Effective and Efficient Visual Stimuli Design for Quantitative Autism Screening: an Exploratory Study

Tri Vu[1], Hoan Tran[1], Kun Woo Cho[1], Chen Song[1], Feng Lin[1], Chang Wen Chen[1], Michelle Hartley-McAndrew[2], Kathy Ralabate Doody[3], and Wenyao Xu[1*]

Abstract— **Autism spectrum disorder (ASD) is one of the most common childhood developmental disorders. Early detection and intervention for ASD are critical for increasing child success. In the past decade, utilizing the abnormal eye gaze characteristics of children with autism in regard to certain visual stimuli is emerging as a screening approach due to its cost-efficiency and promising accuracy. However, the effect of visual stimulus on children with ASD has not been considered as a diagnostic consideration in the past. In this paper, we first create a visual stimuli database based on an extensive literature review, then we examine the impact of picture stimuli and exposure time on the quantitative accuracy of screenings for ASD. This is done by extracting gaze distribution in a 2-D space and comparing children with ASD to typical peers using the 1st Wasserstein distance. A group of 32 participants with ASD and typical development (TD) were recruited for the study. The f-score accuracy results demonstrate the impact of implementing visual stimulus on screening for ASD. Our study demonstrates that the parsing of "social scene" stimulus with 5-second exposure time has the best performance at 98.24%.**

I. Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental syndrome that is defined by impairments in social reciprocity and communication, and by abnormal restricted, repetitive behaviors [1]. ASD results in impairments in the communication of children with ASD and include delays in, or a total lack of the development of spoken language; marked impairment in the ability to initiate or sustain a conversation with others; stereotyped and repetitive use of language [2]. The prevalence of ASD in the US is a concerning situation, with a prevalence of 14.7 cases in every 1,000 children [3]. The economic and social implications of ASD upon a society are staggering and the impact on individual families is also noteworthy. For example, the divorce rate of families with a child with ASD is 23.5%, compared to that of 13.8% of families who are not touched by ASD [4]. The total cost of educating and caring for individuals with ASD was approximately between $11.5 billion and $60.9 billion in the US in 2014 [5].

Early screening of ASD plays a vital role in the beginning services for children. Early intervention can improve the long-term functioning of individuals with ASD and encourage better outcomes [6], [7]. To date, existing methods for diagnosing ASD mainly focus on direct observation and subjective testing, along with family interviews. The subjectivity of these diagnostic procedures contributes to missed or inaccurate diagnostic outcomes. The Autism Diagnostic Observation Schedule (ADOS), the "gold" standard evaluation tool for a diagnosis of ASD, consists of activities with four different modules to formulate a cut-off score for each observation [8]. This method is time-consuming (30-45 minutes per module), requires numerous materials (more than 85 stimulus items, 12 laminated cartoons, 3 color picture cards) and is costly ($3,095 per kit) [8]. The overall limitations of this diagnostic procedure include complex clinical protocols and highly-trained professional staff. These often result in delayed diagnosis, which further delays the onset of early intervention. Previous studies show that when interventions for children with ASD begin before age 5, children's success rate is 67%, compared to the 11% success rate seen when interventions begin after age 5 [9]. The eye gaze patterns of children with ASD have been considered significantly different than those without ASD. [10], [11]. The effect of visual stimulus is critical for gaze attention for children with ASD. Nonetheless, there have been no in-depth exploratory studies examining effective and efficient visual stimuli for screening for ASD.

In this paper, we conduct an exploratory study on effective visual stimuli design. The key factor in our quantitative ASD screening protocol lays in the design of visual stimuli in consideration of its two aspects: stimuli content and exposure time. Stimuli content is the content of the pictures shown for children with ASD. It can belong to a various number of categories: human face, social scene, food, toy, etc. It is important to understand the impact of different stimuli on our ASD screening method. Another quantitative factor to be investigated is the effect of exposure time on visual stimuli construction. Exposure time has the possibility to affect ASD screening's effectiveness in different ranges of duration [12]. Taking these two aspects into account, in our study, we construct a visual stimuli design consisting of different picture stimuli contents (social scene, human face and objects) and exposure time (1s-, 3s- and 5s-duration time). The result shows that "social scene" stimulus in 5-second exposure time achieves the best performance in ASD screening with 98.24% accuracy. In summary, our main contributions in this paper are:

[1]T. Vu, H. Tran, K. Cho, C. Song, F. Lin, C. Chen, and W. Xu are with the Department of Computer Science and Engineering, University at Buffalo (SUNY), Buffalo, NY 14260, USA.

[2]M. McAndrew is with the Department of Pediatrics, University at Buffalo (SUNY), Buffalo, NY 14260, USA

[3]K. Doody is with the Department of Exceptional Education, Buffalo State University (SUNY), Buffalo, NY 14222, USA

[*]Address comments to wenyaoxu@buffalo.edu or +1 (716)-645-4748.

978-1-5090-4179-4/17/$31.00 ©2017 IEEE

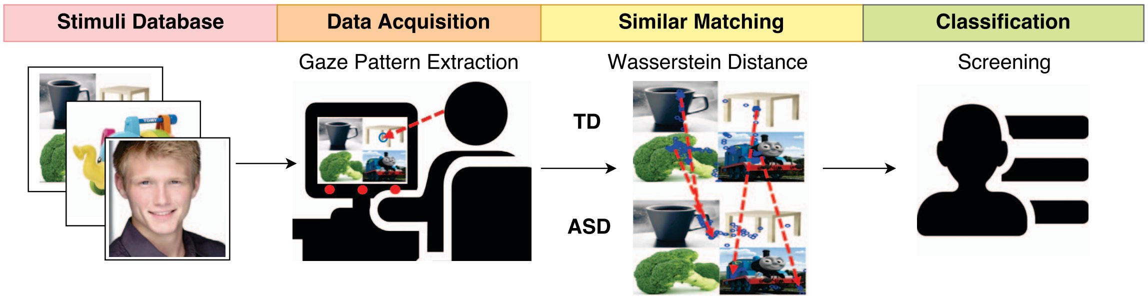


Fig. 1: The framework overview of the proposed ASD screening system.

- Visual stimuli database design for the ASD screening.
- Evaluation of the impact and efficiency of picture stimuli and exposure time on ASD screening accuracy both independently and dependently.

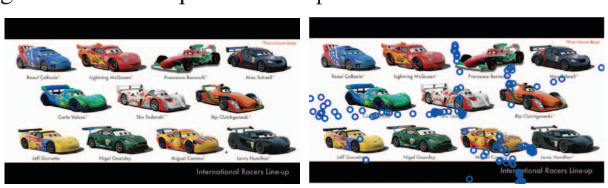
II. ASD Screening System

A. System Overview

Our visual stimuli assessment framework consists of four main parts which are demonstrated in Fig. 1. The first part is to construct a comprehensive visual stimuli design which will be meticulously assessed in Section III. The second part is data acquisition, containing hardware setup for stimulus' data collection. Next, similarity matching step utilizes the 1^{st} Wasserstein distance for assessing distance. Classification based on kNN will extract the recall and precision rate for later f-score accuracy evaluation.

B. Data Acquisition

The setup consists of two main parts: the Tobii EyeX Controller and the monitor. The Tobii EyeX Controller [13] is an eye tracking device that is capable of accumulating eye positions of movement in a duration of time in form of gaze pattern. The monitor for displaying visual stimulus is a Dell P2214H monitor with 20.2" in width and 13.9" in height [14]. The complete setup of the experiment is illustrated in Fig. 1's "Data Acquisition" step.



(a) Original visual stimulus

(b) The corresponding gaze pattern of the subject

Fig. 2: The example of the visual stimulus and the gaze pattern.

C. Similarity Matching

Our gaze point data could be seen as distribution in space as illustrated in Fig. 2. In order to compare the similarity between distributions, we employ the 1st Wasserstein Distance [10] to calculate the distance, which is the minimum effort to make one distribution become another. Overall, Wasserstein distance has short computation time [15] and has better results when the ground distance is relatively meaningful.

D. Classification

In our study, we employed the kth Nearest Neighbors (kNN) algorithm to distinguish the children with and without ASD. The idea of kNN is to classify a new object with the label that has the most number of votes among its kth nearest neighbors in terms of distance. In our case, we use $k = 3$ as this setting achieves the best performance.

III. Visual Stimuli Design

A. Visual Stimuli Categories

The initial database is constructed with a total of 16 pictures with six different stimulus (social scene, human face, food, toy, animal and drawing). Fig. 3 represents the framework for our stimuli database design. This database is based on our extensive literature review. Then it is classified into three categories: social scene, human face and object from our associating experts' knowledge. Therefore, we reduce the dataset to 12 pictures - four for each stimulus to make the numbers of pictures equal to each other for each stimuli.

Social scene: A social scene is chosen since ASD is apparently associated with children' social impairments [1]. "Social scene" stimulus contains four separate pictures: two reverse versions of the same picture; one is a familiar scene of a birthday party and one is an unfamiliar scene of workplace.

Human face: We choose a human face as one of our stimuli because children with ASD typically struggle to process characteristics of a human face [16]. The "human face" stimulus is divided into four pictures: one contains a pair of blurred/normal faces; one contains a pair of direct gaze/averted gaze faces and the last two consist of two pairs of happy/sad faces.

Object: "Object" stimuli include four pictures of food, toys, animals and an abstract drawing. Initially, they are four different stimuli. The reason for this combination is based on Koldewyn *et al.* [17] report that children with ASD appear to show no impairment in processing global information.

B. Exposure Time

The gazing duration for each picture was 5 seconds to prevent both the reduction of the children's attention and the insufficiency of gaze data. To form the exposure time separation, we evaluated each picture with gaze pattern data in 1-second, 3-second and 5-second time range respectively.

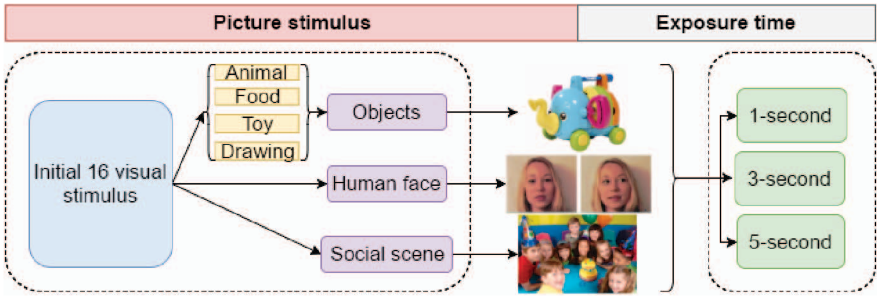


Fig. 3: The visual stimuli design includes picture stimulus and exposure time.

To sum up, each of the four pictures in three picture stimuli categories (social scene, human face, object) will have three ranges of exposure time (1-second, 3-second and 5-second). In total, we have 36 visual stimuli data cases.

IV. Evaluation

A. Experimental Setting

1) Participants: The study was approved by the Institutional Review Board of Women & Children's Hospital, University at Buffalo and Buffalo State University [IRB: 595026]. There were a total of 32 participants between 2 and 10 years old. Among those participants, 16 of the children were diagnosed with ASD by the ADOS or in accordance with DSM V diagnostic criteria and the other 16 presented as children with the absence of any developmental impairments. All participants were obtained through an existing research program. Obligatory parental consents were also conducted before each experiment. The participants, along with their family members, were provided with comprehensive information regarding the research study.

2) Protocol: The participants were instructed to sit in front of a computer monitor with setup mentioned above. A total of 16 pictures was displayed separately for total duration of 5 seconds each, with a 2-second interval between each picture to make stimuli independent and refresh participants.

3) F-score accuracy: For each picture, one data point from the eye gaze was selected, while the remaining data points were used as training data. After kNN is applied 30 times sequentially for each gaze data of that picture, $precision$ and $recall$ are recorded. For more in-depth results, we employed the balanced f-score accuracy (F_1), which is the subcontrary means of precision (P) and recall (R).

$$F_1 = 2 \cdot \frac{P \cdot R}{P + R}$$

B. Evaluation Results

1) Stimuli Comparison: In order to find the most effective stimulus for screening, the results are taken into consideration regardless of screening time. The performance is summarized in Table I. Across all the three stimuli, "social scene" stimulus gives consistently high accuracy at about 98.24%. Human face stimulus has similar accuracy at 97.22%. Exposure times were 1, 3 and 5 seconds respectively. Object yielded the lowest accuracy comparatively at 90.26%. For both groups of children, with and without ASD, the social scene proved to be the most effective stimulus, followed by the human face and the object, respectively. All stimuli provided good results, and proved to be efficient for screening. Details of performance across three stimuli are illustrated in Fig. 4. The red line indicates the median value and the blue box indicate the standard deviation. Overall, the social scene is the most effective stimulus for accurate screening. This result further suggests that "social scene" and "human face" stimuli yield better accuracy, compared to "object" stimulus.

TABLE I: Average accuracy percentage performed by each stimulus.

Stimuli	Object (%)	Social scene (%)	Human face (%)
ASD	88.59 ± 8.84	97.82 ± 1.52	96.36 ± 2.34
TD	91.55 ± 6.74	98.61 ± 1.20	96.87 ± 1.41
Overall	90.26 ± 2.08	98.24 ± 0.56	97.22 ± 1.07

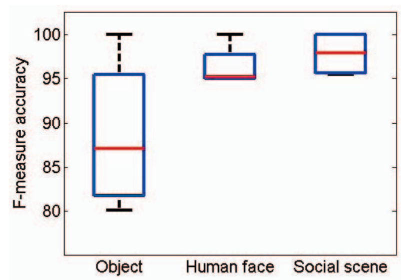


Fig. 4: Boxplot on F-score accuracies (%) of three picture stimuli in 5-second exposure time.

2) Impact of Exposure Time: Fig. 5 compares the result of different exposure times. The 5-second exposure time has the highest result at 95.24%, followed by 3-second and 1-second exposure time with 90.01% and 85.56% respectively. These numbers suggest that the accuracy is proportional to screening time. Overall, the 5-second screening time yielded the best result, and 3-second exposure time also yielded high accuracy with over 90%. The 1-second result also had high accuracy, but 3-second exposure time had a better trade between time and accuracy. This suggests that 3 to 5 seconds screening time will yield highly accurate results.

3) Stimuli and Duration Combination: We compared the result of each stimulus with 1, 3 and 5 seconds, respectively, to find the most effective stimulus. At the same time, we considered the accuracy of each stimulus at different exposure time. Fig. 6 shows the detail accuracy of each stimuli when time is taken into consideration. The 5-second interval shows a high and consistent result, especially with

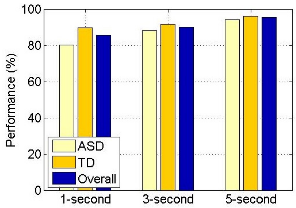


Fig. 5: Accuracy percentage over exposure times among ASD, TD and overall.

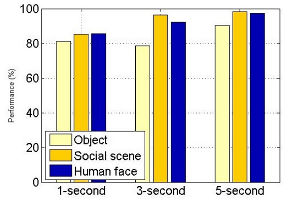


Fig. 6: Accuracy percentage comparison in consideration of visual stimuli and exposure time simultaneously.

social scene stimuli at 98.24%. The lowest result of 5 seconds is 90.27%, which is the accuracy of object stimuli. The 3-second interval also had a high result with social scene and human face at 97.08% and 96.49% respectively. The 1-second screening time showed considerably lower results in the social scene and human face, with the exception of object stimuli. Object has the lowest accuracy in all three exposure times, with the lowest of 78.58% with 3-second exposure time. This figure suggests that using social scene with full duration exposure yields the optimal result at nearly 100% accuracy. One sub-optimal result is using human face stimulus at full duration exposure. With social scene and human face stimuli, 3 seconds is a preferable alternative to reduce screening time, while maintaining high overall accuracy. Using object stimulus, and/or using 1 second exposure time is not recommended.

V. Related Work

Studies on ASD stimulus have been of interest to researchers in many areas of study. Lovaas *et al.* [18] explored into stimuli overselectivity, suggesting that the children with ASD respond to only part of a relevant cue, or even to a minor and, often irrelevant feature of the environment, without learning about the other relevant portions of that environment. Fletcher *et al.* [19] investigated the eye-movements of children with ASD with complex social scene in comparison without social scene circumstances in 3 seconds duration. In addition, besides social scene stimulus, other studies such as Pierce *et al.* [16] investigates into human face stimuli for individuals with ASD and assured their difficulty in face processing. In consideration of normal objects, McCleery *et al.* [20] with concentration on object stimuli suggests that individuals with high-risk of ASD have faster responses to objects than human faces. However, these studies only considered the impact of certain stimulus on children with ASD. In our research, we provide comprehensive visual stimulus analysis and studies including the effect of exposure time in screening for ASD.

VI. Conclusion

In this paper. we explored into the effect of visual stimuli including picture stimuli and exposure time on accuracy for quantitative ASD screening together with the utilization of 1^{st} Wasserstein and kNN method for autism diagnosis. We constructed a stimuli design database of ”social scene”, ”human face” and ”object” for picture stimuli and exposure time range of 1-second, 3-second and 5-second. ”Social scene” stimuli appears to be the most convincing for ASD screening. As for exposure time, 5-second time range has the most promising result of accuracy predictably. Our future work will explore more categories of stimuli for our ASD screening system.

References

[1] M. B. First, “Diagnostic and statistical manual of mental disorders,” *DSM IV-4th edition. APA*, p. 1994, 1994.

[2] C. Lord *et al.*, “Autism diagnostic interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders,” *Journal of autism and developmental disorders*, vol. 24, no. 5, pp. 659–685, 1994.

[3] A. L. Siu, K. Bibbins-Domingo, D. C. Grossman, L. C. Baumann, K. W. Davidson, M. Ebell, F. A. García, M. Gillman, J. Herzstein, A. R. Kemper, *et al.*, “Screening for autism spectrum disorder in young children: Us preventive services task force recommendation statement,” *JAMA*, vol. 315, no. 7, pp. 691–696, 2016.

[4] S. L. Hartley *et al.*, “The relative risk and timing of divorce in families of children with an autism spectrum disorder.” *Journal of Family Psychology*, vol. 24, no. 4, p. 449, 2010.

[5] D. L. Christensen, “Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, united states, 2012,” *MMWR. Surveillance Summaries*, vol. 65, 2016.

[6] D. S. Mandell *et al.*, “Factors associated with age of diagnosis among children with autism spectrum disorders,” *Pediatrics*, vol. 116, no. 6, pp. 1480–1486, 2005.

[7] C. Song *et al.*, “A Comparative Study of Social Scene Parsing Strategies between Children with and without Autism Spectrum Disorder,” *North American Journal of Medicine and Science (NAJMS)*, vol. 9, no. 3, pp. 96–103, July 2016.

[8] M. Rutter *et al.*, “Autism diagnostic interview-revised,” *Los Angeles, CA: Western Psychological Services*, vol. 29, p. 30, 2003.

[9] E. C. Fenske *et al.*, “Age at intervention and treatment outcome for autistic children in a comprehensive intervention program,” *Analysis and Intervention in Developmental Disabilities*, vol. 5, no. 1-2, pp. 49–58, 1985.

[10] K. Cho *et al.*, “Gaze-wasserstein: Exploring a quantitative screening approach to autism spectrum disorder,” in *IEEE Annual Wireless Health Conference*, oct 2016.

[11] L. Zwaigenbaum *et al.*, “Behavioral manifestations of autism in the first year of life,” *International journal of developmental neuroscience*, vol. 23, no. 2, pp. 143–152, 2005.

[12] J. N. van der Geest *et al.*, “Eye movements, visual attention, and autism: a saccadic reaction time study using the gap and overlap paradigm,” *Biological Psychiatry*, vol. 50, no. 8, pp. 614–619, 2001.

[13] Tobii. (2016, oct) Tobii eyex controller. [Online]. Available: https://tobiigaming.com/?v=7516fd43adaa

[14] Dell. (2017, jan) Dell 22 monitor - p2214h. [Online]. Available: http://www.dell.com/en-us/shop/dell-22-monitor-p2214h-21-5-/apd/320-9791/monitors-monitor-accessories

[15] K. Ni *et al.*, “Local histogram based segmentation using the wasserstein distance,” *International journal of computer vision*, vol. 84, no. 1, pp. 97–111, 2009.

[16] K. Pierce *et al.*, “Face processing occurs outside the fusiformface area’in autism: evidence from functional mri,” *Brain*, vol. 124, no. 10, pp. 2059–2073, 2001.

[17] K. Koldewyn *et al.*, “Global/local processing in autism: Not a disability, but a disinclination,” *Journal of autism and developmental disorders*, vol. 43, no. 10, pp. 2329–2340, 2013.

[18] O. I. Lovaas *et al.*, “Stimulus overselectivity in autism: a review of research.” *Psychological bulletin*, vol. 86, no. 6, p. 1236, 1979.

[19] S. Fletcher-Watson *et al.*, “Eye-movements reveal attention to social information in autism spectrum disorder,” *Neuropsychologia*, vol. 47, no. 1, pp. 248–257, 2009.

[20] J. P. McCleery *et al.*, “Atypical face versus object processing and hemispheric asymmetries in 10-month-old infants at risk for autism,” *Biological psychiatry*, vol. 66, no. 10, pp. 950–957, 2009.