Comparison of Webcam and Remote Eye Tracking

Katarzyna Wisiecka
Krzysztof Krejtz
Izabela Krejtz
kwisiecka@swps.edu.pl
kkrejtz@swps.edu.pl
ikrejtz@swps.edu.pl
SWPS University
Warsaw, Poland

Damian Sromek
Adam Cellary
Beata Lewandowska
damian@realeye.io
adam@realeye.io
beata@realeye.io
RealEye, Sp. z o.o.
Poznań, Poland

Andrew T. Duchowski
duchowski@clemson.edu
Clemson University
Clemson, SC, USA

(a) RealEye. (b) Gazepoint GP3. (c) RealEye + GP3.

Figure 1: Heatmap visualization of gaze data collected by: (a) webcam, (b) remote eye tracker, and (c) integrated solution.

ABSTRACT

We compare the measurement error and validity of webcam-based eye tracking to that of a remote eye tracker as well as software integration of both. We ran a study with $n = 83$ participants, consisting of a point detection task and an emotional visual search task under three between-subjects experimental conditions (webcam-based, remote, and integrated). We analyzed location-based (e.g., fixations) and process-based eye tracking metrics (ambient-focal attention dynamics). Despite higher measurement error of webcam eye tracking, our results in all three experimental conditions were in line with theoretical expectations. For example, time to first fixation toward happy faces was significantly shorter than toward sad faces (the happiness-superiority effect). As expected, we also observed the switch from ambient to focal attention depending on complexity of the visual stimuli. We conclude that webcam-based eye tracking is a viable, low-cost alternative to remote eye tracking.

CCS CONCEPTS

• **Software and its engineering** → **Empirical software validation**; • **Human-centered computing** → **Laboratory experiments**.

KEYWORDS

webcam eye tracking, gaze detection, remote eye tracking

ACM Reference Format:

Katarzyna Wisiecka, Krzysztof Krejtz, Izabela Krejtz, Damian Sromek, Adam Cellary, Beata Lewandowska, and Andrew T. Duchowski. 2022. Comparison of Webcam and Remote Eye Tracking. In *2022 Symposium on Eye Tracking Research and Applications (ETRA '22), June 8–11, 2022, Seattle, WA, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3517031.3529615

1 INTRODUCTION

Webcam-based eye tracking is a promising method to record eye movements in natural, ecological settings. Relative low cost with high speed of data acquisition make this method increasingly popular within the eye tracking community. As with any novel method, web-based eye tracking raises concerns about its accuracy and validity, but research on this topic is sparse. The present paper addresses this gap by testing webcam-based eye tracking accuracy, precision and validity against a widely available remote eye tracker.

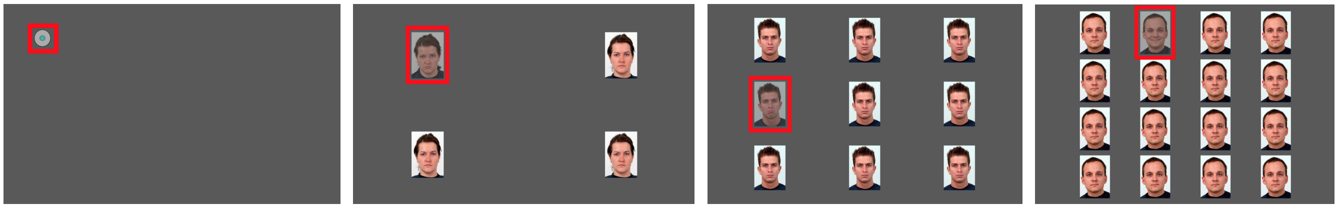
We present results from two experimental tasks supporting the claim that, despite lower accuracy and precision, webcam-based eye tracking is a highly reliable method, similar to remote eye tracking. Additionally, we contribute by validating a protocol that integrates both methods within a web-based interface. This new approach of eye tracking has potential for crossing the boundary between remote (in-lab) and webcam-based (online) empirical protocols.

2 RELATED WORK

Eye movement recording, using optical cameras without infrared light illumination, has for many years been subject to accuracy concerns which often discouraged its use. Among the greatest concerns was the possibility to discern, in real-time, the pupil from the rest of

ETRA '22, June 8–11, 2022, Seattle, WA, USA

ACM ISBN 978-1-4503-9252-5/22/06…$15.00
https://doi.org/10.1145/3517031.3529615



(a) Point target. (b) 2×2 matrix. (c) 3×3 matrix. (d) 4×4 matrix.

Figure 2: Areas Of Interest on displayed (a) point target, and face target with (b)–(d) various-sized matrices.

the iris [Sewell and Komogortsev 2010]. Currently, there is growing demand for real-time webcam-based solutions, which has led to the development of web-based eye tracking applications [Papoutsaki et al. 2017, 2016]. Deployed across the Internet, these applications rely on eye movement detection without additional infra-red light, in the visible spectrum (ambient light). Available solutions exploit such techniques as face landmark detection and machine learning to predict the users' eye positions based on relatively low resolution optical-camera input [Gudi et al. 2020; Meng and Zhao 2017].

Studies increasingly show that utilizing a webcam eye tracker can produce reliable results [Burton et al. 2014; Zheng and Usagawa 2018]. For example, Burton et al. [2014] compared results obtained using the SMI infra-red and Sticky webcam eye tracking technologies. The results of the study showed that use of the SMI infra-red eye trackers yielded an increase in accuracy over the use of the webcam along with Sticky software, particularly for small target images and images near the edges of the screen. However, webcam technology achieved nearly comparable accuracy in detecting fixations over larger images, suggesting that webcam eye tracking is a viable alternative for certain tasks.

Similarly, Zheng and Usagawa [2018] used a webcam as the main device for eye tracking and achieved accuracy of 94% on a screen divided into 9 sections, reduced to 78% when the screen was divided into 25 sections (during simulation). Their study used a webcam with a low resolution of 640×480 with corresponding algorithms to suit the low-quality image. The approach was considered to be a fast eye tracking method suitable for general human-computer interaction.

2.1 Cognitive and Behavioral Studies

Validation of webcam-based eye tracking holds potential for behavioral and cognitive science research. For example, Semmelmann and Weigelt [2018] used a JavaScript-based eye tracking software library and consumer-grade webcams to record eye movements of participants in-lab and online in three tasks: fixations (detecting a point), pursuit, and free-viewing face detection. They reported roughly 200 pixel spatial accuracy. The online data showed higher variance, lower sampling rate, and increased experimental time, but no significant difference with regard to spatial accuracy during face detection compared to the in-lab setting.

Yang and Krajbich [2021] evaluated webcam eye tracking using WebGazer software. They tested the procedure with a decision-making study adjusting the code to reduce calibration/validation and improving the temporal resolution (from 100-1000 ms to 20-30 ms). Findings showed comparable results to previous in-lab findings regarding the relationship between gaze and choice with little degradation in spatial and temporal resolution.

Bott et al. [2018] examined also the relationship between a 30-minute Visual Paired Comparison (VPC) recognition memory task and cognitive composite indices sensitive to a subtle decline related to Alzheimer disease. Eye tracking data for the 30-minute VPC task was collected simultaneously by a commercial-grade eye tracker (Tobii X2-60) and a laptop-embedded camera. In a sample of typical older adults, performance on a 30-minute VPC task correlated modestly and positively with computerized and paper-pencil based cognitive composites that serve as pre-clinical Alzheimer disease cognitive indices. The strength of these relationships did not differ between camera devices.

To investigate the usability of home-based eye tracking, Greenaway et al. [2021] investigated the set-up time, number of calibration failures, and other issues faced by older adults living with and without Alzheimer's disease. They found that home-based eye tracking is feasible with set-up support such as face-meshing that helps to position of the face.

2.2 The Use of Webcam Eye Tracking

There are several development paths suitable for webcam eye tracking, such as informing/controlling gaze-based systems, and assistive technology development. For example, Skovsgaard et al. [2011] showed that a webcam tracker (the ITU Gaze Tracker) can match the performance of two commercial gaze-tracking systems (Tobii T60 and Mirametrix S1) in an interaction task. They showed that the webcam-based eye tracker can yield performance comparable to more expensive systems. The accuracy of the webcam-based gaze tracker (0.88°) was significantly better than the accuracy of the Mirametrix system (1.34°), but not significantly different from the Tobii T60 (0.67°). These results are particularly valuable to the field of control systems, where an eye tracking system using an unmodified webcam can enable severely disabled people to interact with computers without specialized equipment. For example, Juhong et al. [2018] and Wanluk et al. [2016] used eye movements recorded by webcam and customized image processing to control appliances, a wheelchair, and communications with the caregiver.

Elsewhere, Khonglah and Khosla [2015] created a low cost webcam-based eye tracker that requires no calibration as assistive technology for young children with autism in a digital communication medium. With an accuracy of 0.4° and a frame rate of 20 fps, this system was shown to be beneficial during the initial stages of applied behavioral analysis in therapeutic interventions where physical objects are uesd to teach basic skills to the individual.

3 THE PRESENT STUDY

The present study examines the validity of online webcam eye tracking via comparison to a remote video eye tracker in two tasks: a point detection task and an emotional visual search task, the Face-In-the-Crowd task (FIC), where the latter broadens the scope of eye movement comparison, contrasting location-based (e.g., fixations) and process-based eye tracking metrics (e.g., dynamics of ambient/focal attention).

3.1 Face-In-the-Crowd Task

The FIC task is widely used in psychological research to evaluate attentional biases towards emotional stimuli e.g., happy faces suggesting "the happiness superiority effect" [Gilboa-Schechtman et al. 1999]. Happy faces are less ambiguous than other faces and therefore they can be detected faster than other facial expressions [Becker et al. 2011]. In our study, we focused on detecting happy faces among neutral faces and, for comparison, sad faces among neutral faces. We checked to see if participants noticed happy faces faster than sad faces among neutral distractors differing in number and whether results were similar between the different eye movement recording conditions.

3.2 Ambient-Focal Processing

The process of visual search is a dynamic interplay between fixations and saccades. Their characteristics reflect two modes of attentional processing: *ambient* and *focal* processing, with the latter generally more serial rather than parallel in visual search. Krejtz et al. [2016] validated the $\mathcal{K}$ coefficient to capture ambient and focal eye movement patterns when they are expected during visual search tasks. $\mathcal{K} > 0$ indicates relatively long fixations followed by short saccade amplitudes, suggesting focal processing. $\mathcal{K} < 0$ is derived from relatively short fixations followed by relatively long saccades, suggesting ambient processing. For details pertaining to its computation, see Krejtz et al. [2012, 2017].

In the FIC task, we assumed that large crowds would induce a serial search reflected in more focal attention than ambient attention. In small crowds, targets would pop out triggering relatively faster localization of the target evidenced by a long saccade (large amplitude) directed to the target yielding ambient attention.

4 METHOD

We designed the present study as a 3 (RECORDING CONDITION) × 3 (CROWD SIZE) × 2 (FACE TYPE) mixed design, in-lab experiment with three between-subjects experimental conditions of data recording: remote GP3 eye tracking, RealEye webcam eye tracking, and the integrated condition: RealEye software communicating with the GP3 eye tracker.

In the point detection task, the dependent variable was the distance between eye fixation and the displayed point. In the visual search task, the analysis was conducted with the CROWD SIZE (2 × 2 vs. 3 × 3 vs. 4 × 4 matrix) and FACE TYPE (happy vs. sad) as the key independent within-subjects variables. The dependent variables were the time to first fixation and ambient/focal attention dynamics.

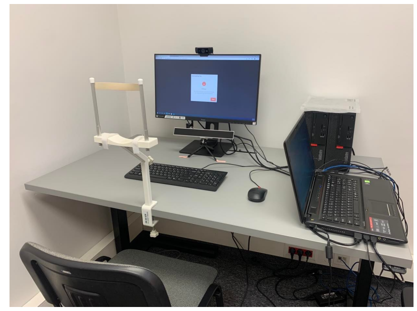


Figure 3: Experimental setting.

4.1 Hypotheses and Design

For the point detection task, we predicted that the webcam eye tracker would yield greater measurement error than the other two eye tracking conditions. We expected that the integration of web-based software with the remote eye tracker would yield similar accuracy to that of the remote condition. For the FIC task, we hypothesized that in all recording conditions we would observe similar effects: (1) time to first fixation would be shorter to a happy face than a sad one, and (2) degree of focal attention would be directly proportional to crowd density.

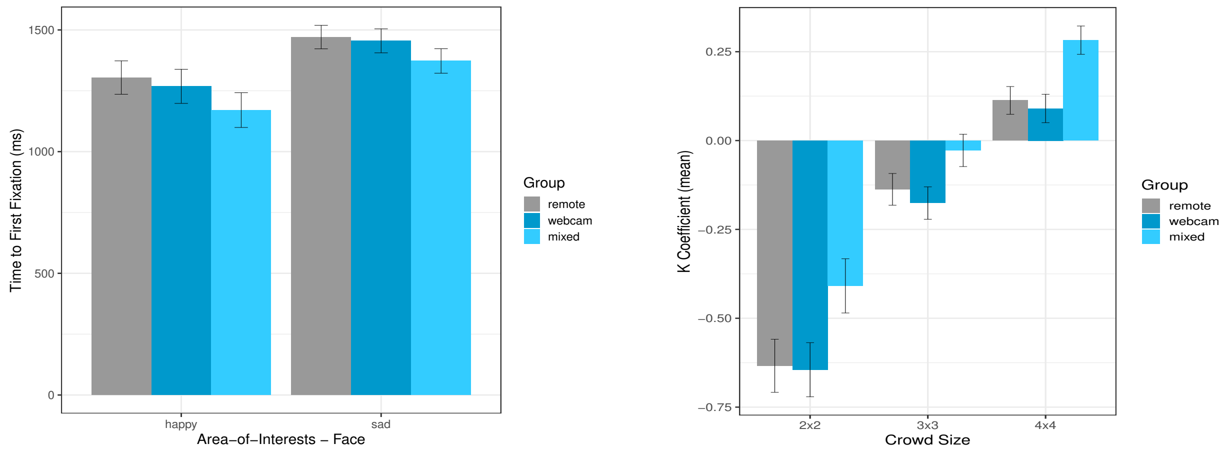
4.2 Hardware and Software

All measurements were made in a lab setting, on a 75 Hz (1900×1820 resolution) screen connected to a laptop.

4.2.1 Webcam Recording Condition. RealEye is an online software that uses a regular webcam and web browser to record gaze position. The eye tracker uses the client machine to perform face landmark and gaze detection. The web browser runs an eye tracking engine written in JavaScript. The software is based on WebGazer [Papoutsaki et al. 2016], improved and customized using TensorFlow.js with a face landmark model (Apache License 2.0). Webcam access (via JavaScript Media Devices API) sets resolution to 640 × 480 at minimum 30 fps and up to 60 fps if the webcam supports it.

RealEye uses an algorithm similar to the I-VT (Velocity-Threshold Identification) fixation filter, assumming data with a sampling rate of over 20 Hz, with minimum fixation duration set to 100 ms by default. A median filter (set to 21 by default) is used for noise reduction. RealEye software provides an online platform for preparation and running of the study. It supports analyzing the data online by real-time gaze/fixation estimation on specific Areas Of Interest (AOIs). In the webcam recording condition, we used RealEye to set up the experimental procedure and to collect data. A LOGITECH HD Pro C920 webcam (1920 × 1080 resolution) was used to record eye movements. We reduced the sampling rate to 30 Hz to gather data most representative of typical recording conditions.

4.2.2 Remote Condition. This condition was prepared in PsychoPy 3 with the ioHub eye tracker interface for connection to the GP3 eye tracker [Peirce et al. 2019]. Eye movements were recorded by



(a) Time to First Fixation.

(b) Dynamics of Ambient-Focal Attention.

Figure 4: Differences between experimental conditions in (a) time to first fixation depending on target face, and (b) dynamics of attention depending on crowd size in the FIC task ($\mathcal{K} < 0$ and $\mathcal{K} > 0$ indicate ambient and focal attention, respectively). *Note*: the height of bars represents estimated means and error bars represent -1 SE and +1 SE.

the GP3 eye tracker with sampling rate of 60 Hz. The raw eye position was pre-processed with the *gazepath* library in R, the computational language for statistical analysis [R Core Team 2017]. Eye movement events, fixations and saccades, were detected using a non-parametric speed-based approach [Mould et al. 2012]. This approach estimated velocity thresholds per individual and used the fixation duration threshold set to 80 ms. Spatio-temporally overlapping fixations were combined.

4.2.3 Integrated Condition. RealEye supports third-party eye tracking hardware (e.g., the GP3) or webcam (e.g., OpenGaze) that utilizes the OpenGaze API. Integration with the GP3 requires: (1) recommended webcam eye tracking system components; (2) recent version of the Gazepoint Control software; (3) the GP3 hardware; (4) a recent version of the RealEye OpenGaze API Adapter. The webcam is still used for facial coding and making sure participants keep their heads in the correct position. Calibration needs to be done using the eye tracker's software, e.g., Gazepoint Control. With the GP3 hardware, the OpenGaze API is provided by GazePoint Control software. Access is available via TCP/IP sockets with a socket bound to a virtual IP address, e.g., localhost.

In this condition we used the same procedure prepared for the webcam recording condition and recorded eye movements with the GP3 eye tracker running at 60 Hz. The procedure was run on the RealEye platform using the Microsoft Edge browser. The data preparation such as fixation detection was performed using the RealEye platform with default settings as described above.

4.3 Participants

A total of 83 students volunteered to participate in the in-lab experiment in exchange for student activity credit points (56 Females, $M_{age} = 24.73$, $SD_{age} = 3.22$). Participants were recruited via an announcement on the University recruitment system, social-media groups or recruited at the University campus. They were randomly assigned to one of three recording conditions: remote (27 individuals), webcam (27 individuals), and integrated recording (29 individuals). Participants declared that their vision was normal or corrected to normal.

4.4 Procedure and Experimental Tasks

After coming to the laboratory, participants were informed about the aim of the study and signed an informed consent form. They were asked to put their chin on a chin rest. The height of the setup was customized to each individual. Participants then proceeded with calibration: a standard 40-point calibration on RealEye software or a 5-point calibration in Gazepoint Control and in the integrated condition. After successful calibration, participants completed two tasks: the point detection task and the visual search task. The tasks were identical in each condition. The procedure lasted approximately 9 min.

4.4.1 Point Detection Task. Participants were asked to look and mouse click on the displayed point as fast and accurately as possible. The points were shown separately on each slide, three times at one of the nine spots on the screen, giving 27 trials. The task was self-paced, meaning that the next trial started whenever participants clicked on the previous point.

4.4.2 Face-In-the-Crowd Task. Participants were asked to find and click on the face expressing a different emotion (happy or sad) from all other neutral faces shown in the crowd matrix [Gilboa-Schechtman et al. 1999]. There were three sizes of the matrices: 2×2, 3×3 and 4×4. The target face (happy or sad one) was shown at each of the in the crowd except for the 3×3 matrix in which the middle face was always neutral. Therefore there were 8 trials for the 2×2 matrix, 16 trials for the 3×3 matrix and 32 trials for the 4×4 matrix resulting in 56 trials. To prepare the matrices, we selected six Caucasian faces from the Warsaw set of emotional

Table 1: Data in rows represent means and standard deviation in each recording condition (*M*: mean value, *SE*: standard deviation).

Time to First Fixation			
Face Type	Remote $M(SE)$	Webcam $M(SE)$	Mixed $M(SE)$
Happy	$1306ms(60)$	$1270ms(60.5)$	$1172ms(61.1)$
Sad	$1472ms(60)$	$1457ms(60.5)$	$1374(61.1)$
Main Effect of Face Type	$F(1,24) = 15.70^{**}$	$F(1,23) = 9.38^{**}$	$F(1,24) = 12.87^{**}$
Ambient/Focal Attention			
Crowd Size	Remote $M(SE)$	Webcam $M(SE)$	Mixed $M(SE)$
2×2	$-0.60(0.05)$	$-0.66(0.06)$	$-0.41(0.06)$
3×3	$-0.18(0.05)$	$-0.18(0.06)$	$-0.03(0.06)$
4×4	$0.11(0.05)$	$0.09(0.06)$	$0.28(0.06)$
Main Effect of Crowd Size	$F(1,39) = 63.18^{**}$	$F(1,32) = 70.96^{**}$	$F(1,39) = 57.79^{**}$

**statistically significant effect at $p<0.01$

facial expression pictures (WSEFEP) [Olszanowski et al. 2015]. Half the facial expressions were female and half male. The task was self-paced, meaning that whenever participants clicked on the target face, the next trial appeared. Between each trial a fixation point was displayed for 1 second. Prior to the analyses, we defined specific AOIs around the target points and sad/happy faces (see Figure 2). This allowed us to calculate the time to first fixation on the AOI.

5 RESULTS

Results are given in two parts: (a) measurement error estimation, i.e., precision and accuracy of point detection in the self-same task and (b) validation of theoretical-based predictions in the visual search task. All statistical analyses were performed in R, the language for statistical computing [R Core Team 2017].

5.1 Measurement Error

To check the differences in measurement accuracy between RECORDING CONDITIONS, one-way ANOVAs were conducted with the distance (in pixels) between the center of the point target AOI and position of the participant's eye fixation as a dependent variable. In line with the first hypothesis, ANOVA of measurement error revealed a significant difference between conditions, $F(2,80) = 9.88,\ p < 0.01,\ \eta^2 = 0.22$. Post hoc comparisons with Tukey correction showed that the average error was significantly higher in the webcam condition ($M = 45.1,\ SE = 2.81$) than in the remote ($M=34.2,\ SE=2.75,\ t=2.78,\ p=0.02,\ \eta^2=0.11$) and the integrated conditions ($M = 27.70,\ SE = 2.81,\ t = 4.39,\ p < 0.01,\ \eta^2 = 0.26$). The difference between remote and integrated condition was not significant ($t=1.66,\ p=0.22$).

We repeated one-way ANOVA with the dispersion (in pixels) of the participant's eye fixations on each target point as a dependent variable to check the differences in measurement precision between RECORDING CONDITIONS. ANOVA of precision error revealed a significant difference between conditions, $F(2,67) = 25.60,\ p<0.01,\ \eta^2=0.43$. Post hoc comparisons with Tukey correction showed that the dispersion of fixations was significantly higher in the webcam condition ($M=58.9,\ SE=2.68$) than in the remote ($M=37.9,\ SE=2.62,\ t=2.78,\ p<0.01,\ \eta^2=0.38$) and the integrated conditions ($M = 33.5,\ SE = 2.68,\ t = 6.69,\ p < 0.01,\ \eta^2 = 0.48$). The difference between remote and integrated condition was not significant ($t=1.18,\ p=0.47$). An example of fixation dispersion in each recording condition is shown in Figure 1.

5.2 Visual Search Task

We ran a three-way mixed-design ANOVA to test the effect of RECORDING CONDITION, FACE TYPE and CROWD SIZE, separately for time to first fixation on target AOI, and for dynamics of ambient-focal attention as dependent variables.

5.2.1 Time to First Fixation on Target Face. In line with theoretical predictions, ANOVA revealed a significant main effect of FACE TYPE, $F(1,69)=48.42,\ p<0.001,\ \eta^2=0.06$. The time to the first fixation on the target face was shorter for happy than sad faces in all three RECORDING CONDITIONS (Figure 4(a), Table 1).

The main effect of CROWD SIZE was also significant, $F(1,95) = 183.25,\ p < 0.001,\ \eta^2 = 0.42$. Search time increased significantly with crowd size (for 2×2: $M = 1007ms,\ SE = 40.3$; 3×3: $M = 1234,\ SE=40.3$; 4×4: $M=1784,\ SE=40.3$). We did not observe a significant main effect of RECORDING CONDITION, $F(2,69)=1.16,\ p=0.32$, nor did we find any significant two-way interactions, either between RECORDING CONDITION and FACE TYPE ($F<1$), or between RECORDING CONDITION and CROWD SIZE ($F(1,95)=2.18,\ p=0.10$). The three-way interaction between RECORDING CONDITION, FACE TYPE and CROWD SIZE was also not significant ($F<1$).

Results support the validity of webcam eye tracking, as well as the integrated condition in the visual search task. In all three conditions, happiness attracted attention and search complexity led to an increase in search time.

5.2.2 Ambient-Focal Attention. Analysis of the dynamics of ambient-focal attention revealed a significant main effect of FACE TYPE, $F(1,69)=23.45,\ p<0.001,\ \eta^2=0.07$. Overall, participants exhibited less ambient search for a happy face ($M=-0.12,\ SE=0.02$) than a sad face ($M=-0.22,\ SE=0.02;\ t=4.84,\ p<0.01$). These results are in line with previous findings about greater focal attention during exploration of happiness [Krejtz et al. 2018].

The main effect of CROWD SIZE was also significant, $F(1,109) = 108.95$, $p < 0.001$, $\eta^2 = 0.48$. In line with the hypothesis about dynamics of visual attention, post hoc comparisons suggested that more focal attention was observed in larger crowds in all RECORDING CONDITIONS (see Figure 4(b)).

No significant interaction was found, either between RECORDING CONDITION and FACE TYPE ($F(2,69) = 1.35$, $p = 0.26$), or between RECORDING CONDITION and CROWD SIZE ($F < 1$) or between all three factors ($F < 1$). Descriptive statistics are given in Table 1.

6 DISCUSSION & CONCLUSION

We compared measurement error and validity of webcam (RealEye) to remote (the GP3) eye tracking as well as to an integrated method in point detection and visual search tasks. We evaluated the webcam eye tracking in the precision and theory based tasks. As predicted, webcam recording showed lower accuracy and greater precision error than the other conditions. Our theory-based hypotheses however were supported in three conditions of data recording, stating that, despite lower precision in the webcam recording, effects of visual search would be similar to those in the remote and integrated conditions, suggesting happiness-oriented visual attention and similar dynamics of visual attention. Therefore, we supported previous results that webcam eye tracking can be used in cognitive and behavioral studies [Semmelmann and Weigelt 2018].

Our contribution is threefold. First, the measurement error was relatively small compared to earlier studies [Burton et al. 2014]. The improvement in precision and accuracy is related to the hardware and webcam platform development.

Second, to the best of our knowledge, this is the first study to investigate the dynamics of visual attention recorded by a webcam eye tracker. Current results may be useful for development of real-time alerting systems of focal processing, as focal attention indicates deeper attentional processing [Krejtz et al. 2016]. Such applications may be beneficial in many fields, including computer-supported learning or assitive technology [Skovsgaard et al. 2011].

Third, we proposed the integration of webcam and remote eye tracker software. Results showed similar, but stronger effects in the integrated condition compared to the remote condition. It is worth emphasizing that the same fixation filters were used as in webcam recording condition. The default filters have high velocity thresholds and noise reduction which likely works better with the webcam camera. These settings may be changed manually during data preparation in the RealEye software. Nevertheless, our aim was to show that using a the GP3 eye tracker with RealEye software may make preparation and analyses easy and fast leading to similar results even with differences in sampling rate and fixation duration.

Finally, considering the in-lab experimental setting used, findings should be replicated outside the lab in in-house conditions. Controlled factors such as head position in the present study may enhance accuracy of the webcam.

ACKNOWLEDGMENTS

Open access was financed by the European Union resources within the European Social Fund no. POWR.03.02.00-00-I054/16-00 and supported by doctoral fellowship ICT&Psychology–Interdisciplinary PhD Study from SWPS University of Social Sciences & Humanities and Polish Japanese Academy of Information Technology.

REFERENCES

D Vaughn Becker, Uriah S Anderson, Chad R Mortensen, Samantha L Neufeld, and Rebecca Neel. 2011. The face in the crowd effect unconfounded: happy faces, not angry faces, are more efficiently detected in single-and multiple-target visual search tasks. *Journal of Experimental Psychology: General* 140, 4 (2011), 637. https://doi.org/10.1037/a0024060

Nicholas Bott, Erica N Madero, Jordan Glenn, Alexander Lange, John Anderson, Doug Newton, Adam Brennan, Elizabeth A Buffalo, Dorene Rentz, and Stuart Zola. 2018. Device-embedded cameras for eye tracking–based cognitive assessment: validation with paper-pencil and computerized cognitive composites. *Journal of medical Internet research* 20, 7 (2018), e11143. https://doi.org/10.2196/11143

Liz Burton, William Albert, and Mark Flynn. 2014. A comparison of the performance of webcam vs. infrared eye tracking technology. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 58. SAGE Publications Sage CA: Los Angeles, CA, Los Angeles, CA, 1437–1441. https://doi.org/10.1177/1541931214581300

Eva Gilboa-Schechtman, Edna B Foa, and Nader Amir. 1999. Attentional biases for facial expressions in social phobia: The face-in-the-crowd paradigm. *Cognition & Emotion* 13, 3 (1999), 305–318. https://doi.org/10.1080/026999399379294

Anne-Marie Greenaway, Slawomir Nasuto, Aileen Ho, and Faustina Hwang. 2021. Is home-based webcam eye-tracking with older adults living with and without Alzheimer's disease feasible?. In *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility*. Association for Computing Machinery, New York, NY, USA, 1–3. https://doi.org/10.1145/3441852.3476565

Amogh Gudi, Xin Li, and Jan van Gemert. 2020. Efficiency in Real-Time Webcam Tracking. In *European Conference on Computer Vision*. Springer, New York, NY, USA, 529–543. https://doi.org/10.1007/978-3-030-66415-2_34

Aniwat Juhong, T Treebupachatsakul, and C Pintavirooj. 2018. Smart eye-tracking system. In *2018 International Workshop on Advanced Image Technology (IWAIT)*. IEEE, New York, NY, USA, 1–4. https://doi.org/10.1109/IWAIT.2018.8369701

Jessy Rimaya Khonglah and Arun Khosla. 2015. A low cost webcam based eye tracker for communicating through the eyes of young children with ASD. In *2015 1st International Conference on Next Generation Computing Technologies (NGCT)*. IEEE, New York, NY, USA, 925–928. https://doi.org/10.1109/NGCT.2015.7375255

Izabela Krejtz, Agnieszka Szarkowska, Krzysztof Krejtz, Agnieszka Walczak, and Andrew T. Duchowski. 2012. Audio Description as an Aural Guide of Children's Visual Attention: Evidence from an Eye-tracking Study. In *ACM Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '12)*. Association for Computing Machinery, New York, NY, USA, 99–106.

Krzysztof Krejtz, Arzu Çöltekin, Andrew Duchowski, and Anna Niedzielska. 2017. Using Coefficient K to Distinguish Ambient/Focal Visual Attention During Map Viewing. *Journal of Eye Movement Research* 10, 2 (2017), 1–13 pages. https://doi.org/10.16910/jemr.10.2.3

Krzysztof Krejtz, Andrew Duchowski, Izabela Krejtz, Agnieszka Szarkowska, and Agata Kopacz. 2016. Discerning Ambient/Focal Attention with Coefficient K. *ACM Transactions on Applied Perception* 13, 3, Article 11 (May 2016), 20 pages. https://doi.org/10.1145/2896452

Krzysztof Krejtz, Katarzyna Wisiecka, Izabela Krejtz, Paweł Holas, Michał Olszanowski, and Andrew T Duchowski. 2018. Dynamics of emotional facial expression recognition in individuals with social anxiety. In *Proceedings of the 2018 ACM symposium on eye tracking research & applications*. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3204493.3204533

Chunning Meng and Xuepeng Zhao. 2017. Webcam-based eye movement analysis using CNN. *IEEE Access* 5 (2017), 19581–19587. https://doi.org/10.1109/ACCESS.2017.2754299

Matthew S. Mould, David H. Foster, Kinjiro Amano, and John P. Oakley. 2012. A simple nonparametric method for classifying eye fixations. *Vision Research* 57 (2012), 18–25. https://doi.org/10.1016/j.visres.2011.12.006

Michał Olszanowski, Grzegorz Pochwatko, Krzysztof Kukliński, Michał Ścibor-Rylski, Peter Lewinski, and Rafał K. Ohme. 2015. Warsaw Set of Emotional Facial Expression Pictures: a Validation Study of Facial Display Photographs. *Frontiers in Psychology* 5 (2015), 1516. https://doi.org/10.3389/fpsyg.2014.01516

Alexandra Papoutsaki, James Laskey, and Jeff Huang. 2017. SearchGazer: Webcam Eye Tracking for Remote Studies of Web Search. In *Proceedings of the ACM SIGIR Conference on Human Information Interaction & Retrieval (CHIIR)*. Association for Computing Machinery, New York, NY, USA, 17–26. https://doi.org/10.1145/3020165.3020170

Alexandra Papoutsaki, Patsorn Sangkloy, James Laskey, Nediyana Daskalova, Jeff Huang, and James Hays. 2016. Webgazer: Scalable Webcam Eye Tracking Using User Interactions. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI'16)*. AAAI Press, New York, New York, USA, 3839–3845. https://doi.org/10.3758/s13428-017-0913-7

Jonathan Peirce, Jeremy R. Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman, and Jonas Kristoffer Lindeløv. 2019. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods* 51, 1 (2019), 195–203. https://doi.org/10.3758/s13428-018-01193-y

R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Kilian Semmelmann and Sarah Weigelt. 2018. Online webcam-based eye tracking in cognitive science: A first look. *Behavior Research Methods* 50, 2 (2018), 451–465. https://doi.org/10.3758/s13428-017-0913-7

Weston Sewell and Oleg Komogortsev. 2010. Real-time eye gaze tracking with an unmodified commodity webcam employing a neural network. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 3739–3744. https://doi.org/10.1145/1753846.1754048

Henrik Skovsgaard, Javier San Agustin, Sune Alstrup Johansen, John Paulin Hansen, and Martin Tall. 2011. Evaluation of a remote webcam-based eye tracker. In *Proceedings of the 1st conference on novel gaze-controlled applications*. Association for Computing Machinery, New York, NY, USA, 1–4. https://doi.org/10.1145/1983302.1983309

Nutthanan Wanluk, Sarinporn Visitsattapongse, Aniwat Juhong, and C Pintavirooj. 2016. Smart wheelchair based on eye tracking. In *2016 9th Biomedical Engineering International Conference (BMEiCON)*. IEEE, New York, NY, USA, 1–4. https://doi.org/10.1109/BMEiCON.2016.7859594

Xiaozhi Yang and Ian Krajbich. 2021. Webcam-based online eye-tracking for behavioral research. *Judgment and Decision Making* 16, 6 (2021), 1486.

Chenyang Zheng and Tsuyoshi Usagawa. 2018. A rapid webcam-based eye tracking method for human computer interaction. In *2018 International Conference on Control, Automation and Information Sciences (ICCAIS)*. IEEE, New York, NY, USA, 133–136. https://doi.org/10.1109/ICCAIS.2018.8570532