A VR-based Training and Intelligent Assessment System Integrated with Multi-modal Sensing for Children with Autism Spectrum Disorder

Yan-Qing Chen
Dept. of Computer Science and Information Engineering; National Central University
yanqingchen0709@gmail.com

Fu-An Lin
Dept. of Computer Science and Information Engineering; National Central University
frank30719@gmail.com

Ting-Yu Yang
Dept. of Computer Science and Information Engineering; National Central University
chris312319@gmail.com

Shih-Ching Yeh
Dept. of Computer Science and Information Engineering; National Central University
shihching.yeh@g.ncu.edu.tw

Eric Hsiao-Kuang Wu
Dept. of Computer Science and Information Engineering; National Central University
hsiao@csie.ncu.edu.tw

James M Poole
FAAP VRCA, Inc
fstbraiin@gmail.com

Charles Shao
FAAP VRCA, Inc
shao8843@yahoo.com

***Abstract*—Autism Spectrum Disorder (ASD) exhibits social communication and social interaction disorders, and abnormal restrictive and repetitive behaviors. However, symptoms of infants less than 1-year-old are difficult to reliably predict subsequent diagnosis. Patients with mild ASD may not be discovered until school age, because schools have more opportunities for social activities. In addition, the therapist also needs to consider the labor cost. To provide effective treatment, it also needs to consume more resources. The current situation in Taiwan is that outlying islands and remote areas often have insufficient manpower for therapists. If VR technology can be applied, some of the problems may be solved. However, due to the global pandemic, COVID-19, early treatments or group treatments in many countries have been forced to stop. If VR technology can provide interpersonal interaction scenes, the training of ASD children can hardly be affected.**

This research uses Virtual Reality (VR) technology, combined with wearable multi-model sensing technology, including EEG, eye tracking, heart rate variability (HRV), and breath-sensing strap. Physiological signals and game performance data are collected while users are training, and integrate multiple evaluation scales such as ADOS, SRS, and CBCL. Statistical analysis of these data is performed to classify them through machine learning models to develop a VR assistance system that can be used to evaluate the diagnosis, severity, and social behavior treatment of ASD. This system presents assessment and therapy in a game-oriented way. In addition to enhancing the incentives for users to participate, it provides better training results than traditional training. It is also an effective and convenient tool for the therapist to use during evaluation and training.

Keywords—Autism Spectrum Disorder, ASD, EEG, Virtual Reality, Training, Intervention, Evaluation, Diagnosis, Machine Learning

I. INTRODUCTION

According to the "Diagnostic and Statistical Manual of Mental Disorders 5, DSM-5" [1], the diagnosis of ASD is mainly focused on two aspects of symptoms, namely, the barriers of social communication and interpersonal interaction, and behavioral patterns of repetition and limitation. There are three main obstacles to social communication and interpersonal interaction, including mutual communication, non-verbal communication, and defects in understanding social relationships. The repetitive and limited behavior patterns mainly have four parts, including stereotyped movement or speech, high persistence, narrow interest, and insufficient or overreaction to feelings.

However, it is a challenge to use DSM-5 criteria to diagnose whether young children have ASD. Patients with mild ASD may not be discovered until school age, because schools have more opportunities for social activities [2]. However, studies have confirmed that if early detection and early correction can be made, it makes up for the defects of the innate learning ability of autistic patients, reduces the appearance of unsuitable and destructive behaviors, and enables them to fully realize their potential and prevent their symptoms. Early behavioral intervention is helpful [3], which encourages us to develop more methods to diagnose ASD as soon as possible.

In recent years, virtual reality (VR) have rapidly emerged. Early treatment, training, and innovative technology have been integrated to make quantitative performance possible. On the other hand, compared to the physical environment, the simulated 3D virtual scene can not only overcome the problem of being unable to teach outdoors in bad weather but also provide a safer environment. In addition, another advantage of using a virtual environment is that the parameters, events, interferences, and so on in VR can be precisely controlled by software combined with a VR head-mounted display and provide users with an immersive interactive experience.

In addition, wearable sensing devices are also used in this study, which conveniently measures various physiological data, including EEG, eye tracking, HRV, and breath-sensing strap. Especially in medical and health care, integrating VR systems with wearable sensing technology has remarkable advantages over traditional diagnostic methods. Compared to traditional methods, it is easier to obtain movements and physiological information through wearable sensing devices. This also greatly reduces the cost of measuring physiological indicators and obtains more diverse medical information, which makes the diagnosis more accurate. At the same time, this research presents evaluation and treatment in a game-based way, which is a more friendly plan for patients. Therefore, it enhances the incentives for case participation and provides better results.

The physiological signals and game performance data that are collected during the process are analyzed with statistics and machine learning technology to create a more accurate diagnosis system.

II. Related Work

Patients with ASD usually have great difficulties in social communication and interaction. To assess whether users have effective social communication, the time that the listener looks at the communicator's face while interacting with each other is one of the most important indicators [4]. Therefore, this research combines eye-tracking technology. At present, many studies on eye-tracking have confirmed that eye movement can be used to detect the difference between ASD patients and normal children. For example, Klin *et al*. [5] found that adolescents and adults with ASD are less likely to look at the eye area of characters in movies so that the main characteristics of ASD can be transformed into more detailed and quantifiable data to further the study. Jones *et al*. [6] used eye-tracking technology to compare the differences between 2-year-old ASD patients and those with normal development. They found that ASD patients were less likely to look at the actor's eyes, but might look at the actor's mouth more. In addition, Babu *et al*. [7] used VR and eye-tracking technology to evaluate and treat ASD patients aged 10−19 years, using different social situations and questionnaires to analyze their fixation time, pupil size, blink frequency, and so forth to reflect the tendency of ASD and the degree of anxiety.

In addition, the techniques and methods for evaluating and treating ASD patients have other aspects, including pupil size, blinking frequency, skin temperature, and ECG. For example, Kuriakose *et al*. [8] combined a variety of sensors in a social situation based on VR simulation to study physiological indicators such as pulse, skin temperature, and ECG to map the user's ASD tendency and anxiety.

However, the past studies still have the following shortcomings. First, the current ASD samples studied using VR technology are old, and there are few studies on preschool children with ASD. Second, although there are currently researches on the application of VR to ASD, there are little researches on neural sensing devices that combine eye-tracking, pupil size, fixation time, HRV, and breath-sensing. In addition, previous studies only used statistical analysis methods to conduct comparative analysis or correlation analysis on individual indicators and did not perform overall modeling analysis on all indicators. At the same time, fewer studies use machine learning-related methods for data analysis, which is a pity, and it also means that more research is needed in this field.

Therefore, there are two purposes in this study. First, because the prognosis of ASD is closely related to the time of intervention, we set the target at children aged between 4 to 6 to achieve the purpose of early diagnosis and treatment. Second, for the assessment and treatment of ASD, this research integrated EEG, heart rate variability, breath-sensing strap, and eye-tracking during the process. The machine learning methods are leveraged to analyze task performance, multi-modal data, and evaluation scales, which include ADOS, SRS, CBCL, and so on, and establish ASD assessment and auxiliary diagnosis system.

It is expected widely used in clinical practice to help reduce resources and costs of the social behavior therapy and assessment of ASD.

III. System Design

The system architecture of the VR-assisted diagnosis system for ASD assessment and social behavior therapy (Fig. 1) includes an evaluation module, a social behavior therapy module, a multi-modal sensing module, and a machine learning module.

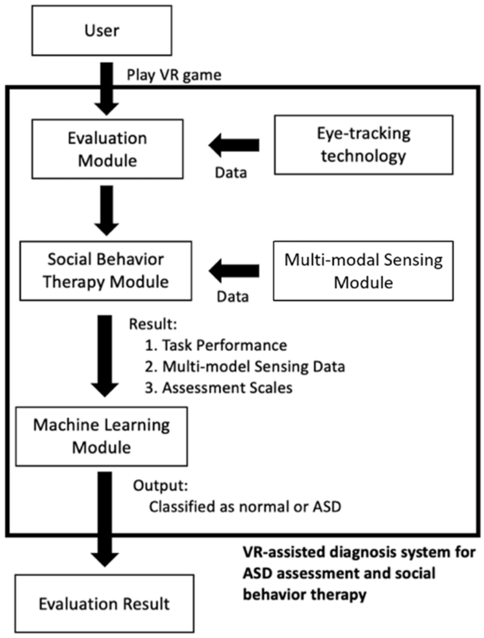

Fig. 1. System Architecture.

A. Evaluation module

The design of the evaluation context refers to the Autism Diagnostic Observation Scales commonly used in clinical and research internationally (ADOS; Lord *et al*., 1999). ADOS is a semi-structured game observation tool that simultaneously evaluates four aspects: social interaction, communication, games, and repetitive behaviors. ADOS has developed four modules to apply to different language proficiency cases. Considering that the participants included in this study are 4−6 years old preschoolers, so the VR evaluation scenario in which the main dimension of evaluation is social interaction and communication is designed according to the ADOS evaluation content of Module 2. The collected measurement indicators include eye gaze time (character eyes, face, and background), responding joint attention, initiating joint attention, reciprocal conversation.

B. Social Behavior Therapy Module

The design of the social behavior therapy module adopts the current international trend of early intervention for autism: Naturalistic developmental behavioral interventions (NDBI, Schreibman *et al*., 2015). Three core essences are emphasized.

(1) The nature of learning goals: children's learning goals are to integrate multiple developmental areas and pay attention to the learning of precursor abilities of various important development milestones. Learning is carried out in a series of interactions to promote the assortment of various abilities.

(2) The nature of the learning situation: following the children's interests or choices, the child initiates the beginning of intervention activities.

(3) The nature of strategies to enhance development: in a highly predictive natural game interaction, combined with behavioral intervention strategies (such as demonstrations), behavior shaping, linking, prompting, differentiation enhancement, systematically expand children's mutuality, communication, social interaction, and other abilities.

C. Multi-modal Sensing Module

To quantify the neural behavior data, different indicators were designed to be applied to the machine learning model. In this study, four wearable devices are used, including EEG, eye-tracking, heart rate variability, and breath-sensing strap to collect real-time physiological data of users.

(1) EEG
Using brain wave measurement equipment, Looxid can be integrated with VR head-mounted displays. Six electrodes were selected including AF3 and AF4, which are attention-related. AF7, AF8, Fp1, Fp2 extract features from time-domain and frequency-domain such as delta, theta, alpha, beta, and gamma band.

(2) Eye-tracking
HTC Vive Pro Eye's built-in Tobii eye-tracking sensor device is used to record pupil size, fixation time, eye openness, and other information.

(3) Heart rate variability (HRV)
Using a heart rate sensor to measure the current state of the participant, the tendency of ASD and the degree of anxiety are monitored.

(4) Breath-sensing strap
A Breath-sensing strap measures the ups and downs of the participant's abdomen when breathing, and applies it in the treatment process.

D. Machine Learning Module

After collecting various physiological data, task performance, and the evaluation scale, we use a t-test to find the difference between ASD and normal children and compare whether there is a significant difference before and after the subjects participate in VR social behavior therapy.

On the other hand, machine learning methods are used to put all the significant feature data sets into the model for training, to be able to classify ASD and classify the differences between subjects before and after an intervention.

IV. Methods

This study is expected to collect data from 100 subjects (aged 4 to 6 years old), 50 children with ASD in the experimental group, and 50 normal children in the control group. There were no significant differences in the average age, grade, and gender between the two groups of participants.

The expected experimental design is shown in the picture (Fig. 2). Before VR intervention, IQ, ADIR, and CASD-C are be evaluated to confirm participant conditions, and let the subjects, their parents, and teachers do ADOS, SRS, CBCL scales to understand the students' social behavior status. After that, the subjects enter the three-month VR treatment stage. The participants is pre-tested for VR through the evaluation module, and then the VR training courses of social behavior therapy is started. After all VR training courses are over, the subjects is asked to participate in the evaluation module again and fill in the ADOS, SRS, CBCL scales as post-tests to observe the subjects' progress after the VR training courses.

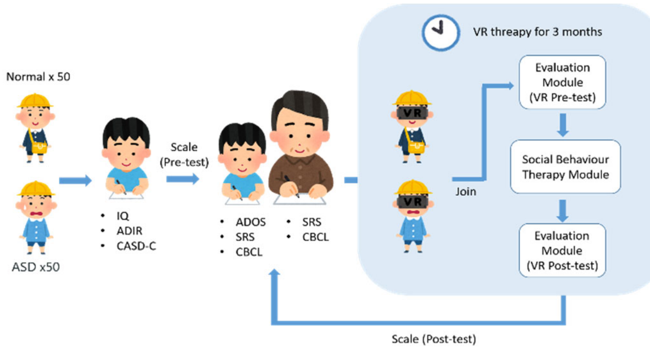

Fig. 2. Experimental Design.

In the evaluation module, the subject plays a puzzle with an NPC (Fig. 3). The NPC and the participant take turns to complete the puzzle. After the NPC starts the topic, it continues the topic or ends the topic according to the children's different oral responses, and then invites the subject to the next social behavior and treatment situation. During the process, the system will record the time the children look at the NPC's eyes, face, and other backgrounds, as well as the conversations they have.

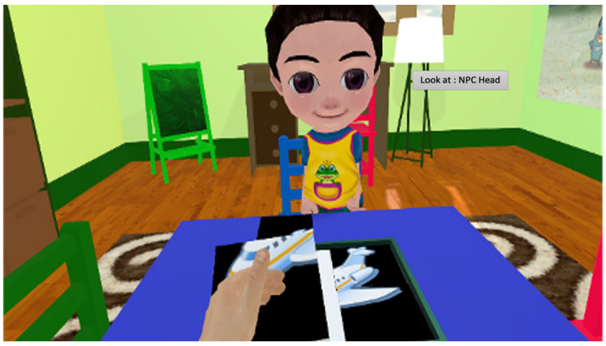

Fig. 3. Puzzle Game

There are two VR training courses in the social behavior therapy module, namely the throwing and catching a ball game and the card game. Each game includes teaching and challenge situations. The teaching mode is to teach the social and emotional behaviors, and the challenge mode is to allow the children to use the behaviors they have learned to carry out different activities and challenges to clarify what they have learned.

The game of throwing and catching the ball (Fig. 4) is played in an outdoor scene. There are three virtual characters on the screen, including a game host and two children. The subject needs to cooperate with two children to throw and catch the ball. Each time they successfully throw and catch the ball, they can obtain coins. In the teaching mode, the host prompts the rules of the game and gradually guides the subject to help the subject make appropriate behaviors. In addition, the design content of throwing and catching the ball takes the user's social activation and social response behavior as the main core axis. Special arrangements are made for emergencies that may trigger emotions to guide the children to learn emotion recognition, awareness, and regulation, including teaching to recognize the emotions of NPCs, perceiving their own emotions, and learning to regulate emotions by deep breathing. Balloons are used here for visual presentations (Fig. 5). The breathing strap is used to sense the subject's breathing, and the balloons are blown up to show the process of emotional relief through deep breathing. In the challenge mode, there is an emergency condition similar to that in the teaching mode. It is necessary to check whether the subject is assimilated to what they have learned. In addition, the completion time of the game is also recorded to continuously track the progress of the user.

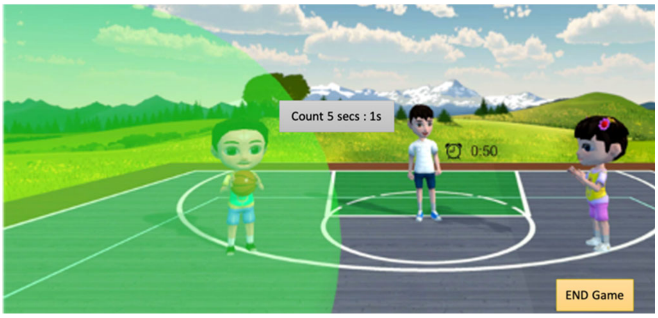

Fig. 4. The Game of Throwing and Catching the Ball

Fig. 5. The Game of Throwing and Catching the Ball – Subject Doing Deep Breath

The card game (Fig. 6) is played in the kindergarten classroom. There are four virtual characters in the scene, including a host and three children. In this game, you need to compare the number of the card to determine who win the game in this round and get coins. This game can be used to observe and cultivate the children's ability to participate in a matter with other children. In addition, since the card game is played after the catch and throw ball game is over, a variety of emergencies is arranged to observe the children's reactions and check whether the children have learned the ability to recognize and resolve emotions. In addition, situational questions, which are about choosing the current emotion of NPC or subjects themselves or choose the behavior when NPC or subject feels angry, are also designed in the card game (Fig. 7). The purpose is for the children to learn how to express their emotions. If the subject chooses the wrong one, the host and NPCs give them a hint, and guide them to choose the right option. In order to cultivate the initiating joint attention of ASD, a display section was designed in the game. If you have the largest number, you pick up the card and show it to everyone, saying that you see my card number is the biggest. Thus, subjects learn to interact with other children and increase the ability of social interaction during the process.

Fig. 6. Card Game – Play Card

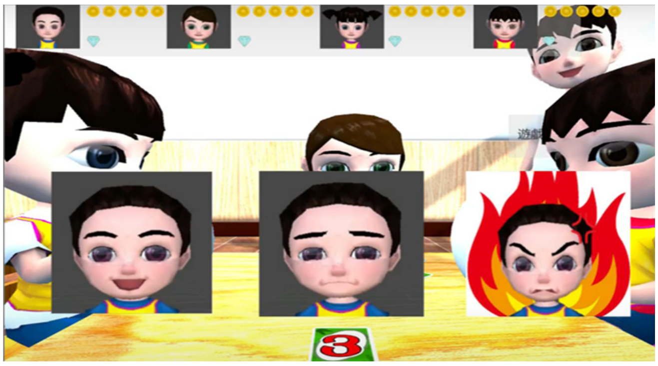

Fig. 7. Card Game – Choose Emotion

V. RESULTS

So far, in the process of evaluating the situation, we have recorded the physiological information of the child's eye-tracking, including the object coordinates, object name, pupil size, pupil diameter, and so on. In addition, since the facial fixation time of ASD is one of the very important indicators, the total time that the subject looks at each object is recorded. Here we take the evaluation module as an example to show the feature and data we have, the eye-tracking data show as follows (TABLE I).

TABLE I. EVALUATION MODULE – EYE-TRACKING FEATURE.

Feature	Content
Eye Position	X, Y, Z
Focus Object	Object name.
Pupil Diameter	Left eye and right eye included.
Eye Openness	Left eye and right eye included.
Focus Time	NPC Head, Puzzle1 to Puzzle4, Puzzle Box, and AOI (Center, Right, Left)
Time Stamp	

In terms of eye-tracking data, the content of the child's speech was recorded, too. Also, the time point of the child's speech was judged whether it was answering the sentence that the NPC communicated with or had nothing to do with the oral communication. It is also recorded whether the children looked at the eyes of the NPC and other important information when they were speaking (TABLE II).

TABLE II. EVALUATION MODULE - THE CONTENT OF THE CHILD'S SPEECH.

Feature	Describe
Recognize Result	The sentence that subject says.
Is Answer	Is the subject answer the question or just say something that is not related to the question?
Contain Keyword	The sentence that the subject says contains keywords.
Conv2Rsp_Start	The time that subjects start speaking.
Conv2Rsp_End	The time that subjects end of speaking.
Is Eye Contact	Look at NPC's eye when speaking.
Time Stamp	

VI. DISCUSSION

Compared with traditional treatment, the therapist needs to manually record every sentence the child utters and gazes at when he speaks. In terms of smart medical care, this VR system records important information and automatically outputs the results after the treatment. For therapists, it is an effective and convenient tool for diagnosis and treatment.

In addition, on the user side, we have invited three 4 to 6 years old children as participants to test the usability of the system in the previous project demonstration, and one of them is ASD. All three participants filled out the questionnaire after the experience. The feedback is that the subjects of this age group are satisfied with the fun and challenge of this system. In the follow-up, we continue to develop and modify according to the feedback of physicians and subjects, improve the entire system, and apply it to ASD assessment and social behavior therapy.

REFERENCES

[1] Diagnostic and Statistical Manual of Mental Disorders (DSM–5). Available from: https://www.psychiatry.org/psychiatrists/practice/dsm.

[2] Kliegman, R., et al., Nelson textbook of pediatrics. 2020.

[3] Dawson, G., Early behavioral intervention, brain plasticity, and the prevention of autism spectrum disorder. Dev Psychopathol, 2008. 20(3): p.775-803.

[4] A. Colburn, M. F. Cohen and S. Drucker, "The role of eye gaze in avatar mediated conversational interfaces", Jul. 2000.

[5] Klin, A., et al., Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. Arch Gen Psychiatry, 2002. 59(9): p. 809-16.

[6] Jones, W., K. Carr, and A. Klin, Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder. Arch Gen Psychiatry, 2008. 65(8): p. 946-54.

[7] Babu, P.R.K., P. Oza, and U. Lahiri, Gaze-Sensitive Virtual Reality Based Social Communication Platform for Individuals with Autism. IEEE Transactions on Affective Computing, 2018. 9(4): p. 450-462.

[8] S. Kuriakose and U. Lahiri, "Understanding the psycho-physiological implications of interaction with a virtual reality based system in adolescents with autism: A feasibility study", IEEE Trans. Neural Syst. Rehab. Eng., vol. 23, no. 4, pp. 665-675, Jun. 2015.