect can be found in figure 3D. A marginally significant correlation between sessions was found for amplitude ($r=.425$, $p<.063$) and no significant correlation was revealed for latency ($r=.337$, $p<.147$). P600 was significantly prolonged in Session 1 compared to Session 2 (latency: μ_2 = 604ms, μ_3 = 550ms).

P300, within sessions. A repeated-measures ANOVA with delay (short vs. long) as repeated measure variable on P300 latency and amplitude for Session 1 stimuli revealed no effect of delay (latency: $F(1,19)=.012$, $p<.914$; amplitude: $F(1,19)=.176$, $p<.681$). Significant correlations were found between delays (latency: $r=.578$, $p<.009$; amplitude: $r=.842$, $p<.001$).

A repeated-measures ANOVA with delay (short vs. long) as repeated measure variable on P300 latency and amplitude for Session 2 stimuli revealed no effect of delay (latency: $F(1,19)=1.259$, $p<.277$; amplitude: $F(1,19)=3.018$, $p<.100$). The results showed a significant correlation between delays for amplitude ($r=.964$, $p<.001$) but no significant correlation for latency ($r=.140$, $p<.556$).

P600, within sessions. A repeated-measures ANOVA with delay (short vs. long) as repeated measure variable on P600 latency and amplitude for Session 1 stimuli revealed no effect of delay (latency: $F(1,19)=1.208$, $p<.286$; amplitude: $F(1,19)=.315$, $p<.582$). A significant correlation between delays was found for amplitude ($r=.758$, $p<.001$) and a marginally significant correlation for latency ($r=.404$, $p<.079$).

A repeated-measures ANOVA with delay (short vs. long) as repeated measure variable on P600 latency and amplitude for Session 2 stimuli revealed no effect of delay (latency: $F(1,19)=1.191$, $p<.290$; amplitude: $F(1,19)=4.019$, $p<.060$). These results revealed a significant correlation between delays for amplitude ($r=.671$, $p<.002$) but not for latency ($r=.323$, $p<.166$).

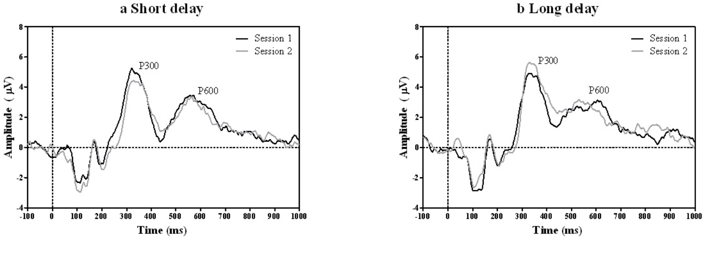

Figure 2. Grand Average of the ERP signal of the different delays and different sessions

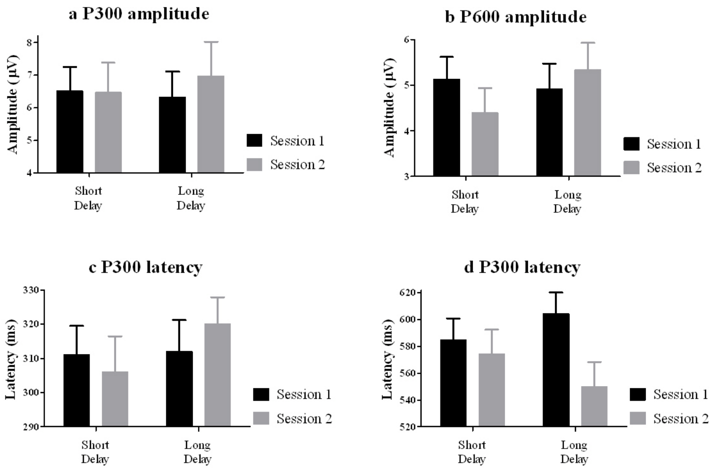

Figure 3. Amplitude (µV) and Latency (ms) for P300 and P600 for long and short delay on session.
Figure 3A-3C: the effects of session on the P300 amplitude, P600 amplitude and P300 latency respectively. No significant effects were present.
Figure 3D:the significant effect of session on the P600 latency for long delay stimuli. Other figures show no significant effects.

DISCUSSION AND CONCLUSION

The aim of this research was to assess stability of visual working memory using the delayed matching to sample task (DMTS). For the first time, a new paradigm was used, which included both short and long delay stimuli, within one paradigm. Comparisons were also made between short and long delay stimuli. Reaction times for long delay stimuli were significantly faster in participants in the second session compared to those in the first session. This difference was also found for P600 latency in response to long delay stimuli, which was longer in the first session. Although behavioral data suggest a significant difference in reaction time regarding long delay stimuli, this difference was accompanied by a high correlation between the first and second session on almost all the other measures. Thus, generally it can be said that the stability of working memory was quite high using this test. As for the comparisons between short and long delay stimuli, participants performed significantly better on short delay stimuli. Only in the second sessions the participants were faster on short delay stimuli. These effects were not found in the EEG signal.

The differences in reaction time between the first and second session for long delay stimuli can be explained in different ways. First of all, it is possible that a learning effect had influence on the reaction time. Learning effects are improvements in performance due to performing a test for a second, third or fourth time (Gluck, Mercado, & Meyer, 2008). A follow-up study can try to eliminate this effect by using two sessions for training before the experimental sessions. A second explanation consists of the development of more suitable strategies, between the first and second session, such as, for example, quickly pressing the button if something is not remembered directly after presentation of the probe trial. In other words, it is possible that participants started to simply quickly guess rather than perform the task properly. This could be avoided by specifically explaining a preferred strategy during instruction. Furthermore, It is possible that participants recognized their relatively long reaction times and tried to find a way to perform faster. This can also explain the fact that accuracy did not improve, but remained stable over sessions. This phenomenon is called speed-accuracy trade-off and is sometimes found in memory studies (Bogacz, Wagenmakers, Forstmann, & Nieuwenhuis, (2010).

Another interesting result is found in the high correlation between reaction time on long delay stimuli between the first and second session. This correlation indicates that all participants improved their reaction time on correct responses, suggesting that there is not a participant-specific change, but a general explanation of the improvement, such as an effect of the novel paradigm itself. This again suggests that task aspects such as additional practice sessions or specific instructions could be modified to reduce these changes between sessions for the long delay stimuli.

The finding that the shorter reaction times are associated with a shorter latency of the P600,.suggests that the P600 is involved in processes that influence reaction time, for example response-selection, executive functioning or decision making. Late positive waves are often associated with response-selection (Friedman, 1990), which is therefore a plausible explanation for the association between reaction time and latency. As compared to the P600, the P300 was very stable over sessions. This corresponds well with previous studies finding high test-retest reliability for P300 amplitude and latency (Hall et al., 2006). The P300, thus, seems to be a stable ERP measure during performance in the DMTS.

From the above, it can be concluded that, responding to short delay stimuli in the DMTS, working memory is stable over sessions, but stability for long delays is absent. Further research could focus on a DMTS with only a long delay, to investigate the idea that this competence improves between two sessions. Changes as compared to the present study should be, e.g., adding an extra practice session, as well as providing specific instructions to all participants. Research could also be done to investigate stability of DMTS-tasks with more different delays. This may explain some of the findings.

Acknowledgement

This research would not have been possible without our supervisor, Anke Sambeth, who gave us the opportunity to analyze behavioral and EEG data.

REFERENCES

Arntzen, E. (2006). Delayed matching to sample: probability of responding in accord with equivalence as a function of different delays. *The Psychological Record, 56*, 135-167. Baddeley, A. (2003). Working memory: looking back and looking forward. N*euroscience, 4*, 829-839.

Bogacz, R., Wagenmakers, E.J., Forstmann, B. U., & Nieuwenhuis, S. (2010). The neural basis of the speed-accuracy tradeoff. *Trends in Neurosciences, 33*, 10-16.

Chamberlain, S.R., Blackwell, A.D., Nathan, P.J., Hammond, G., Robbins, T.W., Hodges, J.R., Michael, A., Semple, J., Bullmore, E.T., Sahakian, B.J., (2011) Differential Cognitive Deterioration in Dementia: A Two Year Longitudinal Study. *Journal of Alzheimer's Disease, 24*, 125-136

Dudchenko, P.A. (2004). An overview of the tasks used to test working memory in rodents. *Neuroscience and Biobehavioral Reviews, 28*, 699-709.

Friedman, D. (1990). Cognitive event-related potential components during continuous recognition memory for pictures. *Psychophysiology, 27*, 136-148.

Gluck, M.A., Mercado, E., & Myers, C.E. (2008). *Learning and memory. From brain to behavior*. New York: Worth Publishers.

Grant, D.S. (1991). Symmetrical and asymmetrical coding of food and no-food samples in delayed matching in pigeons. *Journal of Experimental Psychology, 17*, 186-193.

Hall, M.H., Schulze K., Rijsdijk F., Picchioni M., Ettinger U., Bramon E., Freedman R., Murray R.M. & Sham P.(2006). Heritability and Reliability of P300, P50 and Duration Mismatch Negativity, *Behavior Genetics, 26*, 845-857.

Jasper, H. (1958). The ten-twenty electrode system of the international federation. *Electroencephalography Clinical Neurophysiology*, *10*, 371-375.

Kaestner, E.J. & Polich, J. (2011). Affective recognition memory processing and event- related brain potentials, *Cognitive, Affective & Behavioral Neuroscience, 11*, 186-198.

Klaver, P., Smid, H. G.O.M. & Heinze, H. (1999). Representations in human visual short-term memory: an event-related brain potential study. *Neuroscience letters, 268*, 65-68.

Klein, K., & Fiss, W. H. (1999). The reliability and stability of the Turner and Engle working memory task. Behavior Research Methods, Instruments and computers, 31, 429-432

Luck S.J. (2005). *An Introduction to the Event-Related Potential Technique.* S.J. Luck (ed.). The MIT Press, Cambridge, MA.

Mecklinger, A. (2010). The control of long-term memory: brain systems and cognitive processes. *Neuroscience and Biobehavioral Reviews, 34*, 1055-1065.

Ricker, T. J., & Cowan, N. (2010). Loss of visual working memory within seconds: the combined use of refreshable and non-refreshable features. Journal of Experimental Psychology: Learning, Memory and Cognition. 36, 1355-1368.

Turner D.C., Robbins T.W., Clark L., Aron A.R., Dowson J.H., Sahakian B.J., (2003) Cognitive enhancing effects of modafinil in healthy volunteers, *Psychopharmacology, 165*, 260-269.