sciendo

EUROPEAN
POLYGRAPH
PUBLISHED SEMI-ANNUALLY
2025 VOLUME 19 NUMBER 1 (61)

DOI: 10.2478/EP-2025-0003

You really don't recognise him? The eye-tracker as a forensic tool for concealed knowledge detection

Weronika Celniak*
Dominika Słapczyńska**
Gowthami Krishnan***
Piotr Augustyniak****

Abstract

The Concealed Information Test (CIT), a well-established tool in forensic investigations, has thus far been utilised to measure autonomic nervous system (ANS) changes associated with concealed information. While previous studies have explored the integration of eye-tracking technology in face recognition, the specific application of CIT within a mock crime scenario remains relatively uncharted territory. In this study, we aim to broaden the scope of eye-tracking applications using a mock crime scenario, as well as a machine learning classification method to detect hidden crime-related information.

* Department of Biocybernetics and Biomedical Engineering, AGH University of Kraków, Kraków, Poland; https://orcid.org/0000-0002-1714-7776; wcelniak@agh.edu.pl
** Polish Forensic Sciences Association, Warszawa, Poland; https://orcid.org/0000-001-6197-5029; dslapczynska@gmail.com
*** Department of Biocybernetics and Biomedical Engineering, AGH University of Kraków, Kraków, Poland; https://orcid.org/0000-0002-5444-0875; gkrishnan@agh.edu.pl
**** Department of Biocybernetics and Biomedical Engineering, AGH University of Kraków, Kraków, Poland; https://orcid.org/0000-0001-5986-3247; august@agh.edu.pl

Of the four faces displayed as stimuli, the 'guilty' group volunteers in the test were able to recognise one as they had previously seen it in the context of the mock crime, whereas the 'innocent' group volunteers were all unfamiliar with all four faces. We chose heatmaps depicting the fixation count and fixation durations as the input data for classification. The results obtained with features extracted using ResNet50 and the Support Vector Machine algorithm yielded promising outcomes, achieving an accuracy level of 84.62% for heat maps created using fixation count. These findings suggest the potential development of an innovative tool capable of objectively determining whether an examined person recognises individuals presented in photos, even when denying familiarity with those individuals. The integration of eye-tracking technology and machine learning holds promise for enhancing the accuracy and efficacy of concealed information detection in forensic contexts.

Key words: Concealed Information Test, eye-tracker

Introduction

The detection of concealed crime knowledge in the forensic field can be carried out using a polygraph which measures Autonomic Nervous System (ANS) responses (e.g., changes in breathing, electrodermal, and cardiovascular activity) during the Concealed Information Test (CIT). This method allows for detection of hidden knowledge by measuring responses to stimuli of different meanings. According to CIT theory, only knowledgeable individuals react differently to crime-related items called relevant stimuli (e.g., a stolen ring) presented together with unfamiliar items (e.g., other pieces of jewellery). These temporal changes in physiological reactivity to relevant items reflect the psychological cognitive process of orienting response, caused by the guilty person's attention shifting to the significant items they recognise from committing the crime (Klein Selle, 2022). Additionally, the intention to conceal crime knowledge is associated with an inhibition process involving the examinee's attempt to control their own behaviour (Klein Selle, 2018). Through CIT research, it has been determined that a knowledgeable examinee shows both arousal in electrodermal activity in response to relevant item, and a decrease in heartbeat and breathing rates.

Based on a well-established theoretical foundation, the CIT paradigm offers a framework for detecting crime-related memories using other tools, such as an eye-tracker, which appears to overcome several shortcomings of the polygraph. One of these issues concerns the vulnerability of ANS measures to countermeasures, which can reduce the identification rate of guilty subjects from 80% to 40% when mental countermeasures are used, and down to 10% when physical countermeasures are applied (Honts, 1996).

Current research has shown that eye-tracking, when applied to uncover hidden memories, is a useful method that is resistant to countermeasures (Lancry-Dayan, 2018). Additionally, the incorporation of eye-tracking into the CIT procedure creates the possibility of a fully automatic, contactless memory detection test based on specific gaze behaviour patterns. Considering that such solutions are currently unavailable commercially, the development of this highly effective forensic tool could represent a breakthrough discovery, significantly supporting the daily work of the police or border guard authorities.

To investigate this idea we examined the usefulness of heat maps—one of the eye-tracking data visualisation tools—and a simple machine learning algorithm to detect concealed information by classifying gaze patterns presented in the form of heat maps.

Whereas most studies (Delmas, 2023) using eye-trackers and the CIT focus solely on tasks related to remembering stimuli (e.g., photos, cards, or objects, without the direct involvement of participants in activities resembling those that are penalised), our data were recorded during an experimental design based on a mock theft scenario. The selected experimental design was intended to increase the ecological validity of the study, so, to date, we are among the few researchers who have used an eye-tracker with the CIT in a mock crime scenario.

Related work

An eye tracker can be used in three ways in forensic credibility assessment: during interrogation (Speth, 2021), in the analysis of eye movement during reading activity, or in tests based on recognition of familiar information. The second approach, which is already commercially available, can achieve an accuracy of over 80% (Kircher, 2016).

The latter still requires extensive research to simplify the procedure while achieving high effectiveness. In recent years, several studies that examined the relationship between stimuli recognition and oculomotor features have provided useful to developing the effective CIT protocol. The most important findings showed that:

- eye movements and the human memory system are strongly related (eyes move not only to receive sensory visual information, but also to bring to mind information stored in the memory) (Brockmole, 2005);

– gaze behaviour changes significantly when individuals follow different tasks (Yarbus, 1967);
– gaze is directed towards personally meaningful information (during free viewing of familiar and unfamiliar items, visual attention is directed towards the familiar ones) (Ryan, 2007);
– gaze can be modified to support the observer's goal (Welchman, 2003).

Furthermore, the results of eye-tracking face recognition research seem particularly interesting when considering their implementation in the security field to identify concealed criminal associations. Eye-tracking studies have demonstrated that it is possible to detect recognition of familiar faces, such as criminal associates, when individuals attempt to deny knowing them (fewer fixations and longer fixation durations during face viewing) (Millen, 2019).

It is already known that the face is a particular stimulus analysed and remembered in a certain way depending on the time of face presentation (Iskra, 2016). Additionally, face perception may involve different cognitive processes compared to object or scene perception. When viewing face images, longer fixation durations were recorded than when viewing other types of image content (e.g., a nature scene) (Guo, 2005). Facial features such as the eyes, nose, and mouth provide important information for face recognition.

Further studies have shown that various conditions in CIT experiments with face stimuli elicited different changes in gaze behaviour. The concealed knowledge of familiar faces during free viewing in the experiment with the visual detection task (Nahari, 2019) and the recognition task (Schwedes, 2011) caused the participants to preferentially direct their gaze toward known faces. The analysis indicated an increase in the number of fixations, visits, and the average duration of fixation (Otsuka, 2019). However, when participants took part in a short-term memory task (STM-CIT) (Lancry-Dayan, 2018), which required prior encoding of faces, their gaze towards known faces showed only a brief preference. Subsequently, they stopped focusing on the known stimuli.

These studies deserve particular attention, as in ocular-based CIT involving the simultaneous presentation of crime-related and crime-unrelated items, the location-by-location dynamics of the examinee's visual attention can be monitored (Schwedes, 2011). Moreover, the STM-CIT paradigm is resistant to countermeasures—which is supposed to reflect the interplay between task demands and the ability to voluntarily control gaze behaviour—and also leads to the maximisation of differences between responses to relevant and irrelevant items.

Contrary to the modest accuracy of the classic version of CIT with sequential presentation of stimuli (above 63%), the average efficiency of memory detection in STM-CIT with simultaneous stimuli presentation is 89% (Lancry-Dayan, 2018). It is worth emphasising that multiple-stimuli display in CIT is an interesting solution, made possible only through eye-tracker measurement.

Materials and Methods

This research study received approval from the ethics committee at AGH University of Kraków. Before giving their consent to participate in the study, each participant was briefed on the research procedure. The research sample consisted of 39 volunteers, most of whom were AGH students and employees. Participants' ages ranged from 19 to 60 years (M = 28.05, SD = 9.5) with 56% identified as female, and 44% as male.

Each participant was randomly assigned to the control or experimental group, while maintaining a similar number of individuals in both groups (20 "guilty" and 19 "innocent" participants). All participants in the experiment were motivated by a non-monetary reward instead of the standard procedure, which typically employs money as the motivating factor. The chosen reward consisted of a guided visit to one of limited-access laboratories at AGH, a unique opportunity not available to everyone, as well as the opportunity to participate in a true polygraph test with CIT after an eye-tracking examination. This type of reward is expected to enhance intrinsic motivation by offering an exclusive and meaningful experience.

The motivation of participants was monitored using a six-point scale to gauge their self-reported adherence to instructions (1 = not motivated at all, 6 = very motivated). Those in the "guilty" group were instructed to conceal their recognition of key items, while those in the "innocent" group were directed to prove their innocence by cooperating with the examiner during the tests. The average motivation score was 5.05 for the "guilty" group and 5.0 for the control group. This slight numerical difference was statistically insignificant, indicating that the perceived level of involvement was comparable between the two groups.

Study protocol

The study consisted of two stages. The first involved simulating a crime, while the second stage encompassed an eye-tracking CIT with a visual stimulus in the form of the victim's face. The simulated crime scenario consisted of several steps. Following instructions from the study coordinator, participants in the "guilty" group were asked to secretly steal a test exam from an assistant professor's laptop in his office. They accessed the room by using a key hidden in the assistant's coat, which was placed in a specific location at the university as described in the instructions. A photograph of the assistant—whom they were directed to "rob"—was displayed in the office on the front page of the folder that participants were told to open in order to retrieve the password necessary to access the exam paper. To avoid leaving any traces that could lead to their identification, they were instructed to wear a black cap and latex gloves, both provided to them in an envelope along with the mock crime instructions. The entire mock crime phase lasted an average 15–20 minutes per participant. The individual steps were designed to increase the participants' engagement in the study, and more closely replicate in-field conditions. Innocent participants were unaware of the details of the mock theft details or its course and they did not have the opportunity to see the face of the assistant professor. So in their instructions they were simply asked to take a test to confirm their lack of knowledge regarding the crime. All participants were advised to remain in character throughout the tests and not to disclose to anyone which group they had been assigned to. In addition, the challenge for those in the "guilty" group was to try to outsmart the eye-tracking device by concealing any familiarity with the details of the mock crime, although no specific instructions were provided for doing so.

The Concealed Information Test with an eye tracker incorporated the simultaneous presentation of visual stimuli. Participants viewed a slide containing four faces_one depicting the assistant's face from the theft simulation, while the other three serving as control stimuli, as shown in Fig. 1. It is important to note that all facial images used in the experiment were sourced from the Chicago Face Database (Ma, 2015). Before presenting the slide, a short video featuring an AI-generated avatar was shown on the screen in front of the participants. The avatar asked the participants the following question: "Do you know who uses the room that was broken into?". Following the video, a fixation cross was displayed on the screen to compel participants to focus on the centre, thereby preventing any disturbance to measurements during the stimulus presentation. The decision to use prepared footage instead of having the investigator ask the questions directly was intended to ensure consistent test conditions across all participants.

Fig. 1. Visual stimulus used in the study (face of the "robbed" assistant in the top left corner)

Eye tracking examination set up

Eye movements were recorded using the Tobii Pro Fusion desktop eye-tracker and Tobii Pro Lab software, with visual stimuli displayed on a 27-inch Full HD monitor at 60 frames per second. The screen was positioned at a distance of 70 cm from the participant. This setup facilitated straightforward data collection, analysis, and aggregation of eye-related metrics for further examination and visualisation. The eye-tracker was positioned below the computer screen to ensure an unobstructed view for participants.

Data analysis

In order to automatically classify participants into two groups—"guilty" and "innocent"—heat maps generated using the Tobii Pro Lab software were used. A heat map is a graphical representation of data in which values in a matrix are represented as colours. The purpose of a heat map is to visualise the magnitude of a phenomenon as a colour scale that ranges from cool to warm, with warmer colours indicating higher values and cooler colours indicating lower values. For the heatmap generation, we selected two parameters commonly used to describe fixations in the participants' visual pathways: fixation duration and the number of fixations. Only fixation characteristics were selected as the discriminative parameters in this study, rather than other eye-tracking metrics such as pupil dilation used in previous CIT studies, due to the known temporal delay in pupil responses relative to the eliciting stimulus (Partala, 2003). As a result, the spatial locations indicated on the heatmaps—which reflect gaze distribution—would not accurately correspond to the moments when pupil diameter changes occurred. Therefore, fixation-based metrics offer a more precise spatial representation of attention and cognitive processing during the test. Using heat maps, we were able to present the magnitude of these parameters while maintaining spatial information corresponding to gaze tracking, and thus to displayed stimuli. Examples of a heat maps for 'guilty' and 'innocent' participant used in our study are shown in Fig. 2 and Fig. 3. Following the standard image classification pipeline, our next step was feature extraction. For this purpose, the ResNet50 model (He, 2016), pre-trained on ImageNet and available in the torchvision library, was employed. ResNet, initially designed for image classification, can be repurposed for feature extraction. To do so, the final fully connected (classification) layer must be removed and the images fed through the intermediate, pre-trained convolutional layers. By utilising the output from these layers, one can extract rich, hierarchical feature representations of the images. Before the feature extraction process, each image, containing a full slide with 4 stimuli, was resised to 224×224 pixels upon loading to serve as the input to the aforementioned network. This process yielded an image representation in the form of a vector with a length of 512.

The subsequent classification step involved the use of a simple machine learning classifier, specifically, the Support Vector Machine (SVM). Given the small size of the test sample, leave-one-out cross-validation was employed to evaluate the model performance and ensure its generalisability. This technique assesses the reliability of the classifier by systematically leaving out one observation from the dataset, training the model on the remaining data, and then testing it on the excluded observation. The Support Vector Machine can be used with a number of different kernels (i.e., a function that transforms input data into a higher-dimensional space), each with its unique set of parameters. To identify the optimal configuration, a comprehensive grid search approach was adopted. The tuned parameters included the 'C' value, which controls the trade-off between achieving a low error on the training data and minimising the model complexity, and 'gamma', which defines the influence of a single training example. Additionally, the kernel type (e.g., 'linear', 'poly', 'rbf', 'sigmoid') was selected to find the best decision boundary for the classification task. The same classification procedure was applied to heatmaps generated using both fixation parameters.

Tab. 1. Parameters used in Grid Search

parameter name	parameter range
kernel type	'linear', 'poly', 'rbf', 'sigmoid'
C	21 logarithmically spaced values ranging from 10^{-10} to 10^{10}
gamma	21 logarithmically spaced values ranging from 10^{-10} to 10^{10}, 'scale', 'auto'
degree	2,3,4,5
coef0	0,0.1,0.5,1,2

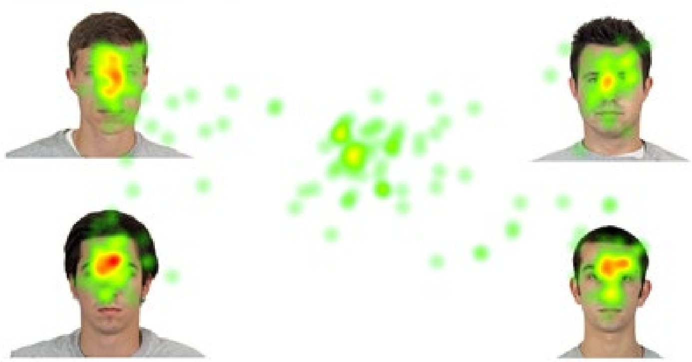

Fig. 2. Example of a generated heatmap depicting spatial distribution in the number of fixations for participant in the 'innocent' group

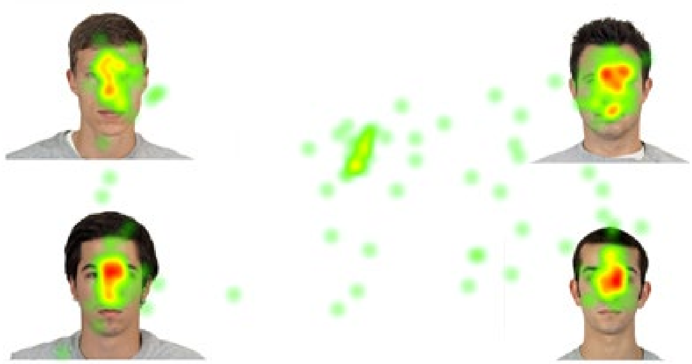

Fig. 3. Example of a generated heatmap depicting spatial distribution in the number of fixations for participant in the "guilty" group

Results

Durations of fixations

After conducting a thorough grid search, the sigmoid kernel with a C parameter of 10 was identified as the optimal choice. The model demonstrated an overall accuracy of 77%, successfully classifying 30 out of 39 participants. Among the participants, the model correctly identified 16 out of 20 deceptive participants, resulting in a sensitivity of 80%. It accurately identified 14 out of 19 truthful participants, achieving a specificity of approximately 74%. However, the model also encountered misclassifications, namely, it incorrectly classified 5 "innocent" participants as deceptive and failed to identify 4 "guilty" participants. These findings are visually represented in Fig. 4.

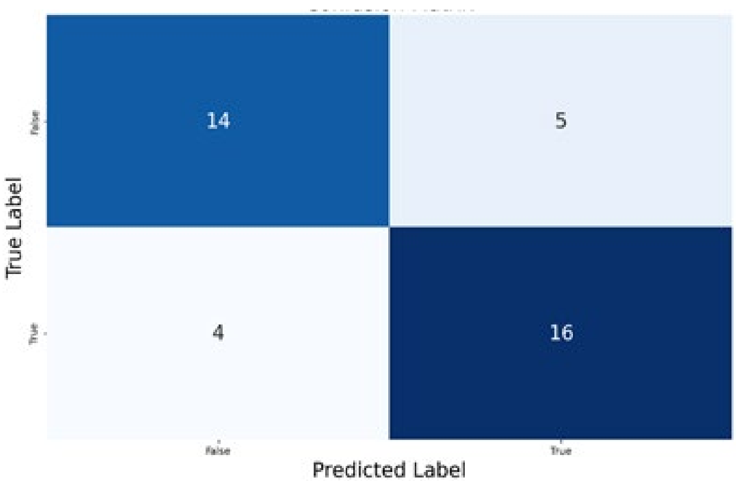

Fig. 4. Confusion matrix for a SVM classification of guilty and innocent participants using fixation duration

Number of fixations

For heatmaps generated using the number of fixations, the outcome of the grid search process was the selection of sigmoid kernel with the C parameter set to 100. The application of this approach resulted in an accuracy level of 84.62%, corresponding to the correct classification of 33 out of 39 study participants. Among the remaining 6 participants, 2 "guilty" individuals were not correctly identified, while 4 "innocent" participants were misclassified as deceptive. Thus, the classification model demonstrated a sensitivity of 90% (correctly identifying deceptive participants) and a specificity of approximately 78.9% (correctly identifying truthful participants) based on true positive (18), true negative (15), false positive (4), and false negative (2) classifications, as seen in Fig. 5.

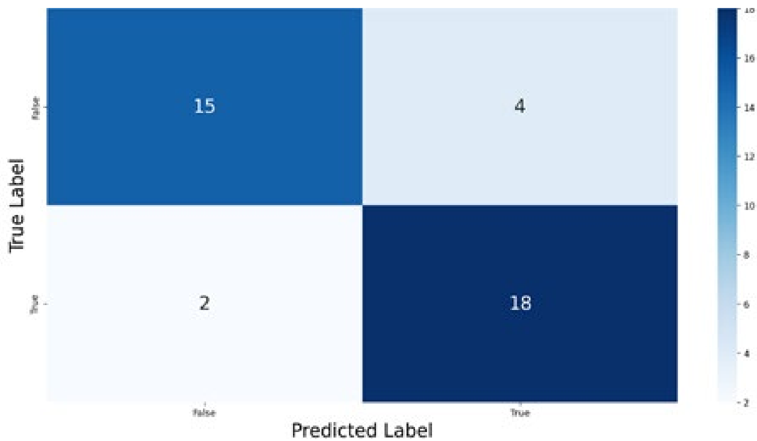

Fig. 5. Confusion matrix for a SVM classification of guilty and innocent participants using number of fixations

The results obtained using SVM trained on heatmaps representing fixation duration and fixations count are higher than those reported for traditional CIT using sequential stimulus presentation. However, better classification results were achieved using the fixation count as a feature rather than their duration. Additionally, it is important to note that similar SVM model parameters were identified as optimal in both cases.

Discussion

Face recognition during the process of identification can have important influences for the work of security institutions. A police lineup, border control, or the interviewing of a person suspected of cooperating with an organised crime group are just a few examples where an automatic, fully independent tool can be indispensable for detecting concealed knowledge.

Contrary to previous research, the current study explored the use of eye-tracker in a mock crime scenario to detect face recognition by the examined individuals. Relevant knowledge is typically acquired as a crime is committed, so it was justified to use a mock crime scenario instead of a memorisation task (such as STM-CIT).

The context in which one learns the relevant crime details during a mock crime is distinct from the context of learning stimuli presented on a computer-screen. Therefore, we assumed that under mock theft conditions, it might be easier to learn and remember crime details or recall the learning context in order to distinguish between familiar and unfamiliar faces.

The second advantage of our experiment is the use of the new version of the CIT procedure with simultaneous stimuli presentation. The aim of this modification was the reduction of the participants' task to merely the free viewing of facial stimuli presented on the screen without additional tasks such as a short-term memory or visual detection task. Previous studies (Delmas, 2023, Nahari, 2019, Lancry-Dayan, 2019) investigating the CIT effect demonstrated high accuracy, but the CIT versions employed were more complex and involved additional tasks with a different degree of difficulty.

In the present study, the CIT phase involved displays comprising only two kinds of stimuli: relevant (one familiar face) and irrelevant items (three unfamiliar faces). This aligns with the practice used in traditional polygraph examinations. The current study did not include the third type of stimulus—referred to as "target" (familiar but not crime related)—which is often used in the "oddball" variant of the CIT.

The version of the CIT presented here, despite its modification, yielded results similar to those obtained by other authors (Lancry-Dayan, 2018). This supports the conclusion that different experimental tasks can achieve high detection accuracy, provided they are specifically designed to reveal familiarity-induced modulation of eye movements (Nahari, 2019).

Furthermore, we found that the CIT protocol with simultaneous photo presentation was effective in detecting knowledge of familiar individuals, even when participants were instructed to conceal this familiarity. Further research is needed to investigate how the voluntary control of gaze interacts with task demands when participants are instructed on how to use countermeasures to mask their familiarity with the person being presented.

Furthermore, although the present study did not specifically focus on evaluating the efficacy of the protocol against countermeasures, we observed that it is feasible for the test administrator to detect whether a participant is cooperating, as some individuals exhibited non-compliant behaviour by fixating solely on the centre of the slide rather than directing their gaze toward the presented stimuli, as illustrated in Fig. 5.

Fig. 6. Example of non-cooperative gaze behaviour during CIT
(most of the fixation were constructed to the centre of the slide)

Conclusions

To sum up, our research demonstrated the potential of eye-tracker heat map analysis combined with a machine learning algorithm in the CIT, revealing an accuracy of 84.62% when using heatmaps generated with the number of fixations. This makes the method a promising alternative to traditional autonomic-CIT.

However, several limitations of this research should be noted. Firstly, the sample size was relatively small. Secondly, the study was conducted in a laboratory setting where participants faced no real consequences based on the test outcomes. Moreover, future research should investigate additional factors that may influence the sensitivity of eye-tracker examination with CIT, such as individual psychological characteristics and the use of countermeasures. Additionally, in future work, we plan to focus on increasing the sample size and exploring the use of additional types of stimuli beyond facial images. This will help assess the generalisability of the method and its applicability across different contexts of concealed information detection.

If further research yields promising results, it could support the development of an objective tool capable of detecting whether an examined individual recognises a person presented in photographs even if they deny it. This could be useful in identifying criminal connections between individuals who act together in a terrorist group for example. To the best of the authors' knowledge, lie detection tests using eye-tracking technology are currently commercially available only for reading-based activities, whereas experimental CIT based on eye movement analysis may offer higher forensic value and investigative utility.

References

Brockmole J.R. & Irwin D.E. (2005), Eye movements and the integration of visual memory and visual perceprion, *Preception&psychophysics*

Delmas H., Ciocan C., Novopashyna M. & Paeye C. (2023), Resistance of a short-term memory concealed information test with famous faces to countermeasures, *Memory & Cognition,* Springer Science and Business Media LLC.

Guo K., Mahmoodi S., Robertson R.G. & Young M. P. (2005), Longer fixation duration while viewing face images. *Experimental Brain Research*, 171(1).

He K., Zhang X., Ren S. & Sun J. (2016), Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*

Honts C.R., Devitt M.K., Winbush M. & Kircher J.C. (1996), Mental and physical countermeasures reduce the accuracy of the concealed knowledge test.

Iskra A. & Tomc H.G. (2016), Eye-tracking analysis of face observing and face recognition. *Journal of Graphic Engineering and Design*, 7(1).

Kircher J.C. & Raskin D.C. (2016), Laboratory and Field Research on the Ocular-motor Deception Test. *European Polygraph, 4(38)*

Klein Selle N. & Ben-Shakhar G. (2022), A new theoretical perspective on concealed information detection. *Psychophysiology*, 60(3).

Klein Selle N., Verschuere B., Ben-Shakhar G. (2018). Concealed Information Test, *Detecting Concealed Information and Deception*. Elsevier.

Lancry-Dayan O.C., Nahari T., Ben-Shakhar G. & Pertzov Y. (2018), Do you know him? Gaze dynamics toward familiar faces on a Concealed Information Test. *Journal of Applied Research in Memory and Cognition*, 7(2).

Ma D.S., Correll J. & Wittenbrink B. (2015), The Chicago face database: A free stimulus set of faces and norming data, *Behavior Research Methods*, 47.

Millen, A.E. & Hancock, P.J. (2019), Eye see through you! Eye tracking unmasks concealed face recognition despite countermeasures. *Cognitive Research: Principles and Implications*, *4*.

Nahari T., Lancry-Dayan O., Ben-Shakhar G. & Pertzov Y. (2019), Detecting concealed familiarity using eye movements: the role of task demands, *Cognitive Research: Principles and Implications*, 4(1).

Otsuka T., Mizutani M. & Yamada F. (2019), Selective attention, not cognitive load, elicited fewer eyeblinks in a concealed information test, *Biological Psychology*, 142.

Partala, T., & Surakka, V. (2003), Pupil size variation as an indication of affective processing. *International Journal of Human-Computer Studies*, *59*(1–2), 185–198.

Ryan J.D., Hannula D.E. & Cohen N.J. (2007), The obligatory effects of memory on eye movements. *Memory*, 15(5).

Schwedes C. & Wentura D. (2011), The revealing glance: Eye gaze behavior to concealed information, *Memory & Cognition*, 40(4).

Speth J., Vance N., Czajka A., Bowyer, K.W., Wright D. & Flynn P. (2021), Deception Detection and Remote Physiological Monitoring: A Dataset and Baseline Experimental Results, *2021 IEEE International Joint Conference on Biometrics (IJCB)*.

Welchman A.E. & Harris J.M. (2003), Task demands and binocular eye movements, *Journal of Vision*.

Yarbus A.L. (1967), Eye Movements During Perception of Complex Objects. *Eye Movements and Vision,* Springer.