4-25-2020

Combining Eye Tracking and Verbal Response to Understand the Impact of a Global Filter

Franceli L. Cibrian

Jazette Johnson

Viseth Sean

Hollis Pass

LouAnne Boyd

Combining Eye Tracking and Verbal Response to Understand the Impact of a Global Filter

Franceli L. Cibrian
University of California
Irvine, CA 92697
fcibrian@uci.edu

Jazette Johnson
University of California
Irvine, CA 92697
jazettej@uci.edu

Viseth Sean
Chapman University
Orange, CA 92866, USA
sean103@mail.chapman.edu

Hollis Pass
Independent Speech-language
Pathologist
Orange, CA. USA
hollispass.slp@gmail.com

LouAnne Boyd
Chapman University
Orange, CA 92866, USA
lboyd@chapman.edu

Abstract
Visual attention guides the integration of two streams: the global, that rapidly processes the scene; and the local, that processes details. For people with autism, the integration of these two streams can be disrupted by the tendency to privilege details (local processing) instead of seeing the big picture (global processing). Consequently, people with autism may struggle with typical visual attention, evidenced by their verbal description of local features when asked to describe overall scenes. This paper aims to explore how one adult with autism see and understand the global filter of natural scenes.

Author Keywords
Autism; assistive technology; sensory processing; global and local processing.

ACM Classification Keywords
Human-centered computing; Accessibility;

Introduction
Seeing the “big picture” in a scene is part of how we process the world. Our senses take in stimulation that contributes to our perception of the environment. For sighted people, visual processing accounts for ~50% of the stimuli received through their senses[3]. Two separate information streams of the brain, the global and local, process the visual information [5] [10].

Figure 1: Sample items similar to items on the Navon test, 1977. The top item is the target. The bottom items are the choices to choose from when a person is asked to find the match.

Global processing allows people to rapidly process the scene to get a holistic understanding of an object, event, or scene (*e.g.,* seeing a forest before seeing the trees) [10] by detecting shapes, proximity, and context [5]. In contrast, local processing allows people to rapidly process the details, known as "analytic processing" [10]. Cognitive processes integrate both the global and local streams of information to produce a complete mental representation of the stimuli, thus filtering the relevant from the irrelevant information in a visual scene [2].

Many researchers claim that the average person first processes the global information and then integrates the local details within a fraction of a second. In artificial conditions, such as Navon's hierarchical letters [9], the default precedence can be determined by contrasting local and global features. For example, Figure 1-top depicts a hierarchical letter that has a global feature that looks like an F, and a local feature is the small letter Zs. People are typically quicker in detecting the F than Z [9].

However, people with autism visually process the world differently, as the global and local streams can be disrupted by the preference for the local details [13]. The lack of attention to the global features results in missed information [1] (e.g., people with autism tend to look at areas with large contrast in luminance, such as a reflection on glass instead of faces [15]). With these multi-dimensional interactions occurring in real-time, people with autism can have challenges with making friends that can lead to social isolation [13].

Recent advances in eye gaze research state that the visual attention for some people with autism is driven by aspects such as high contrast in brightness and color (e.g., a shiny reflection), rather than the semantic or social content of a scene (e.g., looking at people's faces)[16]. In this paper, we aim to filter out local details and highlight the global features of visual scenes to understand if those filters effect how an adult with autism sees the scenes and responds verbally.

Related Work

One manner of measuring visual attention is with eye-tracking technologies, as they measure the motion of the eye and determine how long the eye gazes on an area of an image [12]. By examining fixations, saccades (rapid eye movement as fixation point changes), and other eye activity, researchers can gauge the reaction of the subject to any given stimuli [12].

Existing algorithms for detecting fixations and saccades often used arbitrary and inaccurate eye velocity and acceleration thresholds. For example, Cluster Fix [6] is an algorithm that uses k-means cluster analysis on a combination of distance, velocity, acceleration, and rotation measurements of eye movements to detect fixations and saccades with increased precision. This allowsdetection of small saccades and the precise identification of the start and end of saccades [6].

Eye-tracking also has been used to predict where people look in natural scenes. For example, in [18], researchers built a dataset of 700 images with eye-tracking data of 15 neurotypical viewers and annotation of the image about objects and semantic attributes. Comparing the eye-tracking with the annotation scheme, researchers found that object and semantic information from scenes are the most important aspect that neurotypical people first see in scenes. In [16] the same database was used to predict where people with autism would look. In this study, researchers collected the eye-tracking data of 20 people with autism and compared it with the eye-tracking data from 19 neurotypical people. Results

Filter	Description
Baseline (raw image) 	The image is presented without any filter
Lined edges 	The filter contains black and white line drawings to emphasize the shape and boundary between objects and removed high spatial frequency sections such as shading
Global object 	The filter highlighted the main object of an image and is presented through color and shading, pixilated, and the background was removed.
Global object in blurred context 	The filter highlighted the main objects of the scene with a pixilated background
Animation 	The animation consists of one filter at the time, this is, the same image was presented with one range of information over the time

Table 1. Descriptions of filters

showed that people with autism have a stronger image-center bias regardless of object distribution, reduced saliency for faces, and locations -indicated by the social gaze of people in the images- yet a general increase in pixel-level saliency at the expense of semantic-level saliency [16].

These research findings show that eye-tracking can be useful in understanding what features of an image people visually attend to (i.e., the big picture features or the local details). Therefore, we can leverage these insights regarding global-local processing styles and eye gaze trends to build a system that highlights the socially relevant aspects of images.

The Global Filter Prototype

We prototyped four ways to highlight global aspects of an image (see Table 1):

1) **Lined Edges:** we designed a black and white line drawings filter, as is common in icons used in special-education classrooms [4].
2) **Global Objects:** we designed a filter that highlighted the main object of an image. This filter follows the premise that by removing pixel-level detail from the background, it would convert the object-level into the local level, and therefore, a point of interest.
3) **Global objects in blurred context:** To added more semantic information of the image, we added a monochromatic, pixilated background to the main object.
4) **Animation:** We create an animation consisting of each filter presented one at the time.

We applied each filter to 10 images, taken from the first 50 out of 700 images in the Object and Semantic Images and Eye-tracking (OSIE) data set. This dataset was previously coded for typical and autistic eye tracking as well as for image features [16,18]. The lined edges filter was created using the PhotoScape X program. The global objects were selected using the "Remove Background" tool of PowerPoint 2016. The blurred context was created with the "Artistic effect" tool to blur and make gray the background. The animations were created using the previous filters and then put together as a Gif. We also selected ten original images as the baseline condition. With the images, we created a low-fidelity prototype in PowerPoint that consisted of 10 blocks of images (1 per filter), randomized and separated by a transition. The auto-advance was set for 10 seconds. The prompt after each scene was, "What is this a picture of."

Methods

We ran a pilot study with a 40-year old autistic man (P1). We decided to follow an N of 1 trials with P1 as we want to provide an objective understanding of his eye tracking and verbal responses that can help us to identify the best global filter for a person with autism. As an inclusion criterion, P1 took the Navon test (www.psytool.org). The test revealed P1 processes local over global features, confirming that he struggles to some degree with global processing. P1 also reported having difficulty with social communication, mainly nonverbal communication.

Procedure

P1 put his chin in a chin rest and a technician calibrated the *Eyelink 1000* system. Next, P1 viewed the low fidelity prototype and answered, after each image, the prompt: "What is this a picture of?" Images randomly appeared

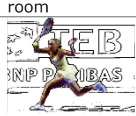
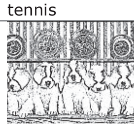
Code	Description	Example
"Local"	Answers only contain one or more local details	 Looks like some phone and a cell phone and a drink
"Local to global"	Answers begins with local details followed by a global feature	 A sofa in a family room
"Global to local"	Answers contain the main objet immediately followed by the action it is doing	 someone playing tennis
	Answers begin with a word that describe the set of objects as a whole	 bunch of dogs
"Global"	Answers with only a global feature (i.e., a word that describe all the scene)	 Basquetball
	If the image only contain one object and answer contains the name of the object	 Toilet

Table 2. Description of the scheme code to score participant's answers

on a display. Finally, we conducted a semi-structure interview with P1.

Data Collection and Analysis

The session was video-recorded, and the answers from P1 were transcribed. Two researchers scored P1 verbal responses following the scheme code of Table 2. With the results of the scores, we ran a 2-tailed ANOVA to test if the filter was significantly different from the baseline.

Eye-tracking data was collected on an Eyelink 1000 system, which gathers gaze point data at a rate of 500 Hz. We then transformed these data into fixations and saccades using Cluster Fix algorithm [6]. Given that in this work, we are interested in the global visual attention, that is thought to occur in the first fixation and saccade [14] and is described to be observed within 40ms, we further analyze first fixations and saccade of P1. There is usually a central bias in the fixation of the gaze as viewers tend to continue looking at a place for a while before a new image appears. To address this bias, we removed the first fixation and computed the second and third fixations to analyze global aspects. From these two early fixation points, we created a vector to indicate the saccade (i.e., a segment of scan path with direction; see Figure 2). We analyzed the direction and intersection of each vector in the scene as initial saccades may uptake global information—especially long saccades and brief fixations [7]. Intersections were according to the hotspots of semantic features of a scene viewed by neurotypical people [18].

We determined the performance of each filter in three manners: (1) by comparing the number of overlaps between the vectors and OSIE hotspots –the overlap demonstrates a shift in attention to global features [7,12]; (2) by classifying the length of the vector below or above the median –saccade length, as it has been associated with global processing [17]; and (3) by computing the combination of the parameters of length and overlap, which we call "combo".

We assigned initial vector of each image first vector a category of "below the median distance" and score of 0 (e.g., local) or "above the median distance" and score of 1 (e.g., global) for the distribution of first saccades across the 50 images. Initial vectors ranged from 23 pixels to 465 pixels with a median length of 222 pixels. Once each vector was labeled, we conducted a paired t-test for each image viewed in baseline compared to each filter's 10 images (see example in Figure 2). Please note that the animation filter begins with a lined edge filter, and the first saccade occurs within the first 1 second. Then, during this time, P1 only saw the lined edge filter. Therefore, this analysis discards the Global to Local Progression via Animation filter.

Results

Both the verbal and the eye-tracking analyses revealed that P1 performed better with filters than in the baseline.

VERBAL RESPONSES

The verbal responses of P1 revealed that he provided more holistic responses with filters in comparison with baseline (Figure 3). In the baselines images, P1 gave both local and global responses. For example, if the scene contains a breakfast (See figures in Table 1), his answer was "a table with a computer and some electronics and someone's meal." This result means that P1 only describes one object at the time, but half of the time, he could not integrate the image as a whole.

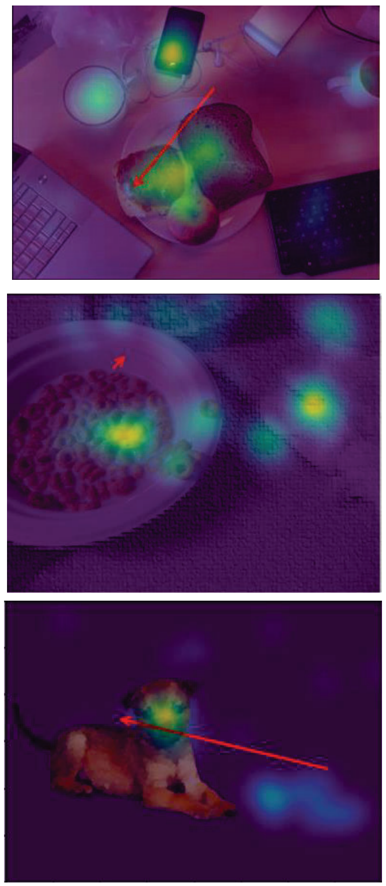


Figure 2. A raw image (top), a global object in blurred context image (center), and a global object filter (bottom), overlapping with the hotspots (in green/yellow) and red arrows showing the first saccade of P1.

P1 gave more global responses with the Global Object (90%; p=0.041), and the Global Object in Burred Context (60%; p=0.1) filters. A possible reason is that some images only highlight the main objects of the scenes where P1 focused all his attention. Also, P1 suggested that both filters made the main object "very obvious." However, the Global filter lacks context details that, for some scenes, could be important. For example, an image with only a toilet could be a "restroom" or a "bathroom" (see figures in Table 1).

In contrast, having the blurred background gave more information about the scene, in this case, P1 started with some of the main details, and then he added in key parts. Perhaps, P1 realized he had to include all the objects on the filters and infer what was happening using some information from the background. In this case, these filters helped him to signal how many details were relevant (to the researchers) to understand the big picture.

With the Lined Edge filter, P1 gave more Global to Local responses (50%; p=0.34), he specifically said that this filter *"depends on the picture, [understand the pictures is] harder with tons of details."* Therefore, he had some difficulties in answering some scenes. A possible reason is that we used the same threshold (i.e., we changed the colors of each image to a black and white silhouette using the same parameters from PhotoScape adaptive threshold filter) to highlight the high spatial frequencies of the images. Therefore, some of the images have more or fewer details according to their illumination level.

With the Local Progression via Animation filter, P1 gave 40% of global responses (p=0.34). He said that this filter allowed him "*to see it more ways than one*" and reminded him of forensic lightening where different aspects of a scene show up in different types of light (suggesting P1's attention shifted). This result could show that although P1 found this filter useful, the animation should be improved to reduce the cognitive load that people with autism may be added by changing images over time. These results suggest that P1 provided more holistic responses with filters —as seen by more global responses when many of the details were filtered. We saw far more global responses when the filter discriminated the global objects from local details.

Eye Tracking

We found that the Global Object filter produced longer saccades vectors more frequently (7 out of 10 images) than the baseline (1 out of 10 images) ($p<0.005$). Also, the blurred background filter performed well compared to baseline (5 of 10 images had long saccades; $p<=.005$). This result shows that with global objects emphasized, more global saccades occurred (saccades tend to be longer); hence, P1 may have scanned more of the image to get global concept.

The overlap between the early saccade and the hotspot was significantly better with the Global Object than the baseline (9 out of 10 vectors overlapping the hotspots compared to 2 out of 10 in baseline; $p < 0.001$; Figure 2). This result might show that the saccade vectors were oriented toward and traveled through the global areas (hotspots) significantly more often when P1 saw the scene with the Global Object filter than the baseline condition. This result suggests that P1 shifted visual attention to global areas by using filters.

When combining saccade length with vectors overlapping the hotspots, we found that the Global

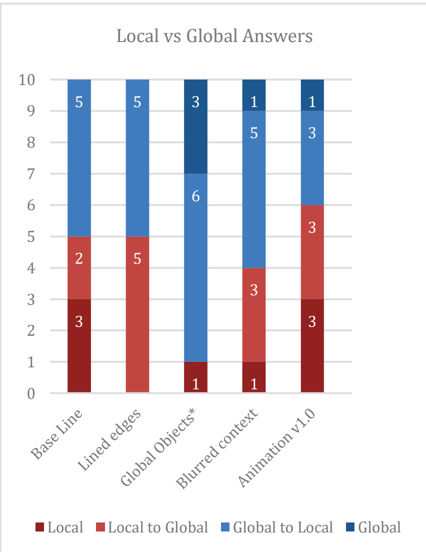


Figure 3. Participant's answers per filter. Please note that * filters has a significant differences comparing with the base line according to a 2-tailed ANOVA with confidence interval of 95%

Object filter performed significantly better than baseline (7 of 10 images with a "combo" compared with 1 out of 10 'combo' p<0.005). This result shows that when P1 saw the scenes with the Global Object filter, his early saccades were longer and oriented to the hotspots, again suggesting that the filter may influence global visual attention. Overall, our results show that the global objects filter quickly redirects the saccades of P1 to important zones of the scenes.

Discussion

Both the verbal and eye-tracking analyses revealed that the global object's filters yielded the best results. These results suggest that the global object filters were successful in shifting eye gaze to important areas of the image and may have helped P1 provide more global responses. Overall, this study produced a rich data set of verbal behavior and eye-tracking—thus enabling us to view global processing from different angles.

From a methodological point of view, we learned that scoring verbal response as global or local is not an easy task. More research needs to be done to understand how can we measure the verbal response of individual in such a manner that we can gain an understanding if they had local or global processing. We also learned that eye-tracking helps to understand where and how people with autism view a scene. We will further explore how this information could be used to infer the visual processing. In this paper, we explored different manners to analyze eye-tracking data by giving more importance to early global processing rather than analyzing all the heat map (or hotspots) or scan paths produced over the full 3 second viewing. We believed this could be a starting point to analyze new ways on how we can measure the global processing of a visual scene.

One limitation of this study is that we only test our prototype with one participant. However, "N of 1 trials" are become more common nowadays, as they show promising results to investigate the effect of drugs, sensors, and digital medicine for health interventions [8,11]. Following this methodological approach, the HCI and assistive technology communities can create personalized pervasive assistive technology in order to maximize the quality of life of people with disabilities. However, it is well known that for "N of 1 trial" is difficult to compare the finding from different studies. Then, as future work, we plan to conduct a more extensive study with more participants to understand if our results can be generalized.

Conclusion and Future Work

This paper explored how manipulating images to promote global processing impacted one autistic adult. Across two analyses, we found that a filter that removed the background appeared effective at shifting the visual attention to global features, which are reflected by his improvement in verbal responses to "what is this picture about." Perhaps this shift in sensory perception paves the way to directly accessing one's cognitive skills. Future work will consist of expanding to group studies, and expanding the eye tracking analysis to examine gaze path patterns across images.

Acknowledgments

We acknowledge P1, and The Brain Institute for allowing us to share their equipment, Alice Wong, Andy Liang, Dr. Aaron Schurger and Tian Lan for help with accessing eye tracking data. We also thanks to the Jacobs Foundation Advanced Research Fellowship and Dr. Gillian Hayes for her feedback.

References

1. Magali Batty and Margot J. Taylor. 2003. Early processing of the six basic facial emotional expressions. *Cognitive Brain Research* 17, 3: 613–620.
2. Jules Davidoff, Elisabeth Fonteneau, and Joel Fagot. 2008. Local and global processing: Observations from a remote culture. *Cognition* 108, 3: 702–709.
3. Susan Hagen. 2012. The mind's eye. *Rochester Review* 74, 4: 32–37.
4. Gillian R. Hayes, Sen H. Hirano, Gabriela Marcu, Mohamad Monibi, David H. Nguyen, and Michael T. Yeganyan. 2010. Interactive visual supports for children with autism. *Personal and Ubiquitous Computing* 14, 7: 663–680.
5. Ronald Hübner and Gregor Volberg. 2005. The Integration of Object Levels and Their Content: A Theory of Global/Local Processing and Related Hemispheric Differences. *Journal of Experimental Psychology: Human Perception and Performance* 31, 3: 520–541.
6. Seth D. König and Elizabeth A. Buffalo. 2014. A nonparametric method for detecting fixations and saccades using cluster analysis: Removing the need for arbitrary thresholds. *Journal of Neuroscience Methods* 227: 121–131.
7. John Liechty, Rik Pieters, and Michel Wedel. 2003. Global and local covert visual attention: Evidence from a bayesian hidden markov model. *Psychometrika* 68, 4: 519–541.
8. E. O. Lillie, B. Patay, J. Diamant, B. Issell, E. J. Topol, and N. J. Schork. 2011. The n-of-1 clinical trial: the ultimate strategy for individualizing medicine? *Personalized medicine* 8, 2: 161–173.
9. David Navon. 1977. Forest before trees: The precedence of global features in visual perception. *Cognitive psychology* 9, 3: 353–383.
10. A Oliva. 1997. Coarse Blobs or Fine Edges? Evidence That Information Diagnosticity Changes the Perception of Complex Visual Stimuli. *Cognitive Psychology* 34, 1: 72–107.
11. Bethany Percha, Edward B. Baskerville, Matthew Johnson, Joel T. Dudley, and Noah Zimmerman. 2019. Designing Robust N-of-1 Studies for Precision Medicine: Simulation Study and Design Recommendations. *Journal of Medical Internet Research* 21, 4: e12641.
12. Rik Pieters and Michel Wedel. 2017. A review of eye-tracking research in marketing. In *Review of marketing research*. Routledge, 143–167.
13. Caroline E. Robertson and Simon Baron-Cohen. 2017. Sensory perception in autism. *Nature Reviews Neuroscience* 18, 11: 671–684.
14. Stine Tanggaard. 2016. Mental Imagery in High-Functioning Autism Spectrum Disorder. Universdity of Oslo.
15. Shuo Wang, Ming Jiang, Xavier Morin Duchesne, Elizabeth A. Laugeson, Daniel P. Kennedy, Ralph Adolphs, and Qi Zhao. 2015. Atypical Visual Saliency in Autism Spectrum Disorder Quantified through Model-Based Eye Tracking. *Neuron* 88, 3: 604–616.
16. Shuo Wang, Ming Jiang, Xavier Morin Duchesne, Elizabeth A. Laugeson, Daniel P. Kennedy, Ralph Adolphs, and Qi Zhao. 2015. Atypical Visual Saliency in Autism Spectrum Disorder Quantified through Model-Based Eye Tracking. *Neuron* 88, 3: 604–616. 2
17. Michel Wedel, Rik Pieters, and John Liechty. 2008. Attention switching during scene perception: How

goals influence the time course of eye movements across advertisements. *Journal of Experimental Psychology: Applied* 14, 2: 129–138.

18. Juan Xu, Ming Jiang, Shuo Wang, Mohan S. Kankanhalli, and Qi Zhao. 2014. Predicting human gaze beyond pixels. *Journal of Vision* 14, 1: 28–28.