University of Applied Sciences and Arts
Northwestern Switzerland

Investigation of visual inspection strategies in detecting threat items in flight passengers´ cabin baggage

- A laboratory experiment with airport security screeners -

MASTER THESIS

Autorin: Simone Günther

Betreuer: Dr. Stefan Michel

Praxispartner: CASRA, Center for Adaptive Security Research and Application

Datum: 17.01.2013

Abstract

The study deals with the analysis of visual search strategies applied by airport security screeners of a European airport while searching for threat items in passenger cabin baggage. In the light of the current threat of terroristic acts and given the fact that airports and planes have recently been subject to criminal assault, the question of how threat items can be best detected, becomes more important than ever.

In this study, we want to draw attention to the process underlying the visual search task and analyze the relation between detection performance and cognitive measures with eye tracking. During a laboratory experiment, screeners from a European airport perform a computer-based test and evaluate x-ray images of passenger cabin baggage. Thereby, quantitative data including eye tracking measures during the visual search task and qualitative data resulting from the think-aloud sessions with the screeners is collected. The results show that there is a significant positive relation between the screener´s detection performance and their fixations in the Area of Interest. No significant relation is found between qualitative speech measures and detection performance. Finally, eye tracking turns out to be an adequate measurement tool for cognitive processes in visual search for security reasons.

Keywords: Eye tracking, visual search, visual inspection strategy, aviation security

1 Introduction

Visual inspection tasks have become an important part of many working environments today as they are often crucial for the quality and performance of goods or services. Schoonard, Gould and Miller (1973) underline the widespread occurrence of visual inspection tasks in different branches and areas such as the food industry, the health industry as well as production or security. Although the application of visual inspection in the working situation can differ strongly from fish or meat inspection in the food industry to x-ray-image interpretation of broken bones in medicine or baggage screening for security reasons, the goals to be reached are usually the same: a maximization of quality and a risk minimization.

In aviation security, the quality of the visual search of a baggage inspection operator and his or her following decision about the baggage is of particular importance as it can affect the lives and security of many people. Recent examples of terroristic acts such as the plane hijacking and crash into the towers of the World Trade Center on 09/11 or the bomb blast in London´s metro station in 2005 demonstrate the catastrophic consequences of what can happen if forbidden objects are not found in passenger´s baggage.

For this reason, many attempts have been made to improve the search quality in this field and research has contributed a lot in terms of finding the key success variables for an effective search (e.g. Wolfe, 2003; Bolfing, Halbherr & Schwaninger, 2008). Thus, certain visual abilities that are essential to become a good x-ray screener could be classified and consequently, tailored pre-employment tests were developed to recruit exactly those candidates for x-ray screening who have the abilities needed for this complex visual search task (Hardmeier, Hofer and Schwaninger, 2005; Hardmeier & Schwaninger, 2008). In addition, individually adaptive training methods like the computer-based training “X-ray-Tutor” (XRT) were developed to improve the individual detection performance. Results show, that screeners who were trained with this method could improve their search and detection time and became more accurate in the detection of threat items (Schwaninger & Wales, 2009).

For a further investigation of the processes underlying the visual search task it is necessary to approach the visual search strategies applied by the screeners. If we

can identify the variables of an effective search, we will get important information about the indications for mistakes being made during visual search and about successful strategies for detecting threat items.

In this study, we want to approach the characteristics of visual inspection in airport security by analyzing eye gaze data of x-ray screening personnel and combining them with information from think-aloud sessions with the screeners. The goal of these methods is to investigate which variables are key to a high performance and how they are interrelated. Furthermore, we want to find out if the eye tracking method is adequate for the research field mentioned and how it can add value.

For this purpose, a controlled laboratory experiment with cabin baggage search operators from a European airport was conducted. Eye gaze data was collected with an SMI eye tracker and subsequent think-aloud sessions with participants gave further insight into the visual inspection strategies of each operator.

We first introduce the research field by discussing related publications and evaluating relevant measurement methods for our examination. From these theories, we deduct our hypotheses. Next, we give an overview of the method design and population before the results are summarized and discussed in detail. Finally, after an overview of the study´s limitations, practical implications are derived and a conclusion is drawn.

Performance measures in visual search tasks

For measuring the efficiency of visual inspection conducted by a human operator and investigating the reasons for high or low performance, it is necessary to define the performance variables of the visual search task first. Since the beginning of visual inspection research, psychophysical models of the human perception process have been developed in order to understand human search behavior and deduce appropriate measures. These so-called “ideal observer analyses” usually base on the Bayesian ideal observer theory and approach the topic by mathematical equations of the human search process (Geisler, 2002).

In the research of detection tasks, the signal detection theory by Green and Swets (1966) has been a milestone and has recently been applied as a model for measuring the detection performance of security screeners evaluating x-ray images at the airport (e.g. Michel, Koller & Schwaninger, 2008). The theory states, that detection performance (d´) can be defined as a function of the z-transformed hit rate and false alarm rate (d´ = z (H) – z (FA)). The hit rate calculation considers the correctly identified items as well as the missed threat items whereas the false alarm rate includes false alarms as well as correctly rejected items. The ROC (Receiver-Operator-Characteristics) - function shows the hit rate in relation to the false alarm rate (see Figure 1). The sensitivity d’ is defined mathematically as interval between the diagonal and the inflection point of the ROC-curve. It can be interpreted as valid measure for a screener´s detection performance as it is a stable unit that can only be changed by continuous training (Schwaninger, 2005b). The figure also demonstrates the relationship between hit rate and false alarm rate by giving an example of two fictional screeners. Screener A has a high detection performance because he has a high hit rate and a low false alarm rate. Screener B also has a high hit rate, but as his false alarm rate is high as well, his overall detection performance becomes much lower.

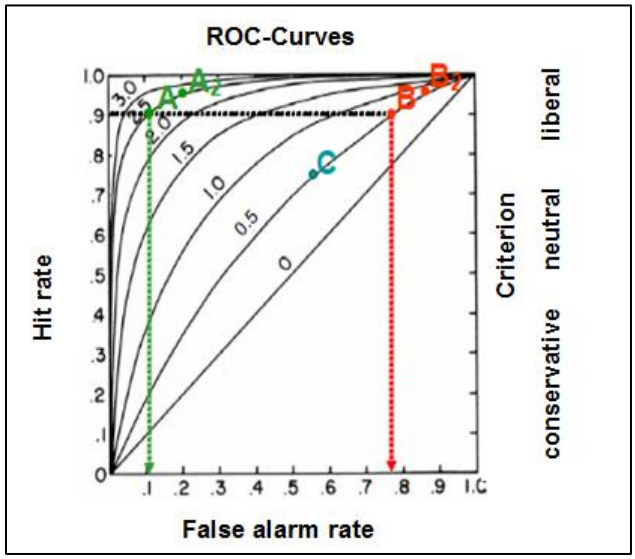

Figure 1. The ROC-function, adapted from Schwaninger, 2005b, (p.115)

Furthermore, the C-criterion (position on the ROC-curve) specifies a person´s response bias. In contrast to d', the C-criterion is not a stable figure but is subject to change with changes in the screener´s environment and gives us an idea about the tendency of the screener to evaluate threat items with a rather conservatory or liberal response style (Schwaninger, 2005b).

In addition to the detection rate, response time is an important indicator of efficiency in inspection tasks- especially if it is viewed in comparison to performance measures. The time a screener needs to check the baggage can have a strong influence on the waiting time for passengers at the checkpoint and is therefore a significant figure in airport security (Schwaninger, 2005a).

Eye tracking and visual search tasks

Apart from the outcome of the visual inspection tasks which can be easily measured by the detection rate and reaction time of the operator, we also want to draw attention to process measures that give insight to the inspector´s cognitive processes during the search.

Duchowski (2007) declares that eye movements "captured during visual inspection provide visualization of the inspector´s process" (p. 251). Thus, eye tracking measures can give important additional information and explanation to the cause of a certain result in a visual search task. Usually, they do not stand alone but are always combined with performance data and analyzed in relation to them.

One of the most influential theories building the basis to research of cognitive processes using eye tracking is the Eye-Mind hypothesis by Just and Carpenter (1976). This hypothesis states, that eye movements can reveal higher, psychological processes as the target a person is looking at will be reflected cognitively by the person at the same time. Based on this assumption, many studies have analyzed the processes behind visual search in different areas using eye tracking measures. Nodine and Kundel (1987) applied eye tracking in x-ray tumor detection and developed a model of visual search demonstrating the different steps from perception to decision making. This made it possible to analyze different sources of error in this complex process. Sadavasian, Greenstein, Gramopadhye & Duchowski (2005) analyzed the eye movements of experts in cargo bay inspection and used it for training novices by visualizing the experts´ scanpath during inspection.

An example of the attempt to define efficiency measures in visual inspection is given by Najemnik and Greisler (2005). They developed a detailed model of an ideal searcher in a visual search task with a target embedded at an unknown location within a random background similar to a natural scene and compared it to human visual search recorded by an eye tracker. Noteworthy, the comparison showed that human searchers were highly efficient and almost reached the performance of the ideal computational searcher.

These examples demonstrate how information taken from eye tracking experiments, like number and duration of fixations or a person´s scanpath, can be used for solving problems in visual search practice.

As a vast number of eye tracking measures exists, it is important to choose the relevant measures for a specific task. Poole and Ball (2005) give an overview of currently existing metrics and areas of application by classifying them into fixation-derived, saccade-derived and scanpath-derived metrics. For our experiment, we want to concentrate on fixation measures as they shed light on the operator´s search focus and intensity of cognitive load at different parts of the search item.

By fixating an object, a person is “stopping the eye to allow further processing of the currently registered visual stimulus” (Mulvey, 2012, p.18). In web-based usability studies, the number of fixations gives hint to the importance of elements perceived by the user (e.g. Dumais, Buscher & Catrell, 2010). Likewise, when searching for an object, a high number of fixations in a certain area (Area of Interest, AOI) indicate that this area is more important to the user than another because the user will fix those objects more often that seem more noticeable to him (Poole, Ball & Phillips, 2004). This means, when evaluating the efficiency of a person´s visual search strategy, we have to examine the distribution of fixations across the objects. Cowen, Ball and Delin (2002) state that an efficient and focused search is characterized by a high concentration of fixations in the AOI, whereas widespread and inefficient searchers usually show evenly spread fixations across the search object.

Hypotheses I

From the findings mentioned, we conclude that an effective searcher who correctly detects the threat item quickly will scan the baggage and then focus his gaze on the area of interest, where he expects to find the threat item. He will then analyze the area of interest in more detail to evaluate the object and then decide upon it. This will result in more fixations in the actual area of interest. We suppose that this effect can be demonstrated in general by analyzing the percentage of correct answers per image as well as on an inter-individual level, depending on the operator´s detection performance. Accordingly, the following hypotheses are derived:

1. There is a positive correlation between number of correctly evaluated images and the percentage of fixations in the AOI.
2. Participants with a higher detection performance will have more fixations in the AOI than participants with lower detection performance.

Detection performance & verbal data

In addition to quantitative measurement instruments like eye tracking, further attempts of rather qualitative nature have been made to investigate the cognitive processes involved in the search procedure. A frequently used, introspective technique is the think-aloud method (Duncker, 1935). During the think-aloud session, the participant is asked to “think aloud” during the problem solving process and to verbalize the steps he takes to identify and solve the problem. Recently, this method has successfully been applied complementary to eye tracking in usability studies to get further insight into the user´s processes during problem solving (e.g. Elling, Lentz & de Jong, 2011). In this context, especially the retrospective think-aloud (RTA) method proves adequate, which is applied after the actual eye tracking session. On the contrary, the concurrent think-aloud method (CTA), which is applied at the same time as the usability test, is not recommended for eye tracking as the participant will be too much distracted from the actual task and his eye movements will not be representative (Hyrskykari, Ovaska, Majaranta, Räihä, & Lehtinen, M, 2008).
A recent analysis of RTA used in combination with eye tracking demonstrated the high validity and reliability for this method (Guan, Lee, Cuddihy & Ramey, 2006). The advantages of think-aloud experiments compared to other methods collecting verbal

data (e.g. interviews) are that there will be no loss of data due to missing memory because the participant immediately tells about his thoughts and that the participant becomes involved in a real task, thus producing more reliable results compared to other, fictitious tasks (Wade, 1990).

An important assumption for the use of think-aloud protocols is that a person´s thinking and speaking style are correlated (Dörner, 1983). Hence, it becomes possible to differentiate efficient and inefficient problem solvers by means of their verbalizations. Roth (1985) analyzed the transliterated verbal data of participants in a complex, computer-simulated problem solving task. He found significant differences in the verbalizations between the successful and unsuccessful problem solvers, showing that the unsuccessful problem solvers used significantly more negations and subjunctives than the successful problem-solvers. This example demonstrates the additional impact of qualitative data for differentiating between efficient and inefficient problem solvers and consequently for determining key success variables of efficient search.

Hypotheses II

We apply the findings mentioned from qualitative research to our area of investigation and postulate that an efficient visual search process can not only be demonstrated by the detection performance and eye gaze data of the operator, but is also revealed by the person´s verbalization. As successful problem-solving in our case will result in a high detection performance, we use this measurement index for this purpose. Consequently, we suppose that we can find the following formal evidence in the verbal data protocols:

3. There is a negative correlation between detection performance and use of negations.
4. There is a negative correlation between detection performance and use of subjunctives.

2 Methods

Eye tracking was chosen as main instrument to analyze the visual search process because it gives the unique possibility to track the searcher´s subconscious information processing in real-time by recording his or her corresponding eye movements. In addition, qualitative data was conducted to further investigate the screeners´ individual inspection strategies (see 2.5 Other measures). A laboratory design was chosen to guarantee optimal conditions for the eye tracking task and to maximize internal validity.

2.1 Stimuli selection

Careful consideration was given to the stimuli selection. First, an item analysis was conducted to choose relevant material from a pool of x-ray images that were developed for a competency assessment test in 2010 and 2011 (Michel, S., Mendes, M. & Schwaninger, A., 2010). Figure 2 shows an example of an image used in the test.

Figure 2. Example of x-ray image used

Overall, 96 x-ray-images of passenger cabin baggage were chosen to be presented to the participants in the eye tracking experiment. Half of the images did not have a threat item in the baggage and half of the images contained either a gun, a knife, an improvised explosive device (IED) or an item from the category "others" (e.g. electronic shock devices). The categories of threat items were uniformly distributed over the stimuli. Another important criterion was the rotation of the threat item. Half of the 48 images showed the object in a rotated view in the baggage and half of them were presented in canonical view.

Apart from the category and rotation, the discriminatory power and item difficulty of each stimulus had to be considered. According to Bortz and Döring (2006), only items with a discriminatory power $r_{it} > .3$ should be selected, because items below this value are bad indicators of the main construct and will not predict correctly the

person´s value in the main construct. In addition, the item difficulty indices must be distributed over the whole range of key so that a differentiation between the participants´ performances is possible. We considered both discriminatory power and item difficulty. Thus, we selected only stimuli with a discriminatory power over .3 and made sure that the item difficulty varied within each item category. In the end, a test difficulty range between .24 and .99 was reached with an average mean test difficulty of .75 of all items to guarantee a medium to rather challenging test.

2.2 Experimental setup

The experiment was conducted in the eye tracking lab of the university Zurich. We used a Belinea 19-in. monitor set at 1280 x 1024 pixels resolution to present the selected stimuli to the participants. A remote eye tracking device by SMI (RED-III pan tilt camera) using the dark pupil system was positioned in front of the monitor. Participants were seated approximately 57cm from the monitor. The SMI eye tracker software “i-View” was configured to collect gaze data at a rate of 50 Hz. The dispersion threshold for the fixation detection algorithm was set to 100 pixels. The minimum duration of a fixation was set to 80 ms. A chinrest was mounted at the table to avoid that the participant´s head and eye focus would shift during the experiment.

2.3 Procedure

When the participants arrived at the laboratory, they were informed about the test procedure by a written instruction, which asked them to evaluate the x-ray cabin baggage items on the screen. By clicking the right mouse button, they answered “OK”, which means that they did not see any threat item in the baggage. Clicking the left mouse button was considered a “Not OK” which means that the participant thought the baggage should be checked again because he or she thinks there is a threat item inside.

The participants had 15 seconds to come to a decision about the baggage. When they clicked the mouse, the next image was automatically displayed on the screen. I-View tracked the response time in the background. If the screeners did not decide about the baggage within the time given, they had 5 more seconds afterwards to click the mouse and were reminded that they have to come to a decision. Like this, the participants evaluated each of the 96 x-ray images. The order of the stimuli was ran-

domized. All in all, the test lasted about 15 to 20 minutes to complete, depending on the response times of the participants.

In order to control for the confounding variables light and distraction, lighting conditions in the laboratory were controlled by darkening the room as much as necessary for optimized camera tracking. During the experiment, there was strict silence and participants were not interrupted during the x-ray image test.

2.4 Eye tracking measures

Whereas the participants performed the x-ray item test and searched for threat items on the screen, their eye movements were tracked. The fixation duration was measured as well as the number of fixations and the reaction time until the decision was made. In addition, the saccades were tracked as well which made it possible to visualize the participants´ scanpath after the experiment.

2.5 Other measures

Besides the eye tracking data, qualitative data was collected after the experiment for an additional insight to the screener´s strategies and screening behavior, using the think-aloud method. The think-aloud method which is often used in usability studies (e.g. Van den Haak, De Jong & Schellens, 2003), is based on the idea that a researcher can partake in the problem solving process of a certain participant if the participant explains his or her thoughts during task completion. For this purpose, 7 x-ray images of passenger cabin baggage (4 of them containing a threat item of each category, 3 of them not) were printed in colour on A4 paper. The participants were assigned to tell what they think when they analyze the baggage and to rate on a 6-point-Likert scale how sure they were with their final decision. In between they were asked questions about their search behavior by the experimenter (e.g. “What makes this item a threat item for you?”, “How do you proceed when analyzing the baggage?”). These short think-aloud sessions lasted between 6 and 16 minutes and were recorded by the experimenter. Finally, the screeners filled out a questionnaire collecting biographic data (gender, age and professional experience).

Figure 3 gives an overview of the methods with the resulting data and the following data analysis methods, which will be discussed in chapter 3.

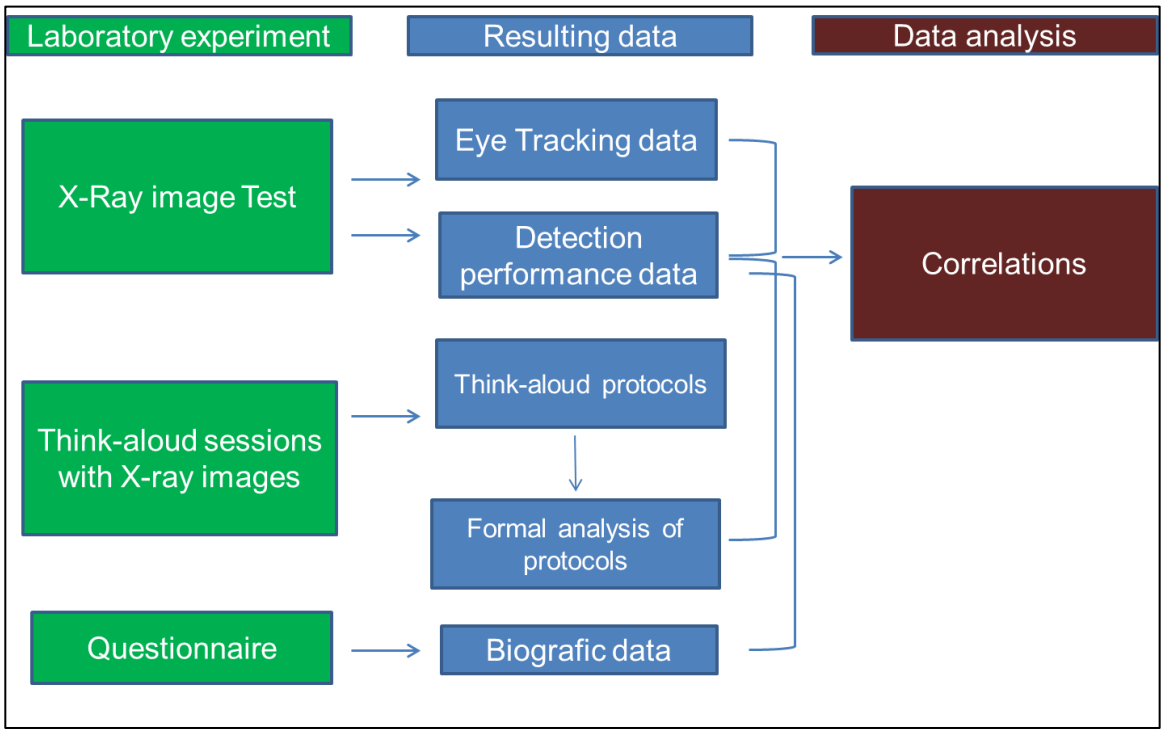

Figure 3. Overview of Methods and Resulting Data

2.6 Participants

The sample consisted of 10 male and 7 female airport security employees working as certified screeners at a European airport. The participants were between 25 and 58 years old (M = 43.24, SD = 10.7) and had an average working experience of 5.7 years (SD = 5.28) as screeners at the airport All participants were used to interpreting x-ray images due to work and training and reported normal or corrected-to-normal vision.

The participation in the study was voluntary and was not compensated. Though, the participants were promised to receive their performance result of how many baggage items they evaluated correctly after the study. In addition, they were promised to receive a journey voucher if they had the best test result.

3 Results

3.1 Data analysis

After the experiment, the detection performance of each screener was calculated using the d'-formula (see chapter 1). Additionally, the participant´s reaction time for each stimulus, the test duration and the personal data (gender, age, screening experience) were collocated. For the eye tracking analysis with the x-ray image test, two data sets had to be excluded due to technical problems with the analyzing software. In addition, one participant had to be excluded totally from data analysis because the order of presented stimuli was not randomized after a change of settings in the eye tracking system.

The screeners´ eye tracking data were analyzed with the software Begaze 3.1 from SMI. For each of the 48 images containing a threat item, an area of interest (AOI) was defined by exactly encircling the shape of the threat item on the screen. Then, after combining the eye tracking data of each participant with the stimuli, fixation measures were exported. We analyzed the fixations in the AOI in contrast to the fixations outside the AOI (percentage fixations in AOI). Finally, we used the statistic program PASW 18 to analyze correlations between the variables. First, the Kolmogorov-Smirnov-test was applied for all variables to test for normal distribution. As all variables showed normal distribution and were interval- or ratio scaled, the Pearson-correlation coefficient was then calculated and tested for significance.

The recorded think-aloud-sessions were transcribed and the resulting protocols were analyzed for the negations and conjunctives used by the participants when explaining their screening methods. Subsequently, the number of negations and conjunctives used was counted for each participant. Furthermore, the logged duration time of the think-aloud-session was gathered.

3.2 Correlations of eye tracking measures

For the first hypothesis, we calculated the percentage of correct answers (Hits + Correct Rejections) per image as well as the average percentage of fixations spent in the AOI per image.

We found a significant correlation between these values with $r(46) = .46$, $p < .01$. This model explains 21% of the variance ($R^2 = .21$).

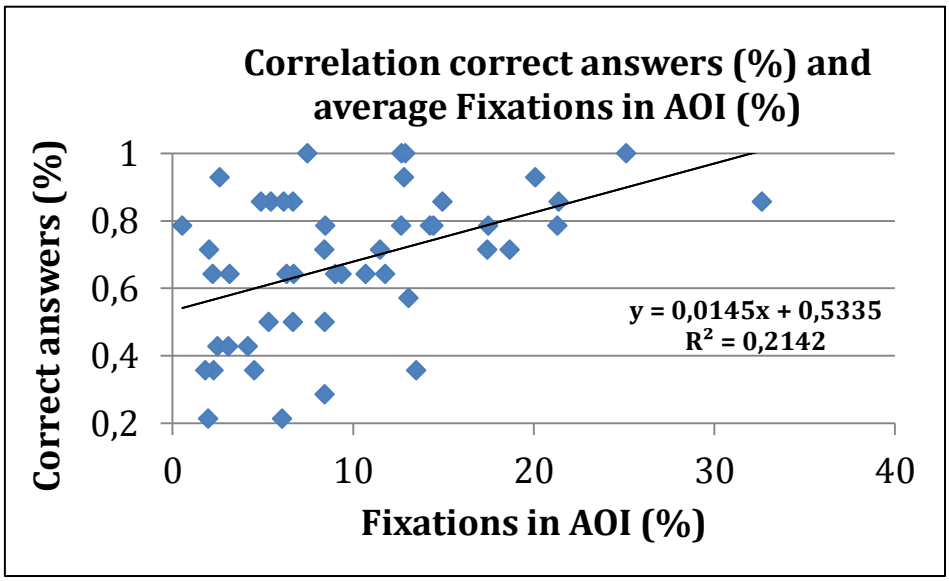

Figure 4. Correlation of correct answers (%) per image and average fixations in AOI (%)

For the second hypothesis, the percentage of fixations spent in the AOI was analyzed per participant. Next, the correlation with d' was calculated. We found a correlation of $r(12) = .32$, $p = .13$ with $R^2 = .10$. The further PASW analysis showed no significance for this correlation.

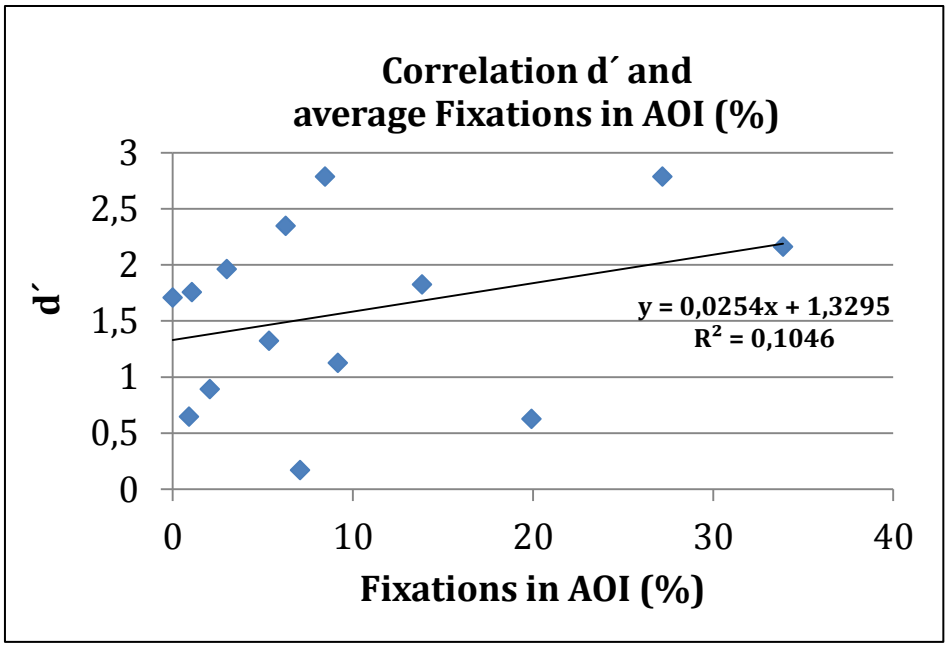

Figure 5. Correlation of detection performance (d´) and average fixations in AOI (%)

3.3 Correlations of verbal measures

For the third hypothesis, the average number of negations used by the participant during the think-aloud session was analyzed per participant. Furthermore, the correlation with d' was calculated. We found a correlation of $r(14) = -.37$, $p = .08$ with $R^2 = .14$. This correlation was not significant.

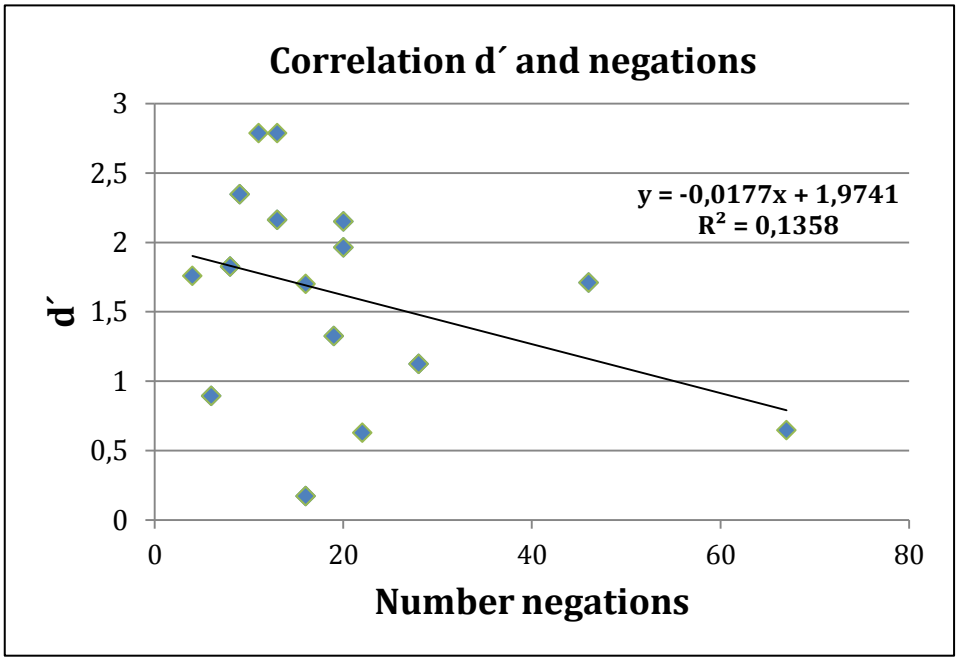

Figure 6. Correlation of detection performance (d´) and negations

The last hypothesis was tested by calculating the average number of subjunctives used by the participant during the think-aloud session. This variable was correlated with d'. The result was a correlation coefficient of $r(14) = -.15$, $p = .29$ with $R^2 = .02$. This correlation neither was significant.

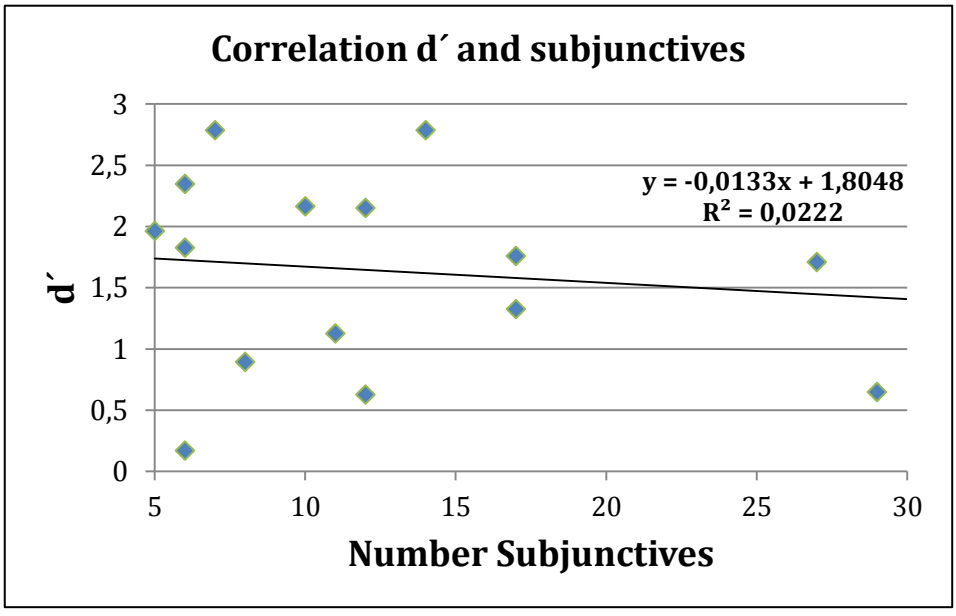

Figure 7. Correlation of detection performance (d´) and subjunctives

3.4 Additional findings

In addition to the hypotheses tested, we found interesting correlations between the variables. As a particularly interesting result for the following discussion, the correlation between average reaction time and d´ should be mentioned. As reaction time, we defined the time until the participant clicked the mouse to decide upon the baggage. The following chart shows the correlation found with $r(12) = -.64$, $p < .05$ and $R^2 = .40$. The PASW analysis confirmed the significance of this correlation.

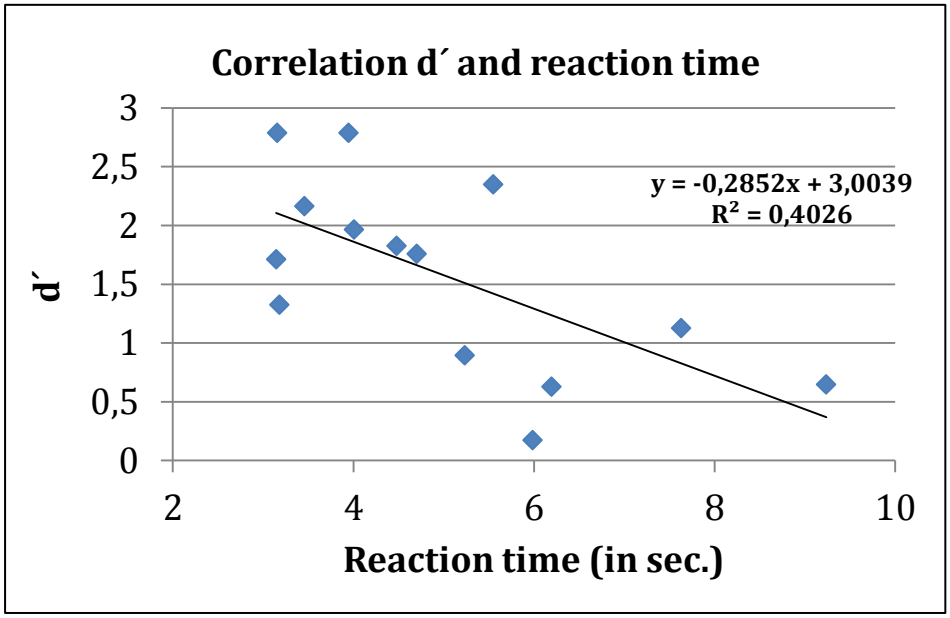

Figure 8. Correlation detection performance (d´) and reaction time (RT)

Another interesting finding was the strong, negative correlation between interview duration and detection performance, $r(12) = -.40$, $p = .15$ with $R^2 = .17$ although the subsequent PASW analysis showed no significance for this correlation.

Regarding the biographic measures, no significant correlations were found. For the individual measure “age”, we found a negative correlation of $r(12) = -.51$, $p = .06$, between the variable and detection performance. For professional experience and d´, the slightly negative correlation was $r(12) = -.02$, $p = .93$. Finally, we calculated the correlation for gender and detection performance which resulted in $r(12) = .02$, $p = .95$.

4 Discussion

From the eye tracking results presented, we can conclude that the first hypothesis is confirmed. A significant, positive correlation between the percentage of correctly evaluated images and the percentage of fixations in the AOI in comparison to fixations in the rest of the image was found in the sample. Regarding the key success variables of visual search in detecting threat items, it can consequently be underlined, that an effective and successful search is linked to an aggregation of fixations in the AOI. This finding fits into the context of research about the ideal searcher, who identifies the search item quickly and focuses quickly on it instead of searching broadly in an unstructured way across the examination object (Duchowski, 2007).

Hypothesis 2 cannot be confirmed. Although there was a positive correlation between detection performance and fixations in the AOI, this correlation was not significant. A problem for this analysis was of course the small number of participants for whom the analysis was possible (N = 14). Another influencing factor could also be the individual screening strategy of each participant. With hypothesis 2 we tried to find common grounds between the screeners´ strategies supposing that a correlation could be found between detection performance and fixations in the AOI. Still, it is possible that the individual search strategies differ more than expected, which makes it more difficult to find similarities in the gaze pattern. A deeper analysis of the individual gaze patterns seems necessary to detect the distinctive success variables of the screeners who show a high detection performance.

Summarizing the first results, we see that there is a link between eye gaze data and the percentage of correct answers given by the participant. Even though this effect could not be demonstrated on an inter-individual level, showing differences between screeners with high and low detection performance, there is still evidence of the fact that a successful search strategy is linked with a structured visual search, including a high concentration of fixations in the AOI. Combined with the additional finding of the significant negative correlation between detection performance and reaction time, we slowly approach the profile of an ideal visual search strategy. The time needed to decide upon a baggage mirrors the speed of all cognitive processes included in this decision. If detection performance and reaction time correlate negatively, it can be deduced, that a successful screener will quickly come to a decision about the bag-

gage item and react accordingly. Seen all mentioned results together, we assume that a person applying an optimized visual search process will quickly get an overview of the baggage items, evaluate if they match his or her knowledge of threat items and then focus on the objects that are most suspicious to him or her to come to a decision. The results imply that an ideal inspector will need less time for visual perception and following cognitive analyses of the objects. This could be due to training and memory of the learned threat items. Future studies could examine this relationship by testing the screener´s long-term memory of the threat items learned during training. Good long-term memory of the threat items could result in better evaluation of the objects and faster decisions. As time for decision making is short in practice, where the screeners have to evaluate each passenger´s baggage within seconds, it could be helpful for the screeners to learn objects of everyday life in addition to the threat items to accelerate the decision making process. In this context, an illuminating method for future research would be a cognitive test including visual perception and declarative memory to find out if inter-personal differences of these constructs can explain differences in reaction time and detection performance.

Regarding the qualitative data analyzed by the verbal protocols, no significance was found for the correlation between use of negations or subjunctives and the individual detection performance. Thus, Hypotheses 3 and 4 regarding the verbal measures cannot be confirmed with the obtained results.

One possibility for the missing significance could be the small number of participants for whom protocols were available (N = 16). Another affecting variable could be the awareness of the personal problem-solving process. As the verbal analysis of the think-aloud protocols based on a description of the individual search method, the subconscious part of the process had to be neglected. For a better understanding of this part, it would be helpful to simultaneously track the participant´s gaze during the think-aloud session and to compare the resulting data to the verbal measures to see if the gaze pattern is consistent with the participant´s description of his strategy.

Apart from the restrictions mentioned, we want to draw attention to the solid, negative correlation between duration of the think-aloud session and detection performance as well as negations used during the think-aloud session and detection performance. Due to the negative correlation between these variables, we can assume that there is a tendency for efficient visual searchers to articulate their problem-solving process

more efficiently, thus needing less time for the explanations. In addition, they seem to concentrate less on excluding items and stating what they DO NOT see in the picture (thus causing less negations) but rather think about possibilities of correctly interpreting the items they perceive. One possibility for the shorter session duration could also be that the participants were more aware of their problem-solving method, hence being more precise in their wording.

5 Practical implications

In the following, we reflect our experiences with using eye tracking in an experimental study of visual inspection and give practical guidelines for future investigations. In this context, the idea of applying eye tracking in training visual strategy is discussed as well.

The study demonstrates that eye tracking can be an interesting additional tool for investigating individual search strategies because it gives the chance to combine measures of detection performance with measures of the individual search process. Our results indicate that a successful visual search which can be measured by correctly evaluated objects in an x-ray image goes along with an efficient search pattern. With analyzing additional variables of the search process, future research in this area can be very illuminating.

Still, some guidelines should be considered as eye tracking is not a self-explanatory method that can easily be applied by everybody in practice. On the contrary, the experimenter has to become acquainted with the technical devices in detail before the study. Pilot experiments are only one prerequisite for an efficient data acquisition process without surprises. Thereby, technical details of the eye tracking setup should be regarded as well as the test duration which can have an influence on the participant´s concentration or detection performance. Another key success factor seems to be the stimuli selection for the eye tracking experiment. For this, not only the number of stimuli that can maximally be presented have to be considered, but also the question which stimuli are best to test for the search object have to be thought about carefully. For most research questions in this context, it might be helpful to define an AOI including the search object before the experiment and to analyze the eye gaze structures in relation to it. If these requirements are given, eye tracking can give added benefit to research of search strategies by tracking the participant´s eye movements in real-time and thus providing detailed, objective data about the person´s mind activity. In order to give additional sense to the tracked data from the operator´s perspective, we suggest applying introspective methods supplemental to eye tracking.

If – as our results imply – performance in visual inspection is shown in a particular search pattern, one could think that it could be possible to improve the individual detection performance by training one´s search strategy. Thus, the idea of using eye

tracking for training the visual search competency might be appealing to practioners in the area of visual search in the future. Yet, some requirements should be met before including eye tracking in training environments. One challenge is the distraction which could occur due to the camera and tracking method during the training session. For a minimal risk of distraction, modern eye tracking technology such as a monitor-based system should be used. Besides, the tracked eye gaze data should only be shown to the participant after the training and not been revealed on the monitor in real-time. Like this, eye tracking in training could be employed as a feedback method for subsequent analysis of one´s own search behavior. Moreover, the data should be used as a focused analysis method for particular cases and not as a general training method which is mandatory for everybody. A specific area of application could be the training of screeners who did not pass the competency assessment test. With the help of eye tracking, they could hence see their gaze data and realize which parts of the baggage have been neglected and where they possibly lost too much time. In summary, the idea of using eye tracking as part of the training of screeners can be interesting if it is used in a target-oriented way.

6 Limitations

Like any experimental study, also this sample is subject to certain limitations. First, the small number of participants should be mentioned. In further studies, it would be interesting to analyze a larger sample resulting in possibly more significant quantitative results. For this purpose, a general compensation of the participants should be considered.

With a larger sample, it would also become possible to apply additional analysis methods. We based our analysis on the calculation of correlations between performance values and eye tracking data, respectively verbal protocols. Still, correlations do not give information about the cause of an effect or the differences between variables. Hence, a larger sample would allow the application of an ANOVA to make further statements about the differences between the screeners. In this context, an analysis of the differences between screeners with high and low detection performance and their corresponding fixation measures would be particularly interesting.

Regarding the experimental design, much effort was put into the choice of the stimuli and the test sequence. The types of forbidden objects have been carefully selected over the categories and the order was randomized. Still, for ideal results, it would be interesting to replicate the study with stimuli containing exactly the same baggage stimulus for those pictures with and without forbidden object. Thus, a sophisticated analysis about the differences in eye gaze data for these two requirements would be possible.

Furthermore, the setup of the eye tracker must be commented as it was not the newest system available on the market. Although the experimental environment provided a good basis for comparable results, a more recent eye tracking technology could be used for ideal tracking results.

7 Conclusion

This study shows a new analysis approach in the area of terror prevention by investigating the individual search process during security inspections. The results have to be interpreted carefully due to the limitations mentioned. Still, the study demonstrates that the combination of quantitative eye tracking data and qualitative introspective data of the participant can give interesting hints to the individual problem solving process that go beyond a mere numerical analysis and hence can reveal new procedures and guidelines for further investigations.

8 References

Bolfing, A., Halbherr, T., & Schwaninger, A. (2008). How image based factors and human factors contribute to threat detection performance in x-ray aviation security screening. *HCI and Usability for Education and Work, Lecture Notes in Computer Science*, *5298*, 419-438.

Bortz, J., & Döring, N. (2006). *Forschungs- und Evaluationsmethoden für Human- und Sozialwissenschaftler* (4th ed.). Heidelberg: Springer Verlag.

Cowen, L., Ball, L.J., & Delin, J. (2002). An Eye Movement Analysis of Webpage Usability. In: *People and Computers XVI - Memorable yet Invisible: Proceedings of the HCI 2002* (pp. 317-335) London: Springer-Verlag.

Duchowski, A.T. (2007). *Eye tracking methodology: Theory and practice*. London: Springer Verlag.

Dörner, D. (1983). Denken, Problemlösen und Intelligenz. In Lüer, G. (Ed.), *Bericht über den 33. Kongreß der Deutschen Gesellschaft für Psychologie*, *Mainz 1982*. Göttingen: Hogrefe Verlag.

Duncker, K. (1935). *Zur Psychologie des Produktiven Denkens*. Berlin: Springer Verlag.

Dumais, S., Buscher, G., & Cutrell, E. (2010). Individual differences in gaze patterns for Web search. *Proceedings of the Third Symposium on Information Interaction in Context, New Brunswick, NJ, USA — August 18-21, 2010* (pp. 185–194). New York: ACM press.

Elling, S., Lentz, L., & De Jong, M. (2011). Retrospective think-aloud method: Using eye movements as an extra cue for participants' verbalizations. In *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems* (pp.1161-1170). New York: ACM press.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Guan, Z., Lee, S., Cuddihy, E., & Ramey, J. (2006). The validity of the stimulated retrospective think-aloud method as measured by eye tracking. In *Proceedings of the SIGCHI conference on Human Factors in computing systems,* April 22-27, 2006, Montréal, Québec, Canada (pp. 1253-1262). New York: ACM press.

Hardmeier, D., & Schwaninger, A. (2008). Visual cognition abilities in x-ray screening. *Proceedings of the 3rd International Conference on Research in Air Transportation, ICRAT 2008, Fairfax, Virgina, USA, June 1-4,* 2008, 311-316.

Hardmeier D., Hofer F., & Schwaninger A. (2005). The x-ray object recognition test (x-ray ort) – a reliable and valid instrument for measuring visual abilities needed in x-ray screening. *IEEE ICCST Proceedings*, 39, 189-192.

Hyrskykari, A., Ovaska, S., Majaranta, P., Räihä, K. J., & Lehtinen, M. (2008). Gaze path stimulation in retrospective think aloud. *Journal of Eye Movement Research, 2*(4), 1-18.

Jacob, R. J. K., & Karn, K. S. (2003). Eye tracking in Human-Computer Interaction and usability research: Ready to deliver the promises. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 573-605). Amsterdam: Elsevier.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, *8*(4), 441-480.

Michel, S., Koller, S., & Schwaninger, A. (2008). Relationship between level of detection performance and amount of recurrent computer-based training. *Proceedings of the 42nd Carnahan Conference on Security Technology, Prague, October 13-16, 2008*, 299 – 304.

Michel, S., Mendes, M., & Schwaninger, A. (2010). Can the difficulty level reached in computer-based training predict results in x-ray image interpretation tests? *Proceedings of the 44th Carnahan Conference on Security Technology, San Jose California, October 5-8, 2010*.

Mulvey, F. (2012). Eye anatomy, eye movements and vision. In P. Majaranta, H. Aoki, M. Donegan, D.W. Hansen, J.P. Hansen, A. Hyrskykari & K.J. Räihä (Eds.), *Gaze Interaction and Applications of Eye Tracking: Advances in Assistive Technologies* (pp.10-20). Hershey, PA: IGI Global publishing.

Namjenik, J., & Geisler, S.W. (2005). Optimal eye movement strategies in visual search. *Nature*, 434, 387-391.

Nodine, C.F., & Kundel, H.L. (1987). Using eye movements to study visual search and to improve tumor detection. *RadioGraphics*, *7*(6), 1241-1250.

Poole, A., Ball, L. J., & Phillips, P. (2004). In search of salience: A response time and eye movement analysis of bookmark recognition. In S. Fincher, P. Markopolous, D. Moore, & R. Ruddle (Eds.), *People and Computers XVIII-Design for Life: Proceedings of HCI 2004*. London: Springer Verlag.

Poole, A., & Ball, L.J. (2005). Eye tracking in human-computer interaction and usability research: current status and future prospects. In C. Ghaoui (Ed.): *Encyclopedia of human-computer interaction* (pp. 211-219). Pennsylvania: Idea Group.

Roth, T. (1985). Sprachstatistisch objektivierbare Denkstilunterschiede zwischen ‚guten' und ‚schlechten' Bearbeitern komplexer Probleme. *Sprache & Kognition 4*, 178-191.

Sadavasian, S., Greenstein, J.S., Gramopadhye, A.K., & Duchowski, A.T. (2005). Use of eye movements as feedforward training for a synthetic aircraft inspection task. In *Proceedings of CHI´2005*, 141-149.

Schwaninger, A. (2005a). Increasing efficiency in airport security screening. *WIT Transactions on the Built Environment*, 407-416.

Schwaninger, A. (2005b). Objekterkennung und Signaldetektion: Anwendungen in der Praxis. In: B. Kersten & M. Groner (Eds.): *Praxisfelder der Wahrnehmungspsychologie* (pp. 106-130). Bern: Huber.

Schwaninger, A., & Wales, A.W.J. (2009). One year later: how screener performance improves in x-ray luggage search with computer-based training. *Proceedings of the Ergonomics Society Annual Conference 2009,* 381-389.

Schoonard, J.W., Gould, J.D., & Miller, L.A. (1973). Studies of visual inspection. *Ergonomics*, *16*(4), 365-379.

Wade, S.E. (1990). Using think alouds to assess comprehension. *The Reading Teacher*, *43*(7), 442-451.

Van den Haak, M. J., De Jong, M.D.T., & Schellens, P.J. (2003). Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue. *Behaviour & Information Technology, 22*(5). 339-351.

Wolfe, J. M. (2003). Moving towards solutions to some enduring controversies in visual search. *Trends in Cognitive Sciences*, *7*(2), 70-76.

9 Redlichkeitserklärung

Hiermit erkläre ich, die vorliegende Master Thesis selbständig, ohne Mithilfe Dritter und nur unter Benutzung der angegebenen Quellen verfasst zu haben.

Datum Simone Günther

10 Appendix

Output of SPSS analyses testing the hypotheses

Hypothesis 1

Result of Kolmogorov-Smirnov-Test: Normal distribution

Kolmogorov-Smirnov-Anpassungstest

		MW_FixAOI_Bild	Prozent_richtige
N		48	48
Parameter der Normalverteilung[a,b]	Mittelwert	9,8737	,6771
	Standardabweichung	6,90949	,21705
Extremste Differenzen	Absolut	,123	,146
	Positiv	,123	,082
	Negativ	-,101	-,146
Kolmogorov-Smirnov-Z		,854	1,009
Asymptotische Signifikanz (2-seitig)		,459	,260

a. Die zu testende Verteilung ist eine Normalverteilung.

b. Aus den Daten berechnet.

Correlation: significant

Korrelationen

		MW_FixAOI_Bild	Prozent_richtige
MW_FixAOI_Bild	Korrelation nach Pearson	1	,463**
	Signifikanz (1-seitig)		,000
	N	48	48
Prozent_richtige	Korrelation nach Pearson	,463**	1
	Signifikanz (1-seitig)	,000	
	N	48	48

**. Die Korrelation ist auf dem Niveau von 0,01 (1-seitig) signifikant.

Hypothesis 2

Result of Kolmogorov-Smirnov-Test: Normal distribution

Kolmogorov-Smirnov-Anpassungstest

		M_Fix_AOI_Person_Prozent	d_prime
N		14	14
Parameter der Normalverteilung[a,b]	Mittelwert	9,8737	1,5799
	Standardabweichung	10,37723	,81353
Extremste Differenzen	Absolut	,241	,134
	Positiv	,241	,089
	Negativ	-,171	-,134
Kolmogorov-Smirnov-Z		,902	,503
Asymptotische Signifikanz (2-seitig)		,390	,962

a. Die zu testende Verteilung ist eine Normalverteilung.

b. Aus den Daten berechnet.

Correlation: no significance

Korrelationen

		d_prime	M_Fix_AOI_Person_Prozent
d_prime	Korrelation nach Pearson	1	,323
	Signifikanz (1-seitig)		,130
	N	14	14
M_Fix_AOI_Person_Prozent	Korrelation nach Pearson	,323	1
	Signifikanz (1-seitig)	,130	
	N	14	14

Hypothesis 3

Result of Kolmogorov-Smirnov-Test: Normal distribution

Kolmogorov-Smirnov-Anpassungstest

		d_prime	Negationen
N		16	16
Parameter der Normalverteilung[a,b]	Mittelwert	1,6232	19,8750
	Standardabweichung	,77090	16,08674
Extremste Differenzen	Absolut	,166	,260
	Positiv	,085	,260
	Negativ	-,166	-,162
Kolmogorov-Smirnov-Z		,662	1,040
Asymptotische Signifikanz (2-seitig)		,773	,230

a. Die zu testende Verteilung ist eine Normalverteilung.

b. Aus den Daten berechnet.

Correlation: no significance

Korrelationen

		Negationen	d_prime
Negationen	Korrelation nach Pearson	1	-,368
	Signifikanz (1-seitig)		,080
	N	16	16
d_prime	Korrelation nach Pearson	-,368	1
	Signifikanz (1-seitig)	,080	
	N	16	16

Hypothesis 4

Result of Kolmogorov-Smirnov-Test: Normal distribution

Kolmogorov-Smirnov-Anpassungstest

		d_prime	Konjunktive
N		16	16
Parameter der Normalverteilung[a,b]	Mittelwert	1,6232	13,6875
	Standardabweichung	,77090	8,66194
Extremste Differenzen	Absolut	,166	,202
	Positiv	,085	,202
	Negativ	-,166	-,158
Kolmogorov-Smirnov-Z		,662	,809
Asymptotische Signifikanz (2-seitig)		,773	,530

a. Die zu testende Verteilung ist eine Normalverteilung.

b. Aus den Daten berechnet.

Correlation: no significance

Korrelationen

		d_prime	Konjunktive
d_prime	Korrelation nach Pearson	1	-,149
	Signifikanz (1-seitig)		,291
	N	16	16
Konjunktive	Korrelation nach Pearson	-,149	1
	Signifikanz (1-seitig)	,291	
	N	16	16

Additional findings

Result of Kolmogorov-Smirnov-Test: Normal distribution

Kolmogorov-Smirnov-Anpassungstest

		d_prime	RT
N		14	14
Parameter der Normalverteilung[a,b]	Mittelwert	1,5799	4992,5647
	Standardabweichung	,81353	1809,84732
Extremste Differenzen	Absolut	,134	,153
	Positiv	,089	,135
	Negativ	-,134	-,153
Kolmogorov-Smirnov-Z		,503	,574
Asymptotische Signifikanz (2-seitig)		,962	,896

a. Die zu testende Verteilung ist eine Normalverteilung.

b. Aus den Daten berechnet.

Correlation: not significant

Korrelationen

		d_prime	RT
d_prime	Korrelation nach Pearson	1	-,635*
	Signifikanz (2-seitig)		,015
	N	14	14
RT	Korrelation nach Pearson	-,635*	1
	Signifikanz (2-seitig)	,015	
	N	14	14

*. Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.

Result of Kolmogorov-Smirnov-Test: Normal distribution

Kolmogorov-Smirnov-Anpassungstest

		Interviewzeit
N		14
Parameter der Normalverteilung[a,b]	Mittelwert	9,9493
	Standardabweichung	3,98365
Extremste Differenzen	Absolut	,231
	Positiv	,231
	Negativ	-,168
Kolmogorov-Smirnov-Z		,865
Asymptotische Signifikanz (2-seitig)		,443

a. Die zu testende Verteilung ist eine Normalverteilung.

b. Aus den Daten berechnet.

Correlation: not significant

Korrelationen

		Interviewzeit	d_prime
Interviewzeit	Korrelation nach Pearson	1	-,402
	Signifikanz (2-seitig)		,154
	N	14	14
d_prime	Korrelation nach Pearson	-,402	1
	Signifikanz (2-seitig)	,154	
	N	14	14

Result of Kolmogorov-Smirnov-Test: Normal distribution

Kolmogorov-Smirnov-Anpassungstest

		Berufserfahrung	Alter	Geschlecht
N		14	14	14
Parameter der Normalverteilung[a,b]	Mittelwert	5,2500	44,0000	1,5000
	Standardabweichung	4,57733	9,68742	,51887
Extremste Differenzen	Absolut	,307	,153	,332
	Positiv	,307	,128	,332
	Negativ	-,239	-,153	-,332
Kolmogorov-Smirnov-Z		1,151	,573	1,244
Asymptotische Signifikanz (2-seitig)		,142	,898	,091

a. Die zu testende Verteilung ist eine Normalverteilung.

b. Aus den Daten berechnet.

Correlation: no significance

Korrelationen

		Berufserfahrung	Alter	Geschlecht	d_prime
Berufserfahrung	Korrelation nach Pearson	1	,394	,364	-,026
	Signifikanz (2-seitig)		,164	,200	,930
	N	14	14	14	14
Alter	Korrelation nach Pearson	,394	1	,015	-,512
	Signifikanz (2-seitig)	,164		,959	,061
	N	14	14	14	14
Geschlecht	Korrelation nach Pearson	,364	,015	1	,017
	Signifikanz (2-seitig)	,200	,959		,954
	N	14	14	14	14
d_prime	Korrelation nach Pearson	-,026	-,512	,017	1
	Signifikanz (2-seitig)	,930	,061	,954	
	N	14	14	14	14