Interaction-based Implicit Calibration of Eye-Tracking in an Aircraft Cockpit

Simon Schwerd[a] and Axel Schulte
Institute of Flight Systems, Universität der Bundeswehr München, Neubiberg, Germany

Keywords: Eye-Tracking, Implicit Calibration, Aviation, Flight Deck.

Abstract: We present a method to calibrate an eye-tracking system based on cockpit interactions of a pilot. Many studies show the feasibility of implicit calibration with specific interactions such as mouse clicks or smooth pursuit eye movements. In real-world applications, different types of interactions often co-exist in the "natural" operation of a system. Therefore, we developed a method that combines different types of interaction to enable implicit calibration in operational work environments. Based on a preselection of calibration candidates, we use an algorithm to select suitable samples and targets to perform implicit calibration. We evaluated our approach in an aircraft cockpit simulator with seven pilot candidates. Our approach reached a median accuracy between 2° to 4° on different cockpit displays dependent on the number of interactions. Differences between participants indicated that the correlation between gaze and interaction position is influenced by individual factors such as experience.

1 INTRODUCTION

Eye-tracking is an important tool to study human factors in aircraft cockpits. In many simulator studies, gaze measurement is used to analyse pilot attention (van de Merwe et al., 2012; Ziv, 2016) or conduct research about pilot cognition (Dahlstrom & Nahlinder, 2009; Schwerd & Schulte, 2020; van de Merwe et al., 2012). With ongoing development, it might be used in real cockpits to complement "black box" flight recorder information (Peysakhovich et al., 2018) or to inform adaptive support systems (Brand & Schulte, 2021; Honecker & Schulte, 2017). Robust and accurate measurement of pilot gaze is fundamental to these undertakings. Therefore, proper calibration of the eye-tracking system plays an important role (Nyström et al., 2013).

2 BACKGROUND

Calibration of eye-tracking systems requires multiple pairs of calibration targets $t_i(x, y)$ and corresponding measured gaze samples $s_i(x, y)$. For *explicit calibration*, a user is required to fixate predefined calibration targets while gaze is measured. This time-consuming process might decrease acceptance of eye-tracking applications in aircraft cockpits. Further, measurement accuracy can deteriorate during operation due to head movements, blinking or a change in the user's relative position to the screen (M. X. Huang et al., 2016; Sugano & Bulling, 2015), which could go unnoticed without re-calibration.

With *implicit calibration*, no explicit user cooperation is required, and calibration targets are determined based on assumptions about probable or required fixation. This idea was originally proposed by Hornof and Halverson (2002), where participants in an experiment had to fixate a specific display position for task-relevant information which was used to monitor gaze accuracy. Following this idea, subsequent studies investigated implicit calibration based on the correlation of gaze and interaction such as mouse clicks (M. X. Huang et al., 2016; Sugano et al., 2008), scene properties like saliency (Kasprowski et al., 2019; Sugano & Bulling, 2015), moving objects (Blignaut, 2017; Drewes et al., 2019; Pfeuffer et al., 2013) or saccade properties (M. X. Huang & Bulling, 2019). Most of this research was conducted in the context of eye-tracking applications for untrained users, such as for webcams of internet users

[a] https://orcid.org/0000-0001-6950-2226

Schwerd, S. and Schulte, A.
Interaction-based Implicit Calibration of Eye-Tracking in an Aircraft Cockpit.
DOI: 10.5220/0011657200003417
In *Proceedings of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2023) - Volume 2: HUCAPP*, pages 55-62
ISBN: 978-989-758-634-7; ISSN: 2184-4321

(Papoutsaki et al., 2017) or for displays in public spaces (Khamis et al., 2016).

One of the most studied implicit calibration methods is the correlation of gaze and mouse clicks. Sugano et al. (2008) and Sugano et al. assumed, that gaze and mouse click position align when the click occurs and presented an algorithm that updated their gaze estimation with every click. According to J. Huang et al. (2012) the proximity of mouse cursor and gaze depends on the current mode of interaction (e.g. read or click) with a median distance of 70-80px for mouse clicks. In a subsequent study, the authors found, that the correlation of gaze and interaction is stronger before than exactly at the moment of interaction peaking at around 0.4s before interaction (M. X. Huang et al., 2016). In a more recent study, Zhang et al. (2018) implemented implicit calibration combining clicks, touches and keypresses over multiple devices. These studies showed the feasibility of interaction-based implicit calibration but pointed out two relevant problems: First, gaze data is usually noisy, which challenges the basic assumption, that gaze *always* aligns with an interaction. Second, task context and individual differences cause inconsistent alignment of gaze and interaction. Both these observations suggest that the interaction-based approach must be able to filter for correct gaze-interaction alignments.

The mentioned research showed the feasibility of implicit calibration. In most studies, the calibration is usually based on one specific type of interaction or scene property. From our review, we identified two open questions: First, how different interactions can be used in the same method for calibration. Second, if the implicit calibration assumptions hold in the use of human-machine-interfaces not specifically designed for this use case. We are convinced that implicit calibration will prove to be useful in operational workplaces to avoid explicit calibration and continuously verify eye-tracking quality. This paper makes the following contributions: First, we propose a method to select calibration target candidates that integrates different types of interactions. Second, we describe an algorithm, which can process these different types in a unified manner. Third, we evaluate the implementation of this system in our aircraft cockpit simulator.

3 METHOD

We propose a method with three stages displayed in Figure 1. The first step “Calibration Candidate Selection” is a classification to identify different types of interaction in natural human-system-interaction that are suitable for implicit calibration. After a calibration target and associated samples has been identified, the second module “Sample Selection” identifies a subset of samples that are likely to be a fixation upon the assumed target. After that, a suitable calibration procedure uses the new and past selected samples to calibrate the system.

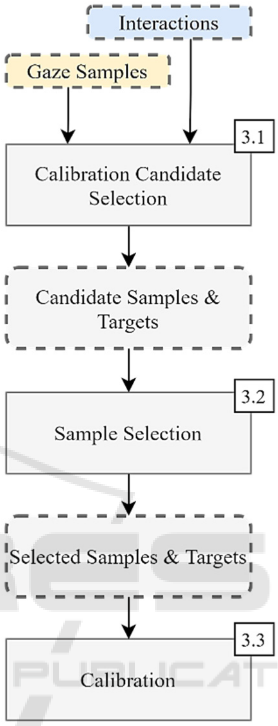

Figure 1: Calibration pipeline. Two stages of selection result in a selection of samples and calibration targets for calibration.

3.1 Selection of Calibration Candidates

To identify suitable calibration targets in a natural human-machine interaction, we propose the following approach: We think about interaction and gaze measurements in terms of two parallel data streams (see Figure 2): First, a gaze data stream, which arrives at high frequency but possibly interrupted (e.g., when the eye-tracker loses the track) and second, interactions, which either happen at distinct moments (e.g., click) or are continuous over a time period (e.g., hold gesture on a touch screen).

Given this perspective, we propose three categories to compute calibration targets and gaze sample candidates: an interaction is either a successor, companion, or precursor of a probable fixation target. We define an interaction to be a

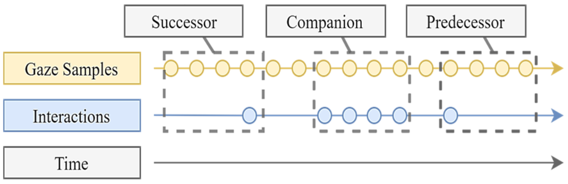

Figure 2: Gaze samples and interactions as parallel data streams with three types of calibration target candidates.

successor when the interaction follows a probable fixation target, for example mouse clicks, which succeed a fixation on the click position. After occurrence of a successor, the preceding gaze samples are selected as calibration candidates. Next, we define an interaction to be a *companion* when it likely occurs together with its associated fixation target. As an example, consider a slider widget in a desktop application that controls a numerical value displayed next to it. A user needs to fixate the changing number when he uses his mouse to slide, thus, it can be considered a fixation target. In this case, all gaze samples measured during an active companion interaction are calibration candidates. Finally, we define an interaction to be a *precursor* when the interaction occurs before a probable calibration target. Many interactions with the interface cause changes that are verified by the user just after interaction. To give a flight deck example, when a pilot moves the gear of their plane by control input, he confirms the position of the gear a few seconds after his input. The action and confirmation of the expected result could be used for implicit calibration, which is especially useful when the interaction is spatially or temporarily dislocated from the location where the change is displayed (e.g., control input and delayed indication of gear).

3.2 Sample Selection

The calibration candidate selection returns an assumed fixation position and associated gaze sample candidates for calibration that were measured either before, during or after the interaction. Since this data set is often noisy, we need to identify the most probable subset of samples representing the fixation on the assumed target. For this, we propose a simple algorithm inspired by RANSAC (Fischler & Bolles, 1981), which selects samples that are in both spatial and temporal proximity to the interaction. Pseudocode is depicted in Algorithm 1.

```
INTERACTION -TIME -CONSENSUS
Input: gaze-samples, calibration-target, interaction-time, threshold-dist, threshold-time,
minimum-samples
Result: Best samples for calibration
    all-consensus-groups ← empty list
    for sample-candidate in gaze-samples do:
        consensus-group ← gaze-samples as selected by candidate selection
        for consensus-candidate in consensus-group do:
            spatial-dist ← SPATIAL_DISTANCE (consensus-candidate, sample-candidate)
            time-dist ← TIME_DISTANCE_ABSOLUTE(consensus-candidate, interaction-time)
            if spatial-dist > threshold-dist or time-dist > threshold-time do:
                REMOVE consensus-candidate from consensus-group

        if SIZE(consensus-group) < minimum-samples do:
            continue
        ADD consensus-group to all-consensus-groups
    return consensus-group with lowest time to interaction
```

Algorithm 1: Algorithm to select most probable subset of fixation samples for calibration target.

3.3 Calibration

Optimally, the sample selection algorithm returns a set of suitable pairs of calibration targets $t_i(x, y)$ and measured gaze samples $s_i(x, y)$ while omitting targets with noisy measurements or insufficient sample size. Then, gaze can be calibrated when enough of these pairs have been identified. The method of calibration is flexible and depends on the use-case and desired fitting attributes. In the following, we present an experiment where we collected data in an experiment and applied a linear calibration function.

4 EXPERIMENT

To evaluate our approach, we conducted an experiment in a research cockpit simulator resembling a generic fast jet cockpit (see Figure 3), which is normally used to evaluate human-autonomy teaming applications (e.g., (Lindner & Schulte, 2020)). The cockpit contains three touchscreens and a Head-Up-Display (HUD) projected into the visuals of the outside world. The participant's gaze was measured with a commercially available four-camera system connected to the simulation software (SmartEye Pro 0.3MP) and gaze was tracked over the three touchscreen displays and the HUD.

In the experiment, we processed the uncalibrated screen positions provided by the eye-tracking system consisting of x- and y-coordinate and the associated screen. The basic idea of the experiment was to collect natural interaction data and uncalibrated gaze measurements in a flight mission, which was then used to perform implicit calibration. Results of this process were compared to the results of a baseline calibration based on a conventional 9-point calibration procedure.

4.1 Calibration Targets

There are a great number of different controls and displays in the cockpit simulator. The participants controlled their aircraft via throttle, stick, and touch interactions. Possible touch interactions were tap, hold, drag, and pinch dependent on the control element. From these different touch interactions, we used all taps as successors, and all holds as companions. Drag and pinch were not used. In the following, we want to give a brief overview of the most important elements on the touchscreen, which are also displayed in Figure 3.

The *tactical map* in the central touchscreen can be used for navigation and creation of mission tasks. It displays the aircraft, all unmanned systems as well as tactical objects and mission information. The map can be dragged and zoomed. It is possible to interact with tactical elements (e.g., building or waypoint) by tapping on the respective element, which is followed by a pop-up context menu used to plan mission tasks

(e.g., reconnaissance). Taps were used as predecessors.

On the left side display, the main interaction element is the *mission plan timeline*. After the creation of a task in the map, it can be inserted into the timeline by tapping a position within the timeline after which a task box moves to its position from the bottom of the screen. Taps were used as successors.

The *autopilot* control can be opened by tapping on the associated button in the side bar of the central display. The control contains buttons to increase and decrease autopilot speed, altitude and heading as well as buttons to engage autopilot modes. The numeric values can be adjusted by either tap or hold interactions on the increase and decrease buttons. The calibration target of each interaction lies on the position of the value display. Taps and holds were used as successor and companions, respectively.

The *HUD* shows several different flight parameters of the aircraft. For implicit calibration, we used the air brake indicator in the centre of the display. The participants controlled the air brake via a throttle button and the indicator in the shows its current position as illustrated by Figure 3. The interaction to control the air brake is considered a predecessor.

Figure 3: Simulator (top center) with relevant interaction elements: Mission timeline (top left), HUD air brake indicator (top right), autopilot control (bottom left), tactical map (bottom center) and context menu (bottom right)

4.2 Experimental Task

The experiment contained different tasks for the participants to trigger interactions on all displays. Before take-off, participants were asked to plan a mission, which required the creation of several tasks for their aircraft and a UAV. This included an interaction with the mission timeline. With their own aircraft, they had to take-off and fly over several way points or tactical objects, while their UAV investigated four different locations on the map. After the take-off, participants received messages with instructions to program the autopilot with given values. During the flight, the UAV sent pictures which had to be classified in two categories. The participant landed again at the airport after completion of all waypoints.

4.3 Participants & Procedure

We conducted the experiment with seven participants (all male, $\mu_{age} = 24.9y$). Four participants were students of aeronautical engineering and three were research assistants. All participants had prior experience in handling the simulator. In the beginning the participants received an introduction to the simulator and the specific tasks necessary to complete the experimental requirements. Then, each participant had to complete three training missions to get familiar with the mission tasks. Before each mission there was a briefing in which the tasks and circumstances were demonstrated. Prior to the experimental mission, we collected validation data based on a 9-point validation on every screen. After this data collection, the participant received a briefing for the experimental mission, which followed on the briefing.

4.4 Data Processing

During the 9-point validation procedure, we collected 100 samples for 9 points on each display. The error was computed by average Euclidian distance of each sample to the calibration target. In the experiment, we logged the following data: (1) position and timestamp of uncalibrated gaze measurements and (2) position, timestamp, type, and associated display element (e.g., autopilot value, map) of interaction. For calibration, we computed the constants that optimized error given the application of the following function to the samples for both x- and y-coordinates:

$$f(x) = a_0 + a_1 x \tag{1}$$

For the HUD, we used the following function since there was only one possible calibration point in the centre of the display:

$$f(x) = a_0 + x \tag{2}$$

Note, that we computed the calibration parameters for each display independently.

5 RESULTS

First, we aimed to identify the optimal set of parameters for the sample selection algorithm. After selecting a set of parameters, we evaluated how many calibration targets and samples were selected for calibration. Finally, we compared calibration performance of implicit calibration with 9-point calibration.

5.1 Optimal Parametrization of the Selection Algorithm

The algorithm presented in section 3.2 expects three parameters: distance threshold, time threshold and minimum number of samples. We performed a grid search to identify optimal parameters. We tested the combinations of the following parameters: distance threshold $[1.0, 1.5, 2.0]°$, time threshold $[0.5, 1.0, 1.5, 2.0]\ s$ and minimum samples $[10, 15, 20, 25]$. We found varying parameters for the participants suggesting individual differences in the gaze-interaction correlation. Also, an individual random component in gaze measurement could influence optimal distance threshold because a low threshold filters fixations with high variance. For the following analysis, we selected a set of parameters returning the best median calibration performance: distance threshold of 2.0°, time threshold of $2.0s$ and minimum samples of 15. These parameters were used for all participants to generate the following results.

5.2 Interactions

Interaction with each display differed in frequency as shown by Table 1. Participants mainly interacted by taps because this was the primary way to operate the system. In addition, some participants did not use a hold interaction to control the autopilot but preferred to tap the buttons multiple times, which is one reason for the high variations of successors and companions in the centre display.

Table 1: Number of occurrences of the different forms of interactions, formatted as “Mean (Standard Deviation)”.

	$n_{successor}$		$n_{companion}$		$n_{predecessor}$	
Display	All	Selected	All	Selected	All	Selected
Center	174.4 (40.6)	147.0 (44.3)	3.7 (3.8)	0.7 (0.5)	-	
Side Left	37.6 (19.0)	37.1 (20.6)	-		-	
Side Right	21.6 (13.9)	20.4 (13.4)	-		-	
HUD	-		-		5.0 (3.0)	4.0 (3.2)

There were very few samples for predecessor interaction on the HUD because was only controlled to land the aircraft. One participant did not use the air brake at all. In average, the sample selection removed 15.8% of successors and 81.5% of companions because no valid sample group was found. Surprisingly, not many successor interactions on the left and right display were removed.

5.3 Calibration Performance

We computed calibration constants with the selected calibration targets and samples from the experiment. Using these constants, we calibrated the data collected in the 9-point validation data of each display. Then, we compared the error of this implicitly calibrated data with uncalibrated and point-calibrated data. Results are displayed in Figure 4.

There are improvements against no calibration for most participants on most displays but there are also instances where our method deteriorated the gaze measurement. Table 2 reports absolute and relative values for implicit calibration. Median accuracy is between 2.1° on the left display to 3.8° on the HUD. For very few cases, improvement by implicit calibration was comparable to point calibration (e.g., see Center, P7 or Side Left P6). Standard deviation is high on the HUD, because of a great implicit calibration error for participant 2 and 4.

Table 2: Median accuracy and relative improvement of implicit calibration.

Display	Median Accuracy (Std)	Median relative improvement (Std)
Center	2.2° (0.96°)	44% (65%)
Side Left	2.1° (1.03°)	22% (41%)
Side Right	2.8° (0.96°)	36% (36%)
HUD	3.8° (3.67°)	0% (136%)

5.4 Discussion

The results showed that the implicit calibration method can improve average accuracy on the 9-point validation but did not reach point calibration quality. Therefore, we conclude that the basic assumptions of implicit calibration do hold for interactions in work environments like an aircraft cockpit, but there are limitations to the presented approach. First, the parameter search revealed that the algorithm had different optimal parameters for different participants which indicates an individual component in implicit calibration. We assume, that two factors are responsible for these individual components. First, participants had different levels of training on the system e.g., the research assistants developed parts of this prototype and students had only experience from prior experiments. When a user is very experienced, he anticipates system reactions and might be very quick in his interaction. Second, individual factors such as computer game experience might also be a confounding factor influencing the correlation between gaze and interaction. Apart from individual difference, another problem is the varying frequency of different interaction types. In our use-case, successor interactions (taps on the central touchscreen) occurred far more often than other forms and therefore had the largest influence on the results. Contrary, the HUD had very few interactions on one possible calibration target, which was problematic in two ways: First, when our method falsely selects samples where the implicit calibration assumption does not apply, there is a great accuracy degradation due to false calibration. The risk of false calibration is especially high when sample size is low. Second, when there are very few targets on the screen (e.g., on the HUD), calibration procedure might overfit areas where most interactions happen.

Our experiment had two main limitations. First, although all participants had prior experience on the simulator, the degree of proficiency differed, which could also explain individual calibration differences. Second, the experimental design did not allow to analyse the problem of continuous deterioration during measurement. We collected validation data before the experiment, so gaze tracking quality change during the experiment could influence implicit calibration performance.

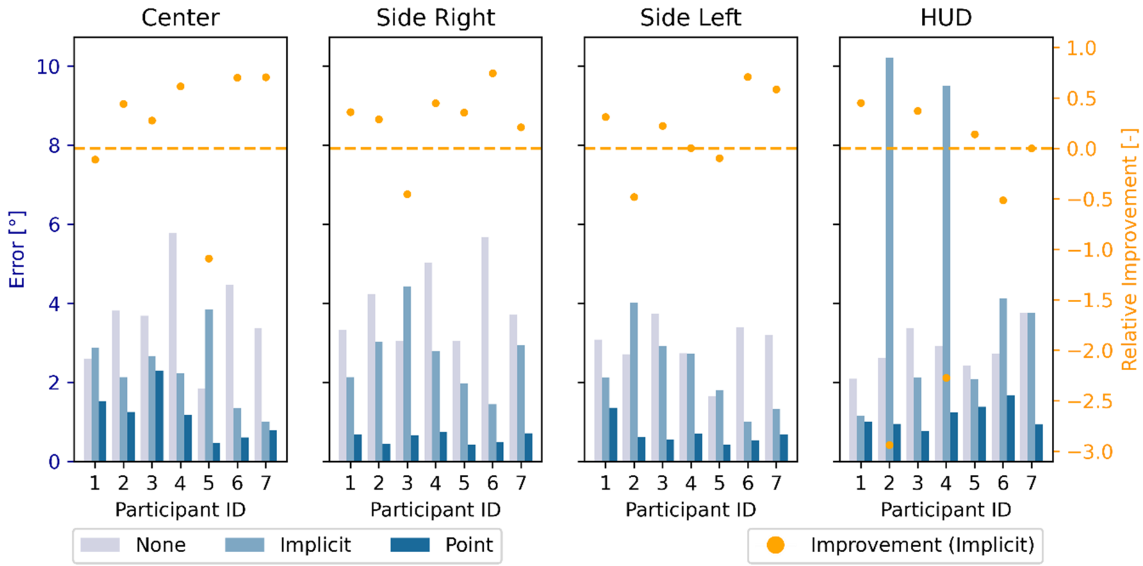

Figure 4: Comparison of no, implicit and point calibration on each display with 9-point validation data (bar, blue/gray). Relative improvement is displayed only for implicit calibration (point, orange). Note: There were no samples for Participant 7 on the HUD.

6 CONCLUSION

Our approach demonstrated the feasibility and the challenges of implicit calibration in work environments, specifically on flight deck. We are planning to improve the current approach by integrating smooth-pursuit calibration, which in our case, could be implemented for moving objects on the map or in the external view. Another possible improvement is to extent the predecessor interactions based on informed use cases of a professional pilot. Since pilot interaction is standardized to a high degree, there are many tasks in the cockpit that could be leveraged for implicit calibration like for example take-off procedures or checklists.

REFERENCES

Blignaut, P. (2017). Using smooth pursuit calibration for difficult-to-calibrate participants. *Journal of Eye Movement Research*, *10*, 1–13. https://doi.org/10.16910/jemr.10.4.1

Brand, Y., & Schulte, A. (2021). Workload-adaptive and task-specific support for cockpit crews: design and evaluation of an adaptive associate system. *Human-Intelligent Systems Integration*, *3*(2), 187–199. https://doi.org/10.1007/s42454-020-00018-8

Dahlstrom, N., & Nahlinder, S. (2009). Mental Workload in Aircraft and Simulator During Basic Civil Aviation Training. *The International Journal of Aviation Psychology*, *19*(4), 309–325. https://doi.org/10.1080/10508410903187547

Drewes, H., Pfeuffer, K., & Alt, F. (2019). Time- and Space-Efficient Eye Tracker Calibration. In *ETRA '19, Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. Association for Computing Machinery. https://doi.org/10.1145/3314111.3319818

Fischler, M. A., & Bolles, R. C. (1981). Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Commun. ACM*, *24*(6), 381–395. https://doi.org/10.1145/358669.358692

Honecker, F., & Schulte, A. (2017). Automated Online Determination of Pilot Activity Under Uncertainty by Using Evidential Reasoning. In D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics: Cognition and Design* (pp. 231–250). Springer International Publishing.

Hornof, A. J., & Halverson, T. (2002). Cleaning up systematic error in eye-tracking data by using required fixation locations. *Behavior Research Methods, Instruments, & Computers : A Journal of the Psychonomic Society, Inc*, *34*(4), 592–604. https://doi.org/10.3758/bf03195487

Huang, J., White, R., & Buscher, G. (2012). User see, user point: Gaze and cursor alignment in web search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1341–1350). https://doi.org/10.1145/2207676.2208591

Huang, M. X., & Bulling, A. (2019). SacCalib. In K. Krejtz & B. Sharif (Eds.), *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications* (pp. 1–10). ACM. https://doi.org/10.1145/3317956.3321553

Huang, M. X., Kwok, T. C., Ngai, G., Chan, S. C., & Leong, H. V. (2016). Building a Personalized, Auto-Calibrating Eye Tracker from User Interactions. In *CHI '16, Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 5169–5179). Association for Computing Machinery. https://doi.org/10.1145/2858036.2858404

Kasprowski, P., Harezlak, K., & Skurowski, P. (2019). Implicit Calibration Using Probable Fixation Targets. *Sensors (Basel, Switzerland)*, *19*(1). https://doi.org/10.3390/s19010216

Khamis, M., Saltuk, O., Hang, A., Stolz, K., Bulling, A., & Alt, F. (2016). TextPursuits: Using Text for Pursuits-Based Interaction and Calibration on Public Displays. In *UbiComp '16, Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 274–285). Association for Computing Machinery. https://doi.org/10.1145/2971648.2971679

Lindner, S., & Schulte, A. (2020). Human-in-the-Loop Evaluation of a Manned-Unmanned System Approach to Derive Operational Requirements for Military Air Missions. In D. Harris & W.-C. Li (Eds.), *Engineering Psychology and Cognitive Ergonomics. Cognition and Design* (pp. 341–356). Springer International Publishing.

Nyström, M., Andersson, R., Holmqvist, K., & van de Weijer, J. (2013). The influence of calibration method and eye physiology on eyetracking data quality. *Behavior Research Methods*, *45*(1), 272–288. https://doi.org/10.3758/s13428-012-0247-4

Papoutsaki, A., Laskey, J., & Huang, J. (2017). Searchgazer: Webcam Eye Tracking for Remote Studies of Web Search. In R. Nordlie (Ed.), *ACM Digital Library, Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval* (pp. 17–26). ACM. https://doi.org/10.1145/3020165.3020170

Peysakhovich, V., Lefrançois, O., Dehais, F., & Causse, M. (2018). The Neuroergonomics of Aircraft Cockpits: The Four Stages of Eye-Tracking Integration to Enhance Flight Safety. *Safety*, *4*(1), 8. https://doi.org/10.3390/safety4010008

Pfeuffer, K., Vidal, M., Turner, J., Bulling, A., & Gellersen, H. (2013). Pursuit Calibration: Making Gaze Calibration Less Tedious and More Flexible. In *UIST '13, Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology* (pp. 261–270). Association for Computing Machinery. https://doi.org/10.1145/2501988.2501998

Schwerd, S., & Schulte, A. (2020). Experimental Validation of an Eye-Tracking-Based Computational Method for Continuous Situation Awareness Assessment in an Aircraft Cockpit. In D. Harris & W.-C. Li (Eds.), *Lecture Notes in Computer Science. Engineering Psychology and Cognitive Ergonomics. Cognition and Design* (Vol. 12187, pp. 412–425). Springer International Publishing. https://doi.org/10.1007/978-3-030-49183-3_32

Sugano, Y., & Bulling, A. (2015). Self-Calibrating Head-Mounted Eye Trackers Using Egocentric Visual Saliency. In C. Latulipe, B. Hartmann, & T. Grossman. (Eds.), *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (pp. 363–372). ACM. https://doi.org/10.1145/2807442.2807445

Sugano, Y., Matsushita, Y., & Sato, Y. (2013). Appearance-based gaze estimation using visual saliency. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(2), 329–341. https://doi.org/10.1109/TPAMI.2012.101

Sugano, Y., Matsushita, Y., Sato, Y., & Koike, H. (2008). An Incremental Learning Method for Unconstrained Gaze Estimation. *European Conference on Computer Vision*, *5304*, 656–667. https://doi.org/10.1007/978-3-540-88690-7_49

van de Merwe, K., van Dijk, H., & Zon, R. (2012). Eye Movements as an Indicator of Situation Awareness in a Flight Simulator Experiment. *The International Journal of Aviation Psychology*, *22*(1), 78–95. https://doi.org/10.1080/10508414.2012.635129

Zhang, X., Huang, M. X., Sugano, Y., & Bulling, A. (2018). Training Person-Specific Gaze Estimators from User Interactions with Multiple Devices. In R. Mandryk, M. Hancock, M. Perry, & A. Cox (Eds.), *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–12). ACM. https://doi.org/10.1145/3173574.3174198

Ziv, G. (2016). Gaze Behavior and Visual Attention: A Review of Eye Tracking Studies in Aviation. *The International Journal of Aviation Psychology*, *26*(3-4), 75–104. https://doi.org/10.1080/10508414.2017.1313096