


RESEARCH

Open Access



# Look me in the eye: evaluating the accuracy of smartphone-based eye tracking for potential application in autism spectrum disorder research

Maximilian A. R. Strobl<sup>1,2\*</sup> , Florian Lipsmeier<sup>3</sup>, Liliana R. Demenescu<sup>3</sup>, Christian Gossens<sup>3</sup>, Michael Lindemann<sup>3</sup> and Maarten De Vos<sup>4</sup>

\*Correspondence:  
maximilian.strobl@gmail.com  
<sup>1</sup> Wolfson Centre  
for Mathematical Biology,  
Mathematical Institute,  
University of Oxford, Radcliffe  
Observatory Quarter, OX2  
6GG Oxford, UK  
Full list of author information  
is available at the end of the  
article

## Abstract

**Background:** Avoidance to look others in the eye is a characteristic symptom of Autism Spectrum Disorders (ASD), and it has been hypothesised that quantitative monitoring of gaze patterns could be useful to objectively evaluate treatments. However, tools to measure gaze behaviour on a regular basis at a manageable cost are missing. In this paper, we investigated whether a smartphone-based tool could address this problem. Specifically, we assessed the accuracy with which the phone-based, state-of-the-art eye-tracking algorithm iTracker can distinguish between gaze towards the eyes and the mouth of a face displayed on the smartphone screen. This might allow mobile, longitudinal monitoring of gaze aversion behaviour in ASD patients in the future.

**Results:** We simulated a smartphone application in which subjects were shown an image on the screen and their gaze was analysed using iTracker. We evaluated the accuracy of our set-up across three tasks in a cohort of 17 healthy volunteers. In the first two tasks, subjects were shown different-sized images of a face and asked to alternate their gaze focus between the eyes and the mouth. In the last task, participants were asked to trace out a circle on the screen with their eyes. We confirm that iTracker can recapitulate the true gaze patterns, and capture relative position of gaze correctly, even on a different phone system to what it was trained on. Subject-specific bias can be corrected using an error model informed from the calibration data. We compare two calibration methods and observe that a linear model performs better than a previously proposed support vector regression-based method.

**Conclusions:** Under controlled conditions it is possible to reliably distinguish between gaze towards the eyes and the mouth with a smartphone-based set-up. However, future research will be required to improve the robustness of the system to roll angle of the phone and distance between the user and the screen to allow deployment in a home setting. We conclude that a smartphone-based gaze-monitoring tool provides promising opportunities for more quantitative monitoring of ASD.

**Keywords:** Gaze tracking, Mental disorders, m-Health, Biomedical monitoring

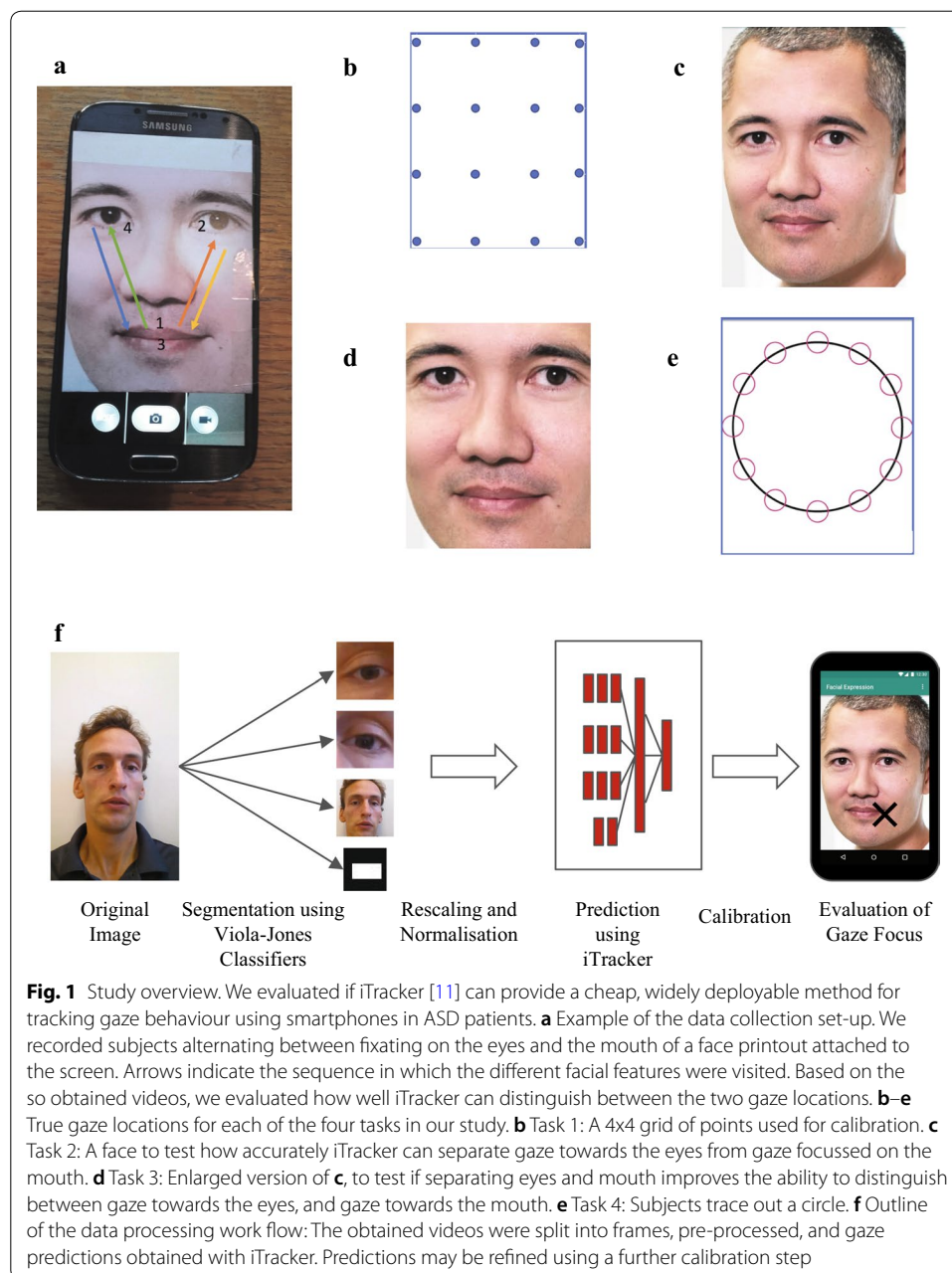


## Background

Autism spectrum disorders (ASD) describes a set of developmental disabilities characterised by “deficits in social communication and social interaction” [1]. As the name “spectrum” suggests, the nature of the impairments and their severities vary from person to person, with some subjects being able to complete university degrees and live independent lives, whereas others need life-long assistance with daily living [2]. Despite this variability there are certain commonalities: Already in one of the earliest accounts of autism, the author noted that the subject “never looked into anyone’s face” [3]. A series of eye-tracking studies since then have established that, when shown the image of a face, ASD patients spend less time fixating on the eyes and more time exploring the mouth or objects in the surroundings (see [4] for an extensive review). In fact, gaze abnormalities are one of the criteria used to diagnose ASD [1].

While abnormal gaze behaviour is not the cause for the difficulties which subjects with ASD experience, it has been proposed that it might provide a quantifiable feature for monitoring the condition over time, and to evaluate the efficacy of treatments [4, 5]. However, a key challenge that has prevented further exploration of this idea so far has been the lack of access to a suitable eye-tracking device. Traditional eye-tracking devices are inconvenient for widespread home use, as they are costly, often not very portable, and require expertise to set up and run [6, 7]. But thanks to recent advances new solutions are emerging that allow eye-tracking using only the camera of a laptop or smartphone. These approaches use machine-learning techniques, such as support vector methods [8], Gaussian processes [9], or neural networks, [10, 11], and hold great potential for longitudinal gaze behaviour monitoring in ASD patients. We envision the development of a mobile-monitoring tool in which patients regularly perform a series of tests on their smartphone that measure their gaze behaviour and so provide insights into the development of their condition over time.

The aim of the present study was to prototype a smartphone-based gaze-monitoring framework, assess its accuracy on healthy volunteers, and identify key challenges to be overcome on the way to the clinic. The test set-up employed imitates that of, for example [12], in which the subject is shown the image of a face and one compares fixation time on the eyes to fixation time on the mouth. To estimate the gaze location, we used the convolutional neural network iTracker [11], which to the best of our knowledge has the best-reported performance on a smartphone in literature. The network was trained on the largest eye-tracking dataset to-date, consisting of around 1.2 million images of 1271 subjects recorded under home-use conditions. The authors report a mean Euclidean Error of 2.04 cm, with further reduction to 1.04 cm through a number of refinement steps (data augmentation, restriction to phone images, and calibration) [11]. We test the set-up on 17 healthy volunteers, and evaluate its accuracy for 2 different types of tasks: (i) separating between gaze towards the eyes and towards the mouth of a face, and (ii) resolving a more fine-grained gaze pattern (see Fig. 1 for an overview of our methods). In addition, we investigated how the accuracy is influenced by the distance between the eyes and the mouth, and how well accuracy can be improved through calibration. We compare two calibration methods: (i) a support vector regression (SVR)-based method proposed in [11], and (ii) a linear transformation-based method proposed by us. The SVR uses 128 features extracted from the final layer of the neural network to adjust the



prediction, whereas the linear transformation merely translates points and rescales distances. We make our eye-tracking pipeline available online (<https://github.com/ms234/iTrackerWrapper>), and hope that our work will serve as a stepping stone to the creation of a tool which will help to improve the understanding and treatment of ASD in the future.

## Results

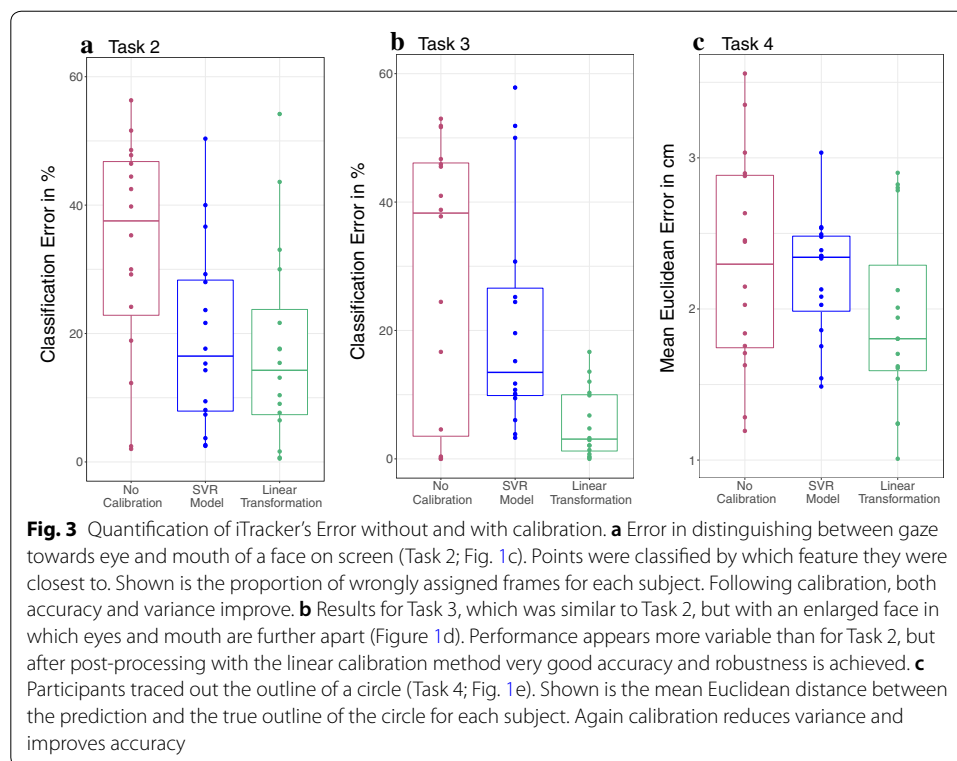
### iTracker captures patterns but its predictions are biased

Figure 2 illustrates the results for one of the subjects in our study (Subject 8). Estimates for eyes and mouth cluster into easily distinguishable distinct patches (Fig. 2d, g). This



separation is more pronounced in Task 3 than Task 2, where separate clusters for the left and right eye become visible. However, for neither task are the clusters centred on the eyes and mouth of the face on the screen. Instead they appear systematically shifted towards the right side of the screen. A similar pattern holds true for all subjects, with the magnitude and direction of the bias being conserved across different tasks for each subject but varying between subjects (for further examples see Additional file 1: Figures S1–S4).

In order to assess the accuracy with which the gaze location can be identified from the collected data, we classified estimates to belong to either the eye or mouth according to which they were closer to. The error of this classification is shown in Fig. 3. In Task 2,



on average 33.23% of a subject's frames are miss-classified (95% CI [24.08%, 42.39%]). In Task 3, we observe a reduction in the mean classification error to 28.65%, albeit the difference is not statistically significant (95% CI [17.23%, 40.08%]; Wilcoxon Signed-Rank Test,  $V = 93$ ,  $p$  value = 0.21). In addition, there is large variation in accuracy between subjects, in particular for Task 3 (Fig. 3a, b).

#### Calibration significantly improves robustness

While the performance of the SVR-based calibration method proposed in [11] reduces the error for Task 2 (Wilcoxon Signed-Rank Test:  $V = 110$ ,  $p$  value = 0.03), it does not provide a statistically significant improvement for Task 3 (Wilcoxon Signed-Rank Test:  $V = 79$ ,  $p$  value = 0.29; see also Fig. 3). Visual inspection of the gaze predictions suggests that this might be due to a tendency to cluster the points around the centre of the screen, thereby losing structure previously visible in the data (see Fig. 2e, h, and k, and Additional file 1: Figures S1–S4).

Motivated by the observation that iTracker's estimates appeared to be linearly shifted with respect to the reference points (e.g. Fig. 2b, g), we evaluated a calibration method based on a linear transformation. As we illustrate for one subject in Fig. 2f, i, the estimates now overlap well with the true gaze locations. Accordingly, compared to the unadjusted iTracker output, the mean classification error for Task 2 is reduced to 17.67% (95% CI [9.38%, 25.96%]; Wilcoxon Signed-Rank Test:  $V = 106$ ,  $p$  value = 0.01). For Task 3, this is even reduced to only 5.44% (95% CI [2.56%, 8.31%]; Wilcoxon Signed-Rank Test:  $V = 77$ ,  $p$  value < 0.01). As such, the linear transformation yields as good of, and for Task 3 an even better, reduction in error than the SVR-based method (Wilcoxon

Signed-Rank Test for Task 2:  $V = 75$ ,  $p$  value = 0.41; Wilcoxon Signed-Rank Test for Task 3:  $V = 130$ ,  $p$  value < 0.01; see also Fig. 3).

Moreover, calibration also improves the robustness of classification. In particular for Task 3, the variance in accuracy between different subjects is significantly lower if calibration is applied (Fig. 3b; Levene Test:  $F$  value = 4.54,  $df = (1.46)$ ,  $p$  value = 0.03851). Whereas not statistically significant, reductions in variance are also noticeable for Task 2 (Fig. 3a; Levene Test:  $F$  value = 0.67,  $df = (1.46)$ ,  $p$  value = 0.42).

### **iTracker can resolve fine-grained movement**

While our main aim was to assess the feasibility of distinguishing between two facial landmarks, it was also of interest how well iTracker could resolve a more fine-grained temporal sequence (Task 4). Figure 2j illustrates for one subject that iTracker's estimates capture the correct shape, and temporal sequence of the trajectory. They are also systematically shifted and contracted, but this can again be addressed with calibration (Fig. 2k, l). For further examples, see Additional file 1: Figures S1–S4.

We show the accuracy of the estimates as the Euclidean distance to the true gaze location in Fig. 3c. Without calibration, the mean Euclidean error is 2.30 cm (95% CI [1.92 cm, 2.69 cm]). Calibration with SVR yields a mean error of 2.21 cm (95% CI [1.99 cm, 2.43 cm]), although the reduction is not statistically significant (paired  $T$  test:  $df = 15$ ,  $t = 0.52$ ,  $p$  value = 0.61). In contrast, the linear transformation yields a statistically significant reduction in the average error to 1.93 cm (95% CI [1.61 cm, 2.26 cm]; paired  $T$  test:  $df = 15$ ,  $t = 2.41$ ,  $p$  value = 0.03).

## **Discussion**

Advancements in gaze-tracking technology for tablets and phones suggest that soon eye tracking could be performed routinely in an everyday setting. Such systems hold the potential for the development of novel, gaze-based digital biomarkers to monitor disabilities, such as ASD. In the present study, we investigated the feasibility of developing such a tool with the current state-of-the-art technology, iTracker, and evaluated the accuracy based on specific biomarker-relevant tests performed with healthy volunteers.

Overall, iTracker allows distinction between gaze towards the eyes and the mouth of a face shown on the screen. While distinction is possible for the smaller of the two faces used in this study ( $17.67\% \pm 8.29\%$ ), performance is significantly more robust when the distance between eyes and mouth is maximised ( $5.44\% \pm 2.88\%$ ; Fig. 3a, b). We, thus, recommend to place the eyes and mouth as far apart as possible on the screen. In addition, we find that more fine-grained temporal sequences can be resolved with acceptable accuracy, as shown for Task 4 (Fig. 2). Our results thereby also independently confirm the error estimates for iTracker reported in [11] (2.04 cm; here:  $2.30 \pm 0.38$  cm), and furthermore indicate that iTracker's performance is relatively robust to the phone type used: iTracker was trained on iPhones whereas we used Android devices.

Moreover, in accordance with [11], we find that accuracy can be improved by post-processing the predictions with an error model trained on a calibration data set collected at the start of the experiment. However, we observed that while the SVR-based method proposed in [11] reduced the distance between true gaze location and the estimates, it tended to do so by clustering the predictions in the centre of the screen, loosing

spatial structure previously visible (e.g. Fig. 2e, h, and k). As an alternative, we tested a simple linear error model to shift the estimates and rescale distances. Under the conditions of our experiment, this outperformed the SVR-based method and preserved the structure in the data better. We hypothesise that this is due to the small number of data points available for calibration (only 16 distinct locations), resulting in over-training of the SVR. In addition, the highly controlled circumstances of our experiments will have helped to reduce the error making a linear error model sufficient. It seems likely that in less-controlled environments (e.g. difficult lighting conditions, tremor when holding the phone), a non-linear model such as the SVR might be required. It would be important in the future to investigate this issue further and trial other statistical techniques, such as generalised linear models, to develop a method that can robustly correct for errors induced by differences in phone hold, lighting conditions, and visual appearance (e.g. skin colour). It might also prove beneficial to explicitly incorporate data such as gyroscope data to allow further personalisation.

Although we did not explore this question in detail, we noticed a strong dependence of iTracker's performance on the distance between the screen and the user's face (Additional file 1). In preliminary experiments, we found that the phone had to be held very close to the face for accurate and consistent performance (at most 20 cm). This distance is shorter than the distance most users would usually hold the phone at, and as such in a practical setting would require a test to ensure this distance is kept. We also noticed an influence of the phone angle on performance, although we were unable to clearly characterise the relationship. Given that the target population includes young children with challenging behaviour and cognitive disabilities, such sensitivity to the phone hold imposes serious restrictions on its use. Research to improve robustness should be a priority in developing this technology further.

To use iTracker, one has to accurately identify the face and the eyes in the image. Despite a long-lasting history of research in this area, a state-of-the-art available implementation failed surprisingly many times. We were only able to segment 74.7% of the frames. Similarly, Krafka et al. [11] who used the inbuilt iOS algorithm reported a 61% segmentation rate (only 1,490,959 of 2,445,604 had both face and eye detection [11]). Considering the large amount of potentially insightful data that are lost, this issue should be given further attention.

A great challenge in developing and benchmarking eye-tracking devices is to obtain accurately labelled validation data. In this study, we aimed to ensure accurate labelling of our data through manual validation. As a result, our data are likely biased towards easy images, since a human observer had to be able to identify the gaze location. Thus, the reported accuracies should be seen as best-case estimates. For further validation, it would be helpful to compare the performance of iTracker with that of a professional eye-tracking device. In addition, it could be interesting to repeat our experiment with other eye-tracking algorithms for phones, such as MPIIGaze [10], to compare their accuracies.

The aim of this study was to evaluate whether smartphone-based gaze monitoring is sufficiently accurate and robust to be employed in ASD research. This is the reason why we worked with healthy volunteers instead of patients. After the observed sensitivity to the phone hold is addressed, this study should be repeated with a cohort of ASD patients to investigate whether it can capture the differences in gaze patterns in practice. It would

also be interesting to test if it can recapitulate the attraction towards non-social extraneous objects in visual scenes, or indeed the reduced time spent looking at the screen during the test, that has been found even more predictive of autism than aversion of direct eye contact [13]. If so, this might provide an alternative marker that could be monitored with a smartphone-based framework. Based on our analyses it seems plausible that such differences might be detectable, as the eyes and mouth in Task 2 were only 3 cm apart. Moreover, while we chose to focus our application on ASD, some of our results might be transferable to other mental disorders characterised by alterations in gaze behaviour. Tasks such as Task 4, for example, could be used to assess smooth pursuit dysfunction in Multiple Sclerosis [14, 15]. By presenting a prototype, and making our code publicly available, we hope to stimulate future research into these directions.

## Conclusions

We simulated a smartphone application in which participants are shown images of faces, and examined the accuracy with which iTracker [11] could distinguish between gaze fixations on eyes and mouth. We conclude that comparing gaze fixations towards the eyes and the mouth of a face shown on the screen is feasible with current technology. A calibration step will be required and care will have to be taken that the phone is held sufficiently close to the face. But, provided this is the case, we confirmed it is possible to obtain accurate estimates of the gaze position. Accuracy can be further improved by maximising the distance between eyes and mouth of the face shown on the screen. Future research should explore how iTracker performs under different recording conditions, to develop solutions to improve its robustness to the distance between the phone and the user and the angle at which the phone is held at, and explore the optimal calibration strategy. Assuming these concerns are addressed, we are confident that it will be soon possible to monitor gaze behaviour using smartphone-based applications.

## Methods

### Data collection

To benchmark iTracker's accuracy, we carried out a proof-of-principle study on 17 healthy volunteers, in which we collected a set of front-camera images from phones for which the subject's true gaze focus was known. The purpose of the study was solely to assess the accuracy of the software, with no medical implications. Eight of the subjects were based in Oxford, UK, and nine in Basel, Switzerland. All subjects were over 18 year old, and gave written consent to participate in the study. The phones used were a Samsung Galaxy S4 (8 subjects), and a Samsung Galaxy S7 (9 subjects). Participants were seated in front of a neutral background (white or grey wall), and were instructed to hold the phone at < 20 cm distance from their head at head height. Spectacle wearers were asked to take off their glasses. Participants carried out four tasks in which they traced out specific patterns on the screen with their eyes (Fig. 1a–e) while a video was taken. The patterns consisted of a set of way points printed on a piece of paper, which was attached to phone screen and was meant to simulate the smartphone application (Fig. 1a). Participants traced out the patterns by focussing on each way point for about 1s, after which they changed their focus to the next way point. Timing was enforced



through use of a metronome. Subjects were allowed to trace out the pattern using their finger and follow their finger with their eyes. The four tasks consisted of:

- *Task 1* Calibration Grid: A  $4 \times 4$  grid of points on the screen which was used for calibration (Fig. 1b).
- *Task 2* Original Face: Participants were shown the image of a face, and alternated their focus between the eyes and the mouth of the face, following the sequence: mouth, left eye, mouth, right eye (Fig. 1a, c). The pattern was repeated 5 times. Distance between eyes and mouth: 3.1 cm.
- *Task 3* Enlarged face: As Task 2, but with a digitally enlarged version of the same face (Fig. 1d). Distance between eyes and mouth: 4.3 cm.
- *Task 4* Circle: A circle on the screen consisting of 12 way points (Fig. 1e).

### Data processing and gaze prediction

Frames were extracted from the videos at 30fps and reviewed manually. Frames in which subjects blinked, or accidentally looked elsewhere on the screen were manually reallocated or excluded. The number of frames extracted for each task varies between subjects (see Additional file 1: Table S1). Next, the images were processed and gaze predictions were obtained with iTracker. The required crops of the face and eyes were extracted using the Viola–Jones detector in OpenCV [16], and rescaled and centred following the instructions in [11] and [17]. One subject (Subject 13) had to be excluded at this point from further analysis, as the face detection algorithm was unable to identify the face and eye regions. A significant number of frames from other subjects were also affected (see Additional file 1: Table S1). In total, we obtained 16,517 labelled images, distributed across 16 subjects, which we based our analysis on. Due to better performance in preliminary benchmarking tests (not shown), we chose the basic version of iTracker (“itracker\_iter\_92000.caffemodel” at [18]), instead of the also provided “25× train-augmented” version. Our pipeline is implemented in Python 2.7 [19], using scikit-video [20], opencv-python 3.1.0. [16], and Caffe 1.0 [21]. Our code is available at: <https://github.com/ms234/iTrackerWrapper>.

### Error metrics

Accuracy of automatic classification was evaluated with the following error metrics: For Tasks 2 and 3, we classified predictions according to which facial feature they were closest to (left, right eye, or mouth), as measured by Euclidean distance to the centre of that feature, and we report the proportion of misclassified frames per subject. Task 4 was mainly assessed visually. However, we also report the mean Euclidean distance between the predicted gaze focus and the true gaze focus for each subject (in centimetres).

### Calibration methods

We compared the improvements achieved by two calibration methods: (1) the SVR-based calibration method proposed in [11], and (2) a linear transformation-based method proposed by us. The methods were trained on the data collected in Task 1, and then applied to correct the predictions from Task 2–4. Accuracy was evaluated as

before. The SVR-based method was implemented following [11], using the e1071 package in R [22]. The linear transformation was motivated by the observation that the gaze patterns captured by iTracker were shifted and stretched compared to the true patterns (see Fig. 2 for an example). Thus, we used a linear transformation to translate the predictions so that the centroid of the predictions coincided with the centroid of the true dots. In order to correct the length scale, we chose the transformation so that so that the variance in distance from the centroid is preserved between the set of predictions and the set of true gaze locations. It can be shown that the transformation which satisfies these requirements is given by,

$$\tilde{x} = a_x \hat{x} + b_x, \quad \text{and} \quad \tilde{y} = a_y \hat{y} + b_y,$$

where  $(\hat{x}, \hat{y})$  is the original prediction,  $(\tilde{x}, \tilde{y})$  is the corrected prediction, and

$$a_x = \sqrt{\frac{\sigma_x^2}{\hat{\sigma}_x^2}}, \quad a_y = \sqrt{\frac{\sigma_y^2}{\hat{\sigma}_y^2}}, \quad b_x = \mu_x - a_x \hat{\mu}_x, \quad \text{and} \quad b_y = \mu_y - a_y \hat{\mu}_y.$$

Here,  $\mu_x$  and  $\mu_y$  denote the mean, and  $\sigma_x^2$  and  $\sigma_y^2$  denote the variance of the x- and y-coordinate values of the grid points in Task 1. Similarly,  $\hat{\mu}_x$  and  $\hat{\mu}_y$  stand for the mean, and  $\hat{\sigma}_x^2$  and  $\hat{\sigma}_y^2$  stand for the variance in x- and y-coordinate values of the predicted positions.

### Statistical analyses

Performance was compared using paired t-tests, if the data followed a normal distribution ( $p$  value from Shapiro Wilk Test  $> 0.05$  for both samples). Otherwise, we used Wilcoxon Signed-Rank Tests. To compare inter-subject variance in performance prior and post calibration, we used a Levene test for homoscedasticity. The post calibration sample for this test was obtained by pooling the results for both calibration methods. All statistical analyses were carried out in R version 3.4.0. [23].

### Additional file

**Additional file 1.** The Appendix contains gaze predictions for four further subjects from the study, and a table with a detailed overview of the results for each subject. Finally, we also present preliminary work on the influence of the distance between the user and the phone on iTracker's accuracy.

### Abbreviations

ASD: autism spectrum disorders; SVR: support vector regression.

### Authors' contributions

MS performed data collection, and carried out the analysis. LD carried out data collection. FL, CG, ML, and MDV developed the initial idea for the study, and oversaw the project. All authors contributed to the design of the study and writing of the manuscript. All authors read and approved the final manuscript.

### Acknowledgements

We would like to thank the volunteers from CIBIM and Roche for helping in collecting data for this project. In addition, we would like to thank two anonymous reviewers for their feedback that allowed us to tie in this paper more closely with the practicalities of ASD research.

**Competing interests**

FL and CG are employees of F. Hoffmann-La Roche AG, Basel, Switzerland. ML and LD are consultants to F. Hoffmann-La Roche AG, Basel, Switzerland. MS and MV declare that they have no competing interests.

**Availability of data and materials**

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

**Consent for publication**

All subjects have consented that their data may be utilised in this publication.

**Ethics approval and consent to participate**

Given the non-intrusive, non-clinical nature of our study, and given that previous work this study was extending ([11]) did not seek ethics approval either, we did not apply for approval for our study. However, all participants were over 18, fully informed about the aims of the study, and gave written consent to participate.

**Funding**

This research was supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC), and by funding from the Engineering and Physical Sciences Research Council (EPSRC) and the Medical Research Council (MRC) [Grant Number EP/L016044/1].

**Publishers' Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Author details**

<sup>1</sup> Wolfson Centre for Mathematical Biology, Mathematical Institute, University of Oxford, Radcliffe Observatory Quarter, OX2 6GG Oxford, UK. <sup>2</sup> Department of Integrated Mathematical Oncology, Moffitt Cancer Center, Magnolia Drive, 12902 Tampa, USA. <sup>3</sup> Roche Pharma Research and Early Development, pRED Informatics, Roche Innovation Center, F. Hoffmann-La Roche Ltd, Basel, Switzerland. <sup>4</sup> Department of Engineering Science, Institute of Biomedical Engineering, University of Oxford, Old Road Campus Research Building, OX3 7DQ Oxford, UK.

Received: 11 February 2019 Accepted: 12 April 2019

Published online: 03 May 2019

**References**

1. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 2013. <https://doi.org/10.1176/appi.books.9780890425596>.
2. Autistic Self Advocacy Network. About Autism. [www.autisticadvocacy.org/about-asan/about-autism](http://www.autisticadvocacy.org/about-asan/about-autism). The National Autistic Society. 2017. <http://autisticadvocacy.org/about-asan/about-autism/>.
3. Kanner L. Autistic disturbances of affective contact. *Nervous Child*. 1943;2:217–50. <https://doi.org/10.1105/tpc.11.5.949>.
4. Black MH, Chen NTM, Iyer KK, Lipp OV, Bölte S, Falkmer M, Tan T, Girdler S. Mechanisms of facial emotion recognition in autism spectrum disorders: insights from eye tracking and electroencephalography. *Neurosci Biobehav Rev*. 2017; <https://doi.org/10.1016/j.neubiorev.2017.06.016>.
5. Vidal M, Turner J, Bulling A, Gellersen H. Wearable eye tracking for mental health monitoring. New York: Elsevier; 2012. <https://doi.org/10.1016/j.comcom.2011.11.002>.
6. Hansen DW, Ji Q. In the eye of the beholder: a survey of models for eyes and gaze. *IEEE Trans Pattern Anal Mach Intell*. 2010;32(3):478–500. <https://doi.org/10.1109/TPAMI.2009.30>.
7. Fujioka T, Inohara K, Okamoto Y, Masuya Y, Ishitobi M, Saito DN, Jung M, Arai S, Matsumura Y, Fujisawa TX, Narita K, Suzuki K, Tsuchiya KJ, Mori N, Katayama T, Sato M, Munesue T, Okazawa H, Tomoda A, Wada Y, Kosaka H. GazeFinder as a clinical supplementary tool for discriminating between autism spectrum disorder and typical development in male adolescents and adults. *Molecular Autism*. 2016;7(1):19. <https://doi.org/10.1186/s13229-016-0083-y>.
8. Xu P, Ehinger KA, Zhang Y, Finkelstein A, Kulkarni SR, Xiao J. TurkerGaze: crowdsourcing saliency with webcam based eye tracking. 2015. <https://doi.org/10.1103/PhysRevD.91.123531>. [arXiv:1504.06755](https://arxiv.org/abs/1504.06755).
9. Nel E-M, Zielinski P. OpenGazer: open-source gaze tracker for ordinary webcams (software). 2015.
10. Zhang X, Sugano Y, Fritz M, Bulling A. Appearance-based gaze estimation in the wild. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition. 2015. p. 4511–20. <https://doi.org/10.1109/CVPR.2015.7299081>.
11. Krafska K, Khosla A, Kellnhofer P, Kannan H. Eye tracking for everyone. In: IEEE conference on computer vision and pattern recognition. 2016. p. 2176–84. <https://doi.org/10.1109/CVPR.2016.239>.
12. Corden B, Chilvers R, Skuse D. Avoidance of emotionally arousing stimuli predicts social-perceptual impairment in Asperger's syndrome. *Neuropsychologia*. 2008;46(1):137–47.
13. Chita-Tegmark M. Attention allocation in ASD: a review and meta-analysis of eye-tracking studies. *Rev J Autism Develop Dis*. 2016;3(3):209–23. <https://doi.org/10.1007/s40489-016-0077-x>.
14. McDonald WI, Halliday AM. Diagnosis and classification of multiple sclerosis. *Br Med Bull*. 1977;33(1):4–8. <https://doi.org/10.1093/oxfordjournals.bmb.a071393>.
15. Mastaglia FL, Black JL, Collins DWK. Quantitative studies of saccadic and pursuit eye movements in multiple sclerosis. *Brain*. 1979;102(4):817–34. <https://doi.org/10.1093/brain/102.4.817>.
16. ITSEEZ. The OpenCV reference manual, 2.4.9.0 edn. 2014. ITSEEZ.

17. Krafska K, Khosla A, Kellnhofer P, Kannan H. GazeCapture Project. [www.gazecapture.csail.mit.edu](http://gaze.capture.csail.mit.edu/index.php). 2017. <http://gaze.capture.csail.mit.edu/index.php>. Accessed 7 Dec 2017.
18. Krafska K, Khosla A, Kellnhofer P, Kannan H. 2017. <https://github.com/CSAILVision/GazeCapture>. Accessed 7 Dec 2017.
19. The Python Software Foundation, Python Language Reference, Version 2.7. [www.python.org](http://www.python.org). The Python Software Foundation. 2017.
20. Scikit-Video. 2017. <https://github.com/scikit-video/scikit-video.github.io>.
21. Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S, Darrell T. Caffe: convolutional architecture for fast feature embedding. 2014.
22. Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F, Chang C-C, Lin C-C. e1071: Misc functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. <https://cran.r-project.org/package=e1071>.
23. R Core Team: R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. R Foundation for Statistical Computing. <https://www.r-project.org/>. 2017.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

