Gaze Behaviour in Adolescents with Obsessive-Compulsive Disorder During Exposure Within Cognitive-Behavioural Therapy

Annika Thierfelder[1(✉)], Björn Severitt[2], Carolin S. Klein[3], Annika K. Alt[3], Karsten Hollmann[3], Andreas Bulling[4], and Winfried Ilg[1]

[1] Hertie Institute for Clinical Brain Research, Section for Computational Sensomotorics, University Hospital Tübingen, Tübingen, Germany
{annika.thierfelder,winfried.ilg}@uni-tuebingen.de
[2] Department of Computer Science, Human-Computer Interaction, University of Tübingen, Tübingen, Germany
bjoern.severitt@uni-tuebingen.de
[3] Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University Hospital Tübingen, Tübingen, Germany
{carolin.klein,annika.alt,karsten.hollmann}@med.uni-tuebingen.de
[4] Institute for Visualisation and Interactive Systems, University Stuttgart, Stuttgart, Germany
andreas.bulling@vis.uni-stuttgart.de

Abstract. Digital health interventions that involve monitoring patient behaviour increasingly benefit from improvements in sensor technology. Eye tracking in particular can provide useful information for psychotherapy but an effective method to extract this information is currently missing. We propose a method to analyse natural gaze behaviour during exposure exercises for obsessive-compulsive disorder (OCD). At the core of our method is a neural network to detect fixations based on gaze patch similarities. Detected fixations are clustered into *exposure-relevant*, *therapist*, and *other* locations and corresponding eye movement metrics are correlated with subjective stress reported during exposure. We evaluate our method on gaze and stress data recorded during video-based psychotherapy of four adolescents with OCD. We found that fixation duration onto *exposure-relevant* locations consistently increases with the perceived stress level as opposed to fixations onto *other* locations. Fixation behaviour towards the *therapist* varied largely between patients. Taken together, our results not only demonstrate the effectiveness of

This work is funded by the German Federal Ministry of Health (BMG) project SSTeP KiZ (2520DAT700) and the European Research Council (ERC SYNERGY Grant RELEVANCE). The authors thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting A. Thierfelder. A. Bulling was funded by the European Research Council (ERC; grant agreement 801708).

D. Salvi et al. (Eds.): PH 2023, LNICST 572, pp. 3–17, 2024.
https://doi.org/10.1007/978-3-031-59717-6_1

our method for analysing natural gaze behaviour during exposure sessions. The fixation analysis shows that patients allocate more attention towards exposure-related objects under higher stress levels, suggesting higher mental load. As such, providing feedback on fixation behaviour holds significant promise to support therapists in monitoring intensity of exposure exercises.

Keywords: mobile eye tracking · obsessive-compulsive disorder · sensor-assisted therapy · exposure exercises · real life gaze behaviour

1 Introduction and Related Work

Digital health interventions are becoming increasingly important to ensure that affected patients have easy access to treatment and to personalize said treatment. Video-based online therapy has proven its effectiveness in treating anxiety [11], including obsessive-compulsive disorder (OCD) [12]. OCD is a psychiatric disorder characterized by a combination of obsessions and compulsions [1]. Obsessions are repeated intrusive thoughts or urges that cause anxiety, whereas compulsions are mental or behavioural repetitions or rituals that the person with OCD feels the urge to do in order to reduce anxiety [33]. In children and adolescents, OCD affects around 0.5–4% of the population [9,10] and if not treated in young age, patients are at high risk to develop chronic symptoms [28].

The state-of-the-art treatment for OCD is cognitive behavioural therapy (CBT) based on exposure and response prevention (E/RP) sessions [1]. In the E/RP sessions, patients confront their obsession whilst refraining from the urge to perform the compulsion and instead enduring the anxiety until it, in the desired case, reduces without the compulsion. Given that obsessions are often linked to specific places or objects, which can be challenging to recreate in clinical environments, moving therapy into patients' homes through video-based CBT allows for a more direct confrontation with the obsession, thereby improving the effectiveness of the treatment [25].

A key limitation of such video-based approaches for CBT is the limited field of view of the web camera. Here, eye tracking devices can add valuable information about what is within sight of the patient, what the patient is focusing on at any point in time, and emerging gaze behaviour throughout therapy.

1.1 Fixations in Mobile Eye Tracking

What constitutes a fixation in mobile eye tracking differs from what is typically considered a fixation in laboratory-based eye tracking research [15,27]. There, fixations are temporal episodes during which the eyes remain stationary. In mobile eye tracking this definition is too narrow given that neither the body nor the head is necessarily still: Gaze may follow a moving object (smooth pursuit) or the head moves while the eyes keep fixating a still object. Thus, a fixation is typically defined as an episode during which gaze stays on a fixed object or location. This difference renders common methods based on eye movement velocity or gaze dispersion unsuitable for mobile eye tracking [15,27]. We therefore adapt the method from [27] that detects fixations based on the similarity of small regions around each gaze location within the visual scene.

1.2 Eye Tracking in Patients with OCD

Eye tracking is being increasingly used to study attention of patients with psychiatric disorders [3], including OCD [4,6,19]. The recording paradigm closest to real-life gaze behaviour is the free-viewing task in which patients are shown a neutral and an OCD-related stimulus at the same time without instructions on where to look. Studies using the free-viewing paradigm have shown evidence for a maintenance bias [4]: Patients with OCD show sustained attention towards OCD-related stimuli resulting in an increased number and duration of fixations on these stimuli [6,19].

An important marker during exposure exercises is the perceived stress triggered by the confrontation with the obsession. In healthy subjects, several studies investigated the effect of stress on gaze behaviour during real-life situations, e.g., flight simulation [30], interview settings [5] or the work day of ICU nurses [2]. These studies have shown that with increased stress the fixation duration drops [5,30] or, similarly, the number of fixations increases [2].

In these studies, stress has usually been induced by a time-pressured increase in mental effort to create anxiety. During exposure sessions, in contrast, stress is induced by confrontation with the obsessive thought without time constraints. Studies in healthy subjects have shown that an increase in mental effort without time pressure leads to longer fixation duration [8,16] and sometimes a concurrent increase in variability of fixation duration [22]. Longer fixations are thought to reflect the narrowed but increased attention allocated to the fixated goal [16] and therefore to be dependent on the task type [8].

Due to the lack of methods to analyse fixation behaviour in patients with OCD during real-life situations, the study of attention in said patients has been constrained to laboratory settings. While the effect of stress and mental effort on fixation behaviour in real-life situations has been well studied for healthy subjects, studies on their influence on real-life gaze behaviour in patients with OCD are still missing.

In this work we propose a method to detect and cluster fixations in mobile eye tracking during real-life therapy sessions, and analyse the effect of the subjective stress levels on fixation behaviour. Based on existing evidence for maintenance bias from laboratory eye tracking studies and the effect of mental effort on fixation behaviour, we hypothesised that attention on exposure-relevant locations increases with rising stress levels, showing in a higher fixation count and longer fixations towards these locations. If the predominant influence on gaze behaviour are stress and anxiety, the number of fixations onto other objects should also increase but fixations should become shorter.

2 Method

2.1 Study Details

Eye tracking data for this work was collected within the SSTeP KiZ study [13]. In this study, different sensor modalities were integrated into video-based CBT for children and adolescents with OCD, allowing patients to receive treatment in their home environments. An overview of the sensor-assisted therapy setup can be seen in Fig. 1. The sensor modalities included eye tracking, heart rate monitoring and hand movements, which have been shown to be promising candidates for measuring stress and compulsive behaviour [29]. The procedure was approved by the local ethics committee (877/2020BO1). In this work, we will focus on the eye tracking recordings of four patients from this study.

Fig. 1. Patients were equipped with a system to record their therapy sessions in their home environment. The data was streamed to the therapist UI, where the therapist had access to the egocentric video including gaze estimation and physiological measures. All icons are attributed to Flaticon.com

Treatment consisted of 14 session of video-based CBT. There was no exposure exercise within the first four sessions, since these were dedicated to building a therapist-patient relationship and psychoeducation in preparation for E/RP exercises. Afterwards, the amount of sessions including an exposure exercise was dependent on condition and therapy progress of the individual patient.

2.2 Data Collection and Labelling

The software architecture to record, transmit and display sensor data was custom designed for the purpose of the study [21]. Sensor data was recorded and synchronised locally and streamed to a therapist User Interface (UI), where the therapist could access the egocentric video of the patient together with the current gaze estimation and physiological parameters.

The therapist UI additionally served as a platform for data labelling where the therapists tagged time points defining the course of the therapy session, including start and end point of the exposure exercise. Throughout therapy, patients were asked to rate their perceived stress level on a scale from 0 to 10, which was also provided as label through the therapist UI.

Gaze data was recorded using the *Look!* head-mounted eye tracking device [14]. It included a scene camera with a resolution of 640 × 280 px and two eye cameras with a resolution of 320×240 px each. All videos were recorded at 30 Hz. Gaze estimation was computed using a convolutional neural network designed to be robust against small movements of the eye tracker to reduce the need for frequent recalibration [23]. Patients were asked to calibrate the system regularly, but at least before the first session and towards the middle of therapy.

2.3 Patients

We investigated a sample of four patients that participated in the SSTeP KiZ study. Symptom severity was assessed with the CY-BOCS score (Children's Yale-Brown Obsessive Compulsive Scale) before (t_0) and after treatment (t_1) [24]. A general reference for therapy success is a reduction of the CY-BOCS score by at least 35% [17], however, numerical symptom reduction can differ from personal experience. CY-BOCS values for all patients as well as demographic information and the amount and duration of exposure exercises can be found in Table 1.

Table 1. Overview of the patients included in this work. The table shows the amount of recorded exposure exercises per patient, mean and standard deviation of their durations, and the CY-BOCS score before and after treatment.

Patient			exposure exercises		CY-BOCS		
	age (year)	sex	n	length (min)	t_0	t_1	reduction
P1	17	f	9	38.6 ± 15.6	28	25	10.71%
P2	16	f	10	30.1 ± 11.9	21	13	38.10%
P3	17	m	7	28.9 ± 9.3	28	0	100%
P4	18	f	8	23 ± 6.4	29	12	58.62%

Manifestations of OCD are very heterogeneous across patients, which reflected in the four patients investigated in this work.

P1 showed a manifestation of OCD caused by the thought that certain objects are contaminated, triggering a strong feeling of disgust. To neutralise, these objects as well as both hands were cleaned excessively. Exposure exercises involved physical contact with "contaminated" objects such as doorknobs or contaminating "clean" objects such as one's bed. Subjectively, P1 reported a beneficial impact of the treatment, even though numerical symptom reduction was minor.

P2 showed the a repetition compulsion, repeating actions until they felt "just right", and a counting compulsion, mentally reciting certain number sequences over and over. Exposure for P2 included performing certain actions only once, e.g., closing the lid of a pen, and writing down a number included in the number sequence several times without mentally reciting the entire sequence.

P3 showed a contamination-based manifestation of OCD accompanied by the urge to perform frequent hand-washing. Exposures mainly consisted of touching "disgusting" objects like glue, the bathroom sink or toilet bowl. P3 started therapy with severe symptoms, but showed a surprisingly successful therapeutic effect, managing to reduce symptom severity by 100% to a minimum.

P4 showed a repetition-based manifestation of OCD with the urge to repeat actions with a positive thought until they felt "just right". Exposure sessions consisted mainly of selecting an item once and then performing the action with that item once (e.g., selecting and wearing the clothing item). The exercises were intensified by having to think of a negative event during the one-time execution.

2.4 Fixation Analysis Pipeline

To analyse the fixation behavior in the introduced patients, we adapted a fixation detection method for mobile eye tracking [27], structured the fixations using unsupervised clustering and assigned each cluster to *exposure-relevant*, *therapist* or *other* locations. The complete pipeline can be seen in Fig. 2.

Fixation Detection. First, we reduced noise in the gaze estimation by applying a moving average filter over a window of five frames. We then adapted an approach specifically designed for mobile eye tracking based on the assumption that image patches around the gaze estimation stay similar during a fixation [27].

We cropped an image patch of 50×50 px around the gaze estimation for every frame in the video. Each pair of consecutive frames served as input for a convolutional neural network (CNN) pretrained to predict patch similarity on the liberty dataset [7,34], resulting in a vector containing the patch similarity for each pair of frames. Sequences with similarities above the threshold of 1.3 were kept as fixation candidates. To remove outliers and ensure validity of fixations, we discarded candidates shorter than 3 frames (i.e., 100 ms) or longer than 95% of the data which corresponded to a maximal fixation length of 1 s.

The pipeline including the architecture of the CNN is shown in Fig. 2. It consists of two parallel 2-channel streams, one processing both full patches ("periphery") and one processing the central crops in a higher resolution ("fovea") that are integrated by two fully connected layers. Details on the network architecture can be found in the original publication [34].

To adapt the approach to our data, we tuned three hyperparameters: (1) image patch size, (2) similarity threshold and (3) the dataset for pretraining the patch similarity network. For tuning, we created a validation dataset consisting of a total video length of 15 min (i.e., roughly 27.000 frames) taken from three different subjects of the SSTeP KiZ study. All videos were labeled by at least two and at most three annotators to form the ground truth labels "fixation" and "no fixation" for each frame. The parameters were tuned by evaluating the fixations with event detection performance metrics [31].

Fixation Clustering. For each detected fixation, we extracted the centre frame as a representative and replaced its patch by a larger 256×256 px image cropped around the gaze estimation to obtain more context information. Since especially

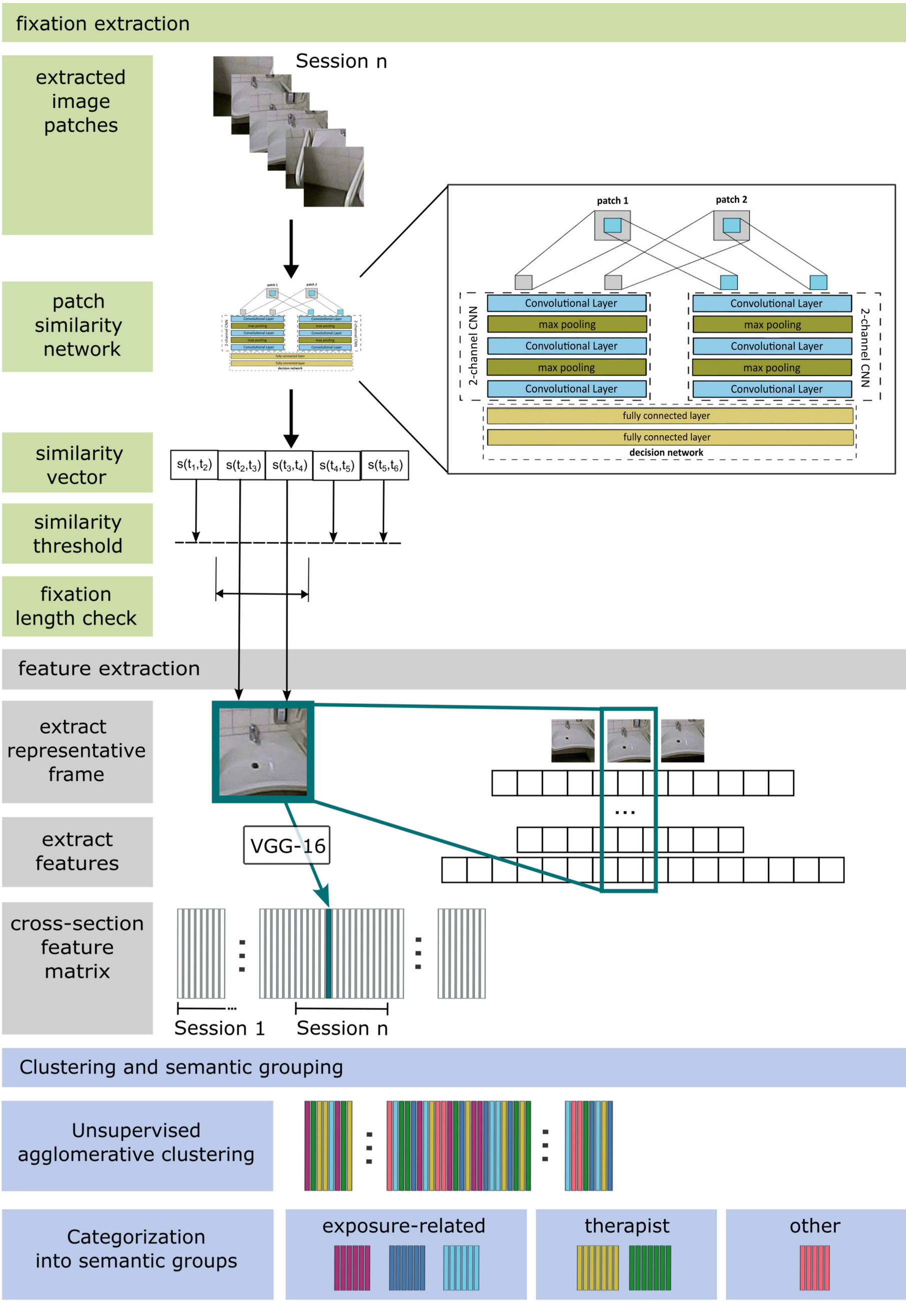

Fig. 2. Visualisation of the fixation processing pipeline, including fixation detection, feature extraction and unsupervised clustering. The network figure is adapted from [34].

in real-life environments the prediction of mobile eyetracking is not always accurate, cropping a larger window for clustering also reduces the effect of small localisation errors.

The image size was chosen as the optimal input size for the VGG-16 network pretrained on image recognition that we used to extract a feature vector for every representative image [26]. We extracted the output of the last fully-connected layer to get a 4096-dimensional vector containing high-level feature information for each image, resulting in a feature matrix of the size $n_{\text{fixations}} \times 4096$ for each session.

In order to find meaningful clusters across all therapy sessions, we appended the feature matrices of all sessions into one feature matrix. After normalization, the features were clustered with the agglomerative clustering algorithm implemented in the scikit-learn toolbox [20]. Agglomerative clustering is a bottom-up hierarchical clustering algorithm that starts with every sample as a single cluster and subsequently merges the closest clusters until either a maximum distance between clusters or a predefined number of clusters is reached.

The distance between single samples was computed as the euclidean distance and extended to distance between clusters with the Ward linkage criterion [32]. We computed distances between clusters for the full hierarchical tree until all clusters were merged. As stopping criterium, we then defined the maximal distance between clusters as the knee point among the largest 2000 distances.

We visually checked the resulting clusters and grouped them semantically into *exposure-related*, *therapist* and *other* locations. Few clusters showed a mix of different groups and were therefore not assigned. Note here, that the *exposure-related* group can contain very distinct clusters, since the conducted exposure exercises within a patient vary between sessions.

For visualising the results, we calculated a lower-dimensional representation of the features using UMAP [18]. Parameters were chosen to capture both the local and global structure of the data.

2.5 Analysis

For every reported stress level within an exposure exercise, we extracted the fixations within 1.5 min before and after the report. We computed fixation metrics for every semantic group separately, including the number of fixations, the median fixation duration and the interquartile range (IQR) of the fixation durations. Pearson correlation was calculated to assess the relation between the different fixation metrics and the corresponding reported stress level.

3 Results

3.1 Clustering Results

The clustering resulted on average in 31.25 clusters per subject of which on average 25 could be assigned to one of the semantic groups (P1: 25 cluster (21 assigned), P2: 25 (17), P3: 47 (40), P4: 28 (22)).

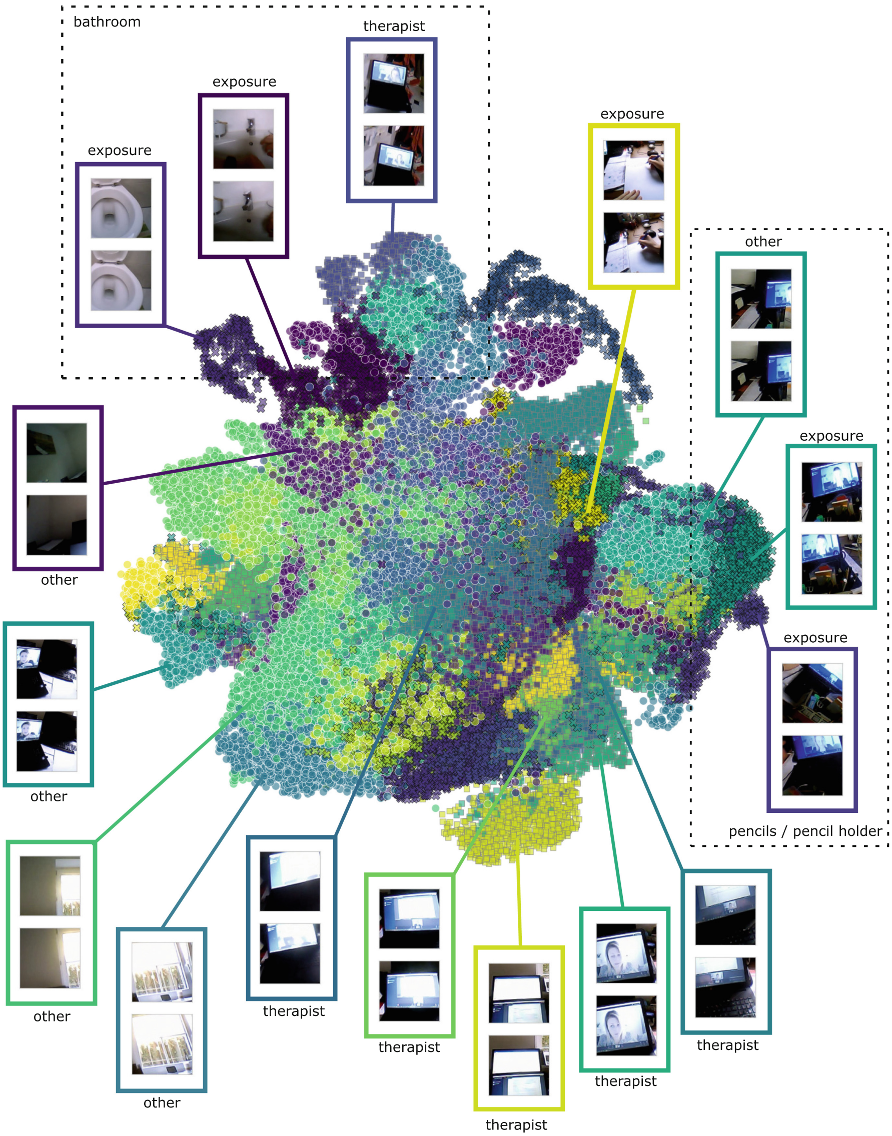

Fig. 3. Visualisation of the clustering results for P3. clusters were projected into a 2D space using UMAP. Clusters are colour-coded, while semantic groups are displayed as different shapes (X-shape: *exposure-relevant*, square: *therapist*, and circle: *other*). For exemplary clusters, the two images closest to the cluster centre are shown for illustration. Semantically meaningful grouped clusters are indicated by dashed lines.

The visualisation of clustering results for P3 along with examples for each cluster is shown in Fig. 3. In the 2D feature space, *therapist* clusters locate at the lower right. The *exposure-relevant* clusters where the pencil holder with the "contaminated" glue was placed in front of the therapist is separated but close to the *therapist* clusters. Clusters connected through the higher semantic level "being in a bathroom" are located at the top left part of the feature space.

3.2 Correlation Results

Results for the correlations of fixation metrics onto *exposure-relevant* locations with the reported stress level are displayed in Fig. 4. There was a significant increase in fixation duration ($p < 0.001, r = 0.41$) and fixation duration variability ($p < 0.001, r = 0.3$) with the perceived stress across all patients, and a small decrease in the number of fixations ($p = 0.01, r = -0.18$). The trend for increased fixation duration was observed in all subjects with statistical significance for P1 ($p = 0.02, r = 0.37$) and P2 ($p = 0.004, r = 0.34$). The increase in the fixation duration IQR did not occur in P1, but was significant for both P2 ($p = 0.016, r = 0.29$) and P3 ($p = 0.01, r = 0.42$). The patient-specific patterns regarding the number of fixations were individually different, and were only significant for P2 ($p = 0.01, r = -0.3$).

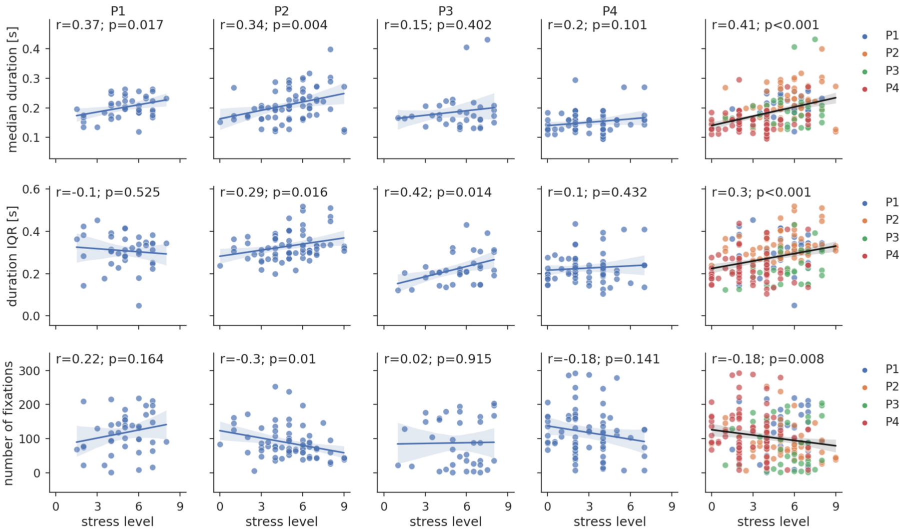

Fig. 4. Correlation of the patients' reported subjective stress levels with the metrics for fixations onto the *exposure-relevant* cluster. The first four columns represent a patient each, while the last column presents the results taken across all patients.

For fixations onto *other* locations, displayed in Fig. 5, we found small correlations of the stress level with fixation duration ($p = 0.04, r = 0.16$) and fixation

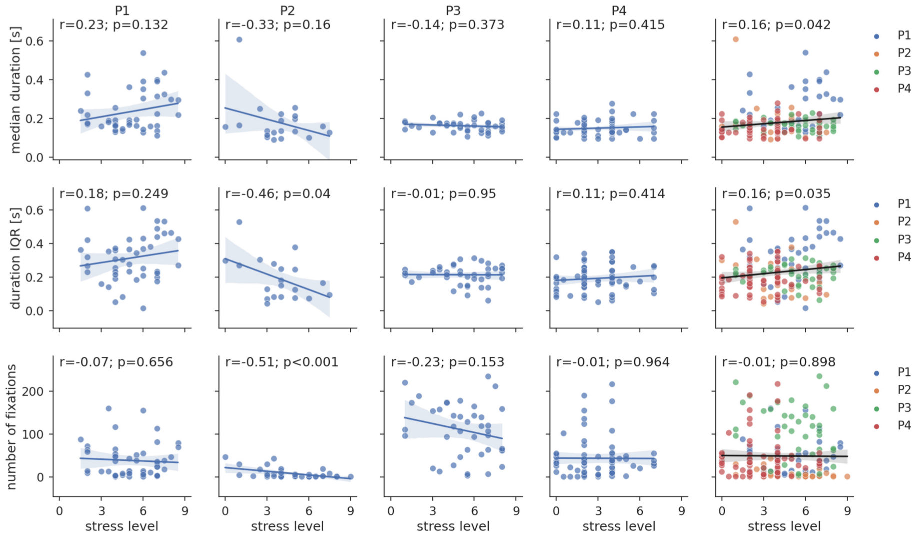

Fig. 5. Correlation of the patients' reported subjective stress levels with the metrics for fixations onto the *other* cluster. The first four columns represent a patient each, while the last column presents the results taken across all patients.

duration variability ($p = 0.04, r = 0.16$), but no correlation with the number of fixations ($p = 0.9$) across all patients. There was no consistent trend of duration increase in single patients and none showed significant results. We observed similar results for duration IQR, which was only significant for P2, where correlation was strongly negative ($p = 0.04, r = -0.46$) as opposed to the positive correlation across all patients. P2 was also the only patient that showed a significant correlation with the number of fixations ($p < 0.001, r = -0.51$).

Correlation results for fixations onto the *therapist* are shown in Fig. 6. For fixation duration and duration variability, there was no correlation with stress levels across all patients ($p > 0.48$) and no common trend across single patients. Correlation was only significant in P2 for fixation duration ($p = 0.01, r = 0.39$) and variability ($p = 0.002, r = 0.48$). The number of fixations across all patients decreased with reported stress level ($p < 0.001, r = -0.25$), which was not consistent across single patients, where the number of fixations decreased with stress for P1 ($p = 0.007, r = -0.35$) but increased for P2 ($p = 0.02, r = 0.34$).

4 Discussion

We proposed a method to analyse the fixation behaviour of children and adolescents with OCD during exposure exercises within video-based CBT. The method was specifically designed for challenges caused by real-life behaviour recorded with mobile eye tracking in home environments. For fixation detection we adapted an approach based on gaze patch similarity that is robust to head

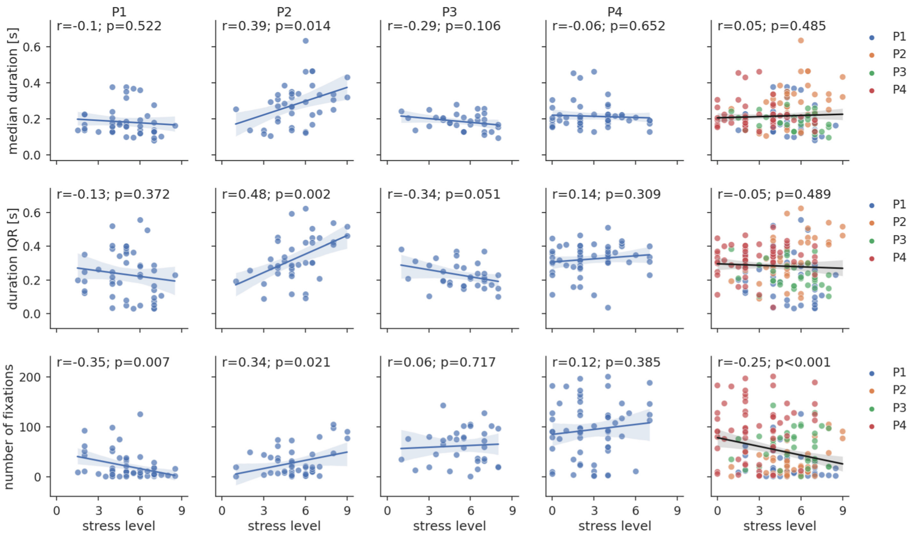

Fig. 6. Correlation of the patients' reported subjective stress levels with the metrics for fixations onto the *therapist* cluster. The first four columns represent a patient each, while the last column presents the results taken across all patients.

movements. We extended the method with unsupervised clustering to automatically identify targets in the real world from the fixated locations. We demonstrated that these clusters have semantic meaning and represent three categories of gaze targets: *exposure-relevant*, *therapist* and *other* locations.

We found that fixation duration onto *exposure-relevant* locations consistently reflected reported stress levels of the patient, i.e., fixation duration correlated positively with higher stress levels. This was accompanied by an increase in fixation variability, suggesting that fixation duration did not change systematically but that only some fixations lasted longer. While similar effects could be observed for *other* locations, these proved smaller and inconsistent across subjects.

Our results reveal that during real-life exposures patients put more attention on exposure-relevant locations if the subjective stress level, and thus the intensity of the exposure, increased. Given that fixation duration has been shown to increase with higher mental load, our results suggest that the patients' mental load and perceived stress level are closely connected.

There was no increase in the amount of fixations, neither onto *exposure-relevant* nor *other* locations, which we would have expected as an effect of the rising stress. In contrast, some patients even showed a decrease in the amount of fixations. Together with the increase in fixation duration, this indicates that reported stress reflects mental effort during exposure rather than anxiety.

Fixation behaviour towards the therapist was highly individual and there was no common effect across patients. This is not surprising given that fixation behaviour towards the therapist can depend on many variables that differ between patients, like the type of conversation during exposure or the patient-therapist relationship.

Our results underline the effectiveness of our method. However, it should be noted that the method assumes that exposure sessions contain a physical exposure to objects or locations. Therefore, this approach is most suitable for manifestations of OCD where obsessive thoughts are connected to a physical counterpart like an action, an object or a specific location.

In general, our findings stress the importance of analysing fixation behaviour during real-life exposure sessions as an extension of controlled eye tracking studies in the laboratory. Especially fixation duration and variability promise to be valuable parameters for therapists to monitor and adapt the patient's stress level and the connected mental load during exposure sessions. Since our analysis is exemplary on four patients, future studies would have to replicate and validate these results with a larger population.

4.1 Conclusion

We proposed a pipeline suitable for analysing gaze behaviour of patients with OCD during exposure sessions in their home environments. Although further research is needed to validate our findings, our work provides a preliminary argument for the usefulness of eyetracking for patients with OCD. Providing feedback about gaze behaviour could therefore support therapists in monitoring stress and mental load of patients, helping them to adapt to the needs of the patient. Next steps will be to use our approach for behavioural feedback to therapists in exposure exercises practised as homework outside of therapy sessions, to ensure correctness and prevent avoidance behaviour. In future research, it will also be interesting to connect the gaze features not only with perceived stress but also with physiological measures of stress such as heart rate.

References

1. Abramowitz, J.S., Taylor, S., McKay, D.: Obsessive-compulsive disorder. Lancet **374**(9688), 491–499 (2009)
2. Ahmadi, N., et al.: Quantifying Workload and Stress in Intensive Care Unit Nurses: Preliminary Evaluation Using Continuous Eye-Tracking. Human Factors (2022)
3. Armstrong, T., Olatunji, B.O.: Eye tracking of attention in the affective disorders: a meta-analytic review and synthesis. Clin. Psychol. Rev. **32**(8), 704–723 (2012)
4. Basel, D., Hallel, H., Dar, R., Lazarov, A.: Attention allocation in OCD: a systematic review and meta-analysis of eye-tracking-based research. J. Affect. Disord. **324**, 539–550 (2023)
5. Behroozi, M., Lui, A., Moore, I., Ford, D., Parnin, C.: Dazed: measuring the cognitive load of solving technical interview problems at the whiteboard. In: Proceedings of the 40th International Conference on Software Engineering: New Ideas and Emerging Results, pp. 93–96. ACM, Gothenburg Sweden (2018). https://doi.org/10.1145/3183399.3183415

6. Bradley, M.C., Hanna, D., Wilson, P., Scott, G., Quinn, P., Dyer, K.F.W.: Obsessive-compulsive symptoms and attentional bias: an eye-tracking methodology. J. Behav. Ther. Exp. Psychiatry **50**, 303–308 (2016)
7. Brown, M., Gang Hua, Winder, S.: Discriminative learning of local image descriptors. IEEE Trans. Pattern Anal. Mach. Intell. **33**(1), 43–57 (2011). https://doi.org/10.1109/TPAMI.2010.54
8. Chen, S., Epps, J., Ruiz, N., Chen, F.: Eye activity as a measure of human mental effort in HCI. In: Proceedings of the 16th International Conference on Intelligent User Interfaces, pp. 315–318. IUI 2011. ACM, New York (2011). https://doi.org/10.1145/1943403.1943454
9. Douglass, H.M., Moffitt, T.E., Dar, R., McGee, R., Silva, P.: Obsessive-compulsive disorder in a Birth Cohort of 18-Year-Olds: prevalence and predictors. J. Am. Acad. Child Adolesc. Psychiatry **34**(11), 1424–1431 (1995)
10. Heyman, I., Fombonne, E., Simmons, H., Ford, T., Meltzer, H., Goodman, R.: Prevalence of obsessive-compulsive disorder in the British nationwide survey of child mental health. British J. Psychiatry J. Mental Sci. **179**, 324–329 (2001)
11. Hollis, C., Falconer, C.J., Martin, J.L., Whittington, C., Stockton, S., Glazebrook, C., Davies, E.B.: Annual research review: digital health interventions for children and young people with mental health problems - a systematic and meta-review. J. Child Psychol. Psychiatry **58**(4), 474–503 (2017)
12. Hollmann, K., et al.: Internet-based cognitive behavioral therapy in children and adolescents with obsessive-compulsive disorder: a randomized controlled trial. Front Psychiatry **13** (2022)
13. Klein, C.S., et al.: Smart sensory technology in tele-psychotherapy of children and Adolescents with Obsessive-Compulsive Disorder (OCD): a feasibility study. preprint, SSRN (2023). https://doi.org/10.2139/ssrn.4395216
14. Kübler, T.: Look! Technical specifications, Blickschulungsbrille (2021)
15. Lappi, O.: Eye movements in the wild: oculomotor control, gaze behavior & frames of reference. Neurosci. Biobehav. Rev. **69**, 49–68 (2016)
16. Marquart, G., Cabrall, C., De Winter, J.: Review of eye-related measures of drivers' mental workload. Procedia Manuf. **3**, 2854–2861 (2015)
17. Mataix-Cols, D., de la Cruz, L.F., Nordsletten, A.E., Lenhard, F., Isomura, K., Simpson, H.B.: Towards an international expert consensus for defining treatment response, remission, recovery and relapse in obsessive-compulsive disorder. World Psychiatry **15**(1), 80–81 (2016)
18. McInnes, L., Healy, J., Melville, J.: UMAP: uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426 (2018)
19. Mullen, M., Hanna, D., Bradley, M., Rogers, D., Jordan, J.A., Dyer, K.F.W.: Attentional bias in individuals with obsessive-compulsive disorder: a preliminary eye-tracking study. J. Behav. Cogn. Ther. **31**(2), 199–204 (2021)
20. Pedregosa, F., et al.: Scikit-learn: machine learning in Python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)
21. Primbs, J., et al.: The SSTeP-KiZ system-secure real-time communication based on open web standards for multimodal sensor-assisted tele-psychotherapy. Sensors **22**(24), 9589 (2022)
22. Recarte, M.A., Nunes, L.M.: Effects of verbal and spatial-imagery tasks on eye fixations while driving. J. Exp. Psychol. Appl. **6**(1), 31–43 (2000)
23. Santini, T., Niehorster, D.C., Kasneci, E.: Get a grip: slippage-robust and glint-free gaze estimation for real-time pervasive head-mounted eye tracking. In: Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications, pp. 1–10. ACM (2019). https://doi.org/10.1145/3314111.3319835

24. Scahill, L., Riddle, M.A., McSwiggin-Hardin, M., Ort, S.I., King, R.A., Goodman, W.K., Cicchetti, D., Leckman, J.F.: Children's yale-brown obsessive compulsive scale: reliability and validity. J. Am. Acad. Child Adolesc. Psychiatry **36**(6), 844–852 (1997)
25. Selles, R.R., et al.: Effects of treatment setting on outcomes of flexibly-dosed intensive cognitive behavioral therapy for pediatric ocd: a randomized controlled pilot trial. Front Psychiatry **12** (2021)
26. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint arXiv:1409.1556 (2015)
27. Steil, J., Huang, M.X., Bulling, A.: Fixation detection for head-mounted eye tracking based on visual similarity of gaze targets. In: Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications, pp. 1–9. ACM (2018). https://doi.org/10.1145/3204493.3204538
28. Stewart, S.E., Geller, D.A., Jenike, M., Pauls, D., Shaw, D., Mullin, B., Faraone, S.V.: Long-term outcome of pediatric obsessive-compulsive disorder: a meta-analysis and qualitative review of the literature. Acta Psychiatr. Scand. **110**(1), 4–13 (2004)
29. Thierfelder, A., et al.: Multimodal sensor-based identification of stress and compulsive actions in children with obsessive-compulsive disorder for telemedical treatment. In: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 2976–2982 (2022). https://doi.org/10.1109/EMBC48229.2022.9871899
30. Tichon, J.G., Wallis, G., Riek, S., Mavin, T.: Physiological measurement of anxiety to evaluate performance in simulation training. Cognition, Technol. Work **16**(2), 203–210 (2014)
31. Ward, J.A., Lukowicz, P., Gellersen, H.W.: Performance metrics for activity recognition. ACM Trans. Intell. Syst. Technol. **2**(1), 1–23 (2011). https://doi.org/10.1145/1889681.1889687
32. Ward, J.H.: Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc. **58**(301), 236–244 (1963)
33. World Health Organization: International classification of diseases for mortality and morbidity statistics (11th revision). https://icd.who.int/browse11/l-m/en. Accessed 9 July 2023
34. Zagoruyko, S., Komodakis, N.: Learning to compare image patches via convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4353–4361 (2015)