Combining Trending Scan Paths with Arousal to Model Visual Behaviour on the Web: A Case Study of Neurotypical People vs People with Autism

DOI:
10.1145/3320435.3320446

Combining Trending Scan Paths with Arousal to Model Visual Behaviour on the Web: A Case Study of Neurotypical People vs People with Autism

Oludamilare Matthews[1], Sukru Eraslan[2], Victoria Yaneva[3], Alan Davies[1],
Yeliz Yesilada[2], Markel Vigo[1] and Simon Harper[1]
[1]University of Manchester, UK
[2]Middle East Technical University Northern Cyprus Campus
[3]University of Wolverhampton, UK

ABSTRACT

People with autism often exhibit different visual behaviours from neurotypical users. To explore how these differences are exhibited on the Web, we model visual behaviour by combining pupillary response, which is an unobtrusive measure of physiological arousal, with eye-tracking scan paths that indicate visual attention. We evaluated our approach with two populations: 19 neurotypical users and 19 users with autism. We observe differences in their visual behaviours as, in certain instances, individuals with autism exhibit a lower arousal response to affective contents. While this is consistent with the literature on autism, we confirm this phenomenon on the Web. We discuss how our modelling method can be used to identify possible UX issues such as the presence of stress, cognitive load and differences in the perception of Web elements in relation to physiological arousal.

CCS CONCEPTS

• **Human-centered computing** → **User studies**; **HCI theory, concepts and models**; **Empirical studies in HCI**; • **Applied computing** → *Psychology*;

KEYWORDS

Arousal, Pupillary response, HCI, UX, Scan path, STA, Autism

ACM Reference Format:

Oludamilare Matthews, Sukru Eraslan, Victoria Yaneva, Alan Davies, Yeliz Yesilada, Markel Vigo and Simon Harper. 2019. Combining Trending Scan Paths with Arousal to Model Visual Behaviour on the Web: A Case Study of Neurotypical People vs People with Autism. In *27th Conference on User Modeling, Adaptation and Personalization (UMAP '19), June 9–12, 2019, Larnaca, Cyprus.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3320435.3320446

Corresponding author: oludamilare.matthews@manchester.ac.uk.

UMAP '19, June 9–12, 2019, Larnaca, Cyprus

ACM ISBN 978-1-4503-6021-0/19/06…$15.00
https://doi.org/10.1145/3320435.3320446

1 INTRODUCTION

We present a novel methodology that can be used to generate a descriptive user model based on users' visual behaviour and their affective response. This can be used to better understand user interaction, and is therefore applicable in adaptive systems and intelligent user interfaces. We have used the case of individuals with autism vs. neurotypical users on the web to suggest that our approach can be used to uncover differences in interest, affective response and visual behaviour.

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterized by differences in communication and social interaction [2]. The overall prevalence of autism is estimated to be between 1.1 and 1.2% of the UK populace [9]. The presence of autism relates to different experiences when interacting with the web and can elicit different affection responses [16, 17]. For example, people with autism often exhibit idiosyncratic visual attention patterns, which have been shown to affect their processing of web pages [27]. Also, the strong preference for structure and familiarity that many individuals on the spectrum have, may be challenged by changes in the structure or the interface of many applications [49]. Furthermore, there are well-documented differences in the way some people on the spectrum interpret affective states, and the presence of common facial expressions on web pages may yield inconsistent outcomes by people with autism [22].

There are many methods of investigating UX problems from both qualitative and quantitative paradigms, including questionnaires, eye-tracking and physiological computing. Qualitative methods of investigating UX problems often require participants to communicate their experience. This may be inaccurate due to many reasons such as lack of self-reflection with regards to one's own difficulties, as well as difficulties with communication in general, which is also one of the diagnostic criteria for ASD [32]. An alternative approach which requires less verbalisation is using self-reported scales to elicit feedback from users. However, recollecting experiences requires cognitive processing and self-reflection and, because of these reasons, it may be highly subjective and inaccurate. Also, some people with autism, ADHD (Attention Deficit Hyperactivity Disorder) and similar developmental disorders may exhibit cognitive impairment [43], thereby limiting the accuracy of their reported scores. Due to these limitations, we argue that methods that require eliciting intentional feedback from individuals with autism are less reliable for understanding the UX challenges that people with autism face. Usability metrics such as error rates and completion times can be used to detect common problems that users experience on the web. Predominantly, these metrics answer the question - *"Is there a problem?"* but, discovering problems is only the first step to improving UX. Analysis of gaze behaviour with metrics like fixation location/duration and saccades (visual transitions) can be used to answer the more advanced question - *"Where is the problem?"* but, this is also limited because knowing where the problem lies may not always lead to a solution. Trying a different approach may alleviate the problem for one user, but, without knowing why the problem exists, we may not have an understanding of which group of users the problem affects. People with autism are not the only atypical groups of users on the web. In addition to the idiosyncracies of typical users, people with learning, developmental, obsessive-compulsive (OCD), and other types of disorders may also require special considerations when carrying out studies to identify UX issues specific to these user groups. The answer to the question - *"Why is there a problem there?"* provides more context so that researchers can offer recommendations to overcome the existing UX issue(s).

Affective computing is one that relates to, arises from or influences emotions [34]. This is useful notably when the physiological state of users can be detected and related to their interaction patterns. We show that the combination of visual behaviour and physiological computing presents a rich methodology for identifying and understanding UX issues. Data collected through physiological means are often noisy and may also lack required sample sizes and data distributions to make finding generalizable. Our approach provides a descriptive method that aids hypothesis genesis about UX issues which can then be followed up by qualitative feedback from users or further inferential statistical analysis.

For our methodology, we combine two different algorithms: 1. The Scanpath Trend Analysis (STA) algorithm, which provides the trending scan path followed by a group of users [18] and 2. Arousal detection through the analysis of pupillary response to identify moments of increased arousal. Gaze analysis is then used to identify the user's visual attention on the screen (i.e., visual element) during moments of increased arousal [25]. The individual arousal scores for each participant are averaged for each visual element. The sequence of each visual element of the trending scan path and the corresponding arousal score is combined to produce a visualisation. The output of our approach is an aggregate of a group of users' scan paths over a visual stimulus, and their arousal response to each element of the scan path. Based on this approach, our research questions are as follows:

- RQ1. Does the combination of scan path analysis and level of arousal reveal differences between people with autism and neurotypical people in web browsing tasks?
- RQ2. Does the combination of scan path analysis and arousal reveal *where* the differences in arousal occur between people with autism and neurotypical people in web browsing tasks?

To investigate our research questions, we use the dataset from an eye-tracking study with 38 participants (19 with ASD and 19 neurotypical users) on a browsing task. We observed a difference in the trending scan paths between both groups and a difference in the arousal score for some of the visual elements that made up those trending scan paths. Our contributions are as follows:

(1) A methodology that combines visual scan path analysis with arousal scores to provide a more holistic understanding of the users' experience.
(2) The analysis and visualisation of differences in individuals with autism vs neurotypical users in web browsing tasks using our methodology.

Our approach and findings present a novel research methodology to identify and improve understanding of user interaction problems of user groups with varied interaction patterns and experiences.

2 BACKGROUND

Eye-trackers are devices that capture the user's gaze at high frequency, usually > 50Hz. They are used to observe gaze behaviours to aid in understanding responses pertaining to temporal aspects of the interaction, or locations mapped to the stimulus by researchers for semantic representation of areas that can be interrogated for analysis known as areas of interest (AOIs). Boraston *et al.* discussed the applications of eye tracking in the study of autism [8]. Some of the metrics used in eye-tracking studies include the user's saccades (visual transitions), scan paths (sequential path of fixations) and fixation duration on an AOI [29]. Studies have also shown that fixation duration can be used as a proxy measure of attention [35]. These eye-tracking metrics have been used in usability [36], accessibility [37] and UX testing [7] to improve the quality of human-computer interaction (HCI).

2.1 Physiological arousal

Arousal is a dimension in the Pleasure-Arousal-Dominance (PAD) scale proposed by Russell *et al.* that defines the strength and intensity of an emotion [39]. There are many arousal detection mechanisms including Galvanic Skin Response (GSR), Electrocardiogram (ECG), Electroencephalogram (EEG), Electromyography (EMG), Heart Rate (HR) and Skin Temperature [1, 3, 10, 14, 40, 56]. Physiological sensors are capable of generating rich, highly granular data. However, they are easily distorted by noise and confounding factors. For example, GSR is sensitive to temperature [26], motion [44], and has known issues with latency [28]. Encephalograms are commonly used in to capture data in control settings due to its complexity and skills to set up while the data is analysed posthoc because the EEG signals are sensitive to cognitive and motor activities, making it complex to analyse. Electrocardiography and heart rates are known for sensing low intensity and long-term changes.

Multi-modal approaches (using multiple sensors) have become a popular approach to sensing arousal [12, 13]. They have been used in lab settings to reduce the effect of confounding variables. Multiple sources are more costly, require more skills to set up, and are likely to be more intrusive, which means that such solutions are not always practical [24]. This implies that they are less suitable in naturalistic settings. Additionally, the majority are not suitable for ubiquitous use due to the cost and skills required to set them up, and they may be of limited use for identifying the context or cause of a change in arousal. For these reasons, pupillary response and gaze detection were given further consideration for use in our methodology.

2.2 Pupillary response

In the context of this study, pupillary response refers to changes in pupil diameter that are related to affect. The primary function of the pupil is to regulate the amount of light from the cornea directed towards the retina. In a controlled experiment where the amount of light could be regulated, the pupillary response is a useful affect discriminator for arousal [31]. When an affective stimulus is perceived, the pupil dilates until it reaches a peak point where the effect begins to decay until it returns to baseline. This cycle of pupillary response takes between two to six seconds to complete [15]. This duration depends on several factors, including the initial state of the pupil (normal, dilated or contracted), the nature of the previous and current stimulus (light qualities, cognitive, affective), individual biological differences etc.

2.3 Potential applications of measuring arousal

As early as 1908, psychologists Yerkes and Dodson established an empirical relationship between arousal and mental performance [54]. The law states that performance increases along with physiological arousal but only up to a point where performance begins to decline as arousal increases. During interaction with visual contents, several cognitive states are of key importance such as, attention, alertness, boredom, calmness, etc. all of which can be plotted on the arousal scale. Additionally, with a level of control over the content and the interaction, it is possible to increase the likelihood of desirable states or ameliorate undesirable ones by altering the stimulus that induces them. Affect detection (especially arousal) has been applied to several domains including, affective gaming [21], psychological interventions [55], group sensing [23], intelligent tutoring systems [41], user interfaces [6], recommender systems [42] and accessibility and usability [30].
To the best of our knowledge, this is the first paper that combines pupillary response, gaze behaviour and scan path analysis to understand users' visual and affective behaviour.

2.4 Autism and the web

Previous work showed evidence that atypical visual attention in individuals with autism may result in unconventional information-searching strategies on the web [17]. Such differences revealed through eye-tracking have also been used to classify users into autistic and neurotypical groups [49]. A more thoroughly researched factor on the web known to affect the two groups differently is textual content. While not specifically presented on web pages, deficits in reading comprehension among people with autism have been widely researched [11, 38], including by means of eye-tracking experiments [46, 47, 52, 53] and in combination with images [51]. This issue has been addressed in readability research by attempting to measure the difficulty of text for readers with autism specifically based on the difficulties they encounter [48, 50]. To the best of our knowledge, the effects of emotive facial expressions on arousal while processing web pages have not yet been studied for this population.

3 METHODOLOGY

In this section we explain the methodology employed to address our research questions.

3.1 Participants

A total of 38 participants, 19 with a formal diagnosis of autism and 19 control-group participants were recruited for this study[1]. None of them had any diagnosed degree of intellectual disability, nor any reading disorders. The mean age for the ASD group was m = 41.05 with SD = 14.04, and m = 32.15, SD = 9.93 for the control group. All participants with ASD were recruited through a UK autism charity, and the student enabling centre at the University of Wolverhampton. All control-group participants were recruited through snowball sampling. To ensure that none of the control group participants had a high incidence of autism-related features, they were required to fill in the Autism Quotient test [5], formally used to screen people at risk before referral for a diagnosis by an expert. Both the participants with ASD and the control-group participants were highly able adults, all of whom were living independently and without relying on a caregiver. From the ASD group, 11 people had completed a higher education degree, six people had a UK equivalent of a high-school degree (GCSE or A-levels), and two people preferred not to answer. From the control group, 15 people had completed a higher education degree, and three people had completed A-levels (equivalent to high school). All participants were native speakers of English except for four control-group participants, who were highly fluent, having lived in the UK for many years. All participants reported that they use the web daily, with only one ASD participant reporting web usage "less than once a month". All participants identified as having normal or corrected to normal vision.

3.2 Apparatus

A Gazepoint GP3 video-based eye-tracker was used to capture pupillary response, fixation location, and fixation duration of the participants at a frequency of 60Hz. All questions and answers were exchanged verbally, hence no mouse or keyboard were used. The stimuli were presented on a 17" LCD monitor. The experiment was run using the Gazepoint experimental environment and the laptop used for the experiments had a Windows 10 operating system.

3.3 Materials and Method

Eight web pages were selected by first exploring the home pages of the top 100 websites listed by *Alexa.com*, excluding those that were repeated more than once. Pages that were not in English and were mainly designed for authentication and/or as search pages were also excluded. We then selected the final eight pages in such a way, as to have a balanced representation of factors such as complexity and space between elements. The complexity values were obtained using the VICRAM algorithm [45]. In our final selection, an equal number of pages had a high complexity (YouTube, Amazon, Adobe and BBC) and low complexity (WordPress, WhatsApp, Outlook and Netflix), as well as small (Outlook, Netflix, Adobe and BBC) and large space (WordPress, WhatsApp, YouTube and Amazon) between their elements. Participants were presented with screenshots of the pages to ensure consistency in the look and feel of the web pages. For images of each page, see Figure 1.

[1]This experiment was approved by the University of Wolverhampton, UK committee on ethics.

There were two types of tasks: browsing and synthesis, presented in a counterbalanced order for each participant. For the browsing task, the participants were free to explore each page for 30 seconds. For the synthesis task, each participant had up to 120 seconds to answer two questions per page, with the possibility move forward earlier if they had answered the questions. In this paper, we selected the browsing task so that the analysis is based on a single web page per website.

All experiments were conducted in a quiet room. First, the consent form and the demographic questionnaire were filled in by the participants. After that, the eye tracker was calibrated using a nine-point calibration, and the experiment commenced. All questions and answers were given verbally, and the participants were all given a break between the tasks. After completing the experiment, all participants were debriefed.

3.4 Analysis

The analysis is carried out in two main stages and the algorithms facilitated are explained below: (1) Scan path analysis using the STA algorithm and (2) Generating arousal scores for the trending scan paths of each group of users (individuals with autism and neurotypical).

3.4.1 Scanpath Trend Analysis (STA). .
The Scanpath Trend Analysis (STA) algorithm identifies the trending path of multiple users on a web page in terms of its AOIs. The STA algorithm is a multi-pass algorithm which is comprised of three core stages: (1) Preliminary Stage, (2) First Pass and (3) Second Pass.

(1) Preliminary Stage: This stage firstly takes a series of fixations for each user on a particular web page and the details of the AOIs of the page. It then matches each fixation with its corresponding AOI to generate the individual scan paths in terms of the AOIs of the web page.
(2) First Pass: Once the individual scan paths are ready for further processing, the First Pass start analysing them to identify trending AOIs by selecting the AOIs which are shared by all the users or catch at least the same attention as the fully shared AOIs based on their total fixation durations and total fixation counts.
(3) Second Pass: After identifying trending AOIs, the Second Pass calculates an overall sequential priority value for each trending AOI based on their positions in the individual scan paths. It then combines these AOIs based their priority values to discover the trending path where the trending AOI with the highest priority will be the first one in the trending path

The detailed description of the STA algorithm can be found in [18]. The STA algorithm was evaluated by comparing its resultant paths with the resultant paths of other similar algorithms by using different AOI detection approaches [18, 19]. The evaluation shows that the resultant path of the STA algorithm is the most similar one to individual scan paths, thus it discovers the most representative path. The detailed results of the evaluation can be found in [18, 19].

3.4.2 Arousal sensing and detection of focal attention. .
This algorithm is used to sense changes in arousal through the analysis of pupillary response data from an eye tracker, as described in the following five steps:

(1) Pupillary response for each participant is collected from an eye-tracker and aggregated into fixed, non-overlapping windows. The most fixated area on the screen (**a**) for this window is also identified.
(2) The aggregated data is transformed into a vector of values that models each participant considering their range and central tendency (median).
(3) The model above is fed into a peak detection algorithm to detect local maxima (peaks). The peak indicates point where there is an increase in arousal level. The magnitude of increase from the lowest point before the increase is measured as (**V**).
(4) The area of the screen that was mostly fixated upon before the peak (**a**) extracted from step 1, as the cause of the peak.
(5) The total fixation duration for (**a**) is denoted as (**t**).

The arousal **(A)** due to the AOI **(a)**, is given by Equation 1 where **(V)** is the magnitude of increase, and **(t)** is the total fixation duration. For a more detailed description of the algorithm, refer to [25].

$$A_a = \left(\sum V_a \right) t_a \tag{1}$$

Equation 1: Arousal score per AOI

We merge the two algorithms by computing the average and standard deviations of the arousal scores for each of the visual segments (AOIs) that make up the trending scan paths from STA algorithm.

3.4.3 Visualizing our visual behaviour model. .
To help us to determine where the primary differences are located, a visualisation was utilised for data exploration (Figure 1). The circles that are superimposed on the AOIs that elicited the arousal contain a letter and number. The letter indicates the sequential order of visualisation that occurred on viewing the web page. The corresponding number represents the arousal levels (**AL**) for each AOI and is rounded to the nearest whole number so that the size of each circle indicates the ordinal level of arousal for each group. The arousal levels on this visualisation can be treated as ordinal measures where 1 to 3 indicate low arousal, 4-6 medium and 7-9 high levels of arousal. This visualisation was used to generate hypotheses that may explain the general behaviour of participants in each group.

4 RESULTS

In relation to RQ1, which concerns detecting differences in arousal between both groups for each task, we computed the mean arousal score per participant for each website. Results from Table 1 indicate that only the *YouTube* website shows a significant difference in arousal between the trending scan paths of the individuals with autism and the neurotypical group.

The mean arousal for the neurotypical group (M=4.00, SD=2.80) for the *YouTube* website was also higher than that of the ASD group

Table 1: Results of Mann Whitney U test comparing arousal between each group (autistc and neurotypical) with Bonferroni correction α = 0.00625

Website	*U*	p-value
WhatsApp	44.0	0.077
Amazon	792.5	0.380
WordPress	755.5	0.233
Netflix	2132.0	0.438
BBC	341.5	0.208
YouTube	388.5	0.002*
Adobe	1118.0	0.211
Outlook	405.0	0.231
Note: * = p < 0.00625		

(*M*=2.41, *SD*=2.24). This is contrary to our hypothesis that people with autism would find browsing tasks more stressful.

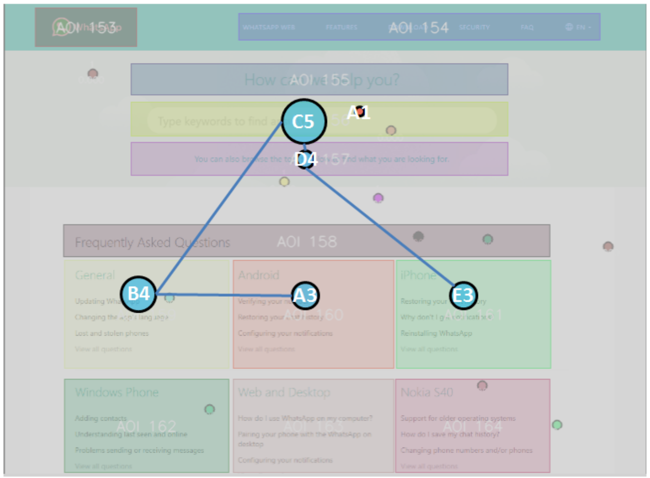
With regards to RQ2, about identifying differences in AOIs between groups, further exploration shows that both groups experience arousal in different degrees from different elements in their respective scan paths.

The average and standard deviations of the arousal scores that make up the trending scan paths for each group is given in Table 2. For example in the first row, 13 participants (out of 19 in the controlled group) experienced a mean arousal score, *M*=3.3, *SD*=2.98 while looking at 'AOI 160' of the *Whatsapp* website whereas, no participant from the ASD group experienced an increase in arousal caused by the same AOI. From our visualisation in Figure 1a, there was a longer trending scan path for the control group, compared to the ASD group, which had only one element, AOI 156 (the *WhatsApp* search bar). The control group experienced a level 5 measure of arousal, whereas the ASD group was level 1. In Figure 1d, the controlled group experienced a more varied range of arousal from level 1 to level 6 over nine different elements compared to the ASD group that ranged from level 3 to 5 over five elements. Similarly, on figure 1f, we can observe that the controlled group has a longer scan path and more common visual coverage over the website whereas, the ASD group have a linear, horizontal scan path across the *YouTube* web page. The arousal response due to AOI 200, from the controlled group (*AL*=6) was more than the ASD group (*AL*=2). Interestingly, AOI 200 pertains to a video about Stephen Hawking and his death. Overall, the ASD group exhibited lower levels of arousal (*AL*=4 vs *AL*=5 by the control group) when looking at the UI element (AOI 199). The image displayed on this AOI was a picture of people laughing in excitement. This difference may potentially be due to differences in the processing of emotive facial expressions between people with and without autism, as previously suggested by other studies [20].

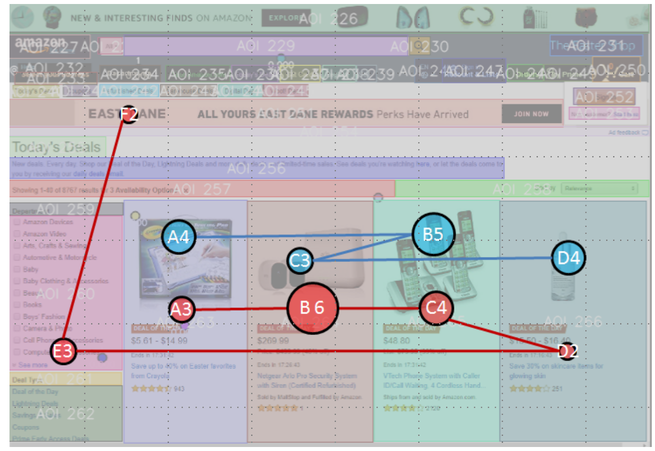
Results show that the fusion of the STA algorithm and the arousal sensing algorithm can be used to summarise behaviours between groups of users. The observations made using our approach can be analysed further using either inferential statistics or qualitative methods to provide additional supporting evidence for research findings.

Table 2: Scan path Sequence (Seq), participants (*n*) with change in arousal level per AOI, mean arousal (*M*) for the participants and standard deviation (*SD*) for the controlled and the individuals with autism (ASD).

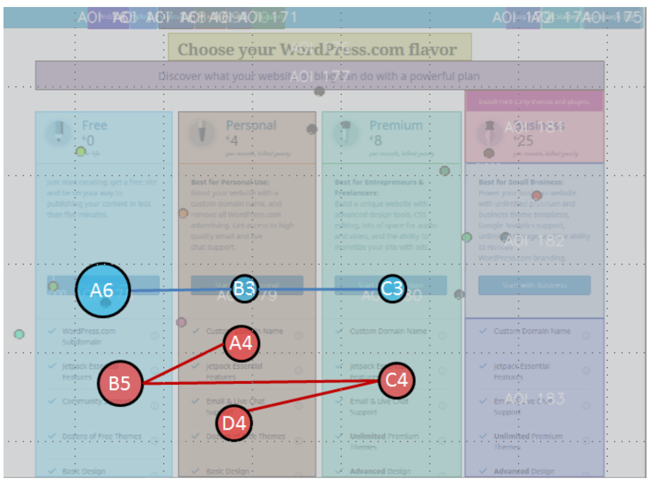
		Controlled				ASD			
Website	**Seq**	**AOI**	***n***	***M***	***SD***	**AOI**	***n***	***M***	***SD***
WhatsApp	A	160	13	3.3	2.98	156	4	1.28	0.33
	B	159	10	3.68	3.09				
	C	156	5	5.37	3.61				
	D	157	3	2.02	1.28				
	E	161	8	2.8	2.7				
Amazon	A	263	8	3.82	2.76	263	8	2.95	2.58
	B	265	6	5.19	3.59	264	12	6.27	3.22
	C	264	9	3.32	3.35	265	11	3.67	3.02
	D	266	10	4.11	3.83	266	5	2.36	2.62
	E					260	7	3.47	3.31
	F					251	7	1.61	0.96
WordPress	A	178	14	5.67	3.38	179	10	4.48	3.03
	B	179	16	2.96	2.83	178	12	4.74	3.45
	C	180	7	2.68	2.93	180	13	3.6	3.27
	D					179	10	4.48	3.03
Netflix	A	126	6	1.67	1.03	277	6	5.31	4.1
	B	125	11	2.8	2.57	279	12	3.97	3.11
	C	277	7	4.77	2.86	280	11	2.99	2.53
	D	280	4	6.44	3.78	279	12	3.97	3.11
	E	279	12	3.83	3.48	126	6	3.25	2.99
	F	277	7	4.77	2.86	281	10	2.71	3.01
	G	280	4	6.44	3.78				
	H	118	1	1	-				
	I	279	12	3.83	3.48				
	J	283	6	1.33	0.82				
	K	281	6	4.17	3.82				
BBC	A	109	17	5.48	3.89	109	17	5.71	3.37
	B	110	9	3.09	3.42	110	13	4.19	3.35
YouTube	A	198	5	5.16	2.83	199	8	3.93	3.64
	B	199	6	5.29	3.11	200	6	2.35	1.31
	C	200	7	6.24	3.15	202	9	1.67	1.33
	D	203	5	3.5	3.21	203	6	1.54	0.67
	E	202	8	3.15	1.88				
	F	210	3	2.47	2.54				
	G	206	7	3.57	2.51				
	H	207	3	1.32	0.33				
Adobe	A	16	7	3.55	1.54	15	6	4.11	3.45
	B	15	6	3.6	2.9	18	8	3.27	2.2
	C	18	11	3.38	2.76	16	8	2.52	2.35
	D	19	7	3.02	2.88	19	9	3.69	2.49
	E	21	13	3.78	3.08	20	5	4.02	3.73
	F	22	5	3.8	3.35	21	8	4.55	3.97
	G	225	7	6.1	3.69				
Outlook	A	137	10	4.21	3.42	137	8	2.56	2.01
	B	138	16	3.53	3.26	138	16	5.42	3.44
	C					136	11	3.47	2.18



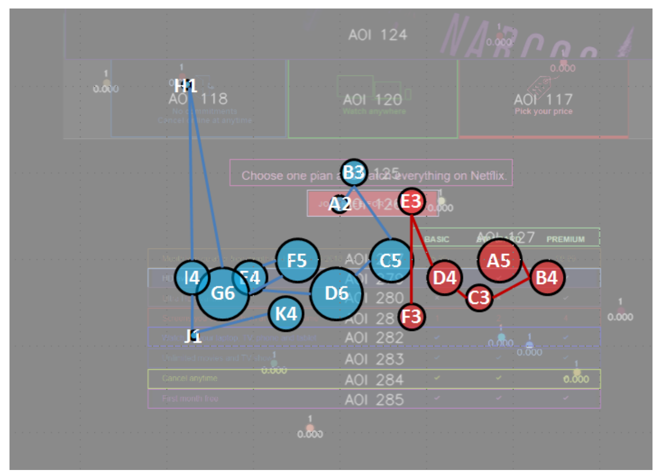
(a) WhatsApp



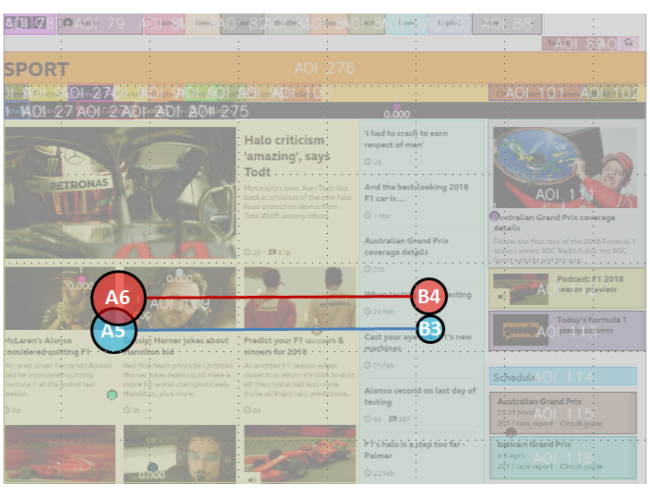
(b) Amazon



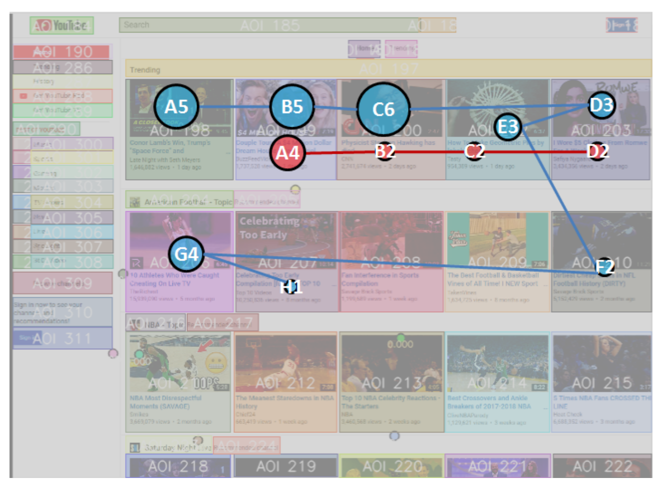
(c) WordPress



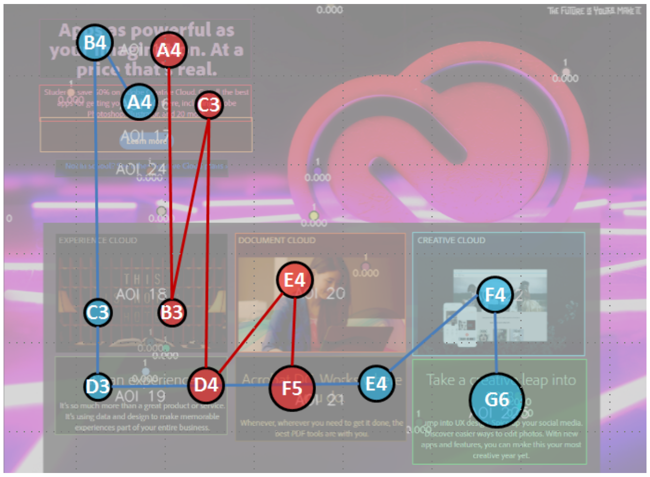
(d) Netflix



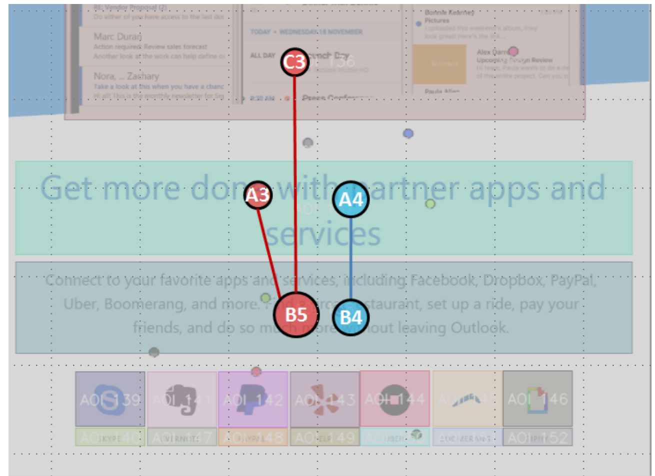
(e) BBC



(f) YouTube



(g) Adobe



(h) Outlook

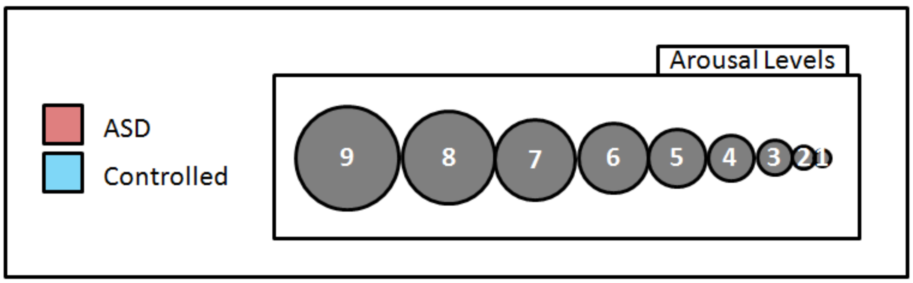


Figure 1: Levels of arousal from each group's trending scan path, overlaid on the AOI's of each website.

5 DISCUSSION

Regarding RQ1, that pertains to determining if there is a difference in arousal between both groups in browsing tasks, we hypothesised that the ASD group would experience more arousal compared to the neurotypical users. However, this was not the case. This could be because an increase in arousal could be caused by diverse reasons. Moreover, a more direct approach would be to test each UI element based on this hypothesis. This was not however possible due to the small and uneven sample size making statistical group comparisons inappropriate. With regards to RQ2, about identifying differences in AOIs between groups, we were able to observe differences in visual and physiological patterns between both groups in our descriptive approach. Take for example our observation that the individuals with autism showed less arousal compared to the neurotypical group for several UI elements on the *YouTube* page. One element shows people in a happy state, while the other contains the thumbnail for a video regarding the death of physicist, Stephen Hawking. A UX researcher may relate this to symptoms where people with autism interpret affective expressions differently from neurotypical people [33]. In research by Baron-Cohen *et al.*, they observed that individuals with autism do not recognise bodily expressions of affect as well as the neurotypical populace [4]. Therefore, an implication could be that, if a UI element contains a link that is a functionally significant aspect of the website, they must present it in a manner that does not rely primarily on facial expressions or affective cues. This is one such usability issue that can be identified using this methodology. Another behaviour that our methodology can help to uncover is that of understanding the affective states of users before leaving a web page. It may be the case that the initial UI elements that the users engage with eliciting higher arousal than the final ones, as is the case with the *BBC* and *YouTube* web page, or that participants experience an increase in arousal in the middle of their interaction compared to the beginning and end, as the case with the *Amazon* website or they experience lower arousal levels at the final UI elements as with the *Adobe* web page. Arousal could indicate positive states such as attraction, neutral ones (i.e. cognitive load), or negative affective states like frustration and stress. When participants experience an increase in arousal towards the end of an interaction, this could imply that they found what they were looking for, i.e., excitement in completing a task/goal, or that they were frustrated. Both of these cases could benefit from optimising the user interface. In the first instance, if users take a long time to find the item of interest on a page, it means that the user experience may be improved by re-positioning the element to a more visually accessible location, or by using a more attractive design to draw the attention of users towards that particular content. When users experience frustration with a UI element before leaving a web page, it could be an indication that the UI element has a usability problem. This type of diagnosis is mainly possible because we combined users visual scan path with a measure of their arousal levels. The implications for design include aggregation of different user groups of users, and potential modelling of their behaviour. For example, on an e-learning website, people with autism can have a different profile that takes account of and adapts to their unique traits and requirements. Eye-tracking is becoming more accessible, and we anticipate that webcameras and mobile phone cameras will one day have eye-tracking capabilities. Our methodology could then be used within social media and mobile applications. Posts and feeds can be treated as atomic UI elements so that the characteristics of different posts (sentiment, object classification, colour etc.) can be investigated against the visual scan sequence and the corresponding affective states that are elicited.

Our work is not without limitations. Due to the limited accuracy of eye-tracking technology, our analysis has been based on group behaviour as opposed to individual behaviour. Another limitation is that our methodology can currently only be used in laboratory settings. Ambient light, inter-colour differences between (and within) stimuli and other environmental variables may introduce confounding factors which may yield different results in the wild. Therefore, our methodology needs to be optimised to handle these factors dynamically in naturalistic settings.

6 CONCLUSIONS AND FUTURE WORK

Traditional metrics for usability evaluations like error rates, completion rates are useful in identifying the efficiency and performance issues of a website. However, of equal importance to UX researchers, is determining the user's affective state during the interaction. Recognising the user's affective state is important because it reveals a richer understanding of why users behave in certain ways in the presence of certain content and tasks. We have demonstrated through our study that it is possible to combine methods that answer both of these questions, 'how do users interact with websites' and 'how do they feel when they interact with web pages. The former was achieved by summarising the users' visual scan paths into a trending scan path using the STA algorithm, while the later was achieved by generating arousal scores that each UI element elicits for each group of users. Furthermore, we created a novel visualisation that can aid researchers in assimilating and generating research questions that can be investigated further using more established statistical or qualitative methods. We have utilised changes in arousal as our affective metric in this study, in future, we recommend an approach that also identifies the valence of the users' affective state. We anticipate that this methodology could be deployed on web cameras with integrated eye-tracking and pupillometry capabilities. In domains such as e-learning and gaming, users often drop out due to undesired affective and cognitive states during their interaction. Our methodology provides context (users' affective state, and visual attention) which can be fed back into the system. Real-time adaptation of user interfaces and contents can then be carried out to improve the quality of their user experience.

REFERENCES

[1] Ljubomir I Aftanas, Anton A Varlamov, Sergey V Pavlov, Viktor P Makhnev, and Natalya V Reva. 2002. Time-dependent cortical asymmetries induced by emotional arousal: EEG analysis of event-related synchronization and desynchronization in individually defined frequency bands. *International Journal of Psychophysiology* 44, 1 (2002), 67–82.

[2] American Psychiatric Association. 2013. Diagnostic and Statistical Manual of Mental Disorders (5th ed.).

[3] Bradley M Appelhans and Linda J Luecken. 2006. Heart rate variability as an index of regulated emotional responding. *Review of general psychology* 10, 3 (2006), 229.

[4] R Peter Baron-Cohen, J Ouston, and Antony Lee. 1988. Emotion recognition in autism: Coordinating faces and voices. *Psychological medicine* 18, 4 (1988), 911–923.

[5] Simon Baron-Cohen, Sally Wheelwright, Richard Skinner, Joanne Martin, and Emma Clubley. 2001. The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians. *Journal of autism and developmental disorders* 31, 1 (2001), 5–17.

[6] Sara H Basson, Dimitri Kanevsky, and Daniel A Oblinger. 2015. Predictive user modeling in user interface design. US Patent 9,165,280.

[7] Jennifer Romano Bergstrom and Andrew Schall. 2014. *Eye tracking in user experience design.* Elsevier.

[8] Zillah Boraston and Sarah-Jayne Blakemore. 2007. The application of eye-tracking technology in the study of autism. *The Journal of physiology* 581, 3 (2007), 893–898.

[9] Terry Brugha, Sally Anne Cooper, Sally McManus, S Purdon, J Smith, FJ Scott, N Spiers, and F Tyrer. 2012. Estimating the prevalence of autism spectrum conditions in adults: extending the 2007 adult psychiatric morbidity survey. *The NHS Informaiton Centre* (2012).

[10] John T Cacioppo, Richard E Petty, Mary E Losch, and Hai Sook Kim. 1986. Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions. *Journal of personality and social psychology* 50, 2 (1986), 260.

[11] Hsu-Min Chiang and Yueh-Hsien Lin. 2007. Reading comprehension instruction for students with autism spectrum disorders: A review of the literature. *Focus on Autism and Other Developmental Disabilities* 22, 4 (2007), 259–267.

[12] Sidney D'Mello and Jacqueline Kory. 2012. Consistent but modest: a meta-analysis on unimodal and multimodal affect detection accuracies from 30 studies. In *Proceedings of the 14th ACM international conference on Multimodal interaction.* ACM, 31–38.

[13] Sidney K D'mello and Arthur Graesser. 2010. Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. *User Modeling and User-Adapted Interaction* 20, 2 (2010), 147–187.

[14] Mehdy Dousty, Sabalan Daneshvar, and Majid Haghjoo. 2011. The effects of sedative music, arousal music, and silence on electrocardiography signals. *Journal of electrocardiology* 44, 3 (2011), 396–e1.

[15] Jan Ehlers, Christoph Strauch, Juliane Georgi, and Anke Huckauf. 2016. Pupil Size Changes as an Active Information Channel for Biofeedback Applications. *Applied psychophysiology and biofeedback* (2016), 1–9.

[16] Sukru Eraslan, Victoria Yaneva, Yeliz Yesilada, and Simon Harper. 2017. Do Web Users with Autism Experience Barriers When Searching for Information Within Web Pages?. In *Proceedings of the 14th Web for All Conference on The Future of Accessible Work (W4A '17).* ACM, New York, NY, USA, Article 20, 4 pages.

[17] Sukru Eraslan, Victoria Yaneva, Yeliz Yesilada, and Simon Harper. 2018. Web users with autism: eye tracking evidence for differences. *Behaviour & Information Technology* (2018).

[18] Sukru Eraslan, Yeliz Yesilada, and Simon Harper. 2016. Scanpath Trend Analysis on Web Pages: Clustering Eye Tracking Scanpaths. *ACM Trans. Web* 10, 4, Article 20 (Nov. 2016), 35 pages. https://doi.org/10.1145/2970818

[19] Sukru Eraslan, Yeliz Yesilada, and Simon Harper. 2016. Trends in Eye Tracking Scanpaths: Segmentation Effect?. In *Proceedings of the 27th ACM Conference on Hypertext and Social Media (HT '16).* ACM, New York, NY, USA, 15–25.

[20] Vittorio Gallese. 2006. Intentional attunement: A neurophysiological perspective on social cognition and its disruption in autism. *Brain research* 1079, 1 (2006), 15–24.

[21] Kiel Gilleade, Alan Dix, and Jen Allanson. 2005. Affective videogames and modes of affective gaming: assist me, challenge me, emote me. *DiGRA 2005: Changing Views–Worlds in Play.* (2005).

[22] Madeline B Harms, Alex Martin, and Gregory L Wallace. 2010. Facial emotion recognition in autism spectrum disorders: a review of behavioral and neuroimaging studies. *Neuropsychology review* 20, 3 (2010), 290–322.

[23] Oludamilare Matthews, Zhanna Sarsenbayeva, Weiwei Jiang, Joshua Newn, Eduardo Velloso, Sarah Clinch, and Jorge Goncalves. 2018. Inferring the Mood of a Community From Their Walking Speed: A Preliminary Study. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers.* ACM, 1144–1149.

[24] Oludamilare Matthews, Markel Vigo, and Simon Harper. [n. d.]. Towards Arousal Sensing With High Fidelity Detection of Visual Focal Attention. *Measuring Behavior 2018* ([n. d.]).

[25] Oludamilare Matthews, Markel Vigo, and Simon Harper. 2018. Sensing Arousal and Focal Attention During Visual Interaction. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI '18)*. ACM, New York, NY, USA, 263–267.

[26] Robert L Maulsby and Robert Edelberg. 1960. The interrelationship between the galvanic skin response, basal resistance, and temperature. *Journal of Comparative and Physiological Psychology* 53, 5 (1960), 475.

[27] Susan Dickerson Mayes and Susan L Calhoun. 2007. Learning, attention, writing, and processing speed in typical children and children with ADHD, autism, anxiety, depression, and oppositional-defiant disorder. *Child Neuropsychology* 13, 6 (2007), 469–493.

[28] Robert A McCleary. 1950. The nature of the galvanic skin response. *Psychological Bulletin* 47, 2 (1950), 97.

[29] Maja Pantic, Alex Pentland, Anton Nijholt, and Thomas S Huang. 2007. Human computing and machine understanding of human behavior: a survey. In *Artifical Intelligence for Human Computing*. Springer, 47–71.

[30] Timo Partala and Aleksi Kallinen. 2011. Understanding the most satisfying and unsatisfying user experiences: Emotions, psychological needs, and context. *Interacting with computers* 24, 1 (2011), 25–34.

[31] Timo Partala and Veikko Surakka. 2004. The effects of affective interventions in human–computer interaction. *Interacting with computers* 16, 2 (2004), 295–309.

[32] Delroy L Paulhus and Simine Vazire. 2007. The self-report method. *Handbook of research methods in personality psychology* 1 (2007), 224–239.

[33] RCM Philip, HC Whalley, AC Stanfield, R Sprengelmeyer, IM Santos, AW Young, AP Atkinson, AJ Calder, EC Johnstone, SM Lawrie, et al. 2010. Deficits in facial, body movement and vocal emotional processing in autism spectrum disorders. *Psychological medicine* 40, 11 (2010), 1919–1929.

[34] Rosalind Wright Picard et al. 1995. Affective computing. (1995).

[35] Rik Pieters and Luk Warlop. 1999. Visual attention during brand choice: The impact of time pressure and task motivation. *International Journal of research in Marketing* 16, 1 (1999), 1–16.

[36] Alex Poole and Linden J Ball. 2006. Eye tracking in HCI and usability research. In *Encyclopedia of human computer interaction*. IGI Global, 211–219.

[37] Luz Rello, Gaurang Kanvinde, and Ricardo Baeza-Yates. 2012. Layout guidelines for web text and a web service to improve accessibility for dyslexics. In *Proceedings of the international cross-disciplinary conference on web accessibility*. ACM, 36.

[38] Jessie Ricketts, Catherine RG Jones, Francesca Happé, and Tony Charman. 2013. Reading comprehension in autism spectrum disorders: The role of oral language and social functioning. *Journal of autism and developmental disorders* 43, 4 (2013), 807–816.

[39] James A Russell and Albert Mehrabian. 1977. Evidence for a three-factor theory of emotions. *Journal of research in Personality* 11, 3 (1977), 273–294.

[40] Yu Shi, Natalie Ruiz, Ronnie Taib, Eric Choi, and Fang Chen. 2007. Galvanic skin response (GSR) as an index of cognitive load. In *CHI'07 extended abstracts on Human factors in computing systems*. ACM, 2651–2656.

[41] K Dmello Sidney, Scotty D Craig, Barry Gholson, Stan Franklin, Rosalind Picard, and Arthur C Graesser. 2005. Integrating affect sensors in an intelligent tutoring system. In *Affective Interactions: The Computer in the Affective Loop Workshop at*. 7–13.

[42] Marko Tkalčič, Urban Burnik, and Andrej Košir. 2010. Using affective parameters in a content-based recommender system for images. *User Modeling and User-Adapted Interaction* 20, 4 (2010), 279–311.

[43] Fred Volkmar, Matthew Siegel, Marc Woodbury-Smith, Bryan King, James McCracken, Matthew State, American Academy of Child, et al. 2014. Practice parameter for the assessment and treatment of children and adolescents with autism spectrum disorder. *Journal of the American Academy of Child & Adolescent Psychiatry* 53, 2 (2014), 237–257.

[44] Tracy Westeyn, Peter Presti, and Thad Starner. 2006. ActionGSR: A combination galvanic skin response-accelerometer for physiological measurements in active environments. In *Wearable Computers, 2006 10th IEEE International Symposium on*. IEEE, 129–130.

[45] Ou Wu, Weiming Hu, and Lei Shi. 2013. Measuring the visual complexities of web pages. *ACM Transactions on the Web (TWEB)* 7, 1 (2013), 1.

[46] Victoria Yaneva. 2016. *Assessing text and web accessibility for people with autism spectrum disorder.* Ph.D. Dissertation.

[47] Victoria Yaneva and Richard Evans. 2015. Six good predictors of autistic text comprehension. In *Proceedings of the International Conference Recent Advances in Natural Language Processing*. 697–706.

[48] Victoria Yaneva, Richard Evans, and Irina Temnikova. 2016. Predicting Reading Difficulty for Readers with Autism Spectrum Disorder. In *Proceedings of Workshop on Improving Social Inclusion using NLP: Tools and Resources (ISI-NLP) held in conjunction with LREC*.

[49] Victoria Yaneva, Le An Ha, Sukru Eraslan, Yeliz Yesilada, and Ruslan Mitkov. 2018. Detecting Autism Based on Eye-Tracking Data from Web Searching Tasks. In *Proceedings of the Internet of Accessible Things*. ACM, 16.

[50] Victoria Yaneva, Constantin Orasan, Richard Evans, and Omid Rohanian. 2017. Combining Multiple Corpora for Readability Assessment for People with Cognitive Disabilities. In *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications*. 121–132.

[51] Victoria Yaneva, Irina Temnikova, and Ruslan Mitkov. 2015. Accessible texts for autism: An eye-tracking study. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*. ACM, 49–57.

[52] Victoria Yaneva, Irina P Temnikova, and Ruslan Mitkov. 2016. A Corpus of Text Data and Gaze Fixations from Autistic and Non-Autistic Adults.. In *LREC*.

[53] Victoria Yaneva, Irina P Temnikova, and Ruslan Mitkov. 2016. Evaluating the Readability of Text Simplification Output for Readers with Cognitive Disabilities.. In *LREC*.

[54] Robert M Yerkes and John D Dodson. 1908. The relation of strength of stimulus to rapidity of habit-formation. *Journal of comparative neurology* 18, 5 (1908), 459–482.

[55] Heng Zhang, Shkurta Gashi, Hanke Kimm, Elcin Hanci, and Oludamilare Matthews. 2018. Moodbook: An Application for Continuous Monitoring of Social Media Usage and Mood. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*. ACM, 1150–1155.

[56] Haihong Zhang, Yongwei Zhu, Jayachandran Maniyeri, and Cuntai Guan. 2014. Detection of variations in cognitive workload using multi-modality physiological sensors and a large margin unbiased regression machine. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*. IEEE, 2985–2988.