A new method to assess pilots' allocation of visual attention using a head-up display

Axel Hillebrand[1], Eva Wahrenberg[1], & Dietrich Manzey[2]
[1]EADS Innovation Works, Hamburg
[2]Berlin Institute of Technology, Berlin, Germany

Abstract

Objectives of the study were the development and validation of a new assessment method for the allocation of visual attention in using head-up displays (HUD). Based on a secondary-task approach, this method specifically allows an evaluation of possible consequences of HUD on pilots' visual behaviour and thus can help to identify inappropriate HUD designs that may lead to cognitive tunnelling. Background: There are studies which use the detection of runway incursions to evaluate the impact of HUD design on pilots' visual attention. This method, however, requires a large number of participants and focusses solely on the final approach. Based on a secondary target-detection task with targets presented in different locations of the field of view, the new method reduces the effort to be made for human factors investigations in this field of research. Furthermore, it allows the validation of pilots' visual attention allocation in all flight phases. Method: Twelve experienced pilots flew six approaches while concurrently performing a target detection task. The study was conducted in a low-fidelity simulation environment and involved two independent variables: HUD design and location of the targets within the field of view. Primary-task performance (e.g. deviation from flight path) as well as performance in the target detection task (detection rate, response times) were evaluated. Data of detection rate, response time and deviation from flight path were collected as well as of subjective workload. Results: The method is capable to reproduce well-known effects of HUD design on the allocation of attention. A significant better detection rate for centrally localised targets is achieved and provides evidence for attentional tunnelling effects.

Theoretical background

Head-up displays (HUDs) provide primary flight, navigation, and guidance information in pilots' forward fields of view. The presentation of essential information on a transparent screen mounted in front of the windshield has many advantages concerning the efficiency and safety of aircraft operation. HUDs can make landings possible in conditions of low visibility, and thus disruptions of landings and take-offs can be reduced by up to 60% (Proctor, 1997). Furthermore, there are HUD concepts, like the T-NASA (Foyle et al., 1996), which increase the

In D. de Waard, K. Brookhuis, F. Dehais, C. Weikert, S. Röttger, D. Manzey, S. Biede, F. Reuzeau, and P. Terrier (Eds.) (2012). Proceedings HFES Europe Chapter Conference Toulouse. Available from http://hfes-europe.org

safety of taxi way operations during low visibility. The presentation of information on the transparent display reduces the amount of head-down times and enables the pilot to keep his or her eyes in the far domain. In general, the use of head-up displays reduces the amount of attention switching between the instrumentation and the far domain and thus can lead to a higher degree of situation awareness.

But despite the obvious performance benefits, HUD design can also lead to inappropriate scan behaviour of the pilot. One effect that has been described in literature involves the tendency to keep attention in solely one domain. Especially if a HUD design is complex, the pilot may tend to focus on information in the near domain while missing safety-critical events in the far domain (i.e. the outside environment). This phenomenon is called cognitive tunnelling and needs to be taken seriously because it can offset the benefits of HUDs by introducing new risks. (Wickens & Long, 1995; Wickens, Ververs & Fadden, 2004). Another effect which may come along in the development of HUD design is information cluttering. If too much information is presented in too small an area of a display, the pilot's information processing can be disturbed. An inappropriate degree of information density can restrict the pilot's ability to differentiate between relevant and irrelevant information and can even overlay necessary information in the far domain (Martin-Emerson & Wickens, 1997). Such clutter effects have been observed in several studies (for example Fadden, May Ververs, & Wickens, 2001; Fischer, Haines, & Price, 1980; Wickens & Long, 1995).

These cognitive-psychological phenomena have to be taken into account in the development process of HUD design. Specifically, the information presented in a HUD should have an appropriate degree of complexity regarding quantity and quality which provide the best compromise to achieve the wanted performance benefits by at the same time minimizing the risk of attentional tunnelling effects. As a consequence assessment methods are needed that can be used to evaluate effects of different HUD designs on attention allocation and to identify risks arising from possible attention tunnelling effects. In addition, such assessment methods preferably should enable the designers to evaluate these effects as early during the HUD development process as possible, e.g. based on low-fidelity flight simulations.

There are several methods that have been used to assess the allocation of visual attention in use of HUD. Often, the ability of HUD design to avoid restricted visual attention allocation is operationalised by the detection rate of runway incursions. Fischer et al. (1980) for example use this method and find out that aircrafts crossing the runway unexpectedly were less detected when using a HUD. Wickens (2005) points out that this methodical procedure is challenged by certain statistical issues. By definition a surprising event cannot be presented to the participants more than once or two times in a single experimental session. Otherwise it would not be surprising any more. If the budget of a study has to be kept constant, the sample size is necessarily limited. Consequently, the variability and thus the statistical power decrease. A researcher and his or her management have to accept a greater likelihood of the error of the second kind by raising the alpha level. In addition, the participant taking part in an experiment including runway incursions will be familiar with this kind of experimental setup and cannot be recruited for a follow-up experiment, and this occurs, of all things, in a group that is very small anyway.

An alternative approach was chosen by Alexander et al. (2005). Instead of runway incursions they used artificial "blimps" that were presented within the field of view as representations of possible airborne hazards. The extent these blimps would be detected by pilots using a HUD while flying in a high-fidelity simulator was analysed. Half of the blimps appeared in the middle of the display and half of them on its periphery. The pilots had to react to the appearance of the blimps by pressing a key on the yoke as fast as possible. Each flight involved three blimps. Visual attention allocation was operationalised by the mean blimp detection time. Although having obvious advantages over the approaches mentioned above, the number of targets presented was still small, thus leading to the same statistical problems as in the use of simulated runway incursions.

The objective of the study presented here is to describe the development of a new assessment tool that enables designers and researchers to investigate the influence of HUD design on human performance and particularly on visual attention behaviour in early stages of its development and through the use of low-fidelity simulations. The tool proposed capitalises on the approach used by Alexander et al. (2005).

Similar to their work target stimuli presented at different locations in the field of view are used to probe the attention allocation of pilots using HUDs. These targets are artificial and are faded in for half a second. Pilots' secondary task is to react to the target stimuli by pressing a key as fast as possible. The dependent variables are the response time and the detection rate. Thus, the idea underlying the new method is an analogy to the secondary task paradigm. While the latter is used for the assessment of workload, the new method consists of a secondary task to assess the allocation of visual attention. Moreover, a much higher frequency of target presentation can be achieved and thus a better statistical power of analysis can be provided. In addition, the spatial resolution of locations where targets can appear is increased considerably. This allows for much more detailed and sophisticated analyses of the influences that different HUD designs have on human performance including both influences on attention allocation between near and far domain and influences on attention allocation within the near domain, i.e. the scanning of the information displayed in the HUD.

In the following chapter the basic paradigm used for the proposed attention-allocation assessment tool (given the name ATTENDO which is Latin for I pay attention) is described, followed by the description of a first experimental study. The main objective of this study was to test the suitability of ATTENDO for the intended purpose. Effects of two different HUD designs on ATTENDO performance were investigated during an approach and landing task. The main research question was whether or not ATTENDO would be able to identify the typical effects of attentional tunnelling which represent frequently reported issues in the use of HUDs (Wickens, May Ververs & Fadden, 2004).

Beyond the suitability test the statistical benefit of ATTENDO should be shown by comparing it to another established method. For this purpose, a runway incursion was included in the experiment. It occurred in one of the flights shortly before landing and that had to be detected by the pilots participating and responded to appropriately.

Paradigm and hypotheses

With respect to its basic principle the approach used for ATTENDO resembles the secondary-task paradigm known from workload research. Yet, instead of workload, the approach presented here is used to investigate the spatial distribution of (visual) attention by presenting stimuli at different locations in the far and near domains of a simulated HUD.

The stimuli used have the form of squares. They are distributed systematically within the HUD as well as on its periphery and are presented to the pilots one after another. Each target is faded in for half a second and they are randomised with regard to appearance time and location. The intervals between the targets vary from 7 to 15 seconds. The pilots' task is to react to the targets by pressing a key as fast as possible while controlling the aircraft. The placement and the number of the targets can be chosen according to the characteristics of the HUD system to be evaluated and of the simulation. Figure 1 shows the design of the assessment tool for visual attention allocation used in this study. While during an experiment the targets are presented one after another, the figure shows all the targets at the same time as well as their systematic arrangement. There are three areas: 1) the centre of the HUD, 2) its edges and 3) its periphery. This arrangement allows for the identification of display areas less observed by the pilots and for the accomplishment of statistical analyses. The allocation of visual attention is operationalised by the target detection rate and the detection time.

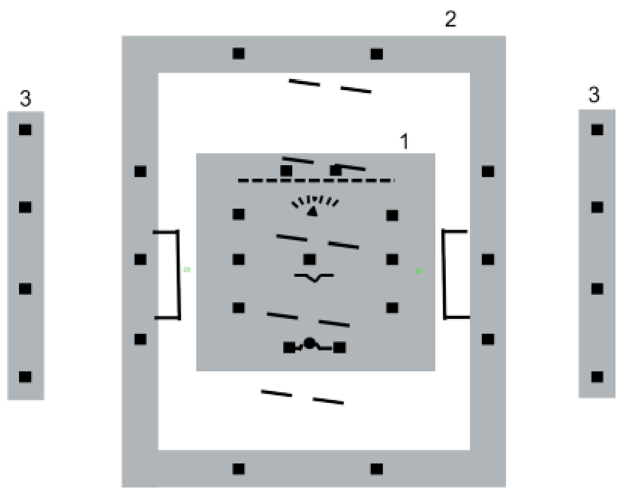

Figure 1. Targets of the assessment tool, HUD symbology included. To get a better overview the three areas of the systematic arrangement are marked in grey colour.

ATTENDO is purposed to the comparison of different HUD variants concerning their tendency to capture visual attention in the centre and should verify the influence of HUD design on cognitive tunnelling (see chapter 2). It has to be evaluated to which extent the new method is able to reproduce these aspects relevant to visual attention using a HUD. To test the tool’s sensitivity the hypotheses H1 und H2 are formulated.

Hypothesis H1:

HUDs lead to concentrating visual attention on central instrument features
$H1_A$: The more centrally targets are located, the higher the detection rate is.
$H1_B$: The more centrally targets are located, the faster the response time is.

Hypothesis H2:

The risk of an inappropriate allocation of visual attention allocation increases with the complexity of the HUD.
$H2_A$: The more complex a HUD is, the lower is the detection rate for peripheral targets.
$H2_B$: The more complex a HUD is, the slower is the response time for peripheral targets.

Method

Participants

In general, pilots are much more experienced in eye-scanning patterns and precise landings than novices (Kasarskis et al., 2001). In order to exclude influences of different scanning behaviour and landing skills, only trained pilots were chosen as participants. Twelve male pilots with instrument flight experience participated in the study. All of them were certified at the Private Pilot level or higher. More precisely, three were certified at the Commercial Pilot level or higher and eight of them were certified at the Airline Transport Pilot level. In addition, three of the participants were also certified Flight Instructors and nine held the Instrument Airplane rating. HUD experience, though on a low level, was stated by two pilots. The participants were between 30 and 62 years old (SD = 9.03) and their flying times ranged from 280 to 20.000 hours. With regard to these parameters the participants were equally distributed over the different experimental conditions.

Task

The main task of the participants was to fly a total of six approaches including landing on the runway in a low-fidelity flight simulator. Primary flight display information was presented on a simulated HUD on the computer screen. Pilots were explicitly instructed to focus on this flying task as their primary task and to try to remain as close as possible to the ideal approach path. The pilots were requested to avoid unsafe situations. The participants were also told to give the flaps control (“flap 1”, “flap 2” and “flap 3”) and the landing gear control to the experimenter like they would do in a real flight situation to the co-pilot.

In addition to the primary-task the pilots had to perform the target detection task. Each target was presented for half a second at randomly varying intervals of 7 to 15 seconds. The detection task included twenty-nine targets (see figure 1) to react to by pressing a key which was placed on the top of the flight stick. The detection task was explicitly introduced as a secondary-task, i.e. a task of lower priority.

Whereas the first five trials included different approach patterns without any critical situations, an unexpected and surprising runway incursion (RI) occurred during the last trial. This RI consisted of an airplane crossing the runway approximately 45 seconds before landing and required the initiation of a go-around manoeuvre. Visibility conditions were always chosen in a way that it was principally possible to detect the occurring RI in due time.

Material

The experiment was accomplished on a Fujitsu Siemens Computer with 4 GB RAM, a 3.33 GHz dual-core processor and a 64-bit system with Windows 7 as operating system. The monitor (Fujitsu Siemens) with a resolution of 1920:1200 was placed at a distance of 60 cm from the pilot. The flight simulation tool 'x-plane 9® 2008 Laminar Research' was presented in a low-fidelity environment on the monitor. 'x-plane®' was controlled by means of the flight stick-throttle combination 'Thrustmaster HOTAS Cougar'. All the pilots had to do was to navigate the aircraft with the joystick and to control thrust with the throttle.

The HUD variants were created with the help of the 'x-plane®' tool 'Plane Maker®' which allows for a customisation of every cockpit and aircraft element. The reduced HUD (figure 2) consisted of the rudimentary symbology a pilot needs for landing an aircraft. The complex HUD (figure 3) is extended by a flight director, presented by a cross and showing the required trajectory, and by a flight path vector which is an indicator for the future position of the aircraft. Moreover, a tunnel-in-the-sky was included in the complex HUD that ensures flight path accuracy because pilots only have to follow the row of rectangles marking the tunnel. The functionality and usability of the HUD variants were confirmed by pilots who are members of the Vereinigung Cockpit e.V. Figures 2 and 3 are screenshots of 'x-plane®' and were produced in nightly scenery to improve the visibility of the HUDs. The experiment was conducted in cloudy, but bright scenery. The variation of the HUD was chosen in order to have a higher amount of information elements within the HUD and not so much to have a higher degree of usability. Thus a higher tendency of the complex HUD variant to evoke a concentration of visual attention in the centre could be achieved.

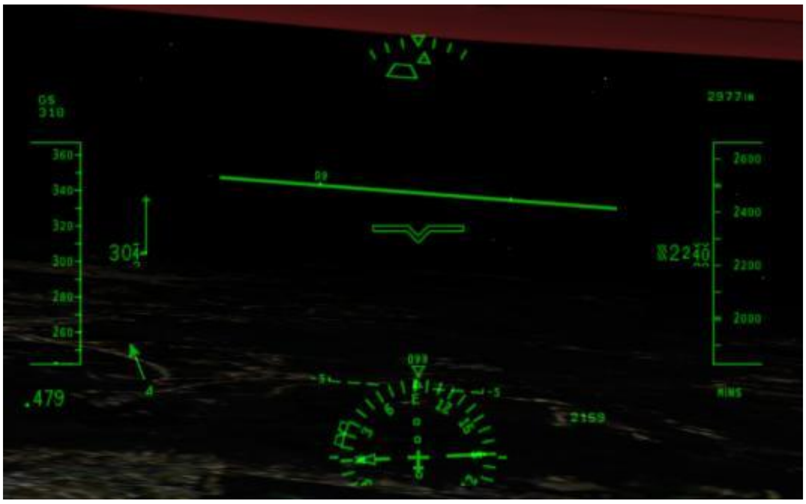

Figure 2. Reduced HUD

Figure 3. Complex HUD

The detection task was integrated into 'x-plane®' by means of a plug-in implemented with C++ and Open GL. Therefore, the 'Software Development Kit' of 'x-plane®' was used for a framework for the development of the plug-in.

Procedure

The experiment took place at three locations from July to August 2011: (1) at the Technical University of Berlin, (2) at EADS Innovation Works in Hamburg and (3) at the Headquarters of the Vereinigung Cockpit e.V. in Frankfurt which is the German airline pilots' association.

One experimental session consisted of six approaches and lasted approximately one hour. The experimental session began with a training period during which the pilots could make themselves familiar with the symbology sets and the aerodynamics of the aircraft as well as with the input devices. Afterwards the experimental period started with the first of a total of six approaches on Hamburg-Finkenwerder of which two were direct approaches and four curved approaches (two from the left, two from the right). All of the approaches started at an intercept point which was ten miles (16.1 km) away from the runway and at an altitude of 3000 ft. (914,4 m). They were randomised between the participants for every HUD condition to avoid influence of the flight task order. Wind speed and direction also varied during the six flight tasks. The approach was stopped by the instructor as soon as the aircraft was in safe mode on the runway so that no parking manoeuvre was needed.

The experimenter was also in the room and sat at a distance of two meters from the participant. In order not to distract the participant the experimenter faced away from the monitor and the participant. The experimenter operated the flaps and landing gear control at the command of the pilot.

After each of the six trials the subjective workload was assessed by filling out the NASA-TLX questionnaire (Hart & Stavelan, 1988). Finally, a questionnaire was filled out to collect demographic data and information about the pilot's flight and HUD experience. Pilots were not paid for participating in this study.

Experimental Design

A 2x3 experimental design was used for the study. The first factor involved to variants of HUD design, i.e. reduced vs. complex HUD, and was implemented as a between-subjects factor. The second factor "target location" was a within-subjects factor and consisted of three levels (see figure 1): (1) targets in the centre of the HUD, (2) targets on the edges of the HUD, (3) targets on the periphery of the HUD (3).

Three sorts of data were collected: (1) Primary task performance was assessed by calculating the absolute orthogonal horizontal and absolute orthogonal vertical deviation from an ideal approach path which can be found in official Instrument Approach Charts (IAC), (2) the allocation of visual attention was operationalised by the detection rate and the response times to detected targets, and (3) subjective workload was assessed by the NASA-TLX questionnaire (Hart & Stavelan, 1988). In addition it was assessed how many pilots correctly responded to the runway incursion.

Results

Primary task performance

In a first analysis the comparability of the six different approach and landing tasks was verified. If the flight tasks are at a comparable level of difficulty it is possible to average the results of the pilot's six flights. Therefore, the data of the six flights of each pilot could be averaged for further analyses. To this end, the lateral and vertical deviation was compared between the six different tasks. An ANOVA showed that there are no significant differences between the lateral deviation, $F(5,55) = 2,12$; $p > .07$, and the vertical deviation, $F(5,55) = 2,13$; $p > .07$.

In order to analyse whether the two different HUD versions had an impact on primary task performance (i.e. absolute lateral and vertical deviation compared to an ideal flight path) an independent sample t-test was performed as well as subjectively assessed workload (NASA-TLX), detection rate and response time. There were no significant differences between the two HUD versions with respect to both measures, (the lateral deviation: $t(10) = .579$, $p > .6$; vertical deviation: $t(10) = 1.029$, $p > .4$). Furthermore, also the subjectively perceived overall workload as assessed by the NASA TLX did not differ between the two HUD conditions, $t(10) = .316$, $p > .7$, neither did any of the six sub dimensions.

ATTENDO performance

Since the allocation of visual attention was operationally defined by the target detection rate, its analysis was of primary interest. Effects on the detection rate were analysed by a 2 (HUD design) x 3 (target location) ANOVA which was applied across the 12 participants.

There was a significant main effect of the target location on the detection rate, $F(1,345, 13,45) = 16.557$; $p < .01$. Yet, the anticipated effect of complexity of HUD, design on detection rate, $F(1,345, 13,45) = 0.08$, $p > .85$, was not found, neither was a significant interaction effect found, $F(1,10) = 0.54$, $p > .82$. Figure 4 shows the differences of target location related to detection rates per HUD variant. The differences between the mean detection rates were considerable, showing that centrally located targets were detected with much higher probability than targets presented in the two other areas. Most noticeable in this respect is that no differences occurred between detection rates for targets presented close to the HUD edges, where other relevant flight data information was presented, and targets presented clearly peripherally, i.e. outside the area where information was presented.

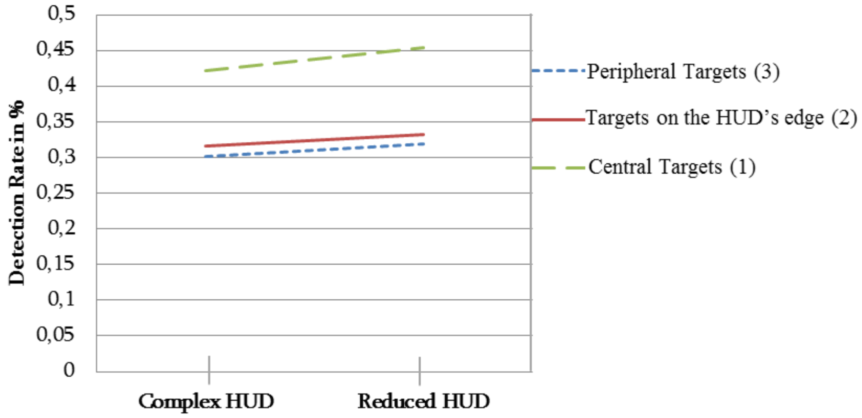

Figure 4. Detection rate per target location and HUD version

Since participants were told to react as fast as possible to the appearing targets, response time was assumed to be an important indicator for how attention is allocated and captured. Effects on response time were analysed by a 2 (HUD design) x 3 (target location) ANOVA, too. However, no significant differences were found in response times for any factor. Neither the HUD design, $F(1,10) = .129$; $p < .80$, nor the target location, $F(2,20) = 1.943$; $p < .20$, nor the interaction of 'HUD design x target location', $F(2,20) = .160$; $p < .90$, showed significant effects on the response time.

Detection of runway incursion

Only one of the twelve pilots did not react to the runway incursion. Instead of initiating a go-around this pilot landed on the blocked runway. Although this low incidence rate does not allow for any decisive conclusions it might be noticeable that the pilot flew with the more complex HUD variant. None of the pilots had participated in other flight simulation studies into which runway incursions were integrated. Therefore all of them could be considered as naive concerning this factor.

Discussion

This study is aimed at testing the sensitivity of a new method to assess the visual attention allocation of pilots using a HUD. The method consists of targets that are faded in for half a second and randomly regarding time and space and that are systematically grouped in the centre and on the periphery of the HUD. Two HUD versions which differ in their complexity are compared to each other.

The results clearly show that targets presented in the centre of the HUD were significantly better detected than targets shown on the edge or the periphery of the display. This effect confirms the hypothesis H1: If visual attention is focussed on the centre of a display, its periphery is less observed and the risk of missing important information provided in more peripheral areas increases. Thus it can be assumed that ATTENDO is able to reproduce the visual attentional effect which has been reported also in other studies (e.g. Alexander et al., 2005). Overall, this provides evidence that the approach introduced here is able to identify attention allocation effects when using HUDs.

However, the hypothesis H2 stating an increased risk of inappropriate allocation of visual attention allocation caused by the more complex HUD could not be confirmed. This is in contrast to findings of earlier studies suggesting that attentional tunnelling effects increase with augmenting complexity of HUD design (Alexander et al., 2005). Yet, other studies (e.g. Williams, 2002) did, either, find nothing but an overall effect of focussed attention on the display information when flying with HUDs, independent of their specific designs. Based on the data from the current suitability study, it is difficult to draw any significant conclusions about the origins of this effect. On the one hand, it might be related to the fact that HUDs in general, i.e. independent of the specific complexity of the design, involve the risk of a centralization of attention. On the other hand, it might be related to the fact that the differences in complexity between of the two designs compared are too small. The choice of the two variants of HUD for the current study was dependent on the somewhat limited range of features provided by the flight simulation 'x-plane®'. Thus it cannot be excluded that the differences in complexity between the two HUDs are simply not large enough to provoke differences in the amount of attentional tunnelling.

No significant effects, neither with respect to target location nor to HUD design, were found for detection times. The explanation for this effect is straightforward. The targets were only presented for comparatively short periods of time (500ms). Obviously, given this, the main issue is to detect a target. However, if a target is detected, the response time does not differ depending on target location. It is to be expected that differences in response time would have occurred only if the targets had been presented for a longer time. The current paradigm with short presentation times was chosen to make detection rates the most sensitive parameter.

Only one of the pilots did not react correctly to the runway incursion. Even though the pilot was from the group flying with the more complex HUD, no conclusions can be drawn from this single case. Even more interesting is the fact that although a centralization of attention was found with both HUDs, this sort of attentional tunnelling obviously is not strong enough to interfere with the detection of an unexpected event on the outside.

One of the advantages of ATTENDO is seen in its higher statistical power because of more reliable data resulting from the comparatively high number of stimuli. Based on the current study this advantage can be described more precisely. The new method produces a large amount of raw data. Twelve pilots participated in the study, all of them flew six approaches and each approach contains twenty-nine targets. Thus, 2,088 single values were produced. In the same experimental constellation, the method used by Alexander et al. (2005) would produce only 216 values, and solely 12 to 24 single values would be gained by using runway incursions for the assessment of visual attention. Due to the large number of raw data produced by the new method, a higher reliability of performance assessment is clearly achieved and a reduction of statistical power is avoided.

The main objective of the current study is to demonstrate the usefulness of ATTENDO for describing the spatial distribution of attention within a display as complex as a HUD. However, lacking a head-down control group, no specific conclusions can be drawn from the current data about HUDs neither with respect to whether the centralization of attention is stronger than with other displays. This might become the objective of more specific studies using ATTENDO in other domains and fields of application in the future.

In summary it can be said, that ATTENDO to assess visual attention allocation using a HUD represents a suitable alternative to existing methods. Through the use of standard hardware and software it is cost-efficient and allows for the testing of innovative HUD design in early stages of development. Because of the high number of raw data the tool produces, statistical shortcomings are avoided. Thus, the new method can increase research efficiency through the use of desktop research.

References

Alexander, A.L., Wickens, C.D., & Hardy, T.J. (2005). Synthetic Vision Systems: The Effect of Guidance Symbology, Display Size, and Field of View. *Human Factors*, *47*, 693–707.

Fadden, S., May Ververs, P., & Wickens, C.D. (2001). Pathway HUDs: Are They Viable? *Human Factors*, *43*, 173–193.

Fischer, E., Haines, R.F., & Price, T.A. (1980). Cognitive Issues in Head-Up Displays. *NASA Technical Paper 1711*. Retrieved from http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19810005125_1981005125.pdf

Foyle, D., Andre, A., McCann, R., Wenzel, E., Begault, D., & Battiste, V. (1996). Taxiway Navigation and Situation Taxiway Navigation and Situation Awareness (T-NASA) System: Problem, Design Philosophy, and Description of an Integrated Display Suite for Low-Visibility Airport Surface Operations. *SAE Transactions: Journal of Aerospace Aerospace Human Factors Research Division*, *105*, 1411–1418.

Hart, S. G., & Stavelan, 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* (pp.139–183). Oxfordshire: Elsevier Science Ltd.

Martin-Emerson, R., & Wickens, C.D. (1997). Superimposition, Symbology, Visual Attention, and the Head-Up Display. *Human Factors*, *39*, 581–601

Proctor, P. (1997). Economic, safety gains ignite HUD sales. *Aviation Week & Space Technology*, *147*, 54–57.

Wickens, C.D. (2005). Attentional Tunnelling and Task Management. *Aviation Human Factors Division Institute of Aviation*. Retrieved from http://www.humanfactors.illinois.edu/Reports&PapersPDFs/TechReport/05-23.pdf

Wickens, C.D., & Long, J. (1995). Object Versus Space-Based Models of Visual Attention: Implications for the Design of Head-Up Displays. *Journal of Experimental Psychology; Applied*, *1*, 179–193.

Wickens, C.D., May Ververs, P., & Fadden, S. (2004). Head-up displays. In D. Harris (ed.), Human factors for civil flight-deck design (pp. 103-140). Aldershot: Ashgate.