Predicting Autistic Traits Using Eye Movement during Visual Perspective Taking and Facial Emotion Identification

Kota Iwauchi[1], Hiroki Tanaka[1], and Satoshi Nakamura[1]

***Abstract*— Autistic traits are broad in severity and difficult to measure quantitatively. Quantitative measurement would be helpful in determining the effectiveness of training and therapy for autistic traits. The development of eye-tracking technology has made it easier to understand autistic traits. Previous works showed that autistic traits can be predicted from eye movements in the facial emotion identification task. It may be possible to measure autistic traits more accurately than the only prediction from facial emotion identification ability. In this study, we used a visual perspective taking task. The results showed that the Social Responsiveness Scale-2 score, which is associated with autistic traits, was predicted at 0.414 in Spearman's correlation coefficient by using eye movements obtained from the two tasks.**

I. Introduction

Autism spectrum disorder (ASD) is a complex and heterogeneous neurodevelopmental disorder. Individuals with ASD have lifelong deficits in social interaction and communication. An ASD diagnosis is made by psychiatrists based on diagnostic criteria [1]. Autistic traits are broad in severity and difficult to measure quantitatively. Quantitative measurement would be helpful in determining the effectiveness of training and therapy for autistic traits. Among various indicators, eye movement, which can be easily obtained, has attracted much attention. The development of eye-tracking technology has made it easier to understand autistic traits. The autistic group has been shown to perform eye scanning differently from the control group during facial emotion identification [2]. Another previous work indicated that communicators' facial expressions could significantly affect the gaze behavior of ASD participants [3]. Many researchers used a binary classification between autistic and control groups [4]; however, autism has a spectrum of symptoms. There are also people with high autistic traits among the general population. Problems with social skills may cause difficulties in social activities. To resolve this, it is essential to undergo social psychological therapy, such as social skills training. Quantifying the degree to which autistic traits are important in order to determine the effectiveness of training and therapy. Against this background, this paper provides the prediction of autistic traits for members of the general population. Previous studies have used eye movement during facial emotion identification to predict autistic traits [5], [6]. However, facial emotion identification is not the only identifiable difficulty related to autistic traits.

People with high autistic traits are known to have difficulty predicting the mental states of others [7]. This ability is called Theory of Mind (ToM). One capability of ToM is perspective taking. People with high autistic traits have difficulty with perspective taking. This makes it difficult for people with high autistic traits to take the other person's point of view through social perspective taking. Social perspective taking is also known to be related to visual perspective taking (VPT) [8]. The relationship between autism and visual perspective taking has been discussed in a review article [9], and the relationship between eye movement and ASD has also been investigated [10]. No study has been conducted to predict autistic traits by machine learning with VPT's eye movements, such as was done with the eye movements of facial emotion identification. By using a task that reflects egocentricity (concerned with the individual rather than society), it may be possible to measure autistic traits more accurately than the only prediction from facial emotion identification ability.

We propose a method to predict autistic traits more accurately by measuring cognitive activities using eye movement during facial emotion identification and eye movement during VPT. Specifically, we compared the prediction results of the Social Responsiveness Scale-2 (SRS-2), which is associated with autistic traits, using eye movement during the facial emotion identification test (FEIT), eye movement during the VPT task, and a composite of eye movement during the FEIT and VPT. We used linear regression and partial least squares (PLS) regression as machine learning models. As a result, the prediction results using the combined eye movement had a coefficient of determination of 0.121 and a Spearman's rank correlation coefficient of 0.414, which were 0.92 higher in root-mean-square error (RMSE) than the prediction results using eye movement during the FEIT only. We summarize the details of these results in this paper.

II. Methodology

We aim to improve the prediction accuracy of autistic traits by combining eye movement during the FEIT and VPT compared to only using eye movement during the FEIT. We used the SRS-2 to measure autistic traits and evaluated the prediction accuracy of the scores.

A. Participants

This study was conducted with ethical approval from the Nara Institute of Science and Technology. Data were collected from 28 participants (11 males and 17 females) between the ages of 22 and 35. Written and oral explanations

*This work was supported by the Core Research for Evolutional Science and Technology (Grant No. JPMJCR19A5).

[1]Kota Iwauchi, Hiroki Tanaka, and Satoshi Nakamura are with Division of information science, Nara Institute of Science and Technology, Ikoma, Japan.

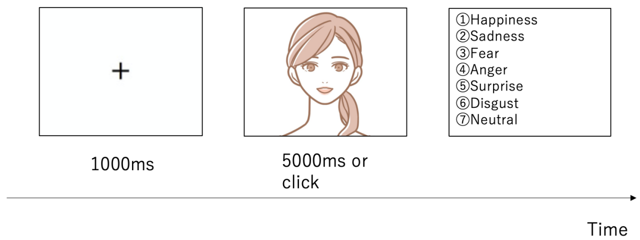

Fig. 1. The FEIT flow. After 5 seconds or by clicking, a choice will appear and you will move to the next question. There are 21 questions in total. For copyright reasons, the images shown here are schematic diagrams, not the actual images.

were given to all participants and their consent was obtained. The SRS-2 [11], Kikuchi Scale of Social Skills:18 (KISS:18) [12], and the new version of the State-Trait Anxiety Inventory (STAI) [13], which are the predictors of this study, were also obtained for all participants. In this study, we report only on the SRS-2. In this experiment, we also obtained data on the evaluation of the reliability of virtual agents according to the degree of their theory of mind and on the feedback evaluation of the social skills training system [14], but these are not included in this report. The second author can provide the eye movement data obtained in this study on request.

B. Social Responsiveness Scale-2

The SRS-2 is an objective assessment instrument containing 65 questions. It was originally designed to assess patients on the autistic spectrum, but it can also differentiate various psychiatric disorders. In addition, its validity has been investigated not only for a disorder group but also for healthy participants [11]. Therefore, the SRS-2 is related to autistic traits in members of the general population as well.

The SRS-2 measures autistic traits using five sub-scales (social awareness, social cognition, social communication, social motivation, and restricted interests and repetitive behavior); the higher the score, the stronger the trait toward autism. The mean and standard deviation of the total SRS-2 raw score is of the 28 participants in this experiment are 56.46±26.03.

C. Tasks

In this study, we used the FEIT [15] and the VPT task [16] developed by Samson et al. We describe each task below.

1) Facial Emotion Identification Test: Fig. 1 shows the flow of the experiment during the FEIT eye movement acquisition. After the participants were instructed to voice their answers aloud, a cross was presented for a second to guide their gaze to the center of the screen. Seven choices (happiness, sadness, fear, anger, surprise, disgust, and neutral) were then displayed by clicking with the mouse after 5 seconds or when the facial emotions were recognized. The participants were asked to select one of these. The test consisted of 21 questions in which three pictures were randomly selected for each of the seven facial emotions.

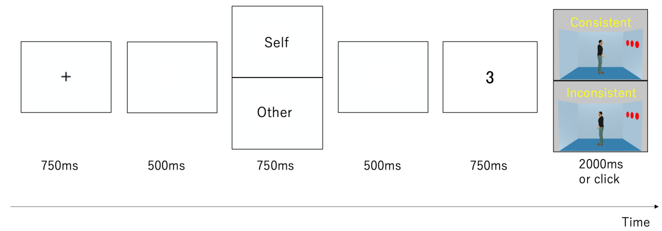

Fig. 2. Flow of VPT. After the "Self" or "Other" instruction, a number is displayed to indicate the number of red dots and an image of the task is shown. Participants are asked to left-click if the instruction matched the last image displayed and right-click if it did not.

2) Visual Perspective Taking: Fig. 2 shows the flow of the VPT experiment. The relationship between this task and the autistic participants has been investigated [10], [17], [9]. The task is to check how many red dots can be seen from the self's or other's point of view and to answer whether the instruction matches the actual image content. After a pause of 500 ms, the word "Self" or "Other" is displayed and, after another pause of 500 ms, a number is shown to specify the number of red dots. Finally, an image with a person in the center and red dots on the left and right walls is displayed and the user is asked to left-click (or 2000 ms elapses) if the situation matched the instruction (Matching) and right-click if it did not (Mismatching). There are four conditions: Self-Consistent, Self-Inconsistent, Other-Consistent, and Other-Inconsistent. In this study, a total of 52 questions were asked with 11 for the Self-Consistent condition, 14 for the Other-Consistent condition, 13 for the Self-Inconsistent condition, 10 for the Other-Inconsistent condition, and 4 for the Filler condition in which no dots are displayed. The tasks were randomly selected. It is recommended that only the Matching case be used in the analysis [16] and, in this study, the data from the Matching case were used in the analysis. The average correct response rate for participants was 93.7%, with all participants resolving the task without difficulty.

D. Eye Movement and Feature Extraction Procedure

We acquired eye movements during these tasks. We used the Tobii Pro Fusion with a sampling rate of 120 Hz and a display resolution of 1920 × 1080 to show the tasks. The participants sat about 65cm from the display. We used the Tobii Pro Lab (Version 1.145) for eye movement analysis [18]. In this study, the areas of interest (AOI) were set to the eyes, mouth, and face for the FEIT images and to the person standing in the center, the right side where the red dot appears, and the left side of the wall for the VPT images. The acquired eye-movement features are the number of fixations (Number of Fixations) and the number of saccades (Number of Saccades). Fixation is the action of stopping the gaze at a specific position. It is a slow, subtle movement to align the gaze with the object and prevent perceptual fading. A saccade is a fast eye movement in which both eyes move in the same direction, induced either spontaneously or involuntarily. We used the Velocity-Threshold Identification (I-VT) fixation classification algorithm, which is a velocity-

TABLE I

LIST OF FEATURES ACQUIRED FOR EACH TASK. WE OBTAINED A TOTAL OF 55 FEATURES.

Task	Feature	Condition	Total features
FEIT	Number of Fixations at eyes Number of Fixations at mouth Number of Fixations at face Number of Saccades Response Time	Happiness Sadness Fear Anger Surprise Disgust Neutral	35
VPT	Number of Fixations at human Number of Fixations at right wall Number of Fixations at left wall Number of Saccades Response Time	Self-Consistent Self-Inconsistent Other-Consistent Other-Inconsistent	20

based classification algorithm that categorizes fixation and saccade based on velocity. If the velocity exceeds 100/s it is classified as saccade; otherwise, it is classified as fixation.

In the FEIT, we calculated the average for each of the seven facial emotions. In VPT, we obtained the Number of Fixations and Number of Saccades for the four conditions and calculated the average for each condition. Response time was also obtained for a total of 55 features. Table I shows a summary of the extracted features.

E. Modeling and Evaluation

Next, we explain the model and algorithm used for the prediction. In this study, we obtained data from 28 participants but were unable to obtain eye movements for one of them. Therefore, we constructed a machine learning model using the data from 27 participants. When the number of features (k=55) is extremely large compared to the sample size (n=27), we need to deal with the curse of dimensionality. Therefore, in this study, dimensionality reduction based on mutual information content was performed before the input. Mutual information $I(x, y)$ is defined as a feature x, an objective variable y, their respective probabilities $p(x)$ and $p(y)$, and the simultaneous probabilities $p(x, y)$, and can be calculated as follows:

$$I(x, y) = \sum_{x,y} p(x, y) \left[\ln p(x, y) - \ln p(x)p(y)\right]. \quad (1)$$

In this study, the number of input features was adjusted according to the training score and cross-validation score to prevent overfitting, and eight features were input in order of mutual information content.

This study is interested in the effect of input features on estimators and interpretability. Therefore, we use linear regression and PLS regression, which are linear models. In addition, all features were standardized to calculate the standard partial regression coefficient. We performed nested leave-one-out cross-validation to adjust and evaluate the parameters. The model was compared under three conditions:

- Only FEIT features are input to the model
- Only VPT features are input to the model
- FEIT and VPT features are input to the model

The output is each SRS score. We used the coefficient of determination R^2, RMSE, and Spearman's correlation coefficient ϱ as evaluation indices. Scikit-learn was used for implementation [19].

TABLE II

RESULTS FOR EACH MODEL PREDICTING SRS. THE BEST RESULTS FOR EACH METRIC ARE IN BOLD.

Model	Feature set	R^2	RMSE	ϱ
PLS regression	FEIT	0.053	25.32	0.400
	VPT	-1.11	37.83	-0.319
	FEIT+VPT	**0.121**	**24.40**	**0.414**
Linear regression	FEIT	-0.009	26.13	0.321
	VPT	-2.56	49.11	-0.440
	FEIT+VPT	**0.077**	**25.00**	**0.400**

III. RESULTS

Table II shows the results of prediction using PLS regression and linear regression. The evaluation metrics of models when both FEIT and VPT eye-movement features were used in PLS regression showed the highest values of 0.121 for the coefficient of determination, 24.40 for RMSE, and 0.414 for the correlation coefficient. An uncorrelated test of the correlation coefficient ϱ between the predicted value and the actual value of the model when all the features are used in the PLS regression, which obtained the best correlation in this study, yielded a result of p=0.0317. This model has a significant difference at the $p<0.05$ level. Scatter plots of the true values and predicted values when the SRS is predicted using only the FEIT feature and all the features are shown in Fig. 3 (left side: used only FEIT feature model, right side: used features of both FEIT and VPT). Since PLS regression is a linear model, the coefficients of the regression equation can be calculated. Table III shows the features with large absolute values obtained by averaging the regression coefficients of each model obtained through leave-one-out cross-validation. The regression coefficients are those of the FEIT and FEIT+VPT models, which were predictable.

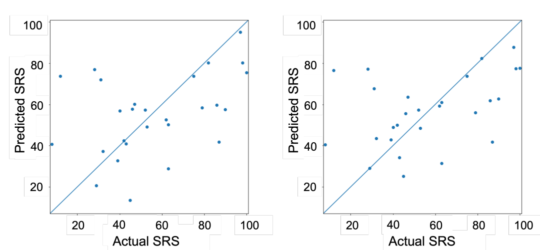

Fig. 3. Scatter plots of the true SRS score and predicted SRS score. The left side is the model that uses only the FEIT feature set. The right side is the model that uses the features of both the FEIT and VPT.

IV. DISCUSSION

In the experiment, the model using the FEIT's and VPT's eye movements outperformed the model using only the FEIT features in all of the evaluation metrics. The model using only the VPT's eye movements failed to predict the results. This trend was also obtained for both the PLS regression

TABLE III

THE BEST 5 REGRESSION COEFFICIENTS OF PLS REGRESSION FOR EACH FEATURE SET.

Feature Set	Feature	Regression Coefficient
FEIT	Number of Saccades Sadness	-13.17
	Number of Saccades Surprise	-6.39
	Number of Fixations at face Sadness	5.43
	Number of Saccades Anger	-4.62
	Reaction time Surprise	3.68
FEIT+VPT	Number of Saccades Sadness	-13.17
	Number of Saccades Surprise	-6.19
	Number of Fixations at face Sadness	5.41
	Number of Saccades Self-Inconsistent	4.85
	Number of Fixations at human Self-Consistent	-4.68

and linear regression models. The regression coefficients in Table III are particularly large for the features obtained from the FEIT. The features contributing to the top three predictions are from the FEIT, which is not different from the model when only FEIT features are used, but the lower two features are from VPT and not from the FEIT. The VPT features are among the top 5 features, indicating that the VPT task contributes to the prediction of SRS. In addition, a positive coefficient is applied to the saccade frequency when the self-viewpoint is asked and a negative coefficient is applied to the number of fixations when the self-viewpoint is asked. This means that a higher number of saccades predicts a larger SRS score and a higher number of fixations predicts a lower SRS score when the self-viewpoint question is asked. When solving this task, it is known that the control group unconsciously counts the number of dots from the other's viewpoint even when they are asked to take the self-viewpoint [16], [20]; this is the "altercentric effect". This suggests that the higher the SRS score, the greater the autistic traits and the less the occurrence of the altercentric effect. In other words, it suggests that the model captures egocentricity, one of the autistic traits.

Our study faced some limitations. The sample size is very small. Further validation with a larger number of data is needed. It will be also interesting to see if these results are consistent with other participant groups. The effects of increasing the number of tasks on participants, e.g., fatigue, should also be considered; for both the FEIT and VPT, the optimal number of tasks to adequately measure autistic traits should be investigated. This study suggests that using both the FEIT and VPT, rather than the FEIT only, improves predictive results. It may be possible that VPT is able to predict aspects of autistic traits that were not predicted by the FEIT alone. A detailed analysis is needed to determine which aspects of autistic traits can be assessed better by VPT than by FEIT alone. Much more work is required before this eye-movement approach reaches a level consistent with clinical use for establishing autistic traits.

V. CONCLUSION

In this paper, we show that the prediction accuracy of autistic traits can be improved by using eye movement during VPT in addition to facial emotion identification. This suggests that tasks that measure different cognitive activities can evaluate autistic traits multidimensionally, resulting in improved prediction accuracy.

VI. ACKNOWLEDGMENT

Funding was provided by the Core Research for Evolutional Science and Technology (Grant No. JPMJCR19A5).

REFERENCES

[1] A. P. Association, *Diagnostic and statistical manual of mental disorders : DSM-5*. American Psychiatric Association Arlington, VA, 5th ed ed., 2013.

[2] W. Jones and A. Klin, "Attention to eyes is present but in decline in 2–6 month-olds later diagnosed with autism," *Nature*, vol. 504, 11 2013.

[3] Z. Akhtar and T. Guha, "Computational analysis of gaze behavior in autism during interaction with virtual agents," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1075–1079, IEEE, 2019.

[4] S. Raj and S. Masood, "Analysis and detection of autism spectrum disorder using machine learning techniques," *Procedia Computer Science*, vol. 167, pp. 994–1004, 2020. International Conference on Computational Intelligence and Data Science.

[5] M. Jiang, S. M. Francis, A. Tseng, D. Srishyla, M. DuBois, K. Beard, C. Conelea, Q. Zhao, and S. Jacob, "Predicting core characteristics of asd through facial emotion recognition and eye tracking in youth," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 871–875, IEEE, 2020.

[6] K. Iwauchi, H. Tanaka, K. Okazaki, Y. Matsuda, M. Uratani, T. Morimoto, and S. Nakamura, "Eye-movement analysis on facial expression for identifying children and adults with neurodevelopmental disorders," *Frontiers in Digital Health*, vol. 5, 2023.

[7] S. Baron-Cohen, "Perceptual role taking and protodeclarative pointing in autism," *British Journal of developmental psychology*, vol. 7, no. 2, pp. 113–127, 1989.

[8] P. Michelon and J. Zacks, "Two kinds of visual perspective taking," *Perception psychophysics*, vol. 68, pp. 327–37, 03 2006.

[9] A. Pearson, D. Ropar, and A. F. de C. Hamilton, "A review of visual perspective taking in autism spectrum disorder," *Frontiers in human neuroscience*, vol. 7, p. 652, 2013.

[10] H. Doi, C. Kanai, N. Tsumura, K. Shinohara, and N. Kato, "Lack of implicit visual perspective taking in adult males with autism spectrum disorders," *Research in Developmental Disabilities*, vol. 99, p. 103593, 2020.

[11] M. J. N. Constantino and P. C. P. Gruber, *Social Respon- siveness Scale, Second Edition (SRS-2) Back*. Western Psychologi- cal Services, 2012.

[12] A. Kikuchi, "The social skills are measured: The handbook of kiss-18," 2007.

[13] C. D. Spielberger, "State-trait anxiety inventory for adults," 1983.

[14] H. Tanaka, T. Saga, K. Iwauchi, and S. Nakamura, "Trustworthiness of virtual agents by effects of theory of mind and social skills training," in *The IEEE conference series on Automatic Face and Gesture Recognition*, 2023.

[15] "Facial emotion identification test for adult," [n.d.]. Accessed on 2021.05.14.

[16] D. Samson, I. A. Apperly, J. J. Braithwaite, B. J. Andrews, and S. E. Bodley Scott, "Seeing it their way: evidence for rapid and involuntary computation of what other people see.," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 36, no. 5, p. 1255, 2010.

[17] T. T. Brunyé, T. Ditman, G. E. Giles, C. R. Mahoney, K. Kessler, and H. A. Taylor, "Gender and autistic personality traits predict perspective-taking ability in typical adults," *Personality and Individual Differences*, vol. 52, no. 1, pp. 84–88, 2012.

[18] Tobii Pro AB, "Tobii pro lab." Computer software, 2014.

[19] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, *et al.*, "Scikit-learn: Machine learning in python," *Journal of machine learning research*, vol. 12, no. Oct, pp. 2825–2830, 2011.

[20] H. J. Ferguson, I. Apperly, and J. E. Cane, "Eye tracking reveals the cost of switching between self and other perspectives in a visual perspective-taking task," *The Quarterly Journal of Experimental Psychology*, vol. 70, no. 8, pp. 1646–1660, 2017.