DOCUMENT CONTROL SHEET

	ORIGINATOR'S REF.		SECURITY CLASS.
	NLR-TP-2004-104		Unclassified

ORIGINATOR

National Aerospace Laboratory NLR, Amsterdam, The Netherlands

TITLE

Pilot workload prediction: NUmber of Display ElementS (NUDES) as a predictor

PRESENTED AT

HPSAA II, Human Performance, Situation Awareness and Automation Technology Conference, at Daytona Beach, Florida

PERMISSION

AUTHORS	DATE	PP	REF
H.G.M. Bohnen and A.J.C. de Reus	March 2004	14	7

DESCRIPTORS

NUDES
MENTAL WORKLOAD
PILOT WORKLOAD
FIGHTER PILOT
WORKLOAD PREDICTION
TASK STRATEGIES
WORKLOAD BALANCING

ABSTRACT

Visual allocation of attention was used as a predictor of workload of military pilots. Pilots flew simulated missions with varying levels of task demand. They also performed a detection task and a memory task. The detection task involved visual sampling of display elements and responding to target symbols.

Missions that included the detection task were reported to be more effortful, but physiological indicators of mental effort did not confirm this. The number of display elements (7 or 28) had an effect on performance but no clear effect on perceived effort. Memory task and flight performance decreased slightly for the condition with 28 display elements.

The results suggested that pilots developed strategies to maximize mission performance and to avoid performance detriments for individual tasks. Efficient strategies can limit mental workload. It appears that predicting workload has more practical value when pilot task strategies are taken into account.

Nationaal Lucht- en Ruimtevaartlaboratorium

National Aerospace Laboratory NLR

NLR-TP-2004-104

Pilot workload prediction: NUmber of Display ElementS (NUDES) as a predictor

H.G.M. Bohnen and A.J.C. de Reus

Customer:	National Aerospace Laboratory NLR
Working Plan number:	AT.1.F.2
Owner:	National Aerospace Laboratory NLR
Division:	Air Transport
Distribution:	Unlimited
Classification title:	Unclassified
	March 2004

Approved by author:	Approved by project manager:	Approved by project managing department:
		11/3

Summary

Visual allocation of attention was used as a predictor of workload of military pilots. Pilots flew simulated missions with varying levels of task demand. They also performed a detection task and a memory task. The detection task involved visual sampling of display elements and responding to target symbols.

Missions that included the detection task were reported to be more effortful, but physiological indicators of mental effort did not confirm this. The number of display elements (7 or 28) had an effect on performance but no clear effect on perceived effort. Memory task and flight performance decreased slightly for the condition with 28 display elements.

The results suggested that pilots developed strategies to maximize mission performance and to avoid performance detriments for individual tasks. Efficient strategies can limit mental workload. It appears that predicting workload has more practical value when pilot task strategies are taken into account.

Contents

(14 pages in total)

1 Introduction

The ability to predict workload is an important step towards the optimal tuning of tasks to the capabilities of the military pilot, leading to better system performance and safety. The visual allocation of attention is one of the most important indicators of mental workload (e.g., see Moray, 1986). Therefore, an explorative study is performed towards the amount of information in the cockpit to be sampled as a predictor of workload.

In this study, the NUmber of Display ElementS (NUDES) defines the amount of information associated with a task. The NUDES on a cockpit display are manipulated in order to determine its effect on workload. For this purpose, a target Detection Task is used which requires the sampling of a number of display elements or symbols. The hypothesis is that pilots experience a higher workload in situations where more symbols need to be sampled. General task demands are not only manipulated by varying the NUDES, but also by the complexity of the flying task itself and by the addition of an auditory Continuous Memory Task (e.g., see Jorna, 1989). This way, the effect of NUDES is tested under different task load conditions.

2 Method

2.1 Participants

Eight male subjects participated in the study. All subjects had a background as former F-16 pilot.

2.2 Apparatus

The study was performed at the National Aerospace Laboratory in Amsterdam, The Netherlands. A fighter mock-up with a simulated out-of-the-window view was used for the piloting task. The Memory Task was presented to the subject by means of a miniature headphone. A switch on the throttle (left hand) was used as response button. The Detection Task was presented on a head-down display in the mock-up. A switch on the stick (right hand) was used as response button.

2.3 Tasks

Flight conditions had a varying level of mental task demand. The first part of the mission (see Figure 1) consisted of six rate-one turns separated by straight legs. These rate-one turns were considered as simple flight maneuvers. The latter part of the mission consisted of a sequence of complex flight maneuvers.

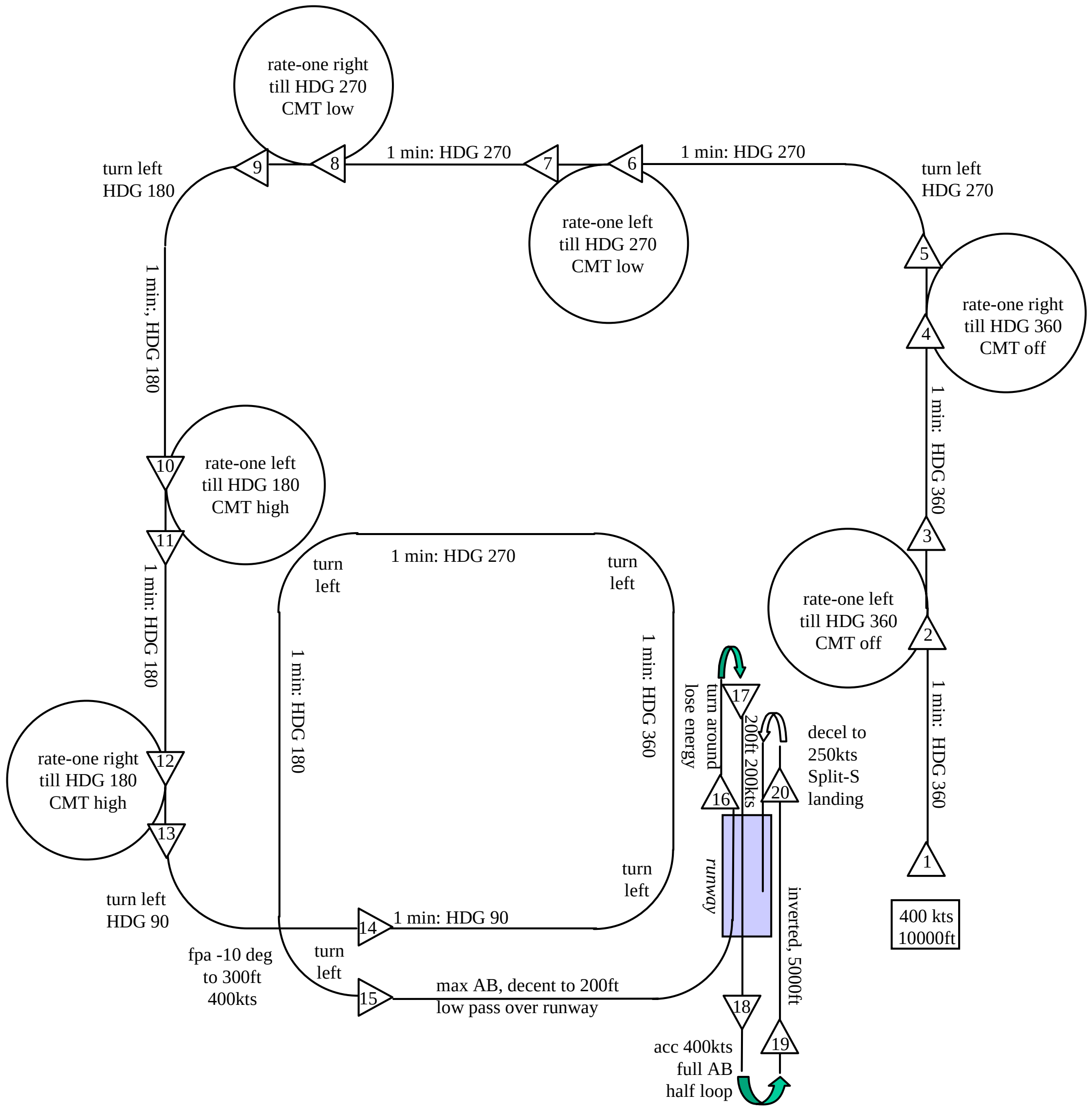


Fig. 1 Workload prediction assessment mission

The symbols (7 or 28 NUDES) for the Detection Task were presented on a fixed position on the screen. Four types of non-target symbols existed and one target symbol (Fig. 2). A symbol changed from non-target to target at random time intervals and at random positions. Only one symbol was the target symbol at any instance. Every 1.5 s one non-target symbol was replaced by another non-target. The first target symbol occurred within 15 s after the beginning of the task. The next target symbol was presented within 15 to 30 s (3 s steps) after the appearance of the previous one. The average interval was 22.5 s. The subjects had to detect the target symbols and to respond as soon as possible. The detected target was then reset to a non-target symbol. Target symbols that were not detected by the subjects automatically became non-target 15 s after their appearance.

To perform the Memory Task subjects had to remember two or four target letters. Thereafter, a series of letters was presented via the headset and subjects had to react to the target letters. Subjects also had to count how often they heard each target letter. After hearing a target letter three times, a specific response was required. The counting for that target letter then started again.

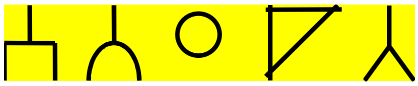


Fig. 2 From left to right: the four non-target symbols and the target symbol

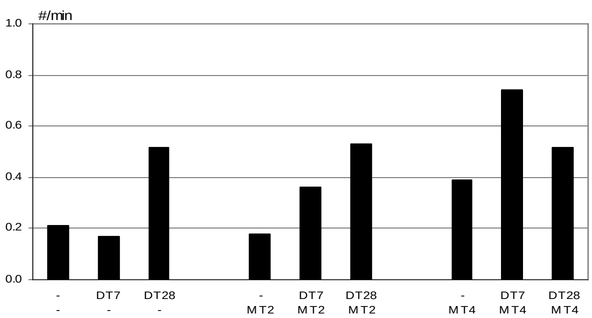
2.4 Procedure

A within-subjects design was used with NUDES (0/7/28) and Memory Task loading (0/2/4) as factors. The order of presentation of NUDES was balanced as far as possible over subjects, while the order of the Memory Task was fixed: 0, 2, and 4. Rate-ones always preceded the complex flight maneuvers. In order to assess mental effort, heart rate data was collected using Vitaport. Visual sampling data were collected using a Gazetracker eye/head tracking system. All data were analyzed by an Analysis of Variance with the aid of Statistica.

3 Results

3.1 Flight task performance

During rate-one maneuvering pilots had to keep their aircraft within 50 ft of the target altitude. Both the number of times and the percentage of time pilots exceeded this limit were analyzed (Fig. 3 and 4). Overall, especially the addition of the Memory Task (MT) increased the number and percentage of exceedings, $F(2,12)=5.52, p=.020$ respectively $F(2,12)=3.53, p=.062$. The addition of the Detection Task (DT) had a less prominent effect. A significant interaction effect between the two tasks (MT and DT) was found: in the highest workload condition (DT 28, MT 4) both measures decreased (exceeding number: $F(4,24)=2.51$, $p=.068$; exceeding percentage: $F(4,24)=4.05$, $p=.012$). This suggests a possible strategy shift.



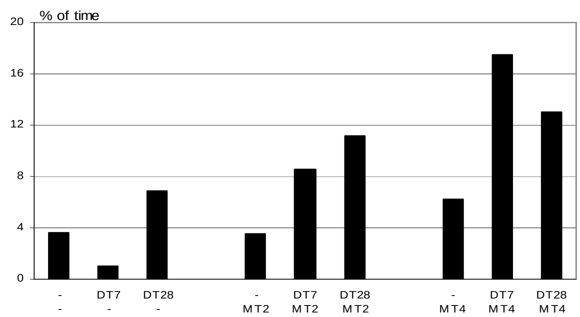


Fig 3 Average number of times the altitude limit was exceeded during rate-ones. DT = Detection Task (7/28 NUDES), MT = Memory Task (2/4 target letters).

Fig. 4 Percentage of time the altitude limit was exceeded during rate-ones.

The complex flight maneuvers involved more freedom for the pilots. Consequently flight task performance data for this mission part were not analyzed.

3.2 Detection Task performance

The percentages of correct responses (hits) for the DT and the reaction times related to those hits are displayed in Fig. 5 and 6. False alarms occurred very rarely. Looking at rate-ones only, no effect was found for the DT load (i.e. NUDES) on the percentage of hits, but the reaction times seemed to increase $F(1,6)=4.22, \underline{p}=.086$. The addition of the MT had a minor influence on the percentage of hits $F(2,12)=2.67, \underline{p}=.110$. The effect of NUDES on reaction time seemed reverse in the highest workload condition (DT 28, MT 4), though this did not reach significance.

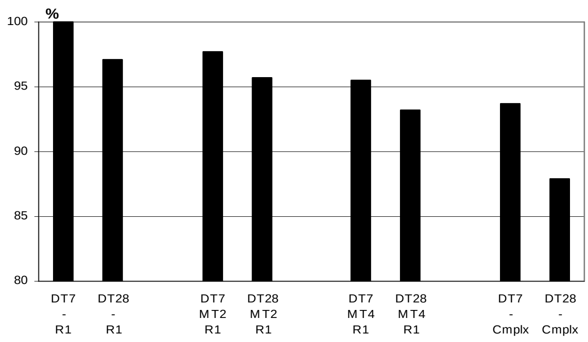


Fig. 5 Percentage of hits in the DT (R1= rate ones; cmplx = complex maneuvering).

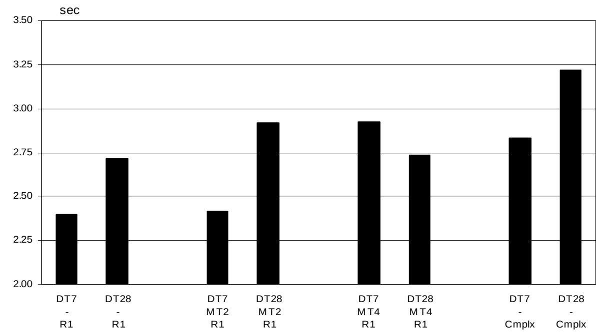


Fig. 6 Reaction times to targets in the DT.

During the complex maneuvers, trends of the NUDES were found on the percentage of hits, $F(1,6)=4.75$,$p=.072$. Comparing the complex maneuvers with the rate-ones without a MT produced a clear difference in the percentage of hits and the reaction time, $F(1,6)=9.18$,$p=.023$ respectively $F(1,6)=5.83$,$p=.052$. In other words, performance on the DT was affected by the complexity of the flight maneuvers.

4 Memory Task performance

There were no false alarms during the Memory Task. In the MT with 2 target letters, hardly any mistakes were made with target letter detection (Fig. 7 and 8). Only when pilots had to remember and count 4 target letters, they started to make more mistakes $F(1,6)=3.98$,$p=.093$, though the largest effect was on reaction times $F(1,6)=12.02$,$p=.013$. The DT had a negative influence on reaction times, $F(2,12)=10.64$,$p=.002$. This effect was not significantly different for the two levels (7, 28) of the DT.

In the target letter counting subtask, a clear effect of the number of target letters was found on the percentage of hits and on the reaction times, $F(1,6)=8.78$,$p=.025$ respectively $F(1,6)=6.21$,$p=.047$. Also much more false alarms were produced with 4 target letters. The effect of the DT was not significant.

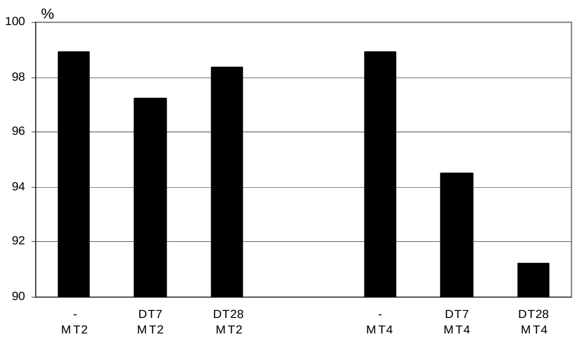


Fig. 7 Percentage of hits of the MT.

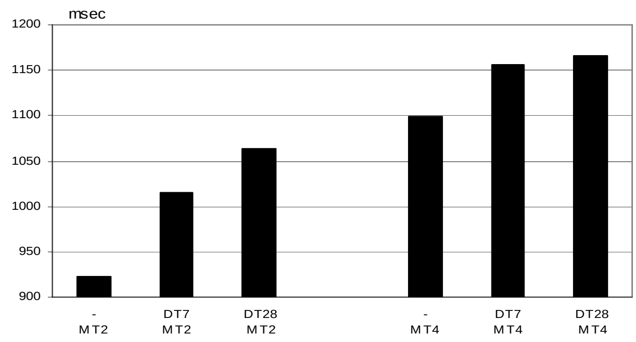


Fig. 8 Reaction times to target letters in the MT.

4.1 Mental effort

The addition of the MT had a clear effect on heart rate during the rate-ones $F(2,12)=40.40, p<.001$ (Fig. 9). No selective effect of the DT was found on heart rate. When the complex maneuvers were compared to the rate-ones without the MT, a significant effect of flight task complexity was found $F(1,6)=28.18, p=.002$. No effect was found on heart rate variability.

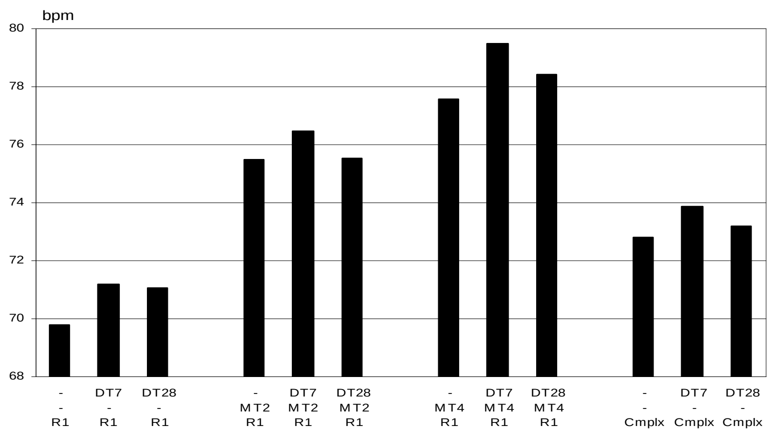


Fig. 9 Heart rate

After each mission subjects used the Rating Scale Mental Effort (Zijlstra, 1993). Without the DT the mission is perceived as nearby “pretty effortful” (rating=70) while with the DT up to “very, very effortful” (rating=90/100), $F(2,12)=6.12, p=.015$. However, pilots did not perceive the DT with 28 NUDES as more effortful than with 7 NUDES. This may be because the difference is “drowned” in the perception of the mission as a whole.

4.2 Visual scanning

Looking at rate-ones only, a trend of the NUDES is found on the dwell time on the display where the DT was presented (Fig. 10). No effect of the MT was found, but an interaction with the detection task is suspected. Especially the DT with 28 NUDES seems to suffer from the addition of the MT, $F(2,12)=9.98$, $\underline{p}=.003$. The time less spent at the DT display is spent on (the center of) the Head-Up-Display (HUD), interaction MT x DT: $F(2,12)=3.38$, $\underline{p}=.068$. During complex maneuvering, more time is spent on the DT display as a function of the NUDES and less on the HUD, $F(1,6)=47.79$, $\underline{p}<.001$ respectively $F(1,6)=12.42$, $\underline{p}=.012$.

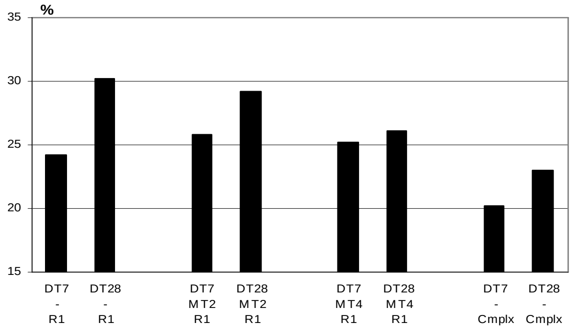


Fig. 10 Dwell time (%) on the DT display.

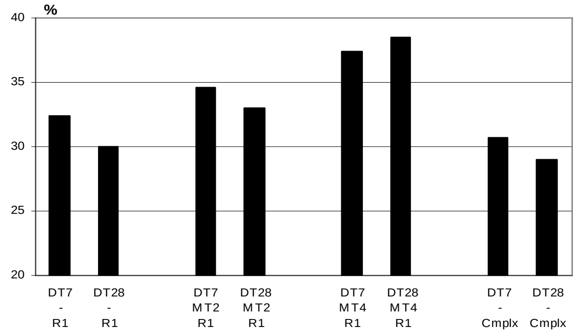


Fig. 11 Dwell time (%) on the HUD.

The NUDES greatly affected the mean fixation duration on the DT display: with 7 elements this was about 0.45 s, while with 28 elements this was about 0.60 s (rate-ones: $F(1,6)=64.17$, $\underline{p}<.001$, complex maneuvering: $F(1,6)=36.00$, $\underline{p}=.001$).

With 7 NUDES, pilots waited about 4 s before they re-fixated at the DT display. With 28 NUDES, this difference was about 5 s (rate-ones: $F(1,6)=6.29$, $\underline{p}=.046$; complex maneuvering: $F(1,6)=7.42$, $\underline{p}=.034$).

5 Discussion

The results show that a mission that includes the Detection Task (DT) costs more mental effort. However, no difference was reported in the perceived effort between the two amounts of information to be sampled in the DT, that is 7 or 28 NUDES. This result was confirmed by the objective mental effort measures, heart rate and heart rate variability. This means that just adding the DT had an effect, regardless the amount of NUDES to be processed in that task. Subjects reacted more slowly to target symbols in case of 28 NUDES instead of 7, and they also missed more target symbols. The addition of the DT slightly affected flight performance. These changes in performance can indicate that subjects were not able to maintain performance because of limited resources, interference in visual processing, or changes in attention allocation. The latter suggests a change in task strategy to actively manage time and resources to accomplish the complete task (i.e., flying the mission and detecting critical symbol, accurately and on time). Strategy shifts can be seen as an index of increasing mental load, because pilots may change their strategy to keep workload at an acceptable level (e.g., see Hart, 1991).

Reaction time to the target symbols in the DT only increased for the high load condition (4 target letters) of the Memory Task (MT). The MT performance itself is slightly affected by the number of symbols in the DT. Flight performance also seems to decrease, although this effect is absent in the MT condition with a high load. This again might reflect a strategy-shift: performance on the MT is ‘sacrificed’ in order to maintain a minimal performance level on the other tasks.

Subjects rated the mission as somewhat above ‘pretty effortful’ while with the DT added to the mission it is perceived nearby ‘very, very effortful’. The mission as a whole, which includes the MT and DT, can thus be described as highly demanding. No effects of the addition of the DT are found in the objective mental effort parameters. This can be caused by a ceiling-effect: the mental effort in the least demanding task situation, in which subjects had to fly and to perform the MT together, is already so high that an increase in resource allocation and hence mental effort is almost not possible anymore.

Effects on heart rate are only found when comparing rate-ones without a MT with complex maneuvers, which suggest an effect of flight maneuver complexity, and for the addition of the MT. As said before, adding the DT results in an increase in perceived effort and hence it is legitimate to state that those subjects had the willingness to spend resources in order to cope with the demands of the combined tasks. However, subjects were unable to cope with those demands, as was evidenced by a decreased performance on for example the MT. This can be another reason for not finding objective mental effort results (Aasman, Mulder, & Mulder, 1987).

The performance on the DT was lower during complex maneuvering than during rate-ones, regardless of the presence of the MT. During complex maneuvers pilots paid slightly less attention to the information presented on the DT display, but paid more attention to the outside world as flying the aircraft always gets the highest priority. Therefore it is not surprising that more targets are missed. The scanning measure that is most influenced by the number of display elements is the mean fixation duration on the DT display, leading to a longer total dwell time. However, the display is less often frequented, with more time between stares when more display elements are present. Furthermore, attention allocation policy can be affected by the NUDES, and has become less efficient, taking into account the performance decrease. Most interesting is that performing the detection task and the MT while flying rate-ones together results in larger dwell times at the central area of the Head-Up-Display (HUD). This can be seen as a simplification of the subject's sampling behavior in situations with a relatively high cognitive load. This decrease is in line with results of Spady and Harris (1983) who found that experienced pilots change scanning behavior in case of an increase in mental load.

Two amounts of visual information had to be processed in the DT (7/28 NUDES). The mental effort parameters did not show any difference for the two different amounts of visual load in this task. May, Kennedy, Williams, Durilap, and Brannan (1990) combined a simple visual task with a task in which subjects had to count a different amount of tones. Their results showed an effect of difficulty of counting on the scanning parameter, which is saccadic extent, used as an indication for mental workload. This suggests a mutual influence of a purely visual and a purely auditory task, both requiring minimal mental effort. In this experiment a slight effect of the amount of cognitive load in an auditory presented memory task (MT) on the duration and the time between stares for the target detection task is found. The fact that both tasks do influence each other can indicate that both tasks use the same cognitive resources that are limited. But considering the results of May, et al., (1990) this effect can also be caused by the fact that the MT load influences the subject's sampling behavior characteristics. This complicates the straight-forwardness of the working hypothesis used for the workload prediction as the level of cognitive load affects a subject's sampling behavior. Hence the assumed clear-cut relation between the predictor (NUDES) and the outcome variable (workload) might not exist.

6 Conclusion

How well do NUDES predict pilot workload? In this study NUDES clearly affect pilot performance measures, though the physiological measures of effort do not seem to be affected. In order to avoid too large performance decrements pilots develop strategies that serve to maximize the overall performance. As said before, research does suggest that a pilot's strategy selection can change workload, with efficient strategies leading to more balanced mental effort and performance. This study indicates that predicting workload by determining visual allocation of attention alone has less practical value than the capability to assess both task strategies (changes) and workload predictors together.

This study was specifically aimed at the effects of the number of display elements on workload disregarding the cognitive processes required after the perception of those display elements. The frequency of occurrence of critical symbols per time period was kept constant, and the number of cockpit displays itself is also not considered in this study. Predicting workload for actual

mission segments, which require multiple task performance and hence multiple display visual sampling will result in more realistic predictor values. The information needed for those tasks is perceptually more complex and meaningful, requires higher order mental processing, and the task load varies over time in a natural way.
To determine the relation between the visual allocation of attention and workload in more detail, more workload scores must be set, preferably in realistic task situations. Definitely needed is a way to uncover pilot task strategies used for balancing workload and performance.

7 References

Aasman, J., Mulder, G., Mulder, L.J.M. (1987). Operator effort and the measurement of heart-rate variability. Human Factors, 29(2), 161-170.

Hart, S. (1991). Pilots' workload coping strategies. AIAA/NASA/FAAMFS Conference on Challenges in Aviation Human Factors: The National Plan Tysons Corner, VA.

Jorna, P.G.A.M. (1989). Prediction of success in flight training by single-and dual-task performance. In: AGARD proceedings on "Human behaviour in high stress situations in aerospace operations". CP-458, Brussels.

May, J.G., Kennedy, R.S., Williams, M.C., Durilap, W.P., and Brannan, J.R. (1990). Eye movement indices of mental workload. Acta Psychologica, 75, 75-89.

Moray, N. (1986). Monitoring behaviour and supervisory control. In K.R. Boff, L. Katitman, and LP. Thomas (Eds.), Handbook of Perception and Human Performance. Vol II. Cognitive Processing and Performance (Chapter 40). New York: Wiley.

Spady, A.A., and Harris, R.L. jr and Sr. (1983). Summary of NASA Langley's Pilot Scanning Research. Second Aerospace Behavioral Engineering Technology Conference Proceedings, Society of Automotive Engineers.

Zijlstra, F.R.H. (1993). Efficiency of work behaviour. A design approach for modern tools. PhD thesis, Delft University of Technology, The Netherlands: Delft University Press.